

BIO-INSPIRED COMMUNICATION THEORIES AND TECHNIQUES FOR
NEXT-GENERATION NETWORKS

BY

BARIŞ ATAKAN

A THESIS SUBMITTED TO THE
GRADUATE SCHOOL OF SCIENCES & ENGINEERING
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
PHILOSOPHY OF DOCTORATE
IN
ELECTRICAL & ELECTRONICS ENGINEERING

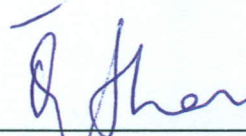
KOÇ UNIVERSITY

DECEMBER 2011

Koç University
Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of the dissertation for the degree of
Doctor of Philosophy by
Barış Atakan
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:



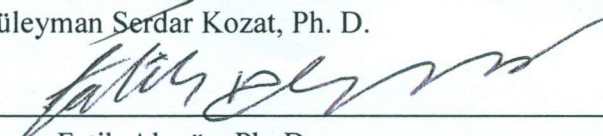
Özgür B. Akan, Ph. D. (Advisor)



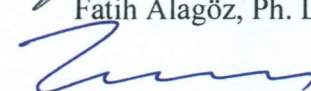
Öznur Özkasap, Ph. D.



Süleyman Serdar Kozat, Ph. D.



Fatih Alagöz, Ph. D.



Tuna Tuğcu, Ph. D.

Date: December 14, 2011

ABSTRACT

BIO-INSPIRED COMMUNICATION THEORIES AND TECHNIQUES FOR NEXT-GENERATION NETWORKS

Atakan, Barış

Ph.D., Department of Electrical & Electronics Engineering

Supervisor : Assoc. Prof. Özgür B. Akan

December 2011, 304 pages

Recent developments in macro, micro, and nanoscale communication technologies stimulate the researchers to devise a next-generation network vision. In this vision, next-generation network architectures are expected to enable the users to efficiently transmit and receive vast amount of information anywhere and anytime through any kind of communication devices from a tiny nanosensor node that operates in a human body for a healthcare application to a base station that is used as a spectrum broker to coordinate the spectrum accesses of unlicensed users. However, the realization of the next-generation network vision brings many crucial challenges that must be addressed. In particular, the centralized control of the next-generation network architectures is not a practical solution. Instead, the next-generation network architectures and communication protocols must have the capabilities of scalability, self-organization, self-adaptation, and survivability. On the other hand, the biological

systems in nature intrinsically have these capabilities. For example, insects in a colony cooperatively accomplish many complex tasks required for their vital functionalities without any need for a central control. Furthermore, natural nanoscale networks such as immune system and bacteria colonies employs incredible nanoscale communication phenomena, e.g., quorum sensing of bacteria, that can be used to devise efficient nanoscale communication paradigms for nanoscale next-generation networks. Hence, the natural biological systems may give great inspiration to develop the communication network models and techniques for the next-generation communication networks. The objective of this thesis is to develop the bio-inspired communication techniques for the different next-generation networks and provide communication theoretical analyses for some of these networks. More specifically, Distributed Node and Rate Selection (DNRS) algorithm is first proposed for wireless sensor networks (WSN) using the principles of natural immune system. Then, Biologically-inspired Spectrum Sharing (BIOSS) algorithm is introduced for cognitive radio networks (CRN), which is based on the adaptive task allocation model in insect colonies. Next, immune system-inspired Evolutionary Opportunistic Spectrum Access (ESA) protocol is presented for cognitive radio ad hoc networks. Homeostasis-inspired Autonomous Communication (HAC) protocol is introduced for wireless audio sensor networks (WASN). Based on the prey model in foraging theory, bio-inspired cross-layer (BIOX) communication and coordination protocol is described for wireless sensor and vehicular actor networks (WSAN). Then, Biological Foraging-inspired Communication (BFC) algorithm is introduced for the energy-efficient and spectrum-aware communication requirements in Intermittently-connected Mobile Cognitive Radio ad hoc Networks (IMCRN). An information theoretical approach is given for capacity of a molecular communication channel between two nanomachines. Using the stochastic model of molecular reactions in biochemical systems, a realistic channel model and a deterministic capacity expression are given for point-to-point, broadcast, and multiple-access molecular communication channels. Then, Mobile Ad hoc Molecular Nanonetwork (MAM-NET) with electrochemical communication is modeled and analyzed. Molecular Array-based Communication (MARCO) scheme is given for nanonetworks. The concept of body area nanonetworks (BAN²) with molecular communication is introduced for future nanomedicine applications. Finally, the concept of Carbon Nanotube-based nanoscale Ad hoc Networks (CANET) is described for future nanotechnology applications.

Keywords: Bio-inspired communication, next-generation networks, nanoscale communications, molecular communication, wireless sensor networks, cognitive radio networks

ÖZET

GELECEK NESİL AĞLAR İÇİN BİYOLOJİK ESİNLİ İLETİŞİM TEORİLERİ VE TEKNİKLERİ

Atakan, Barış

Doktora, Elektrik & Elektronik Mühendisliği

Tez Yöneticisi : Doç. Dr. Özgür B. Akan

Aralık 2011, 304 sayfa

Son günlerde mikro, makro, ve nanoölçekli iletişim teknolojilerindeki ilerlemeler araştırmacıları bir gelecek nesil ağ vizyonu tasarlamaya teşvik etti. Bu vizyonda gelecek nesil ağ mimarilerinin, kullanıcıları herhangi bir yerden, herhangi bir zaman, ve herhangi bir iletişim cihazı ile büyük miktarlardaki bilgiyi iletebilmesi ve alabilmesini olanaklı kılması beklenmektedir. Bu cihazlara örnek olarak insan vücudundaki bir nanotıp uygulamasında çalışan çok küçük bir nanoalgılayıcı ve lisanssız kullanıcılar için tayf acentası olarak kullanılan bir baz istasyonu verilebilir. Bununla birlikte, gelecek nesil ağ vizyonunun gerçekleştirilebilmesi çözülmesi gereken ciddi zorluklar getirmektedir. Bilhassa, gelecek nesil ağ mimarilerinin merkezi kontrolü pratik değildir. Onun yerine gelecek nesil ağ mimarilerinin ölçeklenebilme, kendi kendine organize olabilme, kendi kendini uyarlayabilme ve kalımlı olabilme yeteneklerine sahip olması gerekir. Diğer taraftan doğadaki biyolojik sistemler aslen bu yetenek-

lere sahiptir. Örneğin bir kolonideki böcekler herhangi bir kontrol ediciye ihtiyaç duymadan yaşamsal fonksiyonları için gerekli olan görevleri ortaklaşa yerine getirirler. Ayrıca bağışıklık sistemi ve bakteri kolonileri gibi doğal nanoölçekli ağlar nanoölçekli gelecek nesil ağlar için etkili iletişim paradigmaları geliştirmek için kullanılabilecek çoğunluk algılaması gibi inanılmaz nanoölçekli iletişim fenomenleri içermektedir. Bu nedenle doğal biyolojik sistemler gelecek nesil iletişim ağları için iletişim modeli ve teknikleri geliştirmede büyük esin kaynağı verebilir. Bu tezin amacı gelecek nesil iletişim ağları için biyolojik esinli iletişim teknikleri geliştirmek ve bu ağların bazıları için haberleşme teorik analizler sağlamaktır. Daha açık olarak önce doğal bağışıklık sisteminin prensiplerini kullanarak kablosuz algılayıcı ağları (KAA) için dağıtık düğüm ve hız seçim (DDHS) algoritması önerilmektedir. Daha sonra, böcek kolonilerindeki uyarlamalı görev paylaşım modelini baz alarak bilişsel radyo ağları (BRA) için biyolojik esinli tayf paylaşım (BETP) algoritması verilmektedir. Arkasından, bilişsel radyo düzensiz ağları (BRDA) için bağışıklık sisteminden esinlenen evrimsel fırsatçı tayf erişim (EFTE) protokolü sunuluyor. Kablosuz ses algılayıcı ağları (KSAA) için homeostaz esinli otonom (HEO) iletişim protokolü tanıtılıyor. Ardından, yem arama teorisindeki bir av modelini baz alarak kablosuz algılayıcı ve araçsal aktör ağları için (KAAA) biyolojik esinli çapraz katman (BECK) iletişim ve koordinasyon protokolü tanımlanıyor. Daha sonra kesikli bağlı hareketli bilişsel radyo plansız ağlardaki enerji etkin ve tayf farkında iletişim ihtiyacı için biyolojideki yem aramadan esinlenen iletişim (BYAI) algoritması veriliyor. Arkasından iki nanomakine arasındaki moleküler haberleşme için bir bilgi teoriksel yaklaşım veriliyor. Biyokimyasal sistemlerdeki moleküler reaksiyonların stokastik modellerini kullanarak moleküler iletişimdeki noktadan noktaya, yayın ve çoklu erişim kanalları için gerçekçi kanal modeli ve belirlenimci kapasite ifadeleri veriliyor. Daha sonra, elektrokimyasal iletişim ile hareketli plansız moleküler nanoağlar (HPMN) modellenerek analiz ediliyor. Nanoağlar için moleküler dizi tabanlı (MDTI) iletişim veriliyor. Arkasından ilerideki nanotıp uygulamaları için moleküler haberleşme ile vücut içi nanoağ (VNA) konsepti tanıtılıyor. Son olarak ilerideki nanoteknoloji uygulamaları için karbon nanotüp tabanlı plansız ağ (KNPA) konsepti tanımlanıyor.

Anahtar Kelimeler: Biyolojik esinli iletişim, gelecek nesil ağları, nanoölçekli iletişim, moleküler haberleşme, kablosuz algılayıcı ağları, bilişsel radyo ağları.

To my parents,

Göniül Atakan and Hasan Atakan.

ACKNOWLEDGEMENTS

I would like to first signify my sincere appreciation to my advisor, Dr. Özgür B. Akan, for his invaluable support and guidance in my doctoral study. This dissertation would not have been possible without the guidance and the help of Dr. Akan. He has been my inspiration as I hurdle all the obstacles in the completion of this research work. Due to their constructive comments on my dissertation, I would also like to thank my Ph.D. dissertation committee members, Dr. Öznur Özkasap, Dr. Süleyman Serdar Kozat, Dr. Fatih Alagöz, and Dr. Tuna Tuğcu. I cordially thank to the members of Next-generation and Wireless Communication Laboratory (NWCL). I would also like to express my sincere gratitude to my parents, my brother, and my sister for their endless support. Finally, I also thank to my wife and my son since they give me a great motivation. I am also deeply indebted to my wife for her endless love.

TABLE OF CONTENTS

ABSTRACT	iii
ÖZ	vi
DEDICATON	viii
ACKNOWLEDGEMENTS	ix
TABLE OF CONTENTS	x
LIST OF TABLES	xviii
LIST OF FIGURES	xx
ABBREVIATIONS	xxviii
CHAPTERS	
1 INTRODUCTION	1
1.1 Cognitive Radio Networks	2
1.2 Wireless Sensor Networks	4
1.3 Nanonetworks	6
1.4 Research Objectives and Solutions	9
1.4.1 Immune System-based Distributed Node and Rate Selection in Wireless Sensor Networks	10
1.4.2 Bio-inspired Spectrum Sharing in Cognitive Radio Networks	11
1.4.3 Immune System-inspired Evolutionary Opportunistic Spectrum Access in Cognitive Radio Ad Hoc Networks	11
1.4.4 Homeostasis-inspired Autonomous Communication in Wireless Audio Sensor Networks	12
1.4.5 Biological Foraging-inspired Communication and Coordination in Wireless Sensor and Vehicular Actor Networks	13
1.4.6 Biological Foraging-inspired Communication in Intermittently-connected Mobile Cognitive Radio Ad Hoc Networks	14

1.4.7	An Information Theoretical Approach for Molecular Communication in Nanonetworks	14
1.4.8	Deterministic Capacity of Molecular Information Flow in Nanonetworks	15
1.4.9	Mobile Ad Hoc Nanonetworks with Collision-based Molecular Communication	15
1.4.10	Nanoscale Communication with Molecular Arrays in Nanonetworks	16
1.4.11	Body Area Nanonetworks with Molecular Communications	16
1.4.12	Carbon Nanotube-based Nanoscale Ad Hoc Networks . . .	17
1.5	Thesis Outline	17
2	Immune System Based Distributed Node and Rate Selection in Wireless Sensor Networks	19
2.1	Motivation and Related Works	19
2.2	Immune System and Wireless Sensor Networks	21
2.2.1	Biological Immune System	21
2.2.2	Immune System Based Sensor Networks	23
2.3	Immune System Based Distributed Node and Rate Selection	24
2.3.1	Event Signal Reconstruction	24
2.3.2	Distributed Designated Node Selection	26
2.3.3	Distributed Frequency Rate Selection of Designated Nodes	31
2.4	Performance Evaluation	35
2.5	Conclusion	40
3	BIOlogically-inspired Spectrum Sharing in Cognitive Radio Networks	42
3.1	Motivation and Related Works	42
3.2	Insect Colony and Spectrum Sharing in Cognitive Radio Networks . . .	44
3.2.1	Task Allocation Model of Insect Colony	44
3.2.2	Biological Task Allocation Based Spectrum Sharing Model for Cognitive Radio Networks	45
3.3	BIO-inspired Spectrum Sharing Protocol (BIOSS)	48
3.4	Performance Evaluations	49
3.4.1	Simulation Environment	50
3.4.2	Spectrum Sharing with Single Channel Selection	50

	3.4.3	Spectrum Sharing with Multiple Channel Selection	52
	3.5	Conclusion	54
4		Immune System-inspired Evolutionary Opportunistic Spectrum Access in Cog- nitive Radio Ad Hoc Networks	55
	4.1	Motivation and Related Works	56
	4.2	Immune System and Opportunistic Spectrum Access	58
	4.2.1	Biological Immune System	58
	4.2.2	Immune System and Opportunistic Spectrum Access . . .	59
	4.3	Immune System Based Evolutionary Opportunistic Spectrum Access	61
	4.3.1	Network Model and Assumptions	61
	4.3.2	ESA Overview	62
	4.3.3	Spectrum Opportunity Identification	62
	4.3.4	Spectrum Sharing and Access	63
	4.4	Performance Evaluation	67
	4.4.1	Overview of Existing OSA Mechanisms	67
	4.4.2	Simulation Experiments	68
	4.4.3	Error-tolerance to Inaccuracy in Spectrum Identification .	73
	4.5	Conclusion	74
5		Distributed Audio Sensing with Homeostasis-inspired Autonomous Commu- nication	75
	5.1	Motivation and Related Works	76
	5.2	Network Model and Assumptions	77
	5.3	Biological Homeostasis and Wireless Audio Sensor Networks	78
	5.3.1	Biological Homeostatic System	78
	5.3.2	Homeostasis-inspired Wireless Audio Sensor Networks . .	80
	5.4	Homeostasis-inspired Autonomous Communication in WASN	81
	5.4.1	Spectrum Estimation Phase	82
	5.4.2	N-sensor Clustering Phase	83
	5.4.3	Sampling Probability Determination Phase	86
	5.4.4	Channel Frequency Selection Phase	91
	5.4.5	Signal Reconstruction Phase	92

5.5	Numerical Results	93
5.6	Conclusion	96
5.7	Appendix	97
6	Bio-inspired Cross-layer Communication and Coordination in Sensor and Vehicular Actor Networks	99
6.1	Motivation and Related Works	100
6.2	BIOX Protocol Design Principles	102
6.2.1	Network Model and Problem Formulation	102
6.2.2	Prey Model in Foraging Theory	108
6.2.3	Inspiration from Prey Model	110
6.3	Bio-inspired Cross-layer Sensor-actor Communication Model	111
6.3.1	Next-hop Selection Profitability	111
6.3.2	Channel Access Profitability	112
6.4	Bio-inspired Task Allocation for Actor Nodes	114
6.4.1	Task Allocation Profitability	115
6.4.2	Stability of BIOX Task Allocation	117
6.5	Bio-inspired Cross-layer (BIOX) Protocol Operation	118
6.5.1	Event-driven Master Actor Selection and Positioning	119
6.5.2	Next-hop Selection	120
6.5.3	Channel Access and Rate Control	120
6.6	Performance Evaluation	123
6.6.1	Energy Conservation	123
6.6.2	Timely and Reliable Sensor-actor Communication	125
6.6.3	Efficient Task Allocation	128
6.7	Conclusion	129
7	Biological Foraging-inspired Communication in Intermittently-Connected Mobile Cognitive Radio Ad Hoc Networks	130
7.1	Motivation and Related Works	131
7.2	Network Model and Assumptions	134
7.3	A Prey Model in Foraging Theory	137
7.4	Biological Foraging-inspired Communication in IMCRN	138

7.4.1	Inspiration from Prey Model	139
7.4.2	Relay Selection Profitability	139
7.4.3	Channel Selection Profitability	141
7.4.4	Relay Selection Algorithm in Biological Foraging-inspired Communication	143
7.4.5	Channel Selection Algorithm in Biological Foraging-inspired Communication	144
7.5	Performance Evaluation	147
7.5.1	Relay Selection Performance	147
7.5.2	Channel Selection Performance	150
7.6	Conclusion	154
8	On Channel Capacity and Error Compensation in Molecular Communication	155
8.1	Motivation and Related Works	156
8.2	Molecular Communication Model	157
8.3	Molecule Delivery	160
8.4	An Information Theoretical Approach for Molecular Communication	162
8.5	Adaptive Error Compensation in the Molecular Communication . . .	165
8.5.1	A Selection Strategy for Molecular Bit Transmission Prob- ability	169
8.5.2	An Adaptive Molecular Error Compensation Scheme . . .	169
8.6	Numerical Analysis	171
8.6.1	Effect of Environmental Factors on Molecular Communi- cation Capacity	172
8.6.2	Effect of the Parameters Specific to TN and RN on Molec- ular Communication	174
8.6.3	Adaptive Molecular Error Compensation Scheme	176
8.7	Conclusion	179
9	Deterministic Capacity of Information Flow in Molecular Nanonetworks . . .	180
9.1	Motivation and Related Works	181
9.2	A Realistic Channel Model for Molecular Communication	182
9.3	Deterministic Analysis of Molecular Nanonetworks	187
9.3.1	Point-to-point Molecular Channel	187

9.3.2	Molecular Broadcast Channel	189
9.3.3	Molecular Multiple-access Channel	191
9.3.4	Information Flow in a Molecular Nanonetwork	193
9.4	Numerical Analysis	195
9.4.1	Point-to-point Molecular Channel	196
9.4.2	Molecular Broadcast Channel	199
9.4.3	Molecular Multiple-access Channel	199
9.4.4	Diamond-like Molecular Nanonetwork	202
9.5	Conclusion	202
10	Mobile Ad Hoc Nanonetworks with Collision-based Molecular Communication	204
10.1	Motivation and Related Works	205
10.2	Mobile Ad Hoc Molecular Nanonetwork Communication Model . .	208
10.3	Collision and Adhesion of Nanomachines	209
10.3.1	Collision of Nanomachines	209
10.3.2	Adhesion of nanomachines	211
10.4	Molecular Neuro-Spike Communication	214
10.4.1	Molecular Neuro-Spike Communication	214
10.4.2	Information Theoretical Capacity of Molecular Neuro-Spike Communication	217
10.5	Mathematical Model of Mobile Ad Hoc Molecular Nanonetwork . .	221
10.6	MAMNET Performance Analysis	226
10.6.1	Speed and Size of Nanomachines	226
10.6.2	Speed and Size of Infostations	228
10.6.3	Contact Time	228
10.6.4	Number of Bonds Required for Adhesion	229
10.6.5	Neuro-Spike Communication Parameters	229
10.6.6	Number of Spike Bits in a Message	231
10.6.7	Number of Nanomachines	232
10.7	Conclusion	232
11	Nanoscale Communication with Molecular Arrays in Nanonetworks	233
11.1	Motivation and Related Works	234

11.2	Molecular Array-Based Communication	237
11.2.1	The Relevant Delay Variables in MARCO Channel	241
11.2.2	Probability of ISI in MARCO Channel	244
11.2.3	Derivation of Error Probability in MARCO Channel	246
11.2.4	Rate of Molecular Array-Based Communication	248
11.3	Numerical Results	249
11.3.1	Effects of Inter-Emission Time and Inter-Symbol Time	249
11.3.2	Effects of Inter-Symbol Time and Inter-Node Distance	252
11.3.3	Effects of Inter-Emission Time and Diffusion Coefficient	253
11.4	Performance Comparison With Existing Molecular Communication Schemes	255
11.4.1	Z-channel Approach	256
11.4.2	Binary Channel Approach	257
11.4.3	4-input 2-output Memoryless Channel Approach	258
11.4.4	Performance Comparison	259
11.5	Conclusion	261
12	Body Area NanoNetworks with Molecular Communications in Nanomedicine	262
12.1	Motivation	262
12.2	Nanomedicine and Body Area Nanonetworks	265
12.2.1	A Brief Overview of Nanomedicine	265
12.2.2	Molecular Communications-enabled Body Area Nanonetworks in Nanomedicine Applications	267
12.3	Communication Theoretical Approaches To BAN ² For Nanomedicine	268
12.3.1	Channel Model and Communication Rate	269
12.3.2	Channel Noise	271
12.3.3	Deployment and Topology	272
12.3.4	Computation, Logic Operations, and Decision	272
12.3.5	Medium-Access and Routing	273
12.4	Conclusion	274
13	Carbon Nanotube-based Nanoscale Ad Hoc Networks	275
13.1	Motivation	275

13.2	Hardware Components of a CANET Nanonode	277
13.2.1	Nanotransceiver	277
13.2.2	Carbon Nanotube Antenna	281
13.2.3	Nanobattery	282
13.2.4	Nanomemory and Nanoprocessor	282
13.3	Fundamental Communication Issues in CANET	283
13.3.1	Operating Frequency Band	283
13.3.2	Path Loss and Noise	284
13.3.3	Communication Rate and Range	284
13.3.4	Analog vs. Digital Communication	285
13.3.5	Encoding of Information	285
13.4	Fundamental Networking Issues in CANET	286
13.4.1	Deployment and Topology	287
13.4.2	Multi-hop Reliability	287
13.4.3	Medium Access Control	288
13.5	Conclusion	289
	BIBLIOGRAPHY	290
	VITA	304

LIST OF TABLES

Table 2.1	Relationship between Immune System and Wireless Sensor Networks (WSN).	23
Table 2.2	Simulation Parameters	35
Table 3.1	Mapping Between CRN and Insect Colony	46
Table 3.2	Simulation Parameters	50
Table 4.1	Relation between Immune System and CRN.	60
Table 4.2	Comparison of Distributed OSA MAC Protocols [84].	67
Table 5.1	Evaluation Parameters	93
Table 6.1	Relationship between Prey Model and WSN.	110
Table 6.2	Simulation Parameters	124
Table 6.3	Comparison of average task performance delay and its theoretical bound given in (6.30)	129
Table 7.1	Relationship between Prey Model and IMCRN.	139
Table 7.2	The successful channel utilization probability, β , according to varying m and σ_n	153
Table 7.3	The successful channel utilization probability, β , according to varying m and σ_ω	154
Table 7.4	The successful channel utilization probability, β , for varying m and θ . . .	154
Table 8.1	Simulation Parameters	172
Table 9.1	Simulation Parameters	195
Table 10.1	Simulation Parameters	225

Table 11.1 Rate Comparison With Respect To Varying Values of Diffusion Coefficient	
($D, \mu m^2/s$)	259
Table 11.2 Rate Comparison With Respect To Varying Values of Slot and Inter-symbol	
Time (τ and T)	260
Table 11.3 Rate Comparison With Respect To Varying Values of Inter-Node Distance	
($d, \mu m$)	260
Table 13.1 Estimated circuit, antenna, and system size (m^3) for various radios complied	
from the literatures [203]	281

LIST OF FIGURES

Figure 2.1	t_{dns} for the number of DNs	36
Figure 2.2	t_{dns} for the average total energy consumption.	37
Figure 2.3	D for varying t_{dns} with different r	38
Figure 2.4	D for varying f with different t_{dns}	38
Figure 2.5	D vs. time with different t_{dns}	39
Figure 2.6	Average total energy consumption vs. time with and without DNRS.	40
Figure 3.1	The changes in the used power at the four channels according to the time in the single selection case with the licensed user interruption.	51
Figure 3.2	The changes in the used power at four channels according to the time in the single selection case without the licensed user interruption.	51
Figure 3.3	The channel utilization according to the different number of channels in single channel selection.	52
Figure 3.4	The changes in the used power at the four channels according to the time in the multiple selection case.	53
Figure 3.5	The channel utilization according to the different number of unlicensed users in the single and multiple channel selection.	53
Figure 3.6	Expected spectrum handoff delay due to coordination for varying number of unlicensed users in single channel selection.	54
Figure 4.1	The time slot structure of a secondary user.	61
Figure 4.2	Spectrum access and sharing by the mechanisms of FHS generation using mutation and elitist selection.	65
Figure 4.3	Overall ESA operation is illustrated with SI, SS, and spectrum access. All protocol steps can be easily tracked and it is seen how and which parts of ESA operation gives inspiration from the principles of immune system.	66

Figure 4.4 The throughput and the fairness of ESA algorithm compared with the algorithms making only SI and making SI and SS. The throughput for (a) varying primary user channel idle probability where $N = 40$, (b) varying number of channels used, (c) varying number of nodes in the network, and (d) the fairness for varying number of nodes.	69
Figure 4.5 Collision probability, and the convergence of ESA algorithm.	70
Figure 4.6 The throughput performance with respect to the effect of mobility over time. (a) Mobility index = 0.2 (b) Mobility index = 0.4 (c) Mobility index = 0.6 (d) Mobility index = 0.8.	71
Figure 4.7 H_i^j with respect to varying p_a for different σ_s^2 to show the error tolerance capability of ESA.	74
Figure 5.1 The mapping between the homeostatic system elements and sensor networks.	79
Figure 5.2 A portion of irregularly sampled event signal.	81
Figure 5.3 Illustration of correlated and non-correlated neighbors around a cluster head.	83
Figure 5.4 (a) $\widehat{n}$ is shown for different $\frac{S}{K(S-H)}$ values. (b) By substituting $\widehat{n} = 1, 2, \dots, b$, numerical values for the left side of (5.11) is shown.	85
Figure 5.5 Two clusters, i.e., cluster 1 and cluster 2, with their overlapping area including a number of T-sensor.	92
Figure 5.6 (a) The average number of distinct samples delivered within a second, i.e., N_d , is shown with the event signal bandwidth, i.e., B , values. (b) f is shown with B . (c) α is shown with B . (d) The total number of samples taken by all N-sensors during a second, i.e., $\alpha f M H$, is shown with B	93
Figure 5.7 (a) The average number of distinct samples delivered within a second, i.e., N_d , is shown with the number of members in a cluster, i.e., M . (b) f is shown with M . (c) α is shown with M . (d) The total number of samples taken by all N-sensors during a second, i.e., $\alpha f M H$, is shown with M	94
Figure 5.8 (a) The average number of distinct samples delivered within a second, i.e., N_d , is shown with the number of clusters in the network, i.e., H . (b) f is shown with H . (c) α is shown with H . (d) The total number of samples taken by all N-sensors during a second, i.e., $\alpha f M H$, is shown with H	95
Figure 5.9 Overlapping Samples.	97

Figure 6.1	Illustration of an event area in WSAN. In the environment, sensor nodes take measurements and forward the measurements towards vehicular actor nodes.	102
Figure 6.2	A slot duration consisting of two different intervals named frame transmission and ACK interval.	103
Figure 6.3	A data path from sensor node i to the actor node. Hop distances $(d_{ij}, \forall i, j)$ and effective hop distances $(e_{ij}, \forall i, j)$ are shown in the path.	105
Figure 6.4	Illustration of collision domains for the link between sensor nodes i and j .	122
Figure 6.5	Variance of sensor residual energies with increasing number of events.	124
Figure 6.6	Mean of sensor residual energies with increasing number of events.	124
Figure 6.7	The number of expired nodes with increasing number of events.	125
Figure 6.8	Packet delivery ratio with time for the different number of source nodes.	126
Figure 6.9	(a) Packet delivery ratio with time for the different number of actor nodes. (b) Packet delivery ratio with time for the different reporting frequency of source nodes.	127
Figure 6.10	Performance comparison of BIOX and ESRT in terms of packet delivery ratio.	128
Figure 7.1	An IMCRN scenario is illustrated. A source node intends to deliver a message to a destination node by using the relay nodes in an environment in which a number of primary users and primary base stations coexist and operate on the licensed spectrum bands.	134
Figure 7.2	(a) In channel allocation phase, the slot duration consists of the primary sensing, secondary sensing, frame transmission, and ACK intervals. (b) After the channel allocation, the slot duration includes the primary sensing, frame transmission, and ACK intervals.	135
Figure 7.3	End-to-end delay is shown with varying upper bounds of encounter probability, i.e., p_u , for different number of nodes.	147
Figure 7.4	Total energy consumption of nodes in message delivery is shown with varying upper bounds of encounter probability, i.e., p_u , for different number of nodes.	148
Figure 7.5	Number of relay nodes that selected by BFC is shown with p_u according to different N values.	149

Figure 7.6 End-to-end delay is shown with varying p_u values for different values of energy budget E	149
Figure 7.7 Total energy consumption with varying p_u for different E values.	150
Figure 7.8 Number of relay nodes with varying p_u for different values of E	150
Figure 7.9 Upper bound of channel allocation time, τ_u , is shown with varying μ_n values for different values of σ_ω	151
Figure 7.10 Upper bound of channel allocation time τ_u is shown with varying σ_n values for different σ_ω values.	152
Figure 7.11 Upper bound of channel allocation time τ_u is shown with varying μ_ω values for different σ_n values.	152
Figure 7.12 Upper bound of channel allocation time τ_u is shown with varying σ_ω values for different σ_n values.	153
Figure 8.1 Molecular Communication Model.	158
Figure 8.2 M with varying P_A for different k_1	173
Figure 8.3 M with varying P_A for different T	173
Figure 8.4 M with varying P_A for different α	174
Figure 8.5 (a) M for different S . (b) M for different L_{ex}	175
Figure 8.6 (a) M for different t_H . (b) M for different N	176
Figure 8.7 M for varying k_1 with and without MEC.	177
Figure 8.8 M for varying T with and without MEC.	178
Figure 8.9 M for varying α with and without MEC.	178
Figure 9.1 Molecular communication channel in which TN communicates with RN by emitting molecules. All reactions among the emitted molecules and receptors on RN occur in a virtual reception volume that surrounds the RN.	183
Figure 9.2 Point-to-point molecular channel in which a single TN communicates with a single RN by emitting molecules.	187
Figure 9.3 Molecular broadcast channel with one TN simultaneously communicating with two RNs.	190

Figure 9.4 Molecular multiple-access channel with two TNs simultaneously communicating with a single RN.	191
Figure 9.5 (a) Molecular diamond nanonetwork. A source nanomachine communicates with a destination nanomachine with the help of two relay nanomachines. (b) Schematic diagram of diamond-like molecular nanonetwork.	193
Figure 9.6 Information rate of point-to-point molecular channel for different upper bounds of channel input with time.	196
Figure 9.7 Information rate of point-to-point molecular channel for the different diffusion coefficients of the medium with time.	196
Figure 9.8 Average capacity of point-to-point molecular channel according to varying x_u with different D values.	197
Figure 9.9 Average capacity of point-to-point molecular channel according to varying internode distance (d) with different observation intervals (T).	197
Figure 9.10 Information rate of point-to-point molecular channel for the different receptor concentration with time.	198
Figure 9.11 Information rate achieved by two distinct RNs located to different distances from a single TN in a molecular broadcast channel.	198
Figure 9.12 Information rate achieved by two distinct TNs using different level of upper bound for channel input to access a single RN.	200
Figure 9.13 Information rate achieved by two TNs located in different distances from RN in a molecular multiple-access channel.	200
Figure 9.14 Information flow capacity in diamond-like molecular nanonetwork for different diffusion coefficients.	201
Figure 9.15 Information flow capacity in diamond-like molecular nanonetwork for different upper bound levels of channel inputs.	201
Figure 10.1 Information flow in MAMNET.	207
Figure 10.2 Collision of two nanomachines.	209
Figure 10.3 Nanomachine 1 and Nanomachine 2 adhere with each other by M ligand receptor pairs.	212
Figure 10.4 Neuro-spike transmission between nanomachines.	214

Figure 10.5 Channel model of synaptic transmission. Transmission probabilities are shown for the binary channel.	215
Figure 10.6 Block diagram of the channel model of synaptic transmission [167]	217
Figure 10.7 (a) False alarm, miss and channel error probabilities in molecular neuro-spike channel for a fixed value of $SNR = 10$ with varying θ , ($p_1 = 0.5$). (b) Channel error probability in molecular neuro-spike channel with varying θ for different values of SNR , ($p_1 = 0.5$). (c) Mutual information with varying SNR for different values of θ , ($p_1 = 0.5$). (d) Mutual information with varying p_1 for different values of θ , ($SNR = 10$).	218
Figure 10.8 S-I-R Model. Markov chain model of an epidemic disease spreading. . . .	222
Figure 10.9 (a) Average message delay (b) Average throughput (c) Incurred traffic rate with varying average relative speed of nanomachines for different values of nanomachine radius.	226
Figure 10.10 For varying average relative speed of nanomachine and infostation for different values of infostation radius, the plot of (a) Average message delay. (b) Average throughput. (c) Incurred traffic rate. (d) Average number of infected nanomachines at the instant of message delivery.	227
Figure 10.11 (a) Average message delay and incurred traffic rate with varying contact time. (b) Adhesion probability and average message delay with varying contact time. (c) Average message delay and incurred traffic rate with varying number of bonds.	229
Figure 10.12 (a) Average message delay (b) Average message delay and transmission probability (c) Average throughput (d) Incurred traffic rate with varying threshold value for different values of SNR	230
Figure 10.13 Average message delay and incurred traffic rate with varying (a) Number of spike bits contained in the message. (b) Number of nanomachines.	231
Figure 11.1 Illustration of molecular order-based communication. Here, two distinct molecules, i.e., molecule type a and b , are used to transmit information.	238
Figure 11.2 Transmission and correct reception of two molecules in order $a - b$ which corresponds to the symbol 0.	239

Figure 11.3 Probability of correct symbol reception, i.e., $Pr(0 0)$, with respect to the inter-emission time t_e for the different distances between TN and RN.	240
Figure 11.4 Time diagram for the transmission and correct reception of three consecutive symbols 000 using the molecular order $a - b, a - b, a - b$	241
Figure 11.5 Distribution function for the evaluation of the inter-symbol time ($d = 4 \text{ nm}$).	243
Figure 11.6 The ISI probability values are shown for varying inter-emission time t_e and inter-symbol time T values.	250
Figure 11.7 The error probability P_e values are shown with different inter-emission time t_e inter-symbol time T values.	251
Figure 11.8 MARCO channel capacity $\frac{C}{T}$ is shown with inter-emission time t_e according to the different inter-symbol time T values.	251
Figure 11.9 P_{ISI} is shown for varying T according to the different inter-node distance d values.	252
Figure 11.10 The error probability P_e is shown for varying T according to different d values.	252
Figure 11.11 $\frac{C}{T}$ is shown for varying T according to the different d values.	253
Figure 11.12 P_{ISI} is shown for varying t_e according to the different diffusion coefficient D values.	254
Figure 11.13 P_e is shown for changing t_e according to different D values.	254
Figure 11.14 $\frac{C}{T}$ is shown for varying t_e according to the different D values.	255
Figure 11.15 Z-channel approach to molecular communication.	256
Figure 11.16 Binary channel approach to molecular communication.	257
Figure 11.17 4-input 2-output memoryless channel approach to molecular communication.	258
Figure 12.1 Three kinds of nanomedicine applications with appropriate nanoscale tools are categorized according to their time evolution and complexities.	266
Figure 12.2 Two different body area nanonetworks of artificial cells and nanorobots with molecular communication to realize a sophisticated drug delivery and immune system support task.	267

Figure 12.3 Molecular communication channel with passive transport of information carrying molecules.	268
Figure 12.4 Molecular communication channel with active transport of information carrying molecules.	268
Figure 12.5 Communication rate provided by binary channel and 4-input 2-output channel with respect to varying inter-node distance.	270
Figure 12.6 Communication rate provided by n – ary erasure channel with varying channel error probability p_e for varying number of distinct molecules used in the channel.	271
Figure 13.1 Illustration of nanotube radio circuit [195].	278
Figure 13.2 Illustration of integrated hardware required for nanoscale communication in CANET, including nanotransceiver, nanopower, nanoprocessor and nanomemory units.	280
Figure 13.3 Illustration of a CANET application example, i.e., nanoscale bio-sensor network.	286

ABBREVIATIONS

DNRS	Distributed Node and Rate Selection
WSN	Wireless Sensor Networks
BIOSS	Biologically-Inspired Spectrum Sharing
CRN	Cognitive Radio Networks
ESA	Evolutionary Opportunistic Spectrum Access
HAC	Homeostasis-inspired Autonomous Communication
WASN	Wireless Audio Sensor Networks
BIOX	Bio-inspired Cross-Layer
WSAN	Wireless Sensor and vehicular Actor Networks
BFC	Biological Foraging-inspired Communication
IMCRN	Intermittently-connected Mobile Cognitive Radio Ad Hoc Networks
MAMNET	Mobile Ad hoc Molecular Nanonetwork
MARCO	Molecular Array-based Communication
BAN ²	Body Area Nanonetworks
CANET	Carbon Nanotube-based Nanoscale Ad hoc Networks
ATP	Adenosine Triphosphate
CRAHNs	Cognitive Radio Ad hoc Networks
OSA	Opportunistic Spectrum Access
SU	Secondary User
PU	Primary User

MAC	Medium Access Control
RSP	Relay Selection Profitability
CSP	Channel Selection Profitability
TN	Transmitter Nanomachine
RN	Receiver Nanomachine
MEC	Molecular Error Compensation
NM	Nanomachines
CNT	Carbon Nanotubes
INS	Iterative Node Selection
WSS	Wide Sense Stationary
AIS	Artificial Immune System
DN	Designated Node
MMSE	Minimum Mean Square Error
xG	NeXt Generation
DSA	Dynamic Spectrum Access
ADP	Asynchronous Distributed Pricing
SI	Spectrum Opportunity Identification
SS	Spectrum Sharing
CSM	Clonal Selection Mechanism
FHS	Frequency Hopping Sequences
RTS	Request-To-Sent
MRCC	Multiple Rendezvous Control Channel
RFE	Radio Front End

WMSN	Wireless Multimedia Sensor Networks
ACK	Acknowledgment
RTC	Request-To-Cluster
ESRT	Event-to-Sink Reliable Transport
ICN	Intermittently Connected Networks
CR	Cognitive Radio
SAF	Spectrum-Aware Flooding
CCC	Common Control Channel
VRV	Virtual Reception Volume
EPSP	Excitatory Postsynaptic Potential
CDF	Cumulative Distribution Function
ISI	Inter-Symbol Interference
AHL	Acyl-Homoserine Lactone
FM	Frequency Modulation
NBSN	Nanoscale Bio-Sensor Network
CDMA	Code Division Multiple Access
FDMA	Frequency-Division Multiple Access
TDMA	Time-Division Multiple Access

CHAPTER 1

INTRODUCTION

Rapid growth in network technologies results in the emergence of the next-generation network vision. Next-generation networks are envisioned to enable billions of people to receive any kind of information anywhere and anytime through any kind of communication devices and services. However, this vision poses significant challenges for communication networks. Since next-generation network architectures may include diverse set of communication devices from nanosensor nodes to base stations and various network architectures from molecular nanonetworks to cognitive radio networks, centralized control of the next-generation network architectures is not a practical solution. Instead, the next-generation network architectures and communication protocols must have the capabilities of scalability, self-organization, self-adaptation, and survivability. Furthermore, due to inefficiency of the traditional wireless communication paradigms in nanoscale, nanoscale next-generation networks, e.g., nanosensor networks in nanomedicine applications, requires novel nanoscale communication paradigms. In nature, the biological systems intrinsically have these capabilities such that billions of blood cells which constitute the immune system can protect the organism from the pathogen without any central control of the brain. Furthermore, natural nanoscale networks such as immune system and bacteria colonies employ incredible nanoscale communication phenomena, e.g., quorum sensing of bacteria, that can be used to devise efficient nanoscale communication paradigms for nanoscale next-generation networks. Therefore, the natural biological systems may give great inspiration to develop the communication network models and techniques for the next-generation communication networks. This thesis aims to develop the bio-inspired communication techniques for the different next-generation networks and provide communication theoretical analyses for some of these networks. These networks can be categorized into three groups and their subgroups according to their scale and types.

1. Macroscale Next-generation Networks

- *Cognitive Radio Networks*
 - Cognitive Radio Ad Hoc Networks
 - Intermittently-connected Mobile Cognitive Radio Ad Hoc Networks
- *Wireless Sensor Networks*
 - Wireless Sensor and Vehicular Actor Networks
 - Wireless Audio Sensor Networks

2. Micro and Nanoscale Next-generation Networks

- Molecular Nanonetworks
- Mobile Ad Hoc Molecular Nanonetworks
- Carbon Nanotube-based Ad Hoc Networks
- Body-Area Nanonetworks

Next, we briefly introduce these three main next-generation networks, i.e., Cognitive Radio Networks, Wireless Sensor Networks, and Nanonetworks and discuss why bio-inspired communication techniques can be appropriate for the realization of these networks.

1.1 Cognitive Radio Networks

The spectrum access of the wireless networks is generally regulated by governments via a fixed spectrum assignment policy. Thus far, the policies have satisfactorily met the spectrum needs of the wireless networks. However, in the recent years, the dramatic increase in the spectrum access of the pervasive wireless technologies has caused scarcity in the available spectrum bands. Therefore, the limited available spectrum and underutilization of the spectrum necessitate a new communication paradigm, referred to as Cognitive Radio Networks (CRN), to exploit the entire existing wireless spectrum opportunistically [1].

Based on interaction with the environment, cognitive radio enables the users to communicate over the most appropriate spectrum bands which may be licensed or unlicensed. Since the most of the spectrum is already assigned to the licensed users by the governmental policies,

cognitive radio mainly strives for the communication over the licensed spectrum band. In the licensed spectrum bands, cognitive radio exploits the temporally unused spectrum which is defined as the spectrum hole or white space [2]. If cognitive radio encounters the licensed user at the licensed spectrum band, it moves to another spectrum hole or stays in the same band without interfering with the licensed user by adapting its communication parameters such as transmission power or modulation scheme. However, in the unlicensed spectrum bands, all licensed and unlicensed users have the same priority to access the existing unlicensed spectrum bands.

The cognitive radio technology enables the users to opportunistically access the available licensed spectrum bands through the following functionalities [1]:

- **Spectrum sensing:** Detecting the available spectrum holes and the presence of licensed users.
- **Spectrum management:** Capturing the best available spectrum which can meet the user demand.
- **Spectrum mobility:** Maintaining seamless communication requirements during the transition to better spectrum.
- **Spectrum sharing:** Providing the fair spectrum scheduling method among coexisting cognitive radios.

The functionalities above provide the cognitive radio with the ability of spectrum-aware communication while bringing the new challenges to be addressed. Since CRN architectures may include diverse set of communication devices and various network architectures, centralized control is not a practical solution. Instead, the CRN architectures and communication protocols must have the following capabilities.

- **Scalability:** To meet the large number of users' demand over a large geographical area, the developed algorithms and protocols for the CRN must be scalable.
- **Self-adaptation:** To capture the best available channel, cognitive radio must adapt its communication parameters according to the environment conditions such as the available spectrum band, channel parameters, the existence of the licensed users and the interference level at the licensed user receiver.

- **Self-organization:** To collaboratively access the available licensed or unlicensed spectrum band, cognitive radios must organize themselves.

In nature, the biological systems intrinsically possess these capabilities. For instance, in the insect colonies, insects can collaboratively allocate certain tasks according to the sensed information from the environment without any central controller. In this thesis, we explore the existing surprising similarities and mapping between communication challenges posed by the CRN architectures and the potential approaches and solutions in the natural biological systems [3].

1.2 Wireless Sensor Networks

Recent developments in micro-electronics have enabled low cost and low powered sensor nodes with limited wireless communications capability [4]. These multi-functional sensor nodes have enabled Wireless Sensor Networks (WSN) including large number of densely deployed sensor nodes for any monitoring task. Due to spatial deployment of sensor nodes in an environment, WSN provide great monitoring capabilities beyond traditional monitoring mechanisms in this environment. Together with the developments in WSN domain, Internet is getting to include more information obtained from physical world via small sensor nodes. Therefore, WSN are expected to enable connection between physical world and the Internet to provide access to vast amount of information from anywhere and anytime through any kind of communication devices and services. However, this vision poses significant challenges for WSN. Due to the pervasion in its nature, centralized control of WSN is not a practical solution. Instead, WSN and its communication protocols must have the capabilities of scalability, self-organization, self-adaptation, and survivability.

WSN are event-based systems, which enable the sensor nodes to communicate the properties of observed event to the sink. This communication from event to sink has to be reliable and timely to enable the sink node to reliably estimate the event properties and to timely perform the appropriate action if required. However, in WSN, it cannot be easily ensured that observed event information can be timely and reliably communicated to the sink node because most of the information may be lost on the forward path from the event to the sink node. Therefore, centralized solutions are not practical for WSN since coordination between the sink node and

sensor nodes cannot be ensured. Furthermore, due to the pervasion of sensor nodes over large geographic observation areas, it cannot be possible to engineer the centralized controllers to reliably and timely communicate the observed event information. Thus, self-organized protocols that do not need any coordination with central controller are imperative for WSN.

The notion of the self-organizations brings about some important capabilities such as self-adaptation, survivability, and scalability.

- **Self-Adaptation:** Self-organization allows WSN to adapt any change in the environment or network by regulating communication parameters of the sensor nodes.
- **Survivability:** Self-organization allows WSN to survive in any state of the network. For example, when some node failures are experienced, self-organization provides the capability of survivability such that it enables sensor nodes to organize with each other and to survive for reliable and timely communications of event information.
- **Scalability:** Self-organization allows each sensor node to locally interact with its neighbor nodes to communicate the event information. Therefore, as size of the network increases, the self-organization allow WSN to pursue their normal operations, that is, growing network size does not negatively affect the normal operations of the network.

As needed for WSN, self-organization is imperative almost all of infrastructureless network architectures. Especially, ad hoc network architectures heavily need self-organization for several network-wide tasks such as topology formation and routing. WSN are event-based systems, which trigger the event-to-sink data communications as soon as an event is detected and this communication is heavily affected by properties of the event. For example, size of the event area in which source nodes are selected is critical to determine how many sensors are selected as source nodes to cover the event area and which sensor nodes should be selected as source nodes. This event based communication paradigm in WSN imposes unique challenges that necessitate great self-organization capability to WSN beyond the traditional self-organization mechanisms in ad hoc network architectures. This self-organization capability can enable the network to organize according to the properties of the event, dynamically changing environment, and network conditions.

In nature, biological systems are intrinsically self-organized systems and most of the biological systems can organize to react to biological events without any need to a central controller.

For example, in the immune system, when a pathogen enters the body, this event is detected by T-cells in immune system, and T-cells trigger B-cells in immune system. Then, these triggered B-cells organize to determine which B-cells are most appropriate to react to the pathogen. This way, the pathogens are reliably eliminated within a specific time delay before the pathogen results in an infection in the organism.

In the homeostatic system, nervous, endocrine, and immune systems organize to bring the organism from an unstable state to a stable state. Gland cells in endocrine system, neurons in nervous system and blood cells in immune system collaboratively react to an unstable state of the organism to bring the organism back to the stable state. During this operation, without any need of a central controller, homeostatic system enables the organism to reliably bring into a stable state within a specific time delay before some functions of the organism are corrupted due to the unstable state.

In this thesis, we introduce potential solution avenues from the biological systems toward addressing the challenges of WSN such as scalability, self-organization, self-adaptation, and survivability. After the existing biological models are first investigated, biologically-inspired communication approaches are introduced for WSN. The objective of these communication approaches is to serve as a roadmap for the development of efficient scalable, adaptive, and self-organizing bio-inspired communication techniques for WSN [5].

1.3 Nanonetworks

Nanotechnology envisages the practical realization of very low-end nanomachines that have tiny components to accomplish a simple specific task such as communication, computation, and sensing in a scale ranging from 1 to 100 nanometers. For example, molecular machines, a particular type of nanomachines, consist of organized molecular components changing their relative positions to achieve a specific task as a response to an external stimulus. Nanomachines are considered as a part of the potential solution approaches to many crucial problems such as pollution, scarce food resources, and cellular repair [6]. However, the design and construction of nanomachines is one of the most attractive challenges in nanotechnology [7].

The construction of nanomachines is currently pursued by the top-down approach that progressively manipulates more and more tiny nanomachine components by photolithography

and related techniques. However, as Moore's law races towards 50 *nm* feature sizes on chips, it is not yet clear whether smaller nanomachines can be manufactured by using a top-down approach. Moreover, the top-down approach is subject to a drastic cost limitation as the size of the nanomachine components comes closer to nanometer scale [7], [8].

On the other hand, the bottom-up approach follows a more attractive strategy in the construction of frontier nanomachines. Rather than combining individual nanomachine components as in the top-down approach, the bottom-up approach entails the self-assembly of nanoscale building blocks with controlled size, shape, surface structure, and chemical functionality. In fact, by considering nature-made nanomachines constructed by an incredible self-assembly process, it can be easily inferred that the bottom-up approach has many promising advantages over the top-down approach [8]. For example, a chloroplast in a plant cell stands as a nanomachine including arrays of molecules that act as tuned optical antennas for absorption and transformation of solar energy. Mitochondrion can be envisioned as a nanomachine used for controlled combustion of organic molecules to generate adenosine triphosphate (ATP) for fulfilling energy needs of cellular activities. A flagellar motor attached on the membrane of many bacterial cells is a highly structured combination of proteins for providing cellular movement [9].

Biologically-inspired design or genetic adaptation from Nature is currently adopted to construct new nanomachines and nanomaterials through a large set of methods used in biochemistry and supramolecular chemistry [7], [10]. For example, biochemists already tackle with the production of functional cell components such as ribosome that can be used for the synthesis of beneficial proteins. Inspired by the flagella rotary motor in bacteria, a hybrid bio-nanochemomechanical rotary device has been assembled from ATP synthase and a nickel propeller [11]. Moreover, bacteria are also genetically programmed to generate principle vital components ranging from human growth hormone to beneficial enzymes for chess production [6].

While a single nanomachine can perform a very simple specific task, a network of communicating nanomachines can also be engineered to share nanoscale information over a nanonetwork so as to fulfill more complex tasks such as collaborative drug delivery, health monitoring, and biological or chemical attack detection [12], [13]. In fact, such nature-made nanonetworks already exist and are indispensable for nature to cooperatively share nanoscale

information in order to accomplish a specific task. For example, the quorum sensing mechanism allows bacteria to communicate with each other by producing, emitting, and receiving hormone-like messenger molecules called autoinducer. This natural process permits bacteria to synchronize all colony activities and to change the colony state as a response to an external stimulus [14]. Similar to bacteria, in natural immune system, the white blood cells called B-cells and T-cells communicate with each other using molecules to cooperatively sense and eliminate the hazardous pathogen [15]. The molecular communication among white blood cells forms the biological immune network that is an excellent defense mechanism of organism.

As in nature, nanomachines can also be artificially interconnected via molecular communication to form a nanonetwork for a specific task [13]. For example, two artificial cell-cell communication systems for yeast are developed in [16]. In the first system, the engineered yeasts, considered as sender cells, synthesize and emit the plant hormone cytokinin and the emitted cytokinin molecules diffuse into the environment and activate a target signaling pathway in nearby engineered yeasts, considered as receiver cells. In the second system, like quorum sensing in bacteria, the molecular communication between sender and receiver cells provides population density-dependent gene expression process for the synthesis of a functional biochemical material. These artificial cell-cell communication systems are expected to provide new biotechnology applications such as biomaterial fabrication and tissue engineering. Moreover, in [17], using nitric oxide signaling elements, an artificial cell-cell communication system in mammalian cells is introduced in order to govern a population of cells with the synthetic genetic circuits. This system may also serve as a functional tool for many applications such as gene therapy and artificial gene regulatory networks.

Beside the molecular communication paradigm, carbon nanotube based communication, and neuro-spike communication are also promising nano-scale communication paradigms. However, these new nano-scale communication paradigms bring many additional difficulties and challenges over the existing wireless communication and networking paradigms. Therefore, these difficulties and challenges need to be solved to effectively and efficiently realize the nano-scale communication and networking paradigms in many applications from nano-scale body area networks to nano-scale molecular computers.

The aim of this thesis is to theoretically investigate and solve the communication problems in

the nanoscale and molecular communication domains to develop efficient and reliable communication techniques for these domains. For this aim, it is imperative to first understand the channel characteristics in the nanoscale communication and to model these communication channels. After the investigation of the nanoscale channel characteristics using the concept of biology, bio-physics, and information theory, we aim to develop information theoretical approaches for the nanoscale communication channels. Based on these information theoretical approaches, we also aim to develop efficient nanoscale communication algorithms.

1.4 Research Objectives and Solutions

In this thesis, we develop the bio-inspired communication techniques for the above mentioned next-generation networks. More specifically, the following twelve works have been accomplished within the context of this thesis.

1. Immune system-based distributed node and rate selection algorithm has been developed for WSN.
2. Inspiring by the task allocation phenomenon in insect colonies, a spectrum sharing algorithm has been developed for CRN.
3. Immune system-inspired evolutionary opportunistic spectrum access protocol has been provided for cognitive radio ad hoc networks.
4. For wireless audio sensor networks, a homeostasis-inspired autonomous communication algorithm has been devised.
5. Biological foraging-inspired communication and coordination algorithm has been devised for wireless sensor and vehicular actor networks.
6. Biological foraging-inspired communication algorithm has been developed for intermittently-connected mobile cognitive radio ad hoc networks.
7. An information theoretical approach has been given for molecular communication in nanonetworks.
8. Deterministic capacity of molecular information flow in nanonetworks has been formulated.

9. Mobile ad hoc nanonetworks with collision-based molecular communication have been modeled.
10. Nanoscale communication with molecular arrays has been modeled and analyzed.
11. The concept of body area nanonetworks with molecular communication has been introduced.
12. The concept of carbon nanotube-based nanoscale ad hoc networks and the existing research challenges towards the realization has been provided.

Next, we briefly introduce these works without elaborating the details. We will elaborate these works at the following chapters.

1.4.1 Immune System-based Distributed Node and Rate Selection in Wireless Sensor Networks

Wireless sensor networks (WSN) are event-based systems that rely on the collective effort of densely deployed sensor nodes. Due to the dense deployment, since sensor observations are spatially correlated with respect to location of sensor nodes, it may not be necessary for every sensor node to transmit its data. Therefore, due to the resource constraints of sensor nodes it is needed to select the minimum number of sensor nodes to transmit the data to the sink. Furthermore, to achieve the application-specific distortion bound at the sink it is also imperative to select the appropriate sampling frequency of sensor nodes to achieve the minimum energy consumption. In order to address these needs, many efficient data gathering algorithms has been previously provided for WSN [44], [45], [46], [47], [48]. However, most of these algorithms provide the centralized and impractical approaches to the data gathering in WSN.

In Chapter 2, we introduce the Distributed Node and Rate Selection (DNRS) algorithm which is based on the principles of natural immune system. Based on the B-cell stimulation in immune system, DNRS selects the most appropriate sensor nodes that send samples of the observed event and these sensor nodes are referred to as *designated nodes*. The aim of the designated node selection is to meet the event signal reconstruction distortion constraint at the sink node with the minimum number of sensor nodes. DNRS enables each sensor node to

distributively decide whether it is a designated node or not. In addition, to exploit the temporal correlation in the event data DNRS regulates the sampling frequency rate of each sensor node while meeting the application-specific delay bound at the sink. Based on the immune network principles, DNRS distributively selects the appropriate sampling frequencies of sensor nodes according to the congestion in the forward path and the event signal reconstruction distortion periodically calculated at the sink node.

1.4.2 Bio-inspired Spectrum Sharing in Cognitive Radio Networks

Cognitive radio is the promising radio technology, which aims to detect and utilize the temporally unused spectrum bands by sensing its radio environment in order to enhance spectrum utilization. However, these objectives bring significant challenges and required functionalities such as spectrum sensing, sharing, management and mobility for the realization of Cognitive Radio Networks (CRN). In particular, efficient spectrum sharing problem in cognitive radio communication is one of the most important problem which must be addressed in order to enhance the overall spectrum utilization in dynamic spectrum access environments. In the literature, most of the spectrum sharing solutions strictly require coordination via message exchange among the users [18], [19], [21]. In Chapter 3, we introduce a new BIOlogically-inspired Spectrum Sharing (BIOSS) algorithm which is based on the adaptive task allocation model in insect colonies. Without need for any centralized controller or coordination among the unlicensed users, BIOSS enables each unlicensed user to distributively determine the appropriate channel(s) over which it can communicate.

1.4.3 Immune System-inspired Evolutionary Opportunistic Spectrum Access in Cognitive Radio Ad Hoc Networks

Underutilization of wireless spectrum stimulates cognitive radio ad hoc networks (CRAHNs) aiming to enable unlicensed ad hoc users to opportunistically access unused licensed spectrum [22]. The term Opportunistic Spectrum Access (OSA) [23] is broadly used to identify a such kind of overlay spectrum access, in which unlicensed users, i.e., secondary users (SU) detect and access the temporarily unused spectrum bands legitimately licensed to primary users (PU). The realization of CRAHN with OSA includes several difficulties such as self-organization of unlicensed users, self-regulation of communication parameters, and self-

adaptation to time-varying radio environment.

In nature, biological systems intrinsically have these great abilities that can be modeled and adopted to overcome the difficulties posed by opportunistic spectrum access. In Chapter 4, a new immune system-inspired Evolutionary opportunistic spectrum Access (ESA) protocol is introduced. Based on the *self-nonself detection and clonal selection principles in immune system*, ESA allows unlicensed users to separately detect, share, and access the available spectrum bands without interfering the licensed users. The overall ESA operations do not need for any priori information about the access statistics of licensed users and also do not need for any coordination and message exchanges among the network nodes. In addition, unlike the existing works [24], [25], [26], [27], ESA does not require any dedicated control channel in the entire network. Furthermore, ESA also exploits the contention among the nodes and their mobility, if exists, towards accelerating the evolution in the system, and hence, yielding higher overall spectrum utilization.

1.4.4 Homeostasis-inspired Autonomous Communication in Wireless Audio Sensor Networks

Emerging applications of wireless sensor networks (WSN) requiring wide-band event signal communication such as multimedia surveillance sensor networks impose additional challenges including high communication bandwidth requirement and energy cost. Besides their partially or fully dependency on feedback messages from sink node, the existing protocols designed for WSN do not address the communication of wide-band event signals. Furthermore, the feedback messages may not reach in time to provide reliable communication of event information and save scarce network resources. In the literature, an extensive set of research has been already conducted for the communication of event signal in WSN [28], [29], [30] [31]. However, these works always need feedback message from the sink node to provide a reliable and energy-efficient communication of the event signal. Hence, an autonomous communication protocol is imperative in order to provide wide-band event signal communication without any feedback from the sink. In nature, biological systems have self-organization capability, i.e., *homeostasis*, as they autonomously maintain a relatively stable equilibrium state for operation of vital functions. Hence, this natural phenomenon clearly gives promising inspirations in order to develop autonomous and efficient communication models and protocols for WSN

domain. In Chapter 5, the homeostasis-inspired autonomous communication (HAC) protocol is introduced for wireless audio sensor networks (WASN). Using the spectral properties of the wide-band event signal, i.e., audio signal, HAC enables WASN to maintain a relatively stable state in which sensor nodes reliably and energy-efficiently communicate the event signal to the sink node. Furthermore, with its self-organization capability, HAC does not rely on any feedback message from the sink node.

1.4.5 Biological Foraging-inspired Communication and Coordination in Wireless Sensor and Vehicular Actor Networks

In wireless sensor and actor networks (WSAN), sensors gather information about the physical world, while vehicular actors take decisions, and then, perform appropriate actions. Coordination and coordinated efficient communication among sensor and vehicular actor nodes are the main challenges for the realization of WSAN. In the existing literature, there are many studies on energy-efficient communication protocols for wireless sensor networks (WSN) and WSAN [32], [28], [29], [33], [34]. However, the major common drawback of these proposals is the lack of autonomy in the operations of the network nodes

In Chapter 6, based on *the prey model in foraging theory*, BIO-inspired Cross-layer (BIOX) communication and coordination protocol is introduced for WSAN. BIOX incorporates the medium access control (MAC), routing, and transport layer functionalities to provide reliable sensor-actor communication with minimum and fair energy consumption. It also enables actor nodes to efficiently allocate and perform the required tasks associated with the event that occurs in the environment. BIOX permits each sensor node to autonomously determine its next-hop selection and channel access strategy using bio-inspired *next-hop selection and channel access profitability measures*. Based on these profitability measures, BIOX provides optimal performance in energy-efficient and reliable sensor-actor communication. Furthermore, using *task allocation profitability measure*, BIOX also guarantees the stable allocation of available tasks in a way that each task is accomplished by a vehicular actor node within a bounded time delay.

1.4.6 Biological Foraging-inspired Communication in Intermittently-connected Mobile Cognitive Radio Ad Hoc Networks

Intermittently-connected Mobile Cognitive Radio ad hoc Networks (IMCRN) are promising wireless networks in which mobile unlicensed nodes use their temporarily available contacts and vacant licensed channels for end-to-end message delivery. In IMCRN, the lack of permanently available end-to-end paths renders the flooding-based communication more relevant for the interconnection of network nodes. However, due to energy wastefulness of flooding, the aim of an efficient communication in IMCRN is to provide minimum energy consumption without sacrificing end-to-end delay performance. Such an energy-efficient communication must be also spectrum-aware to effectively use idle licensed channel opportunities while minimizing interference to primary system.

There is an emerging body of literature on routing protocols in cognitive radio ad hoc networks [35], [36], [37], [38]. All of these protocols are also based on the assumption that at least one end-to-end path always exists between any two nodes in the network. However, this assumption is invalid for IMCRN that is usually lacking in complete end-to-end paths. In Chapter 7, we propose Biological Foraging-inspired Communication (BFC) algorithm for the energy-efficient and spectrum-aware communication requirements in IMCRN. BFC is based on two profitability measures called *relay selection profitability* (RSP) and *channel selection profitability* (CSP). RSP and CSP provide an autonomous decision-making mechanism that does not need any a priori information on node mobility and spectrum availability patterns in relay and channel selection. This decision-making mechanism also leads to an optimization procedure to determine optimal relay and channel selection rules.

1.4.7 An Information Theoretical Approach for Molecular Communication in Nanonetworks

Molecular communication is a novel paradigm that uses molecules as an information carrier to enable nanomachines to communicate with each other. Controlled molecule delivery between two nanomachines is one of the most important challenges which must be addressed to enable the molecular communication. Therefore, it is essential to develop an information theoretical approach to find out communication capacity of the molecular channel. In Chap-

ter 8, we develop an information theoretical approach for capacity of a molecular channel between two nanomachines. Using the principles of mass action kinetics, we first introduce a molecule delivery model for the molecular communication between two nanomachines called as Transmitter Nanomachine (TN) and Receiver Nanomachine (RN). Then, we derive a closed form expression for capacity of the channel between TN and RN. Furthermore, we propose an adaptive Molecular Error Compensation (MEC) scheme for the molecular communication between TN and RN. MEC allows TN to select an appropriate molecular bit transmission probability to maximize molecular communication capacity with respect to environmental factors such as temperature and distance between nanomachines.

1.4.8 Deterministic Capacity of Molecular Information Flow in Nanonetworks

Molecular communication enables nanomachines to exchange information with each other by emitting molecules to their surrounding environment. Molecular nanonetworks are envisioned as a number of nanomachines that are deployed in an environment to share specific molecular information such as odor, flavour, or any chemical state. In Chapter 9, using the stochastic model of molecular reactions in biochemical systems, a realistic channel model is first introduced for the molecular communication. Then, based on the realistic channel model, we introduce a deterministic capacity expression for point-to-point, broadcast, and multiple-access molecular channels. We also investigate the information flow capacity in a molecular nanonetwork for the realization of efficient communication and networking techniques for the frontier nanonetwork applications. The results reveal that the molecular point-to-point, broadcast, and multiple-access channels are feasible with satisfactorily high molecular communication rate, which allows molecular information flow in nanonetworks. Furthermore, the derived molecular channel model with input-dependent noise term also reveals that unlike the traditional Gaussian communication channel, achievable capacity is affected by both lower and upper bounds of the channel input in molecular communication channels.

1.4.9 Mobile Ad Hoc Nanonetworks with Collision-based Molecular Communication

For the realization of future nanonetworks, it is essential to develop novel and efficient communication and networking paradigms. In Chapter 10, the first step towards designing a mobile ad hoc molecular nanonetwork (MAMNET) with electrochemical communication is

taken. MAMNET consists of mobile nanomachines and infostations that share nanoscale information using electrochemical communication whenever they have a physical contact with each other. In MAMNET, the intermittent connectivity introduced by the mobility of nanomachines and infostations is a critical issue to be addressed. An analytical framework that incorporates the effect of mobility into the performance of electrochemical communication among nanomachines is presented. Using the analytical model, numerical analysis for the performance evaluation of MAMNET is obtained.

1.4.10 Nanoscale Communication with Molecular Arrays in Nanonetworks

Molecular communication is a promising nanoscale communication paradigm that enables nanomachines to exchange information by using molecules as communication carrier. Up to now, the molecular communication channel between a transmitter nanomachine (TN) and a receiver nanomachine (RN) has been modeled as either concentration channel or timing channel [39], [40], [41]. However, these channel models necessitate exact time synchronization of the nanomachines and provide a relatively low communication bandwidth. In Chapter 11, the Molecular ARray-based COmmunication (MARCO) scheme is proposed, in which the transmission order of different molecules is used to convey molecular information without any need for time synchronization. The MARCO channel model is first theoretically derived, and the inter-symbol interference and error probabilities are obtained. Based on the error probability, achievable communication rates are analytically obtained.

1.4.11 Body Area Nanonetworks with Molecular Communications

Recent developments in nano and biotechnology enable promising therapeutic NanoMachines (NM) that operate on inter or intracellular area of human body. The networks of such therapeutic NM, i.e., Body Area NanoNetworks (BAN²), also empower sophisticated nanomedicine applications. In these applications, therapeutic NM share information to perform computation and logic operations and make decisions to treat complex diseases. Hence, one of the most challenging subjects for these sophisticated applications is the realization of BAN² through a nanoscale communication paradigm. In Chapter 12, we introduce the concept of BAN² with molecular communication, where messenger molecules are used as communication carrier from a sender to a receiver NM. The current state-of-the-art of molecular communication

and BAN² in nanomedicine applications is first presented. Then, communication theoretical efforts are reviewed and open research issues are given. The objective of this work is to introduce this novel and interdisciplinary research field and highlight major barriers toward its realization from the viewpoint of communication theory.

1.4.12 Carbon Nanotube-based Nanoscale Ad Hoc Networks

Recent developments in nanoscale electronics allow current wireless technologies to function at nanoscale environments. Especially, due to their incredible electrical and electromagnetic properties, carbon nanotubes (CNT) are promising physical phenomenon that are used for the realization of a nanoscale communication paradigm. This provides a very large set of new promising applications such as collaborative disease detection with communicating in-vivo nanosensor nodes and distributed chemical attack detection with a network of nanorobots. Hence, one of the most challenging subjects for such applications becomes the realization of nanoscale ad hoc networks. In Chapter 13, we define the concept of Carbon nanotube-based nanoscale Ad hoc NETWORKS (CANET) for future nanotechnology applications. CANET can be perceived as the down-scaled version of traditional wireless ad hoc networks without downgrading its main functionalities. The objective of this work is to introduce this novel and interdisciplinary research field and highlight major barriers toward its realization.

1.5 Thesis Outline

This thesis is organized as follows. Chapter 2 introduces the Distributed Node and Rate Selection (DNRS) algorithm for WSN using the principles of natural immune system. Chapter 3 presents a new BIOlogically-inspired Spectrum Sharing (BIOSS) algorithm for CRN which is based on the adaptive task allocation model in insect colonies. Chapter 4 introduces an Immune system-inspired Evolutionary Opportunistic Spectrum Access (ESA) protocol for cognitive radio sensor networks. Chapter 5 presents the homeostasis-inspired autonomous communication (HAC) protocol is introduced for wireless audio sensor networks (WASN). Based on the prey model in foraging theory, Chapter 6 describes BIO-inspired Cross-layer (BIOX) communication and coordination protocol for wireless sensor and vehicular actor networks. Chapter 7 describes Biological Foraging-inspired Communication (BFC) algorithm

for the energy-efficient and spectrum-aware communication requirements in intermittently-connected mobile cognitive radio ad hoc networks. Chapter 8 demonstrates an information theoretical approach for capacity of a molecular channel between two nanomachines. Using the stochastic model of molecular reactions in biochemical systems, Chapter 9 introduces a realistic channel model for the molecular communication and presents a deterministic capacity expression for point-to-point, broadcast, and multiple-access molecular channels. Chapter 10 describes a mobile ad hoc molecular nanonetwork (MAMNET) with electrochemical communication. Chapter 11 describes the Molecular ARray-based COmmunication (MARCO) scheme for nanonetworks. Chapter 12 introduces the concept of body area nanonetworks with molecular communication. Chapter 13 defines the concept of Carbon nanotube-based nanoscale Ad hoc NETworks (CANET) for future nanotechnology applications.

CHAPTER 2

Immune System Based Distributed Node and Rate Selection in Wireless Sensor Networks

In this chapter, we introduce the Distributed Node and Rate Selection (DNRS) method which is based on the principle of natural immune system. Based on the B-cell stimulation model in Immune System, DNRS selects the most appropriate sensor nodes that send samples of observed event, are referred to as designated nodes. The aim of the designated node selection is to meet the distortion constraint for the event signal reconstruction at the sink with the minimum number of sensor node. DNRS enables sensor nodes to distributively decide whether it is a designated node or not according to its correlation with its neighbors and the event source. In addition, DNRS distributively regulates the sampling frequency rate of each sensor node to provide a certain reconstruction distortion with minimum energy consumption. DNRS algorithm was first introduced in [42].

The existing related works and the motivation are discussed in Section 2.1. In Section 2.2, after we briefly introduce the Biological Immune System, we introduce the relationship between Wireless Sensor Networks (WSN) and Immune System. In Section 2.3, we present the DNRS method. We firstly introduce the Distributed Node Selection scheme and then, introduce our Distributed Rate Selection mechanism. In Section 2.4, we evaluate the performance of DNRS. Finally, we conclude this chapter with Section 2.5.

2.1 Motivation and Related Works

Due to the dense deployment of sensor nodes in WSN, sensor observations are spatially correlated according to the location of sensor nodes. This results in transmission of highly redun-

dant sensor data which is not necessary to reconstruct the event signal at the sink. Therefore, due to the energy constraint of sensor nodes it is necessary to select the minimum number of sensor nodes that send samples of the observed event, are referred to as *designated nodes*. Furthermore, the nature of the energy-radiating physical phenomenon constitutes the temporal correlation between each consecutive observation of a sensor node [43]. Therefore, it is necessary to regulate the sampling frequency of sensor nodes to achieve minimum energy consumption while achieving a certain reconstruction distortion.

There has been some research efforts to study about correlated data gathering in WSN according to the different methods [44], [45], [46], [47], [48]. These works provide the great deal of capability to exploit the correlation at the sensor data. In [49], the model for point and field sources are introduced and their spatio-temporal characteristics are derived along with the distortion functions. In [50], exploiting spatial correlation at the MAC layer is achieved by collaboratively regulating medium access so that redundant transmissions from correlation neighbors are suppressed. In [50], the Iterative Node Selection (INS) algorithm is given to select the representative nodes which represent the set of nodes generating spatially correlated data. Since INS selects the representative nodes with centralized manner at the sink and it computes the distortion according to the Wide Sense Stationary (WSS) assumption on the sink which is neither scalable nor realistic. These works consider and capture the spatio-temporal model of physical phenomenon observed by sensor nodes for the realization of advanced efficient communication protocols. However, in these works since the event data is assumed to be Gaussian and wide-sense stationary and the most process encountered in practice are non-stationary, these works are not realistic. Moreover, in the literature, there exist some distributed solutions [51], [52] to exploit the spatio-temporal correlation in WSN. However, these studies neither select the sensor nodes which communicate with the sink node nor regulate the data rate to the sink node according to the estimation distortion at the sink node.

On the other hand, the natural Immune System has given the great inspiration for several researches from robotic to network security [53], [54], [55]. Because of its ability to self and non-self discrimination, it has been used for the data clustering and the computer security as an inspiration. In [53], an effective artificial immune system is presented, which is used as a simple classification tool. Using the clonal selection mechanism in Immune System, it attempts to group similar data items according to relationship between them. In [56], the

problem of protecting computer systems is addressed as the problem of learning to distinguish self from other. It proposes a method for change detection which is based on the generation of T-cells in the immune system.

2.2 Immune System and Wireless Sensor Networks

In this section, we first briefly introduce the immune system and its basic operation principles. Then, we discuss the similarities and the relation between the Immune System and Wireless Sensor Networks.

2.2.1 Biological Immune System

The human immune system is a complex natural defense mechanism. It has the ability to learn about foreign substances (pathogens) that enter the body and to respond to them by producing antibodies that attack the antigens associated with the pathogen [53]. The adaptive immune system is made up of lymphocytes which are white blood cells, B and T cells. Each of B-cells has distinct molecular structure and produces 'Y' shaped antibodies from its surfaces. The antibody recognizes antigen that is foreign material and eliminates it. This antigen-antibody relation is innate immune response [54]. Most antigens have various antigen determinants that are called epitope. In order to grab and latch onto antigen, antibody possesses a structure called paratope. Furthermore, each antibody has a unique epitope called idiotope. Note that different antibody types may have many epitopes in common. An epitope that is unique to a given antibody type is called an idiotope, hence the name idiotypic network for any scheme of regulation that works through the recognition of idiotopes [57]. An antibody type is thought to be stimulated when its paratopes recognize other types, and suppressed if its epitopes are recognized by other antibody paratopes.

The surface of a B-cell contains antibodies for that B-cell. When an antibody for a B-cell binds to an antigen, the B-cell becomes stimulated. The level of B-cell stimulation depends not only on the success of the match to the antigen, but also on how well it matches other B-cells in the immune networks. As the response of the antigen, the stimulated B-cells secrete the antibody to eliminate the antigen.

Immune system has the ability to process information, to learn and memorize, to discriminate between self and non-self, and to keep up harmony of the whole system. Because of these abilities, it gives the great inspiration to many kind of researches in computer sciences, robotics and signal processing. Therefore, Artificial Immune System (AIS) models have been introduced. In [57], an AIS model is proposed, which consists of a set of B-cells and the links between those B-cells. Each B-cell object can represent a data item which is being used for learning. According to this model, each B-cell is capable of responding to an antigen specified by a data item when it is stimulated by that antigen specific data item. In the natural immune system, the level of the B-cell stimulation relates that how well its antibody binds to the antigen and its affinity (or enmity) to its neighbors in the network. In [53], B-cell stimulation is modeled in terms of these three factors. The primary stimulus is defined as the affinity between the B-cell and the pathogen i.e. how well they match. The second stimulus for a B-cell is the affinity to its neighbors. Third is the suppression (enmity) factor from the loosely connected neighbor of this B-cell.

The stimulated B-cells start to produce more lymphocytes (i.e. to clone) and to secrete free antibodies. To model the antibody secretion of the stimulated B-cell, in [58], idiotypic network hypothesis is proposed, which is based on mutually stimulus and suppression between antibodies. In [57], an analytical immune network model is proposed using the idiotypic network hypothesis introduced in [58] as follows

$$\frac{dS_i(t+1)}{dt} = \left(\frac{\alpha \sum_{j=1}^N m_{ij} s_j(t)}{N} - \frac{\alpha \sum_{k=1}^N m_{ki} s_k(t)}{N} + \beta g_i - k_i \right) s_i(t) \quad (2.1)$$

where m_{ij} is mutual stimulus coefficient of antibody i and j , g_i is affinity between antibody i and antigen, k_i is the natural extinction of the antibody i , α, β, η are constants. $s_i(t)$ is the concentration of antibody i at time t and given as follows

$$s_i(t+1) = \frac{1}{1 + e^{(0.5 - S_i(t+1))}} \quad (2.2)$$

In Section 2.3, we adopt the immune network model given in (2.1) and (2.2) to derive an effective communication model between sensor nodes and the sink. In the next section, we introduce some analogies between Immune System and Wireless Sensor Networks (WSNs).

2.2.2 Immune System Based Sensor Networks

Although Immune System seems different from the Wireless Sensor Networks (WSN), they have the great deal of analogies when considered the tasks they accomplish. For example, when the Immune System encounters a pathogen, some B-cells are stimulated and secrete antibodies in different densities to eliminate the antigens produced by the pathogen. Similarly, when an event occurs in WSN environment, some sensor nodes referred to as designated node (DN) sense the event information and send this information to sink with a sampling frequency to achieve a certain event signal reconstruction distortion at the sink.

Table 2.1: Relationship between Immune System and Wireless Sensor Networks (WSN).

Immune System	Wireless Sensor Networks (WSN)
B-cells	Sensor nodes
Antibody	Sensor data
Antibody density	Sampling frequency rate f
Pathogen	Event source
Antigen	Reconstruction distortion
Natural extinction	Packet loss

In WSN, since multiple sensors record information about a single event in the sensor field, sensor observations are spatially correlated with respect to spatial location of sensor nodes. This results in transmission of highly redundant sensor data which is not necessary to reconstruct the event signal at the sink node. Therefore, it may not be necessary for every sensor node to transmit its data to the sink; instead, a smaller number of DN measurements might be adequate to communicate the event features to the sink within a certain reliability/fidelity level [59]. DNRS use the B-cell stimulation model given in [53] to determine the minimum number of DNs while meeting the application specific distortion bound at the sink.

For a fixed number of DNs, the minimum distortion can be achieved by choosing these nodes such that (i) they are located as close to the event source as possible and (ii) are located as farther apart from each other as possible [59]. Similarly, as given in [53], the stimulation of a B-cell depends on the distance between the pathogen and the B-cell and distance between the B-cell and its neighbors which stimulates or suppress it. That is, the B-cells which is nearest to the pathogen and suppressed as least as possible by their neighbors become the stimulated B-cells. The stimulated B-cells start to secrete the antibodies in different densities to keep the antigen densities in a desired level. Similarly, after the selection of DNs they send the event

information to sink with a sampling frequency to achieve the distortion constraint at the sink node. In Table 2.1, we summarize this relationship between Immune System and Wireless Sensor Networks (WSN). According to this relationship, DNRS distributively selects the DNs by using the B-cell stimulation model and distributively regulates the sampling frequency of the DNs based on immune network principles given in (2.1) and (2.2).

2.3 Immune System Based Distributed Node and Rate Selection

In this section, we first introduce a event signal reconstruction scheme used in DNRS. Then, we introduce DN selection scheme and the sampling frequency regulation mechanism in DNRS.

2.3.1 Event Signal Reconstruction

According to the application, the physical phenomenon can be modeled by a point source or by a field source. For example, while the case of object tracking can be modeled by a point source, the case of monitoring magnetic field and seismic activities can be modeled by field source. Here, in order to exploit the spatio-temporal correlation with Immune System based DNRS mechanism, we adopt the point source model given in [49] for event signal, and then we introduce a model for the event signal reconstruction. Note that, in this chapter, we use the point source based event signal reconstruction scheme, however, this scheme can be directly extended to the field source case.

In WSN applications, the event signal modeled by a point source is assumed to generate a continuous random process $f_S(t)$ having a variance σ_S^2 and a mean μ_S . Assuming that the point source is at the center of the coordinate axis [49], the sensor node situated at location (x, y) receives the signal given by

$$f(x, y, t) = f_S\left(t - \frac{\sqrt{x^2 + y^2}}{v}\right)e^{-\frac{\sqrt{x^2 + y^2}}{\theta_s}} \quad (2.3)$$

where v denotes the diffusion velocity of $f_S(t)$ and θ_s denotes the attenuation constant of event signal $f_S(t)$. Since $f(x, y, t)$ is the delayed and attenuated version of $f_S(t)$, the variance and mean of $f(x, y, t)$ can be expressed by

$$\mu_E(x, y) = (\mu_s e^{-\frac{\sqrt{x^2+y^2}}{\theta_s}}) \quad (2.4)$$

$$\sigma_E^2(x, y) = (\sigma_s e^{-\frac{\sqrt{x^2+y^2}}{\theta_s}})^2 \quad (2.5)$$

Using the observations which are attenuated and delayed version of the event signal, each sensor node generates its data packets. The k^{th} data packet generated by sampling the received signal $f(x_i, y_i, t)$ by sensor node n_i at location (x_i, y_i) , i.e., $S_{i,k}$, is defined as

$$S_{i,k} = [f(x_i, y_i, t_{kp}) \ f(x_i, y_i, t_{kp+1}) \ \dots \ f(x_i, y_i, t_{(k+1)p})] \quad (2.6)$$

where $S_{i,k}$ is $p \times 1$ vector, p and k denote the packet length and packet number, respectively, $f(x_i, y_i, t_{kp})$ is a sample of event signal $f(x_i, y_i, t)$, taken at time t_{kp} .

In WSNs, after the detection of an event, sensor nodes sample the event signal and send the generated data packets ($S_{i,k}$) to the sink. During this process, sensor circuitries add noise to these packets ($S_{i,k}$) as

$$S'_{i,k} = S_{i,k} + N_{i,k} \quad (2.7)$$

where $S'_{i,k}$ is the noisy version of the sensor packet $S_{i,k}$ and $N_{i,k}$ is the observation noise, where $N_{i,k} \sim N(0, \sigma_N^2)$.

In addition to the additive noise, due to the constraint in cost, power, and communication, in WSNs, packet losses often arise during the communication between the sensor nodes and the sink. In this chapter, to model the packet losses, we assume an average packet loss probability λ_i which indicates the packet losses of sensor node i due to channel errors, collision, and congestion. The packet loss probability λ_i is computed by the sink by dividing the number of loss packets of sensor node i to the total number of packets transmitted by sensor node i . Hence, we define the lossy and noisy version of the sensor packet ($X_{i,k}$) as

$$X_{i,k} = \begin{cases} 0 & \text{with probability } \lambda_i \\ S_{i,k} + N_{i,k} & \text{with probability } (1 - \lambda_i) \end{cases}$$

Since uncoded transmission considerably outperforms any approach based on the separation paradigm when the measured physical phenomenon is reconstructed within some prescribed distortion level [60], we assume that the sensors perform uncoded transmission. When the sensor nodes transmit the uncoded observations, minimum mean square error (MMSE) estimation is the optimal decoding technique. Therefore, the decoded version of the sensor packet $S_{i,k}$, i.e., $Z_{i,k}$, can be expressed as follows

$$Z_{i,k} = \frac{(\sigma_E^2(x_i, y_i) + \mu_E^2(x_i, y_i))(X_{i,k})}{(1 - \lambda_i)(\sigma_E^2(x_i, y_i) + \mu_E^2(x_i, y_i) + \sigma_N^2)} \quad (2.8)$$

According to the decoded sensor packets ($Z_{i,k}, \forall i, k$), sink generates the reconstructed event signal, i.e., $\hat{S}(\tau, f, M, p)$, as follows

$$\hat{S}(\tau, f, M, p) = \frac{1}{M} \left[\sum_{i=1}^M Z_{i,1} \sum_{i=1}^M Z_{i,2} \dots \sum_{i=1}^M Z_{i,(\frac{M}{p})} \right] \quad (2.9)$$

where τ, M, f denote the reconstruction interval, the number of DNs and the sampling frequency of sensor nodes, respectively. The sampling frequency of a sensor node is the number of samples taken from the sensed event signal. However, since in WSNs each data packet includes a number of samples (p), reporting frequency which is defined as the number of transmitted data packet per unit time is directly proportional with sampling frequency (f). Therefore, throughout this chapter, we mainly use the sampling frequency.

2.3.2 Distributed Designated Node Selection

As discussed in Section 2.2, the stimulation of a B-cell depends on three factors. First factor is the affinity between the B-cell and the pathogen. Second factor is the affinity between the B-cell and its neighbors which make it stimulated. Third factor is the affinity between the B-cell and its neighbors which suppress it [53]. According to the B-cell stimulation model given in [53], for a B-cell if the summation of these three factors specified by some equations is greater than a certain threshold value, this B-cell becomes stimulated. In WSN, the minimum distortion can be achieved by choosing the DNs such that (i) they are located as close to the event source as possible and (ii) are located as farther apart from each other as possible [59]. In fact, as B-cell stimulation, the selection of DNs depends on the three factors given as

follows:

1. First factor is the affinity between the sensor node (B-cell) and event source (pathogen).
2. Second factor is the affinity between the sensor node (B-cell) and its uncorrelated neighbor nodes (stimulating B-cells).
3. Third factor is the affinity between the sensor node (B-cell) and its correlated neighbor nodes (suppressing B-cells).

To determine the correlated and uncorrelated neighbors of a sensor node, we define the application specific correlation radius r . For a sensor node, while the neighbor nodes in its correlation radius r are the correlated neighbors for this sensor node, the neighbor nodes which are not in correlation radius r are uncorrelated neighbors for this sensor node. Since the amount of the spatial correlation between the sensor nodes depends on the variance and the mean of the observed event signal, environment size, node density, we assume that correlation radius r is an application specific value, which can be determined based on the environment size, node density and statistical properties of the physical phenomena observed in a given sensor network application [59]. We also assume that each sensor node is aware of its position and event source location by means of existing localization techniques [61]. Each sensor knows the relative position of its neighbors, as all nodes periodically broadcasts its position to neighbors.

For the determination of the affinity between the sensor node and its correlated and uncorrelated neighbor nodes, the correlations between the nodes and event source are needed. To express the correlations between event source and the nodes, we use the correlation coefficients, $\rho_{s,i}$ and $\rho_{i,j}$. $\rho_{s,i}$ states the correlation between the sensor node i sending information and the event source S . $\rho_{i,j}$ states the correlation between the sensor node i and j . We use the power exponential form [62] to model the correlation coefficients $\rho_{s,i}$ and $\rho_{i,j}$ and give these coefficients as follows.

$$\rho_{s,i} = K_{\theta}(d_{s,i}) = e^{(-d_{s,i}/\theta_1)^{\theta_2}}; \theta_1 > 0, \theta_2 \in (0, 2] \quad (2.10)$$

$$\rho_{i,j} = K_{\theta}(d_{i,j}) = e^{(-d_{i,j}/\theta_1)^{\theta_2}}; \theta_1 > 0, \theta_2 \in (0, 2] \quad (2.11)$$

where $d_{s,i}$ and $d_{i,j}$ denote the distances between event source S and sensor node i and between sensor node i and j , respectively. For $\theta_2 = 1$, the model becomes exponential, while for $\theta_2 = 2$ squared exponential. The covariance function is assumed to be nonnegative and decrease monotonically with the distance, with limiting values of 1 at $d = 0$ and of 0 at $d = \infty$.

Now, as the B-cell stimulation model given in [53], we mathematically model the DN selection. B-cell stimulation model given in [53] depends on three factors (i) the matching between the pathogen a B-cell (ii) the affinities between the B-cell and its neighbor B-cells (iii) the enmity between the B-cell and its neighbor B-cells. According to this model, when the summation of these three factors exceeds certain threshold for a B-cell, this B-cell becomes stimulated. Using this B-cell stimulation scheme, we model the DN selection depending on the similar three factors as follows:

1. We model the first factor which is the affinity between the sensor node i (B-cell) and event source s (pathogen) as $\rho_{s,i}$. Here, $\rho_{s,i}$ indicates the correlation between event source s and sensor node i . As the distance between sensor node i and event source s decreases, $\rho_{s,i}$ increases. Hence, it is more possible to become a DN for a sensor nodes nearest to the event source.
2. We model the second factor which is the affinity between the sensor node i (B-cell) and its uncorrelated neighbor sensor nodes (stimulating B-cells) $j, \forall j$ as follows. Sensor node j is selected as the neighbor node which is not in the correlation radius of sensor node i .

$$\sum_j (1 - \rho_{i,j}) \quad (2.12)$$

Here, $\rho_{i,j}$ indicates the correlation between sensor node i and j . As the distance between sensor nodes i and j increases, $\rho_{i,j}$ decreases. Therefore, $\sum_j (1 - \rho_{i,j})$ is big for a sensor node having the more uncorrelated neighbors (out of r). Hence, it is more possible to become a DN for such sensor nodes.

3. We model the third factor which is the affinity between the sensor node and its correlated neighbor sensor nodes (in r) $k, \forall k$ such that k is in the correlation radius of sensor node i and give as

$$\sum_k (-\rho_{i,k}) \quad (2.13)$$

where $\rho_{i,k}$ indicates the correlation between sensor node i and sensor node k . As the distance between sensor node i and k decreases, $\rho_{i,k}$ increases. Therefore, $\sum_k (-\rho_{i,k})$ is small for a sensor node having the more correlated neighbors. Hence, it is the least possible to become a DN for such sensor nodes.

Here, we define the DN selection weight of sensor node i (T_i) as the combination of these three factors as follows:

$$T_i = \rho_{s,i} + \sum_j (1 - \rho_{i,j}) + \sum_k (-\rho_{i,k}) \quad (2.14)$$

When T_i exceeds the certain threshold t_{dns} called the DN selection threshold, sensor node i becomes a DN. It is clear that while t_{dns} increases, the number of DN decreases because the number of nodes whose weight (T_i) exceed t_{dns} decreases. t_{dns} is determined at the sink node by using reconstruction distortion periodically computed by the Adaptive LMS Filter. Adaptive LMS Filter is typically applied in environments where signals with unknown or non-stationary statistics are involved and appear for this reason particularly suited to be used in highly dynamic systems such as sensor networks [63]. Using the Adaptive LMS Filter, the sink node predicts the reconstructed version of the event signal sample $\hat{S}(k, \tau, f, M, p)$ given in (2.1) from the previous samples $\hat{S}(k - r, \tau, f, M, p)$ $r = 1 \dots m$, where m is the order of the filter and $\hat{S}(k, \tau, f, M, p)$ denotes the k^{th} sample of the reconstructed event signal. Hence, the event signal reconstruction error in the Adaptive LMS Filter is expressed by

$$e[k] = \hat{S}(k, \tau, f, M, p) - \bar{\hat{S}}(k, \tau, f, M, p) \quad (2.15)$$

where $\bar{\hat{S}}(k, \tau, f, M, p)$ denotes the prediction of $\hat{S}(k, \tau, f, M, p)$ with Adaptive LMS Filter. Depending on this error, $e[k]$, we give the reconstruction distortion D as follows.

$$D = \frac{1}{\tau f} \sum_{k=1}^{\tau f} e[k]^2 \quad (2.16)$$

To determine the minimum number of DN which can be sufficient to represent the all sensor nodes in the WSN environment, sink node periodically increases t_{dns} because the number of DNs decreases while t_{dns} increases. To increase t_{dns} , the sink node adds a number (u) to t_{dns}

in every decision interval t and broadcasts the updated designated node selection threshold $(t_{dns} + u)$ to sensor nodes. As the t_{dns} increases in each decision interval t , the number of DN decreases. Unless the distortion level of Adaptive LMS Filter at the sink is increased, the sink continues to increase t_{dns} because the increase in the distortion means that henceforth, the selected DNs can not represent the event data gathering with all sensor nodes in the event area. Therefore, when the distortion is increased, the sink node stops increasing the t_{dns} and it sets t_{dns} to the last updated value of t_{dns} before the increase of the estimation distortion. After the determination of maximum allowable t_{dns} , the sink node broadcasts this t_{dns} to all sensor nodes. Therefore, each sensor node can distributively decides whether it is a DN or not. If $T_i > t_{dns}$, sensor node i is a DN. If $T_i < t_{dns}$, sensor node i is not a DN.

Now, we give outline the entire DN node selection scheme of DNRS with the algorithmic manner as follows.

- When an event occurs, each sensor node computes its DN selection weight $(T_i, \forall i)$ according to its correlated and uncorrelated neighbor nodes and the distance to the event location.
- Sink initially sets the DN selection threshold t_{dns} as low as possible value such that at the beginning all nodes in the event area become DN $(T_i > t_{dns}, \forall i)$.
- Sink node increases t_{dns} at each decision interval t without disturbing the reconstruction distortion D at the sink. It broadcasts the updated threshold value to all sensor nodes in each t . Thus, at each t , the number of DN decreases because the number of nodes which exceed the updated threshold t_{dns} decreases.
- Until D is increased, sink node continues to increase t_{dns} and to broadcast the updated t_{dns} .
- When the distortion is increased at a decision interval, the sink node stops increasing t_{dns} at this decision interval. Thus, sink node sets t_{dns} to the last updated value of t_{dns} before the increase of the reconstruction distortion and it broadcasts this t_{dns} to all sensor nodes.
- According to the final t_{dns} , the all sensor nodes distributively decide whether it is a DN not or not as follows.

Algorithm 1: DN Selection Algorithm in DNRS

```
1 foreach sensor i do
2   computes  $T_i, \forall i$ 
3 end
4 while  $D$  is increased do
5   increases  $t_{dns}$ 
6   foreach sensor i do
7     if  $t_{dns} < T_i, \forall i$  then
8       sensor  $i$  is a DN
9     end
10  end
11 end
```

- If $T_i > t_{dns}$, sensor node i is a DN.
- If $T_i < t_{dns}$, sensor node i is not a DN.

Thus, when an event occurs in WSN environment, DNRS distributively selects the DNs by using above algorithm. The selected DNs represent all nodes in the event area for this event. However, when the location of the event changes, DNRS again employs the DN selection algorithm given above to determine the new DNs. In Algorithm 1, we give a pseudo-code for DN selection algorithm in DNRS.

After the determination of the DNs, it is important to regulate their sampling frequency f for exploiting the temporal correlation in the event data. In the next section, we give our distributed frequency rate selection scheme in DNRS.

2.3.3 Distributed Frequency Rate Selection of Designated Nodes

The sampling frequency of a sensor node f is defined as the number of samples taken and hence packets sent out per unit time by that node for a sensed phenomenon. Hence, the sampling frequency f controls the amount of traffic injected to the sensor field while regulating the number of temporally-correlated samples taken from the phenomenon. This, in turn, affects the event signal reconstruction distortion [59]. For the temporally correlated data regions,

less number of samples is sufficient to reconstruct the event signal while more samples are needed to reconstruct the event signal region which is not temporally correlated. Therefore, while smaller f is needed to reconstruct the temporally correlated regions in the event signal, higher f is needed to reconstruct the event signal regions which are not temporally correlated. However, since while f increases, the contention at the forward path increases and packet losses arise at higher f . Since packet losses prevent to deliver the sufficient information for the reconstruction, it also affects the reconstruction performance such that while the packet loss rate increases, the reconstruction distortion increases. Therefore, to exploit the temporal correlation in the event signal the appropriate frequency rate selection scheme have to consider the packet losses at the forward path to provide the most appropriate sampling frequency to the sensor nodes such that the DNs having higher λ_i use small f while the DNs having less λ_i use higher f . This kind of frequency rate regulation scheme provides a new congestion control mechanism to sensor networks such that while at the data paths in which the packet losses arise, f is decreased, at the non-congested paths f is increased.

As discussed in Section 2.2.1, in human immune system when a pathogen enters a body, some B-cells are stimulated and they start to secrete the antibody with the appropriate density to eliminate the antigens produced by the pathogen. This natural mechanism is modeled with the immune network models given in (2.1), (2.2). Similarly, in WSN when an event occurs, some sensor node should be selected as the DNs and they should send the sensed information with the appropriate sampling frequency to sink node to reconstruct the event signal without the congestion in the forward path. In order to address the sampling frequency regulation and the congestion in the forward path, we exploit the relationship between the immune system and Wireless Sensor Networks (WSN) and we adopt the immune network equations given in (2.1), (2.2) to provide the sensor nodes a distributed rate control mechanism. According to the relation between Immune System and WSNs, which is summarized in Table 2.1, we give the adaptation as follows:

- We consider the each sensor node as a B-cell which secretes only one kind of antibody.
- We consider the sensor data as the antibodies which are secreted by B-cells and we model the sampling frequency of the sensor nodes (f_i) as the antibody concentration given by s_i . f_i denotes the sampling frequency identifier of sensor node i and it is determined by mapping the actual sampling frequency of sensor node i to a number

between 0 and 1.

- In (2.1), (2.2), the stimulus value of antibodies (S_i) manages the concentration of the antibody i according to the antigen concentration and the natural extinction of the antibodies. Based on the similarity summarized in Table 2.1, to manage the sampling frequency rate of DNs according to the reconstruction distortion (D) and the packet loss rates (λ_i), we consider S_i as the rate control parameter $F_i(t_k)$ and give as follows.

$$F_i(t_{k+1}) = F_i(t_k) + \left(\frac{1}{K} \sum_{j=1}^K f_j(t_k) + aD - b\lambda_i\right)f_i(t_k) \quad (2.17)$$

where

$$a = (1 - \text{signum}(\lambda_i - \lambda_c))(1 - \text{signum}(D - D_c))/2$$

$$b = (1 + \text{signum}(\lambda_i - \lambda_c))/2$$

λ_c and D_c are the application specific packet loss rate and reconstruction distortion constraints, respectively, K is the number of DNs, t_k is the k^{th} decision interval, $f_i(t_k)$ is the sampling frequency identifier of sensor node i at the decision interval t_k , and $F_i(t_{k+1})$ is the rate control parameter at the decision interval t_{k+1} . Here, a and b are the coefficients which controls the respective influence between D and λ_i . If $\lambda_i > \lambda_c$, $a = 0$ and $b = 1$. In this case, to reduce λ_i , DNRS neglects the reconstruction distortion ($a = 0$) such that after DNRS decreases λ_i , it deals with decreasing D if $D > D_c$. If $\lambda_i < \lambda_c$ and $D > D_c$, $b = 0$ and $a = 1$. In this case, since $\lambda_i < \lambda_c$ ($b = 0$), DNRS deals with decreasing D .

- We consider the reporting frequency of DN i at t_k ($f_i(t_k)$) as the concentration of antibody i (s_i), which is secreted by B-cell i , given in (2.2). Based on the rate control parameter $F_i(t_k)$ given in (2.17), the sampling frequency identifier $f_i(t_{k+1})$ can be given as follows

$$f_i(t_{k+1}) = \frac{1}{1 + e^{(0.5 - F_i(t_k))}} \quad (2.18)$$

In each decision interval (t_k), each sensor node distributively evaluates these mechanism given in (2.17), (2.18) by means of the coordination between DNs and sink as follows:

- In each t_k , each DN broadcasts its sampling frequency identifier ($f_i(t_k)$) to all DNs.
- In each t_k , each DN computes its packet loss rate (λ_i) by the feedback from the sink node.
- D is computed at the sink by Adaptive LMS Filter and it is broadcasted to all sensor nodes in each (t_k).

The operation of immune System based distributed rate control mechanism given in (2.17), (2.18), which is based on the coordination among the DNs, can be given as

- When an event occurs, some sensor nodes are selected as the DN through the DN selection scheme given in Section 2.3.2.
- The selected DNs start to sample the event signal with a sampling frequency.
- When for DN i , $\lambda_i > \lambda_c$, $a = 0$ and $b = 1$. In this case, F_i starts to decrease and this result in decrease in the sampling frequency of DN i (f_i). Since the decrease in the sampling frequency of DN i reduce the contention at the forward path of DN i , λ_i decreases.
- When for DN i $\lambda_i < \lambda_c$, $b = 0$. In this case, if $D > D_c$ ($a = 1$), DN i increases its sampling frequency ($f_i(t_k)$) to decrease D .
- In addition to rate regulation scheme, immune system based DNRS algorithm provides a kind of combinatorial behavior to DNs such that (i) when the over all packet loss rates increases, all sampling frequencies of DNs (f_i , $\forall i$) decrease more faster to reduce the overall packet losses (ii) when D becomes immensely large, all sampling frequencies of DNs (f_i , $\forall i$) increase more faster to reduce D .

Beyond the operation of the DNRS algorithm, starting from any initial state, DNRS also enables DNs to collaboratively converge a stable state in which the smaller packet loss rate and event signal reconstruction distortion and significant energy conservation can be provided to WSN by collaboratively regulating sampling frequency of DNs. In Algorithm 2, we give a pseudo-code for the rate selection algorithm in DNRS.

Algorithm 2: Rate Selection Algorithm in DNRS

```
1 foreach  $t_k$  do
2   sink broadcasts  $D$  and  $\lambda_i, \forall i$  to all DNs
3   foreach DN  $i$  do
4     broadcasts its sampling frequency  $f_i(t_k)$  to all DNs
5     computes  $F_i(t_{k+1})$  using (17)
6     computes  $f_i(t_{k+1})$  using (18)
7   end
8 end
```

2.4 Performance Evaluation

In this section, we give the numerical simulation results of DNRS. In order to evaluate the performance of DNRS, we develop a simulation environment using MATLAB. In this environment, 100 sensor nodes in $100 \times 100m$ sensor field were randomly positioned. An event was generated with a point source randomly positioned in this environment. The parameters used in our simulation are given in Table 2.2.

Table 2.2: Simulation Parameters

Area of sensor field	$100 \times 100m$
Number of sensor nodes (N)	100
Packet length	100 samples/sec.
Queue size of sensor nodes	20 packets
MAC	802.11
Routing	DSR
Radio range of a sensor node	20 m
Correlation radius r	3-15 m
Correlation coefficients θ_1	0.3
Correlation coefficients θ_2	0.3

DNRS determines the DN selection threshold t_{dns} at the sink node and the sink node broadcasts t_{dns} to all sensor nodes whereby each sensor node distributively decides whether it is a DN or not. To show the effect of the t_{dns} on the number of DN, we increase the t_{dns} by starting from a certain minimum value. In Fig. 2.1, the number of DNs is shown for varying t_{dns} with different r . While the t_{dns} increases, the number of DNs whose weights (T_i) can exceed t_{dns} decreases. Therefore, the number of DNs decreases. As r increases, the number of correlated

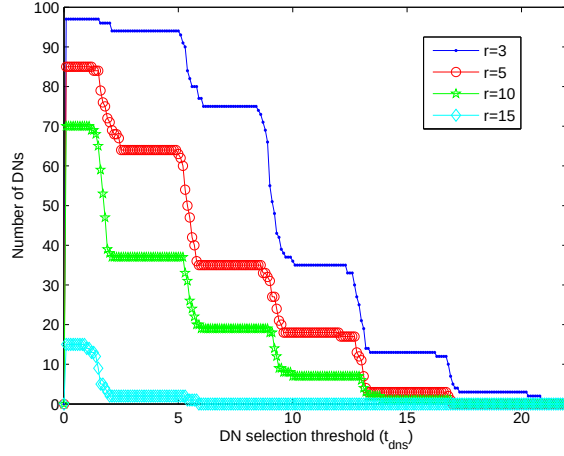


Figure 2.1: t_{dns} for the number of DNs

neighbors of DNs increases and the weights of DNs (T_i) decreases. Therefore, as observed in Fig. 2.1, the number of DNs decreases while r increases.

In WSNs, energy consumption mainly depends on the number of source nodes (DNs) and the total number of samples injected to sensor field by DNs. Since sink can regulate the number of DNs by changing t_{dns} , it can also regulate the average total energy consumption by changing t_{dns} . To investigate the energy consumption of the DNRS, we assume that on average each DN consumes $20mW$ cumulatively for sensing the event and communicating the observed information to the sink node during the overall event monitoring period. In Fig. 2.2, the average total energy consumption is shown for varying t_{dns} with different r . As observed, while t_{dns} increases, the number of DNs decreases and therefore, the average total energy consumption decreases. Similarly, since increase in r results in decreasing the number of DNs, the average total energy consumption decreases as r increases. Hence, maximum r and maximum t_{dns} seem imperative to provide the minimum energy consumption to sensor networks. However, selected r and t_{dns} must also provide the application specific event signal reconstruction reliability to sensor networks. Therefore, t_{dns} and r should be selected, which can provide the application specific reconstruction reliability and minimum energy consumption. In Fig. 2.3, D is shown for varying t_{dns} with different r . As observed, while t_{dns} increases, D decreases. However, after some certain values of t_{dns} , D starts to increase. This is mainly because the number of DNs starts to be insufficient to deliver the sufficient information needed for the reconstruction. The DN selection algorithm of DNRS uses these points in which D starts

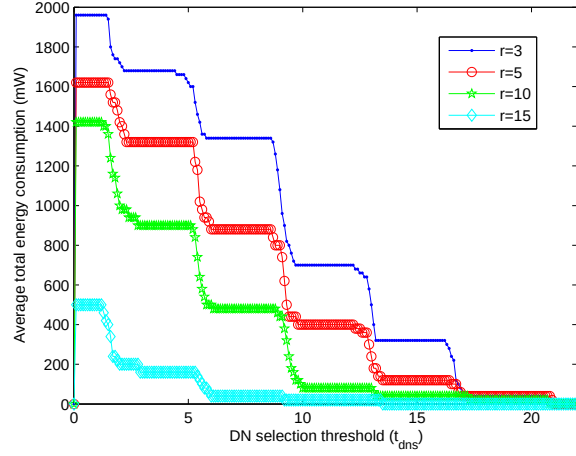


Figure 2.2: t_{dns} for the average total energy consumption.

to increase to determine the most appropriate value of t_{dns} . Furthermore, since increase in r results in smaller number of DNs, as r increases, D starts to increase at the smaller t_{dns} . This prevents to decrease D even at the smaller t_{dns} . Therefore, while $r = 3$ and $r = 5$ can be selected as the correlation radius for the scenario shown in this simulation, $r = 10$ and $r = 15$ cannot.

Apart from the number of DNs, the sampling frequency of DNs affects the reconstruction performance. To show this effect over the reconstruction performance, in Fig. 2.4 D is shown for varying sampling frequency with different t_{dns} . Common method for increasing the reconstruction accuracy is the over sampling such that increasing sampling frequency, reconstruction accuracy can be improved. As observed in Fig. 2.4, while sampling frequency of DNs increases, D decreases. However, after the certain values of f , D starts to increase as f increase. Higher values of f increase the contention at the forward path and packet loss rate of DNs (λ_i). Therefore, for higher values of f , the sufficient information for the reconstruction cannot be delivered to the sink and this increases D as f increases. As t_{dns} decreases, the number of DNs and the contention at the forward path increases. This results in increase in λ_i . Therefore, as observed in Fig. 2.4, while t_{dns} decreases, D increases. Hence, the appropriate rate control mechanism must deal with decreasing the packet loss rate of DNs to provide minimum reconstruction distortion as well as minimum energy consumption.

In addition to the designated node selection, DNRS regulates the sampling frequency of the designated nodes to exploit the temporal correlation in the event data. As discussed in Section

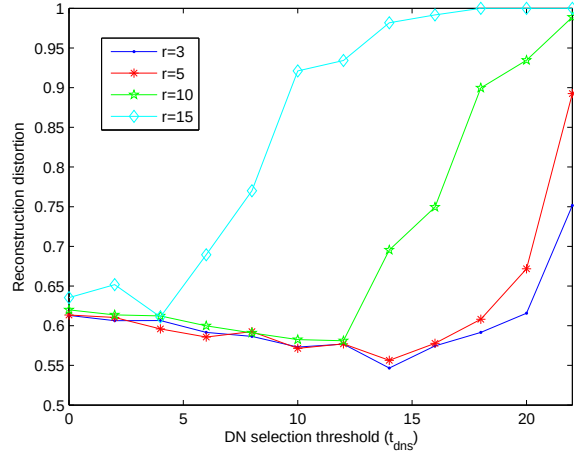


Figure 2.3: D for varying t_{dns} with different r .

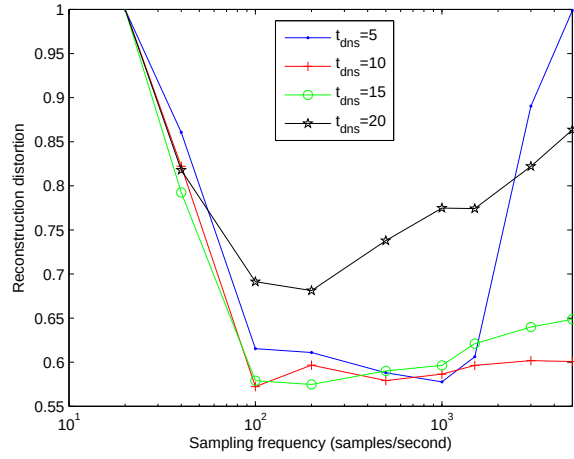


Figure 2.4: D for varying f with different t_{dns} .

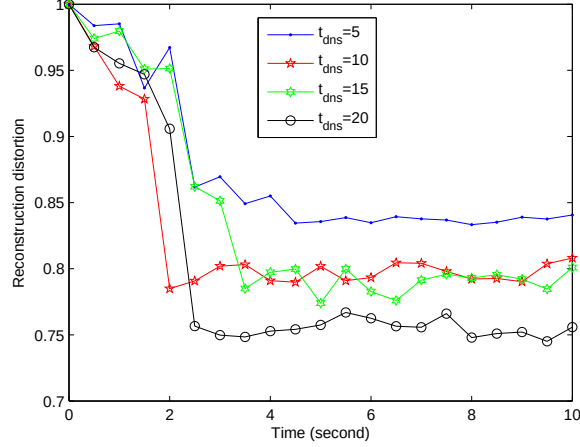


Figure 2.5: D vs. time with different t_{dns} .

2.3.3, temporally correlated event data can be inferred from its previous sample, less number of samples delivered to the sink by DNs is sufficient to reliably reconstruct the event signal at the sink. Therefore, DNRS enables the DNs to regulate their sampling frequency to exploit the temporal correlation in the event data such that when the temporally correlated event data is observed, DNRS reduce the sampling frequency of DNs and when the temporal correlation can not be observed, it increases the sampling frequency of DNs. Furthermore, DNRS reduces the sampling frequency of a DN when it has the higher packet loss rate (λ_i).

To evaluate the performance of the distributed rate control mechanism in DNRS we consider a scenario in which DNs sample the event signal and transmit the sampled event signal to sink during 10 *seconds* time period and each DN updates its sampling frequency in every decision interval ($t = 0.5$ *seconds*). We assume that $D_c = 0.85$ and $\lambda_c = 0.2$ and each DN initially sets its sampling frequency as 200 ($f_i(t_0) = 200$ *samples/sec.* $\forall i$) and each DN can change its sampling frequency between 100 *samples/sec.* and 1000 *samples/sec.* such that 100, 200, 300, ..., 1000. According to the scenario, in Fig. 2.5, D versus time is shown with different t_{dns} . As observed, for all t_{dns} , DNRS provides the reconstruction distortion which is below $D_c = 0.85$ and slightly constant after 2.5 *seconds*. Furthermore, while t_{dns} increases, D decreases. As t_{dns} increases, the contention at the forward path decreases together with the number of DNs. Therefore, for higher t_{dns} , DNRS manages and regulates the sampling frequency of *DNs* to provide minimum D more easily than the case for smaller t_{dns} , which provides more DNs.

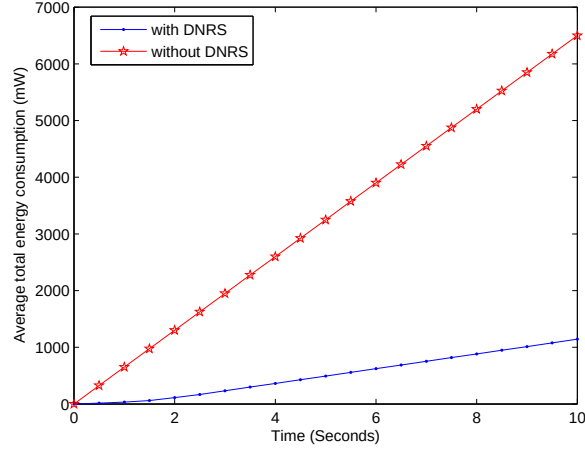


Figure 2.6: Average total energy consumption vs. time with and without DNRS.

To show the energy efficiency of the distributed frequency regulation mechanism in DNRS, in Fig. 2.6, using the same simulation parameters in Fig. 2.5, average total energy consumption versus time is shown with and without DNRS. For the case without DNRS, all DNs use the same sampling frequency ($f = 500 \text{ samples/sec.}$). We assume that for all case, on average, each DN consumes $0.01mW$ cumulatively to deliver a sample of the event signal to sink node. As observed in Fig. 2.6, in terms of energy consumption, the case with DNRS completely outperforms the case without DNRS. This is mainly because DNRS can regulate the sampling frequency of DN according the temporal correlation in the event signal and the packet loss rate of DNs.

2.5 Conclusion

In this chapter, we introduce our method Distributed Node and Rate Selection (DNRS) algorithm which based on the principles of natural Immune System. Based on the B-cell stimulation in immune system, DNRS selects the most appropriate sensor nodes that send samples of observed event, are referred to as designated node (DN). Using the method, each sensor node can distributively decide whether it is a designated node or not according to its correlation with its neighbors and event source. In addition to the determination of the DNs, to exploit the temporal correlation in the event data based on the immune models, DNRS selects the appropriate sampling frequencies of sensor nodes according to the congestion at the forward

paths and the event signal reconstruction distortion periodically calculated at the sink node by Adaptive LMS Filter. With the selection of the minimum number of designated nodes and the regulation of the sampling frequency of designated nodes, DNRS provides the significant energy saving to Wireless Sensor Networks (WSN).

CHAPTER 3

BIOlogically-inspired Spectrum Sharing in Cognitive Radio Networks

In this chapter, we introduce BIOlogically-inspired Spectrum Sharing (BIOSS) algorithm for cognitive radio networks which is based on the adaptive task allocation model of an insect colony. Without need for any coordination among the unlicensed users, BIOSS enables each cognitive radio in the same environment to distributively share the available licensed or unlicensed spectrum bands over which it can effectively communicate. Furthermore, BIOSS enables cognitive radios to distributively decide the multiple spectrum bands over which it can simultaneously communicate without any coordination among cognitive radios. BIOSS was first introduced in [64].

The related works and the motivation are presented in Section 3.1. In Section 3.2, we discuss some relation between biological task allocation and spectrum sharing in cognitive radio. In Section 3.3, we introduce the BIOSS protocol algorithm and its detailed operation principles. Performance evaluation results and discussions are provided in Section 3.4, which is followed by Section 3.5 with concluding remarks.

3.1 Motivation and Related Works

The spectrum access of the wireless networks is generally regulated by governments via a fixed spectrum assignment policy. Thus far, the policies have satisfactorily met the spectrum needs of the wireless networks. However, in the recent years, the dramatic increase in the spectrum access of the pervasive wireless technologies has caused scarcity in the available spectrum bands. Therefore, the limited available spectrum and the inefficiency in the spec-

trum usage necessitate a new communication paradigm to exploit the entire existing wireless spectrum opportunistically [1]. This new networking paradigm is referred to as NeXt Generation (xG) Networks as well as Dynamic Spectrum Access (DSA) and Cognitive Radio Networks (CRN) [1].

Cognitive radio has the ability to capture or sense the information from its radio environment. Based on interaction with the environment, cognitive radio enables the users to communicate over the most appropriate spectrum bands which may be licensed or unlicensed. Since the most of the spectrum is already assigned to the licensed users (primary user) by the governmental policies, cognitive radio mainly strives for the communication over the licensed spectrum band. In the licensed spectrum bands, cognitive radio exploits the temporally unused spectrum which is defined as the spectrum hole or white space [2]. If cognitive radio encounters the licensed user at the licensed spectrum band, it moves to another spectrum hole or stays in the same band without interfering with the licensed user by adapting its communication parameters such as transmission power or modulation scheme. As for the unlicensed spectrum bands in which the licensed users cannot exist, all users have the same priority to access the existing unlicensed spectrum bands. The cognitive radio technology enables the users to opportunistically access the available licensed or unlicensed spectrum bands through four main functionalities required: (i) spectrum sensing, (ii) spectrum managing, (iii) spectrum mobility and (iv) spectrum sharing [1]. Among these, *spectrum sharing* is one of the most important function in cognitive radio, which allows cognitive radio to fairly share the available spectrum bands among the coexisting cognitive radios.

There exist some research efforts on the problem of spectrum sharing in cognitive radio. In some of these studies, the conventional approaches based on the centralized control [18], [19] are introduced. However, due to the heterogeneous and dynamic nature of cognitive radio, centralized approaches are not practical. Instead, in some studies, the distributed approaches which do not need any central controller are suggested [20], [21]. In [20], a distributed coordination which does not need any central controller is proposed. In [21], Asynchronous Distributed Pricing (ADP) scheme is proposed, based on the signal exchange via coordination between users to compensate the ascendant interference level. Due to their distributed approaches, these studies provide better adaptation capability to cognitive radio in the dynamically changing heterogeneous environment. However, the coordination among cognitive radios results in significant amount of coordination delay. Since cognitive radio should ex-

perience minimum delay to provide seamless communication capability during the spectrum handoff, the distributed spectrum sharing must provide minimum delay. In [65], decentralized cognitive medium access control protocols are proposed which enable secondary users to share the available spectrum according to their partial observations. However, the protocols do not consider the fairness of the spectrum sharing. In [66], a device-centric spectrum management scheme with low communication costs is proposed. According to this scheme, users observe their local interference patterns and select the appropriate spectrum and communication parameters. This scheme enables cognitive radio to non-cooperatively and fairly share the available spectrum. However, it does not consider the spectrum sharing for the multiple spectrum sharing case.

3.2 Insect Colony and Spectrum Sharing in Cognitive Radio Networks

As a result of its ability to interact with its environment, cognitive radio enables the users to communicate over the most appropriate spectrum bands which may be licensed or unlicensed. Furthermore, cognitive radio can use multiple spectrum bands simultaneously. Moreover, the selected spectrum bands are not necessary to be contiguous [1]. This provides less performance degradation during spectrum handoff to cognitive radio [1]. This is mainly because, when cognitive radio vacates a spectrum band (spectrum handoff) due to any reason such as appearing licensed user, cognitive radio communication is not interrupted since cognitive radio continues the communication over the rest of spectrum bands. Although the multi-spectrum transmission promises for cognitive radio applications, a spectrum sharing model which can effectively share the all available spectrum bands is a challenging issue.

Similar to cognitive radio, in insect colonies, individuals sense the environment to detect the tasks and then, the detected tasks are performed simultaneously by individuals which are better equipped for the task. Here, we adopt this natural mechanism to introduce an effective spectrum sharing model for CRN.

3.2.1 Task Allocation Model of Insect Colony

In the ant colonies, each ant has relatively little intelligence, while the collaborative behavior of the colony provides a great deal of global intelligence which is capable of optimizing

certain tasks. In social insect colonies, different activities are often performed simultaneously by individuals which are better equipped for the task. This phenomenon is called division of labor [67].

Here, for task allocation problem in insect colony, every individual has a *response threshold* for every task [67]. Response threshold refers to the likelihood of reacting to a task-associated stimuli. Task-associated stimuli s , is defined as the intensity of an activator associated with a particular task and it can be a number of encounters such as a chemical concentration, or any cue sensed by individuals. For example, if the task is larval feeding, the task-associated-stimuli s , may be larval demand expressed through the emission of the pheromone, i.e., a chemical substance deposited by real life ants [67]. Individuals perform the task when the level of the task-associated stimuli s , exceeds their threshold. Therefore, a response threshold θ determines the tendency of an individual to respond to the stimuli s , and perform the associated task.

Thus, based on the above definitions, the probability of performing task as a function of stimuli intensity s , and response threshold θ , is given by

$$T_{\theta}(s) = \frac{s^n}{s^n + \theta^n}, \quad (3.1)$$

where $n > 1$ determines the steepness of the threshold. As observed in (3.1), for $s \ll \theta$, the probability of performing the task is close to zero and for $s \gg \theta$, this probability is close to 1. Therefore, performing the task is likely for $s \gg \theta$ and unlikely for $s \ll \theta$.

In the following, we apply the task allocation model of insect colony to establish an effective dynamic spectrum sharing model for CRN.

3.2.2 Biological Task Allocation Based Spectrum Sharing Model for Cognitive Radio Networks

Spectrum sharing in cognitive radio has great similarities with task allocation in insect colony. For example, cognitive radio senses the environment for the available spectrum bands and then, simultaneously transmits its packets to available spectrum bands. Similarly, in an insect colony, individuals sense the environment through the pheromone for available tasks and, then available tasks are accomplished by individuals better equipped for the tasks. In biological

model, each task is shared to individuals better equipped for this task through the task performing probability given in (3.1). Similarly, in CRN, the available spectrum bands should be fairly shared to cognitive radios by the effective spectrum sharing model. According to this analogy, we adopt the task performing probability given in (3.1) to introduce the *channel selection probability* which enables cognitive radios to effectively share the best available spectrum bands. To this end, we map the task allocation model given in Section 3.2.1 to spectrum sharing in CRN as follows:

- We consider an insect in a colony as a cognitive radio.
- We consider a task as an available channel.
- We consider the task associated stimuli (s) as the estimated permissible power [68] to channel j (P_j), where j denotes channel j .
- We consider the response threshold of an insect (θ) as the required transmission power (p_{ij}) for cognitive radio i to channel j . Cognitive radio i determines p_{ij} as the power which can meet the user requirements¹.

To summarize this mapping between CRN and insect colony, it is also outlined in Table 3.1.

Table 3.1: Mapping Between CRN and Insect Colony

Insect Colony	Cognitive Radio Networks
Insect	Cognitive radio
Task	Available channel
Task associated stimuli (s)	Permissible power to channel (P_j)
Response threshold (θ)	Required transmission power p_{ij}

According to the mapping discussed above and following (3.1), we introduce the channel selection probability T^{csp} as

$$T_{ij}^{csp} = \frac{P_j^n}{P_j^n + \alpha p_{ij}^n + \beta L_{ij}^n}, \quad (3.2)$$

where $n > 1$ determines the steepness of the channel selection probability T_{ij}^{csp} and α and β are the positive constants for respective influences of L_{ij} and p_{ij} .

¹ The required transmission power (p_{ij}) can be determined according to the user requirements, acceptable interference level, acceptable path loss and channel link error by the effective spectrum management model.

L_{ij} is a learning factor which enables cognitive radio i to learn or forget the channel j . If channel j provides the QoS requirement of cognitive radio i , channel j is learned by cognitive radio i and L_{ij} is updated as $L_{ij} - \xi_0$. If channel j cannot provide the QoS requirement of cognitive radio i , channel j is forgotten by cognitive radio i and L_{ij} is updated as $L_{ij} + \xi_1$. ξ_0 and ξ_1 are the learning and forgetting coefficients, respectively.

For sharing the available spectrum bands, each cognitive radio computes the channel selection probability for every available channel ($T_{ij}^{csp}, \forall j$). Then, the channel selection probabilities ($T_{ij}^{csp}, \forall i, j$) are used by the cognitive radios to effectively capture the available spectrum band in the following way:

- When the estimated permissible power to channel j (P_j) is immensely larger than the required transmission power of cognitive radio i ($P_j \gg p_{ij}$), T^{csp} is close to one and it is most probable that cognitive radio i transmits to channel j .
- When the estimated permissible power to channel j (P_j) is much less than the required transmission power of cognitive radio i ($P_j \ll p_{ij}$), T^{csp} is close to zero and it is almost impossible that cognitive radio i transmits to channel j .
- For a constant p_{ij} , as the permissible power to channel j (P_j) increases, the probability that cognitive radio i transmits to channel j increases.
- For a constant P_j , as the required transmission power of cognitive radio i (p_{ij}) increases, the probability that cognitive radio i transmits to channel j decreases.
- If channel j has the small P_j , it is highly likely that channel j is allocated by the cognitive radios having the small required transmission power (p_{ij}). This provides that the channels having less permissible power are allocated by cognitive radios having less user demand.
- If channel j has the large P_j , channel j may be allocated by the cognitive radios having the small or big required transmission power (p_{ij}). This provides that the channels having more permissible power are allocated by cognitive radios having less or more user demand.
- If the spectrum handoff is triggered by the appearance of licensed user or any location change, (3.2) becomes zero for channel j due to $P_j = 0$ for all cognitive radios ($\forall i$)

until channel j is again available. Furthermore, in this case, for channel j , all learning coefficients $(L_{ij}, \forall i)$ are set to initial values.

(3.2) is computed by each cognitive radio when there is any change in the time-varying environment or user requirements. Thus, (3.2) enables cognitive radio to easily adapt to the instantaneous changes in the environment or user requirement, such as small holding time, user mobility, appearing the licensed user and hence to distributively share the available spectrum bands without any central controller and any coordination.

3.3 BIO-inspired Spectrum Sharing Protocol (BIOSS)

According to the model given in Section 3.2.2, the detailed operation of BIOSS protocol algorithm is outlined here. In order to distributively conduct spectrum sharing by following the steps below, each cognitive radio i :

- determines the available spectrum bands by means of the appropriate spectrum sensing algorithms.
- estimates the permissible powers (P_j) to the all available channels $(P_j, \forall j)$.
- sets the learning factor for every available channel $(L_{ij}, \forall j)$ as the same initial values and sets the learning coefficients ξ_0, ξ_1 to appropriate values (e.g., $\xi_0 = \xi_1 = 10$ assumed for simulation experiments in Section 3.4 to have equal learn and forget rates).
- computes the channel selection probability T^{csp} for all available channels $(T_{ij}^{csp}, \forall j)$
- selects the channel having maximum T_{ij}^{csp} to transmit its packets. Similarly, to use the multiple spectrum bands, cognitive radio selects the channels having larger T_{ij}^{csp} . If the number of channels is not adequate to meet the QoS requirements, cognitive radio increases the number of channels to transmit its packets.
 - If the selected channel meets the QoS requirements, cognitive radio using this channel *learns* this channel by updating learning factor $(L_{ij} = L_{ij} - \xi_0)$.
 - If the selected channel can not meet the QoS requirements, cognitive radio using this channel *forgets* this channel by updating learning factor $(L_{ij} = L_{ij} + \xi_1)$.

- When the spectrum handoff is triggered by the licensed user appearance or any location change, the permissible power to vacated channel is set to zero ($P_j = 0$) and all learning factors for this channel ($L_{ij}, \forall i$) is set to initial values.
- Apart from the spectrum handoff, when any change occurs at the environment, all channel selection probabilities ($T_{ij}^{csp}, \forall i, j$) is distributively computed by all cognitive radios.

Thus, BIOSS enables cognitive radios to distributively share the available spectrum bands without any coordination. This provides the better adaptation capability to instantaneous changes in the dynamically changing radio environment. It also provides cognitive radio the seamless communication capability. Furthermore, the learning capabilities of BIOSS enables cognitive radio to learn the slowly changing radio environment.

According to BIOSS, while the interference increase, the permissible power to channel decrease. Therefore, the channel having small permissible power due to the interference can be preferred by the cognitive radio which needs the small power. This enables the cognitive radios to control the interference in the channels such that the channel having more interference can be preferred by the cognitive radio which needs the less transmission power.

BIOSS has the spectrum efficiency such that according to the biologically inspired channel selection probability, while the cognitive radio which needs the more transmission power for its Qos demands prefers the channel which permits the more transmission power, the cognitive radio which needs the less transmission power prefers the channel which permits the less transmission power.

Consequently, BIOSS allows cognitive radio to effectively share the available spectrum bands and provides the seamless communication capability.

3.4 Performance Evaluations

In this section, we present the simulation results on the performance of BIOSS protocol. In order to evaluate the performance, we develop a simulation environment using MATLAB. Firstly, we give some assumptions for the simulation environment and then, we present the results on the performance of BIOSS. Our first results are related to the case that each cognitive radio uses only single channel among the available channels. The second set of results

are for the case that each cognitive radio uses multiple channel simultaneously.

3.4.1 Simulation Environment

We assume that the randomly deployed licensed and unlicensed users with cognitive radio coexist in the environment. The number of users changes from 5 to 50 while the number of channel changes from 5 to 25. Each unlicensed user can access all available channel while the licensed users can access any channel to transmit with a fixed probability. We also assume that licensed users give only the instantaneous interrupt to the channels.

3.4.2 Spectrum Sharing with Single Channel Selection

In this section, we give the performance evaluations of BIOSS when each cognitive radio selects and communicates over one channel. We give simulation parameter in Table 3.2.

Table 3.2: Simulation Parameters

Number of channels	5 – 25
Number of cognitive radios	5 – 50
Initial permissible power interval (P_j)	0 – 400mW
Transmission power interval (p_{ij})	10 – 40mW
Initial learning factors interval (L_{ij})	10 – 40
Learning coefficient (ξ_0)	10
Forgetting coefficient (ξ_1)	10
α	10
β	5

Firstly, we conduct the simulation to observe that how the used power at the channels changes according to time. In Fig. 3.1, it is observed that $0 < t < 25 \text{ seconds}$, the used power at the channels oscillates because of the available channels are being captured by cognitive radios. However, after $t = 25s$, since each cognitive radio learns its appropriate channel, the channel shows stationary behavior although the channels are interrupted by licensed user. BIOSS enables the cognitive radio to eventually learn the appropriate spectrum band. This provides the convergence to appropriate spectrum band. However, since the system is dynamically interrupted by the licensed user, it is not possible to completely converge the proposed algorithm. In Fig. 3.2, similar to Fig. 3.1 we show that the used power at the channels without the interruption of the licensed user. As observed, after $t = 25s$, each cognitive radio finds and learns

its appropriate channel, that is, the channel allocation process converges.

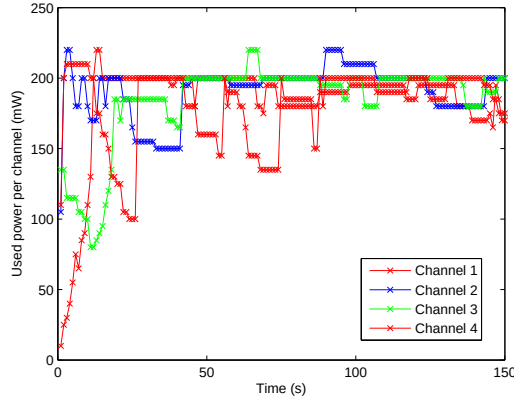


Figure 3.1: The changes in the used power at the four channels according to the time in the single selection case with the licensed user interruption.

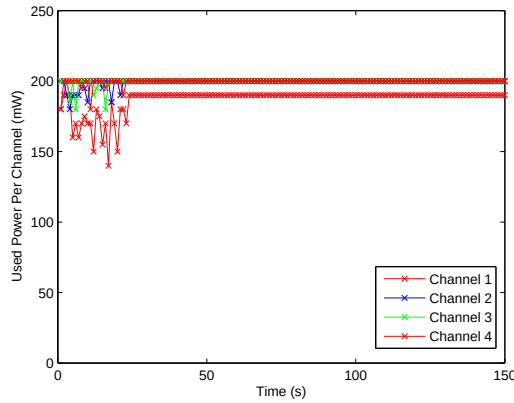


Figure 3.2: The changes in the used power at four channels according to the time in the single selection case without the licensed user interruption.

The overall spectrum utilization as a function of the number of channels is shown in Fig. 3.3. The number of channels is set to be in between 5 and 25. Here, the number of cognitive radios is fixed at 10. As shown in Fig. 3.3, the overall spectrum utilization is slightly above 50% regardless of the number of channels used. Therefore, the utilization is not affected by the number of channels. For this, the main reason is that BIOSS enables cognitive radios to learn their best spectrum which keeps utilization per user even the number of channel increases. We note that the main aim of the spectrum utilization part in the performance analysis is to show that BIOSS provides the great spectrum utilization while providing the seamless communication to cognitive radio.

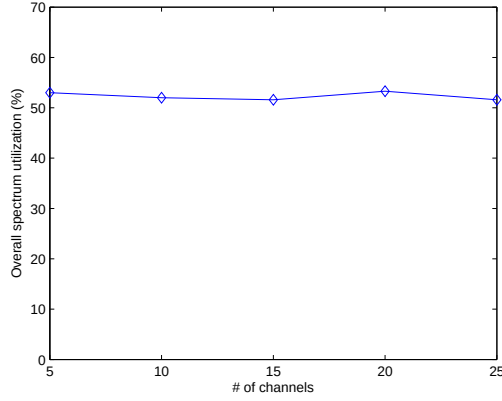


Figure 3.3: The channel utilization according to the different number of channels in single channel selection.

3.4.3 Spectrum Sharing with Multiple Channel Selection

In this section, we give the performance evaluations of BIOSS when each cognitive radio selects and communicates over multiple channel. According to the same simulation parameters given in Table 3.2, we assume that each cognitive radio selects three channels to communicate and transmits the same amount of power to the selected channels. Again, we conduct the simulation to observe that how the used power at the channels changes according to time. In Fig. 3.4, it is shown that in $0 < t < 15$ seconds, the used powers at the channels oscillates because of a lot of spectrum handoff which result from sharing of available spectrum bands. However, after $t = 15s$, since each cognitive radio allocates the available channels, the used power at the channels shows nearly static behavior although the channels are interrupted by the licensed users.

In the multiple channel selection, the licensed user interrupts slightly affect the channel usages of unlicensed users with respect to the spectrum sharing with single channel selection. Moreover, while in single spectrum case channel allocation process converges after $t = 25s$, in multiple case, channel allocation process converges faster (e.g., after $t = 15s$). This is mainly because that in the multiple spectrum selection, the power load of all unlicensed users is shared to the available channels.

The overall spectrum utilization as a function of the number of users is shown in Fig. 3.5 for both single spectrum and multi-spectrum cases. As observed in Fig. 3.5, while the number of unlicensed users increases from 0 to 30, the channel utilization increases for both cases from 0

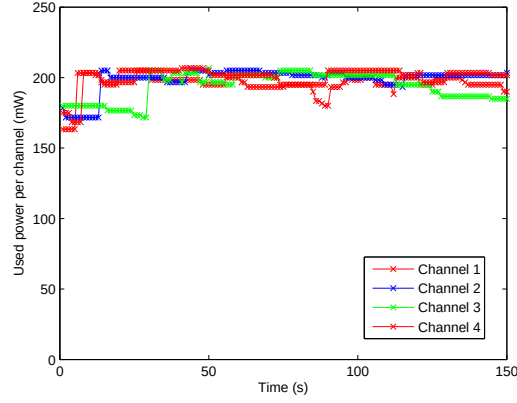


Figure 3.4: The changes in the used power at the four channels according to the time in the multiple selection case.

to 95 percent. However, there is a slightly higher increase in the utilization of multi-spectrum case, which is mainly due to the fact that it intrinsically enables better utilization of the overall spectrum. Note also that as the number of users further increases, the improvement achieved by the multi-spectrum case over single-spectrum is also amplified due to the same reason.

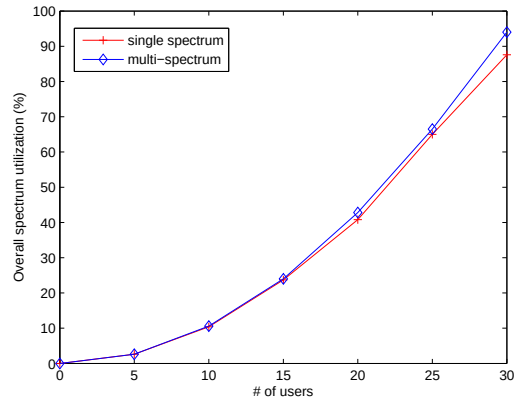


Figure 3.5: The channel utilization according to the different number of unlicensed users in the single and multiple channel selection.

On the other hand, in Fig. 3.6, we observe the expected handoff delay due to coordination with different number of unlicensed users in a single channel selection case. Here, recall that BIOSS does not require any coordination among unlicensed users. Therefore, the objective of this experiment is to show that how much BIOSS saves in terms of potential spectrum handoff delay compared to the existing spectrum sharing techniques which achieve distributed spectrum sharing with a cost of coordination incurred handoff delay. Here, we assume that every

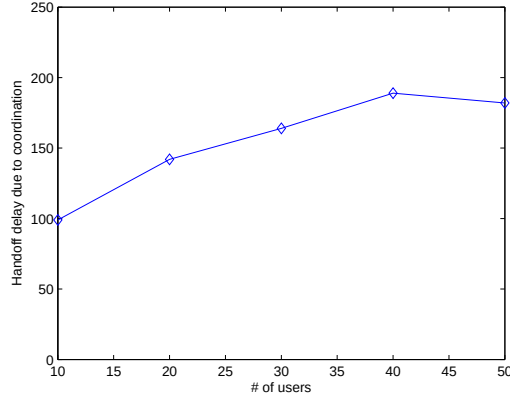


Figure 3.6: Expected spectrum handoff delay due to coordination for varying number of unlicensed users in single channel selection.

spectrum handoff attempt incurs 1 unit of additional coordination related latency. Therefore, as observed from Fig. 3.6, as the number of users increases, the latency overhead of coordination based methods increases as well. Consequently, BIOSS outperforms the existing methods in terms of seamless spectrum mobility support by avoiding additional spectrum handoff delay.

3.5 Conclusion

Efficient spectrum sharing is an important and challenging problem in cognitive radio networks which must be addressed in order to enhance the overall spectrum utilization. In this chapter, we introduced a new BIOlogically-inspired Spectrum Sharing (BIOSS) algorithm which is based on the adaptive task allocation model in insect colonies. BIOSS does not need any coordination among the unlicensed users and achieves distributively determination of the appropriate channel(s). Simulation experiments show that BIOSS realizes efficient dynamic spectrum sharing with high spectrum utilization and without any coordination among the users and hence yielding no spectrum handoff latency overhead due to coordination.

CHAPTER 4

Immune System-inspired Evolutionary Opportunistic Spectrum Access in Cognitive Radio Ad Hoc Networks

In this chapter, immune system based Evolutionary Opportunistic Spectrum Access (ESA) protocol is introduced for cognitive radio ad hoc networks (CRAHNs). ESA was presented in [69]. Based on the principles of immune system, ESA concurrently provides efficient SI, SS and spectrum access operations for SUs in CRAHNs. The salient features of ESA are outlined as follows:

- Unlike all the existing OSA mechanisms using single dedicated control channel that may hamper all SU communications, ESA enables SUs to use all available channels to exchange control information.
- ESA is error-tolerant to provide the resilient identification of all spectrum opportunities despite potentially erroneous spectrum sensing measurements.
- ESA allows SUs to employ optimal spectrum sharing and access strategies to maximize overall spectrum utilization.
- ESA quickly incorporates the changes in the network like node mobility or new node arrivals adaptively.
- ESA is an evolutionary algorithm that can surprisingly improve its performance as contention and also interference among SUs increase. This is a fascinating feature for OSA domain such that undesirable situations such as interference or node mobility can be harnessed to improve network robustness via the evolution of network nodes.
- ESA is an autonomous algorithm that does not need any coordination among network

nodes for SI and SS and it also does not need any priori information about the channel access statistics of primary users.

In Section 4.2, we discuss the analogies and similarities between immune system and OSA. Based on these similarities, in Section 4.3, we introduce an evolutionary OSA model and ESA protocol operations. In Section 4.4, we present the performance evaluation results, which are followed by the concluding remarks given in Section 4.5.

4.1 Motivation and Related Works

Underutilization of wireless spectrum stimulates cognitive radio ad hoc networks (CRAHNS) aiming to enable unlicensed ad hoc users to opportunistically access unused licensed spectrum [22]. The term Opportunistic Spectrum Access (OSA) [23] is broadly used to identify a such kind of overlay spectrum access, in which unlicensed users, i.e., secondary users (SU) detect and access the temporarily unused spectrum bands legitimately licensed to primary users (PU). OSA mainly includes three operations:

- *Spectrum opportunity identification* (SI) aims to identify and intelligently track temporally unused spectrum bands to allow each SU to decide whether a spectrum band is temporally accessible or not.
- *Spectrum sharing* (SS) enables SUs to intelligently allocate the spectrum opportunities to minimize perceived interference level so as to maximize utilization.
- *Spectrum access* allows SUs to access the identified spectrum opportunities to maximize the throughput achieved.

Beside its promising spectrum utilization capabilities, OSA incurs many additional challenges on the conventional access technologies. For example, SI includes scanning and sensing the available channels to determine whether a channel is accessible. However, this process brings significant overhead into the secondary communication in terms of time and energy consumption. Moreover, SI is prone to some inaccurate observations of SUs about channel availability. These wrong observations may impose excessive interference on the licensed communication. Unshared spectrum opportunities may also yield high interference among

the SUs and the overall spectrum utilization inevitably decreases. Thus, the efficient SI, SS, and spectrum access mechanisms are imperative to allow SUs to effectively communicate over the temporarily unused spectrum bands [1].

In nature, biological systems maintain their normal operations in absolutely random and continuously changing environments by means of their *self-organization*, *self-adaptation*, and *self-regulation* capabilities. For example, immune system is a natural defense mechanism to protect organism against pathogens. The aim of immune system is to detect and eliminate the pathogens. Immune system first reliably detects and discriminates the pathogens from self-molecules of the body, e.g., other blood cells or beneficial vitamins and minerals. These operations are intrinsically exposed to many detection and discrimination faults as well. However, immune system can amazingly tolerate the possible detection and discrimination faults via *self-nonself identification mechanism* [56].

Immune system has also an evolutionary instrument called *clonal selection mechanism* (CSM) allowing its cells called B-cells and T-cells to dynamically change their genetic shapes to genetically recognize and eliminate the pathogens. CSM can also improve its genetic recognition talent such that as diversity of infections increases, the organism and immune system become more robust toward a large set of new infections.

Clearly, the capabilities of immune system give great inspiration for the development of many efficient evolutionary algorithms such as network intrusion detection [70], multi-objective function optimization [71], [72] and fault diagnosis [73]. Moreover, based on natural immune system principles, in [42], distributed node and frequency rate selection (DNRS) algorithm is proposed for energy-efficient and reliable communication in wireless sensor networks. In this chapter, to harness the great natural capabilities of immune system we develop an evolutionary opportunistic spectrum access (ESA) algorithm.

Various OSA protocols in current literature are generally categorized according to their control information exchange strategies divided mainly to two different types. In [24], [25], [26], and [27], a dedicated control channel is used to exchange control information. However, a single dedicated control channel may be a bottleneck when primary users frequently use this channel. In [74], [75], time slots are divided into control and data phases. In the control phase, control information are exchanged using a single dedicated channel. However, in the data phase, data transmission is carried out using the channels negotiated in the control

phase. In this approach, the dependency to a single dedicated channel is alleviated because of exchange of only the control information on the single control channel.

In addition to control information exchange strategies, the existing OSA protocols also differ in terms of SI strategies. In fact, some can identify spectrum opportunities [24], [25], [26], [74], while others acquire this information from an external entity. However, dependency on an external entity is a major barrier for the realization of OSA in CRAHNs.

Beside the existing OSA algorithms, some bio-inspired mechanisms are also developed for some spectrum sharing and management problems in cognitive radio based OSA scenarios. In our previous works, some biological systems are investigated to discover their promising properties that can be harnessed to develop efficient algorithms in cognitive radio networks (CRN) [3]. Based on the task allocation phenomenon in an insect colony, a spectrum sharing mechanism is proposed to enhance spectrum utilization [64]. In [76], similarly, some resource allocation mechanisms adopted from insect colonies are used for cognitive wireless networks.

None of the existing OSA protocols simultaneously provides SI, SS, and spectrum access. Lack of a spectrum sharing mechanism severely reduces the spectrum utilization. These protocols are also subject to possible congestion in the single control channel. Moreover, they consider highly static network models in which a fixed number of immobile nodes opportunistically access the available spectrum bands which is not practical for CRAHNs.

4.2 Immune System and Opportunistic Spectrum Access

Here, we first introduce the immune system, then, discuss its relation with OSA.

4.2.1 Biological Immune System

The natural immune system is a complex defense mechanism that protects the organism from hazardous molecules or micro-organisms. This complex task can be accomplished by specific blood cells called B-cells and T-cells in immune system. When an organism is exposed to antigen molecules, B-cells and T-cells cooperatively detect and eliminate the antigens through *self-nonsself detection* and *CSM* [71].

Self-nonsel self detection mechanism can be considered as a distributed detection mechanism in which B-cells are envisaged as small independent detectors that diffuse through the organism in the blood [56]. Detection of antigens can be achieved when the genetic shape of an B-cell chemically matches with an antigen genetic shape. However, due to the huge diversity in B-cell genetic shapes, it is highly likely for B-cell to match with the self-cells of the organism. Although this is a fatal dilemma for the organism, immune system has the great error-tolerance capability to avoid this fatal dilemma. This is accomplished through thymus that is an organ and includes most self-molecules of the organism and in which B-cells genetically mature. During this maturation process, if an B-cell matches with some self-molecules in the thymus, it is eliminated to avoid eliminating self-cells of the organism.

While the self-nonsel self detection mechanism discriminates antigens from self-molecules of the organism, CSM concurrently allows B-cells to recognize genetic shape of antigens by means of *affinity maturation* and *hypermutation* phenomenon. Hypermutation separately enables each antibody to change its genetic shape. As the genetic change can improve the matching between the antibody and antigen, affinity maturation process allows the B-cells having the best matching with the antigen to proliferate so as to genetically recognize and eliminate the antigen. The genetic shape of B-cells that having the best matching with the pathogen can be considered as an *elitist genetic shape* that is no longer mutated and is memorized by continuously proliferating in the course of life time. This *elitist selection mechanism* provides great robustness for organism. Moreover, this mechanism gives great inspiration for the development of immune system based multi-objective function optimization algorithms [71], [72]. Next, we introduce some advantageous analogies between immune system and OSA.

4.2.2 Immune System and Opportunistic Spectrum Access

Although OSA seems different from the natural immune system, they have great deal of intrinsic analogies in terms of operation principles. When immune system encounters a molecule, it determines whether this molecule is a nonself harmful pathogen by means of *self-nonsel self identification mechanism*. If the molecule is determined as a pathogen, immune system recognizes and eliminates the pathogen using *clonal selection mechanism*. Similar to self-nonsel self identification in immune system, OSA needs a detection mechanism to identify spectrum

opportunities. Moreover, similar to clonal selection in immune system, OSA also needs a selection mechanism to enable each SU to choose and access a set of spectrum opportunities so as to maximize spectrum utilization.

In addition to the operational analogies, some attributive analogies between immune system and OSA are as follows:

- Immune system is a highly robust system since it is diverse and error-tolerant. For OSA, it is also essential to provide robust opportunistic spectrum access via diversity and error tolerance. Diversity in channel availability improves the OSA performance in terms of achieved throughput and interference level. Error tolerance in channel sensing provides great robustness in terms of interference level incurred on primary users.
- Immune system is adaptive to recognize and learn new infections. Immune system achieves this adaptivity by improving useful genetic shapes and discarding useless or dangerous ones. Similarly, OSA should be adaptive to recognize and learn channel access pattern of primary users. This can be accomplished by discarding useless channels, on which primary user activities are high, and selecting other fruitful spectrum channels on which primary user activities are minor.
- Immune system is a fully autonomous system which does not need any central control to perform its fascinating steps. Similarly, OSA needs autonomous operations to alleviate the coordination latency overhead and to make spectrum access more robust.

In order to harness the operational and attributive advantages of immune system, we establish the following analogies between immune system and OSA outlined in Table 4.1.

Table 4.1: Relation between Immune System and CRN.

Immune System	Opportunistic Spectrum Access
<i>B-cell</i>	Secondary user
<i>Genetic shape of antibody</i>	Frequency hopping sequence of SU
<i>Antigen</i>	Spectrum opportunity
<i>Self-nonsel self identification</i>	Spectrum opportunity identification
<i>Clonal selection via somatic hypermutation and affinity maturation</i>	Spectrum sharing and access

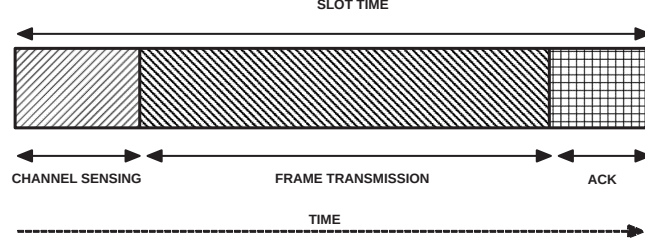


Figure 4.1: The time slot structure of a secondary user.

4.3 Immune System Based Evolutionary Opportunistic Spectrum Access

In this section, we present Immune System Based Evolutionary Opportunistic Spectrum Access (ESA) protocol based on the analogies between immune system and OSA. Before giving the details of ESA, we introduce the network model and assumptions and also give a brief overview for ESA.

4.3.1 Network Model and Assumptions

We consider a network model in which entire spectrum is divided into number of N non-overlapping orthogonal channels having the same bandwidth. Number of M mobile SUs may randomly arrive or depart without any need for a negotiation among SUs. Slot timing is broadcast to SUs by the primary system. Each slot lasts for τ seconds. ESA splits a time slot into three phases as shown in Fig. 4.1. In the channel sensing phase, each SU randomly selects a number of channels less than N and senses them using the energy based channel sensing [77]. In the data transmission phase, each SU intending to transmit makes data transmission. After the data transmission phase, each SU sent data packet waits for an ACK frame from its receiver SU. Each of these three phases consecutively occur in every τ seconds.

All SUs use fixed length frequency hopping sequences (FHS). FHS_i denotes the FHS of SU_i composed of K distinct frequency channels where $K \leq N$. In each time slot, SU_i consecutively hops on FHS_i to transmit or receive data frame. As will be detailed in Section 4.3.4, each transmitter SU can handshake with a receiver SU to tune its FHS.

4.3.2 ESA Overview

ESA concurrently perform SI, SS and spectrum access operations. It enables each SU to resiliently identify all spectrum opportunities based on the entropy based channel availability information acquired by channel sensing. This information can be reinforced to tolerate possible channel sensing errors. Based on the channel availability information, each SU generates its FHS including the channels on which PU activities are minor maximizing overall spectrum utilization.

For spectrum sharing and access, ESA allows each SU to mutate its FHS separately to generate an optimal spectrum sharing and access performance based on a clonal selection based distributed immune algorithm. For this optimization, each receiver SU first interacts with the environment by means of channel sensing and listening and receiving. This interaction lasts for a *mutation interval* that is a specific number of slots and preselected by ESA. During the mutation interval, each receiver SU gathers information from the radio environment to infer which of its frequency hops are exposed to interference. Based on this information, it mutates its FHS at the end of the mutation interval. If a receiver SU does not receive any interference, it stops the mutation operation on its FHS and its FHS becomes an *elitist* FHS. When all receiver SUs have an elitist FHS, the spectrum sharing process converges to a global optimum point maximizing the spectrum utilization.

Next, we detail the ESA operations for SI, SS and spectrum access, which can also be seen in Fig. 4.3.

4.3.3 Spectrum Opportunity Identification

SI in ESA is based on the principles of self-nonself identification in immune system. Immune system enables each B-cell to separately discriminate a harmful pathogen from self-cells of the organism. This discrimination is based on natural maturation process in which each B-cell is matured to infer whether it is useful to eliminate pathogens. This provides great error-tolerance for the identification errors by reinforcing the information about self and non-self cells. Similar to immune system, ESA allows each SU to separately infer whether each spectrum channel can be useful for transmission or not. This is accomplished by the *spectrum opportunity maturation process*. This process empowers each SU_i to reinforce channel avail-

ability information for each channel j as follows.

- Whenever $S U_i$ senses channel j , it keeps a channel availability information about channel j using its *idle and busy counters*, i.e., I_i^j and B_i^j , respectively.
- If $S U_i$ senses channel j as idle, it updates the idle counter I_i^j as $I_i^j = I_i^j + 1$.
- If $S U_i$ senses channel j as busy, it updates the busy counter B_i^j as $B_i^j = B_i^j + 1$.
- Based on the updated I_i^j and B_i^j , $S U_i$ computes the probability, i.e., p_i^j , that channel j is available for $S U_i$, as follows

$$p_i^j = \frac{I_i^j}{I_i^j + B_i^j} \quad (4.1)$$

- To characterize the uncertainty in the availability of channel j , $S U_i$ also computes *channel identification entropy*, i.e., H_i^j , for channel j as follows

$$H_i^j = -\left(p_i^j \log_2(p_i^j) + (1 - p_i^j) \log_2(1 - p_i^j)\right) \quad (4.2)$$

- $S U_i$ computes H_i^j for all available channels and sorts H_i^j , $\forall j$ in ascending order and accordingly, generates the set T_i that includes the channel IDs arranged according to the sorted channel identification entropies.
- T_i includes all channel IDs for $S U_i$ from the most accessible to the least accessible such that $T_i(1)$ is the ID of a channel that is the most useful and accessible for $S U_i$ to maximize spectrum utilization.

Next, ESA spectrum sharing, access method is presented.

4.3.4 Spectrum Sharing and Access

Based on the principles of CSM in immune system, ESA allows each SU to separately share and access the spectrum opportunities that are identified by SI given above. In immune system, to recognize and eliminate a pathogen, each B-cell mutates its genetic shape in a way that it improves some genetic patterns having best matches with the genetic shape of the pathogen. Genetic mutations are based on the some interactions between B-cells and the pathogen. As these interactions increase, the rate of the genetic mutations also increases to recognize and

eliminate the pathogen. Similarly, ESA allows each SU to generate a frequency hopping sequence (FHS) like the genetic shape of B-cells. Let FHS_i is the frequency hopping sequence of SU_i . Each FHS is a fixed length string composed of some channel IDs. Similar to a B-cell in the immune system, each SU_i can mutate FHS_i based on the some interactions with the radio environment such as listening it or transmitting to an intended receiver.

ESA provides the receiver centric spectrum sharing such that each receiver SU mutates its FHS to avoid possible interferences and to share the spectrum opportunities efficiently. The interference avoidance can be only possible for each SU if it can gain a FHS not overlapping with other SU's FHSs via mutations. In Fig. 4.2, mutations for M receiver SUs are observed. The mutations can be modeled as an optimization process subject to

$$\text{minimize } \sum_{i=1}^M \sum_{j \neq i}^M \sum_{l=1}^K \left[1 - \text{sgn}(|FHS_i(l) - FHS_j(l)|) \right] \quad (4.3)$$

where $\text{sgn}(\cdot)$ is the sign function M is the number of receiver SU in the environment, K is the number of channels in each FHS and $FHS_i(l)$ denotes the l^{th} channel of the FHS sequence of SU_i . ESA performs this optimization process similar to the clonal selection based optimization procedures [71], [72] and each SU_i separately performs the following operations,

- During a *mutation interval*, SU_i senses the radio environment and receives the transmissions from the transmitter SUs to infer the interferences on its frequency hops.
- SU_i determines its frequency hops on which it perceives interference from the radio environment. This means some overlapping frequency hops among some SUs.
- If SU_i receives interference on some of its frequency hops, it randomly mutates these hops on FHS_i by randomly flipping these frequency hops.
- If SU_i does not receive any interference on its all frequency hops, it sets its frequency hopping sequence FHS_i as an elitist FHS. Similar to clonal selection based optimization algorithms [71], [72], elitist set selection allows the convergence of the spectrum sharing algorithm.
- Until all FHS become elitist, the mutations continue.

After each FHS becomes the elitist, the spectrum sharing operations are naturally terminated. At the optimal point, since all frequency hops do not overlap, SUs can no longer collide

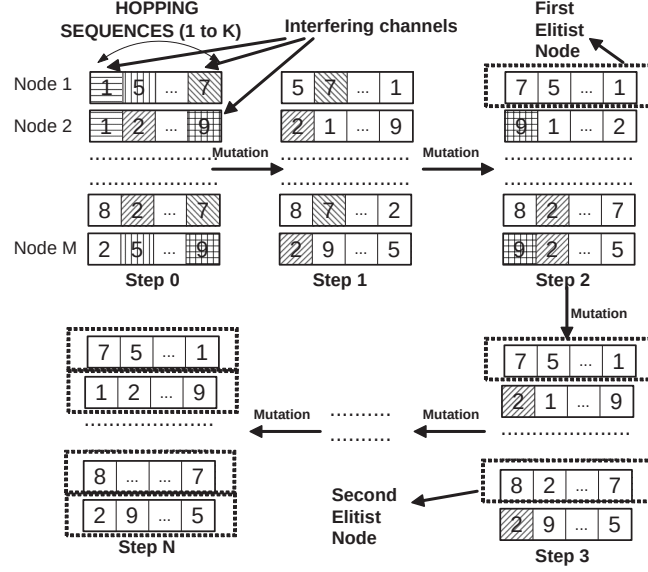


Figure 4.2: Spectrum access and sharing by the mechanisms of FHS generation using mutation and elitist selection.

and interfere with each other providing significant spectrum utilization for the opportunistic communication. All mutation and elitist FHS selection operations are illustrated in Fig. 4.2.

In addition to the simulation analysis given in Section 4.4 for the convergence of the spectrum sharing, here, we also briefly prove it as follows. Let us assume that we concatenate all FHS sequences of receiver SU to form a single large FHS sequence, i.e., $\overline{\text{FHS}} = (\text{FHS}_1 \text{FHS}_2 \dots \text{FHS}_L)$, where L is the number of receiver SUs. Since each FHS has the length K , the length of $\overline{\text{FHS}}$ is KL . We also assume that there exists an optimal sequence $\overline{\text{FHS}}_{opt}$ with length KL that can optimize the objective function given in (4.3). Hence, if there are c mismatches between $\overline{\text{FHS}}_{opt}$ and $\overline{\text{FHS}}$, the probability of convergence, i.e., P_c , can be given as [78]

$$P_c = \frac{c!}{K(K^2 - K)^c} \quad (4.4)$$

P_c shows that the spectrum sharing algorithm is convergent. However, elitist set selection is imperative for the convergence of all immune system based optimization algorithms [78]. Since the mutations are separately performed and elitist FHSs are dynamically generated by each SU, immune system based algorithm strictly converges to the global optimal point. Convergence of artificial immune system algorithms requires a strict elitist selection mechanism [19] such that the artificial immune system algorithm is convergent if an elitist set selection mechanism is used. Hence, due to its elitist FHS selection mechanism ESA is also a conver-

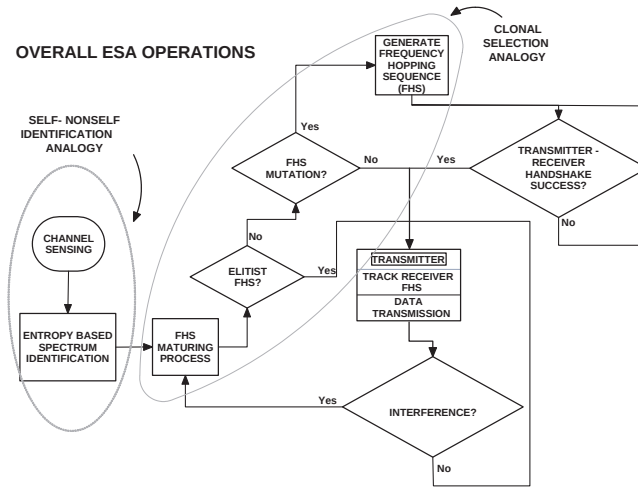


Figure 4.3: Overall ESA operation is illustrated with SI, SS, and spectrum access. All protocol steps can be easily tracked and it is seen how and which parts of ESA operation gives inspiration from the principles of immune system.

gent algorithm. Although its convergence is not analytically proved in the chapter, by means of the simulation results in the performance evaluation part, it is easily shown that ESA is convergent.

For the spectrum access, each transmitter SU also generates FHS including some channels on which PUs do not frequently transmit. In the first phase, each transmitter SU struggles to handshake its intended receiver SU. For this handshake, each transmitter user transmits a Request-To-Sent message (RTS) in a way that it hops on its FHS. When the transmitter SU achieves to deliver a RTS packet, its intended receiver immediately sends a CTS frame including its FHS. This allows the transmitter SU to handshake with its receiver and to track its FHS. However, as presented above, the receiver SUs mutate their FHSs. Therefore, the previously achieved handshakes between the transmitter and receiver SUs are corrupted in the case of the mutations in the FHS of receiver SUs. In this case, the transmitter SU retries to handshake with its receiver SU. However, once the receiver SUs complete the spectrum sharing operation such that all FHS become elitist, SU handshakes no longer suffer from the mutation in receiver SUs.

Table 4.2: Comparison of Distributed OSA MAC Protocols [84].

Protocol	Spec. ID.	Spec. Shar-ing	CC Type	Scan	# of RFE	MRCC
<i>DOSS</i> [24]	√	-	Dedicated	√	3	-
<i>SCA-MAC</i> [79]	-	-	Dedicated	√	2	-
<i>C-MAC</i> [26]	√	-	Dedicated	√	1	-
<i>AS-MAC</i> [25]	√	-	Dedicated	√	1	-
<i>ESCAPE</i> [80]	-	-	Dedicated	√	1	-
<i>DC MAC</i> [81]	-	-	Dedicated	√	1	-
<i>HC-MAC</i> [27]	-	-	Dedicated	√	1	-
<i>BB-OSA</i> [82]	-	-	Dedicated	-	1	-
<i>OS-MAC</i> [83]	-	-	Dedicated	-	1	-
<i>HD-MAC</i> [74]	√	-	Split Phase	√	1	-
<i>SRAC</i> [75]	-	-	Split Phase	-	1	-
<i>Evolutionary Spectrum Access (ESA)</i>	√	√	MRCC	√	1	√

4.4 Performance Evaluation

In this section, the performance of the ESA algorithm is investigated. We first present an overview of the existing OSA techniques in the current literature, and discuss the advantages of ESA with respect to them. Then, we present simulation experiment results of ESA protocol in terms of wide range of performance metrics. Finally, we perform numerical error tolerance analysis of ESA to the inaccuracy in SI process.

4.4.1 Overview of Existing OSA Mechanisms

The performance analysis is made for comparison with other OSA protocols. In [84], a comprehensive performance comparison of the existing solutions are presented. In Table 4.2, the existing OSA algorithms are compared with ESA according to their various properties such as dedicated channel type and number of radio-front-ends ¹.

As observed in Table 4.2, ESA has a number of advantages with respect to others. First of all, ESA is a unique protocol concurrently performing SI and SS, in addition to default CR functionalities such as channel scanning for spectrum sensing. ESA has also a completely

¹ A more comprehensive performance comparison of these existing mechanisms can be found in [84].

different control information exchange strategy from other OSA such that ESA enables the control information exchanges using all available channels, i.e., multiple rendezvous control channel (MRCC) [84], while others use a single control channel, which would be a major barrier under congestion by PU transmissions. Furthermore, ESA requires only a single radio-front-end (RFE) since it splits the channel sensing and data transmission phases in time domain. Consequently, ESA unifies several salient features.

4.4.2 Simulation Experiments

Here, we show the performance of ESA and compare it with generic OSA methods capturing the main features, i.e., SI and SS. The ones making only SI identify the available spectrum channels ordered with respect to entropy and choose K of them. The ones additionally making SS choose the spectrum channels collaboratively with each other but without changing the resulting sequences over time. Simulation experiments are performed in MATLAB. Unless otherwise explicitly stated, in total $N = 10$ channels are used among 10 nodes, half of which are ad hoc opportunistic transmitters and each transmitter is assumed to have a specific target receiver with a single-hop networking topology. The nodes use $K = 6$ hops in FHS and listens to 6 channels in order to collect the entropy information as explained in Section 4.3. The channels are assumed to be independent with bandwidth $B = 1$. Throughput performance of ESA is measured in terms of PU channel access probability, number of channels available, number of SUs. In addition, the fairness analysis of ESA, its collision probability as well as its performance under high node mobility are also presented.

4.4.2.1 Throughput vs. PU channel usage

We vary the PU channel access on-off probability and observe the effect on throughput. It is expected that the nodes using the PU channels are affected much more than the ones utilizing the channels that PU does not reside in. As shown in Fig. 4.4(a), throughput achieved with ESA is much higher than both cases that do not intelligently share the available channels. In addition, ESA is not affected by increasing PU channel access probability due to its evolutionary mechanism which converges to an optimal spectrum access as explained in Section 4.3. The observed advantages are because of adaptive use of channels in order to utilize spectrum holes by using entropy information and efficiently distributing the acquired resources.

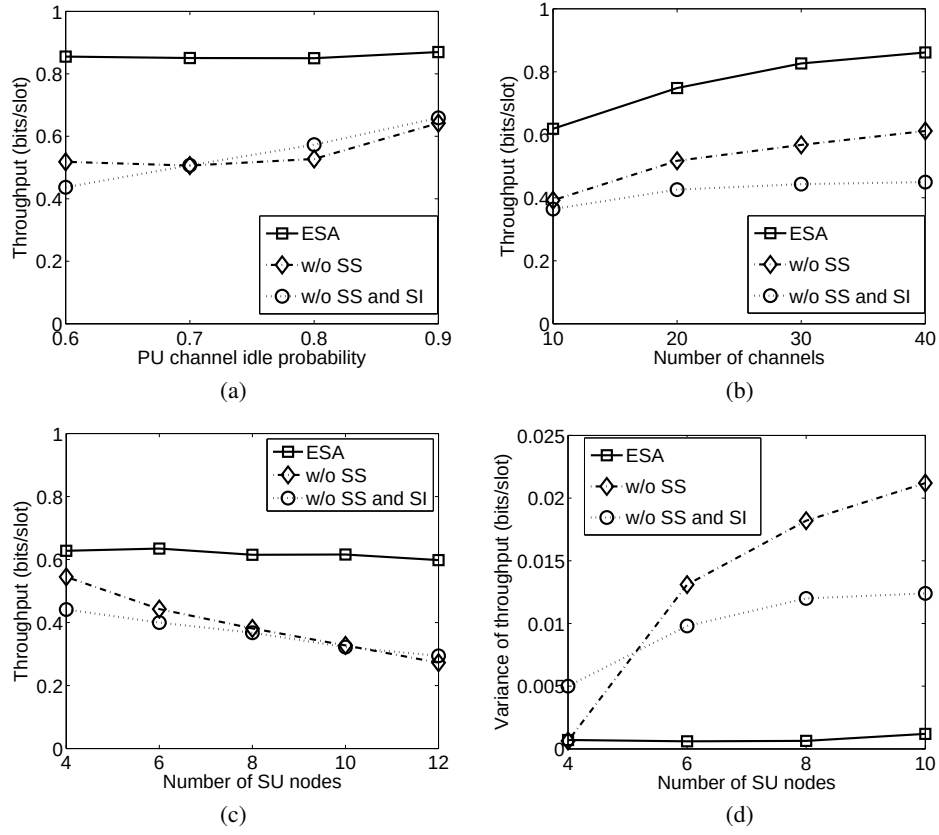


Figure 4.4: The throughput and the fairness of ESA algorithm compared with the algorithms making only SI and making SI and SS. The throughput for (a) varying primary user channel idle probability where $N = 40$, (b) varying number of channels used, (c) varying number of nodes in the network, and (d) the fairness for varying number of nodes.

4.4.2.2 Throughput vs. Number of available channels

Secondly, the effects of changing the number of channels on the throughput are observed. As it can be clearly seen in Fig. 4.4(b), overall throughput increases with the number of channels as expected. However, in all cases, the full utilization of the channel resources is best exploited by ESA algorithm with SI and SS incorporated. The reason is that with a balanced and adaptive use of resources overall throughput is leveraged.

4.4.2.3 Throughput vs. Number of SUs

Throughput performance of evolutionary and bio-inspired nature of ESA algorithm is observed by varying the number of nodes. In any biological process the self-adaptation achieves

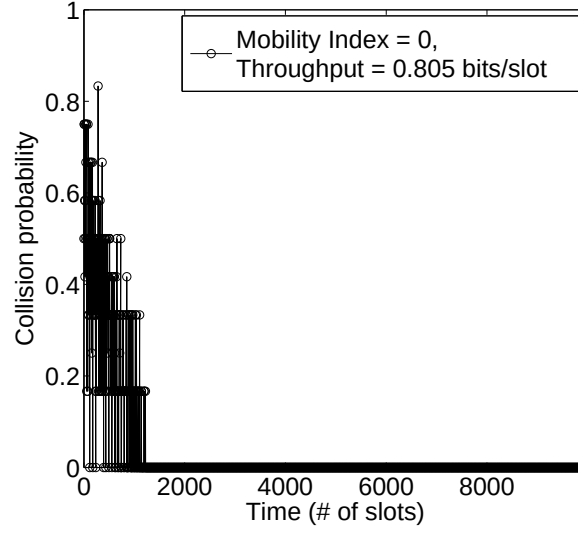


Figure 4.5: Collision probability, and the convergence of ESA algorithm.

efficient response to the dynamically varying size of the system autonomously. As it is observed in Fig. 4.4(c), ESA without SI and SS are severely affected by the increasing number of contending SUs. However, due to CSM, the collisions among the new SUs are mitigated, and the high throughput is maintained regardless of the increasing demand.

4.4.2.4 Fairness

Here, we investigate the fairness performance of ESA protocol in terms of opportunistic distribution of spectrum resources. The variance of the throughput achieved by individual SUs is used as the fairness metric. As shown in Fig. 4.4(d), if SS and SI were not used, the throughput variance significantly increases with the number of SUs. As also outlined in Table 4.2, this is, in fact, the case of majority of the existing OSA mechanisms in the literature. However, ESA protocol, with its immune system-inspired SI and SS mechanisms, maintains relatively constant and very low throughput variance. This reveals that ESA provides fair share of spectrum as well.

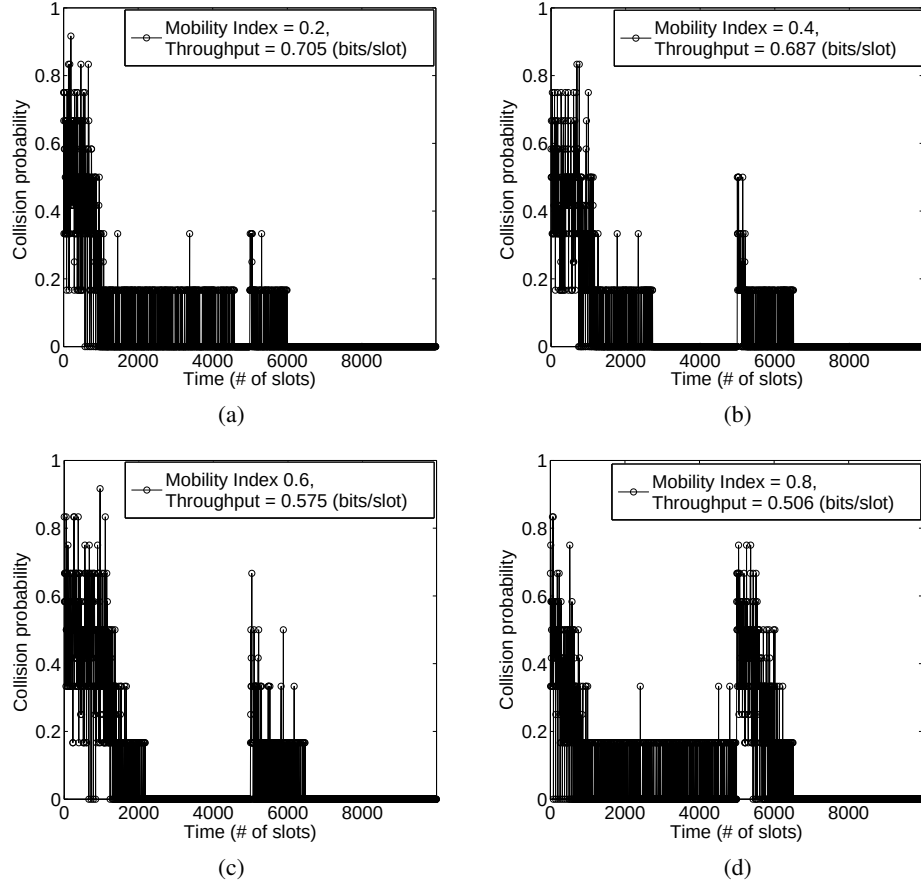


Figure 4.6: The throughput performance with respect to the effect of mobility over time. (a) Mobility index = 0.2 (b) Mobility index = 0.4 (c) Mobility index = 0.6 (d) Mobility index = 0.8.

4.4.2.5 Collision Probability

In Fig. 4.5, overall collision probability is observed during a course of ESA operation. As explained in Section 4.3, ESA enables receiver SUs to hypermutate their FHS to optimally share the available spectrum bands with the minimum collision probability. This optimal point can be easily observed in Fig. 4.5 such that ESA completely avoids all possible collisions and reaches to the optimal point in which SUs no longer observe collisions. Note also that in Fig. 4.6, all x-axes implicitly correspond to time since the duration of each time slot is fixed to a τ seconds. For example, τ is selected as 1ms, the convergence of ESA can be achieved within approximately 5 seconds. Hence, based on the evolutionary mutation mechanisms over FHS of SUs ESA significantly improves overall collision probability regardless of initial state of the network.

4.4.2.6 Node Mobility

Here, we investigate the effects of mobility in the CRAHNs on the performance of ESA protocol. To this end, we define an evaluation variable, i.e., *mobility index* representing the ratio of the mobile SUs in the network. Here, a number of SUs randomly change their locations with a random velocity. Hence, as the mobility index increases, the network becomes more mobile. In Fig. 4.6(a),(b),(c), and (d), the variation of the collision probability with time for the case of mobile nodes with a varying mobility index value is shown.

In Fig. 4.6(a), the network is initially set as immobile and the mobility index is set as 0 for a specific number of slots. In this interval, ESA completely avoids overall collision probability after approximately 5000 time slots. After the convergence, the 20% of the nodes start moving, i.e., the mobility index 0.2. At this point (approximately between slot 4000 and 5000), previous FHS are invalidated by the SU mobilities and collision arises. However, ESA can easily recover the network from the collisions by re-initiating evolutionary hypermutations as observed Fig. 4.6(a).

In Fig. 4.6(b), (c), and (d), the same immobile and mobile network states are consecutively set using the mobility index 0.4, 0.6, and 0.8, respectively. As the mobility index increases, the achieved throughput decreases, due to inevitable but temporary collision. However, in all cases, the ESA protocol operation converges to very low collision probability case regardless of the mobility level in the networks.

Moreover, an evolutionary feature of ESA can be observed in the convergence behavior in Fig. 4.6(b), (c), and (d). If the convergence time lasts for relatively a long time in the immobile phase of the network, the network converges more quickly after the mobile phase of the network. Conversely, if the convergence lasts for a small amount of time, it struggles for a relatively longer time to converge to a low collision state. This is an evolutionary mechanism such that the robustness against the undesirable situations such as node mobility and interference can be improved with more exposition to these undesirable situations for a relatively long time.

4.4.3 Error-tolerance to Inaccuracy in Spectrum Identification

Since spectrum sensing process is prone to false alarm and mis-detection probabilities, SI needs to be resiliently performed in order not to interfere PUs and to maximize spectrum utilization. Similar to error-tolerant self-nonself identification mechanism in immune system, ESA allows each SU to separately tolerate the identification errors by reinforcing the information about SI. Here, we analytically investigate this case using the Gaussian channel model.

When a SU senses a channel, the received signal can be evaluated as [77]

$$x(t) = \begin{cases} n(t), & H_0 \\ s(t) + n(t), & H_1 \end{cases} \quad (4.5)$$

where H_0 and H_1 are the hypotheses corresponding to idle and busy channel cases, respectively. $s(t)$ is the communication signal and $n(t)$ is a zero-mean Gaussian noise.

The output of the energy detector, i.e., Y , has the Chi-square distribution that can be approximated as a Gaussian distribution [77]

$$Y \approx \begin{cases} \mathcal{N}(n\sigma_n^2, 2n\sigma_n^4), & H_0 \\ \mathcal{N}(n(\sigma_n^2 + \sigma_s^2), 2n(\sigma_n^2 + \sigma_s^2)^2), & H_1 \end{cases} \quad (4.6)$$

where n is the number of samples, σ_s^2 and σ_n^2 are the variance of signal $s(t)$ and noise $n(t)$. If we assume that a PU accesses the sensed channel with probability p_a , using (4.6) false alarm p_f and mis-detection p_m probabilities can be given as

$$p_f = p_a \left(1 - Q \left(\frac{\lambda - n(\sigma_n^2 + \sigma_s^2)}{\sqrt{2n(\sigma_n^2 + \sigma_s^2)^2}} \right) \right) \quad (4.7)$$

$$p_m = (1 - p_a) Q \left(\frac{\lambda - n\sigma_n^2}{\sqrt{2n\sigma_n^4}} \right) \quad (4.8)$$

where $Q(\cdot)$ is the Q -function and λ is the detection threshold of the energy detector. In order to show the error tolerance of ESA to potential inaccuracies in spectrum identification process, in Fig. 4.7, we plot the spectrum identification entropy for a channel, i.e., H_i^j in (4.2), with respect to channel access probability of PUs for this channel, i.e., p_a . Despite increasing p_a and channel noise variance σ_s^2 , ESA provides the resilience SI. For the case in which p_a is high, H_i^j becomes also high. However, as p_a decreases, H_i^j becomes smaller. This allows each SU to separately identify the spectrum opportunities with smaller H_i^j in which

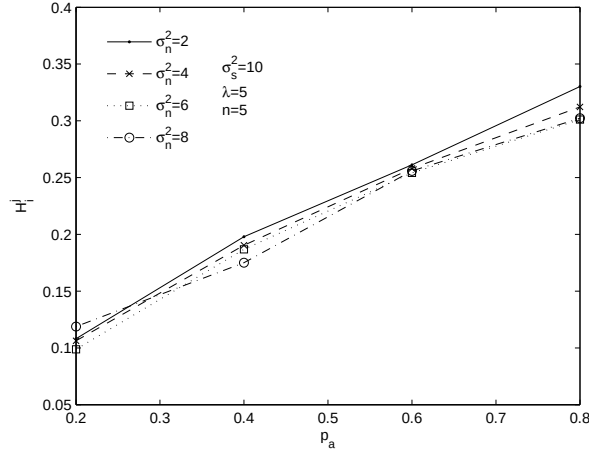


Figure 4.7: H_i^j with respect to varying p_a for different σ_s^2 to show the error tolerance capability of ESA.

PU activities are minor. Entropy-based SI is not affected by the probabilities of false alarm and mis-detection in channel sensing although noise variance increases. This shows the error tolerance capability of ESA to the inaccuracy in SI.

4.5 Conclusion

In this chapter, we introduced a new bio-inspired OSA mechanism, i.e., immune system-inspired evolutionary spectrum access (ESA), for CRAHNS. We have shown that immune system-inspired opportunistic access mechanism has many advantages coming naturally from the inspired process. The efficient, fair and high-level usage of resources are succeeded by ESA with the mechanisms of inspired self-nonself detection and clonal selection. These natural methods make the system self-organized, self-adapted and self-regulated resulting in an evolutionary OSA system. Furthermore, unlike the existing OSA mechanisms, ESA does not require a single dedicated control channel. The performance analysis results show that ESA provides more fair and effective utilization of communication channels where SUs try to use the channels in a cognitive sense without disturbing PU. In addition to this, performance of ESA is not affected by the high traffic load incurred by the nodes, instead, the system evolves faster with the presence of high contention and converges to an optimal state earlier especially in the case of high node mobility.

CHAPTER 5

Distributed Audio Sensing with Homeostasis-inspired Autonomous Communication

In this chapter, the homeostasis-inspired autonomous communication (HAC) protocol is devised for wireless audio sensor networks (WASN). HAC was introduced in [85]. The salient features of HAC are:

- HAC aims to communicate a sufficient number of data samples from the event signal to enable the exact reconstruction of the wide-band event signal with minimum energy expenditure.
- HAC aims to provide real-time communication of the wide-band event signal.
- HAC regulates its communication parameters according to the spectral characteristics of the event signal.
- HAC does not need any feedback message from the sink node.
- HAC allows the self-organization capability of the WASN, inspired by biological homeostasis, maintains the sensor network in a relatively stable operating state in which the sensor nodes autonomously, reliably and, yet energy-efficiently communicate wide-band event signals to the sink.

In Section 5.1, the motivation and related works are discussed. In Section 5.2, the network model and assumptions used in HAC protocol operation are introduced. In Section 5.3, biological homeostasis is reviewed and then, a connection between biological homeostasis and WASN is introduced. HAC algorithm is introduced in Section 5.4 based on the analogies be-

tween WASN and homeostasis given in Section 5.3. In Section 5.5, performance evaluation results of HAC are presented. Finally, concluding remarks are given in Section 5.6.

5.1 Motivation and Related Works

Most of the current wireless sensor network (WSN) applications do not have high bandwidth demands, and are usually delay-insensitive. More recently, the technological advances in CMOS cameras and microphones have fostered the development of wireless multimedia sensor networks (WMSN) for many emerging applications such as multimedia surveillance sensor networks, advanced health care delivery, traffic avoidance, and person locator services [86]. They will also increase the capabilities of traditional monitoring and surveillance systems by virtue of the distributed system of multiple cameras and sensors. Among these, wireless audio sensor networks (WASN), i.e., networks of wirelessly interconnected audio sensors, are also promising for the realization of efficient audio surveillance applications [87], [88].

Communication of wide-band event signal such as audio signal containing high frequency components requires high transmission bandwidth. Once the event signal is sensed by a number of sensor nodes within the event radius, significant amount of traffic is injected to the network, which may easily cause severe packet losses and quick depletion of scarce network resource. Therefore, in order to realize promising WASN applications, the additional communication challenges posed by the unique features of wide-band event signal, i.e., audio signal,¹ must be addressed. In the existing literature, there are many research efforts on reliable communication protocols that are devised for WSN domain and aim to efficiently communicate the event information to the sink [28], [29], [30]. Moreover, recently, a cross-layer rate control scheme is also introduced in order to fairly maximize the quality of the communicated multimedia streams by minimizing distortion [31]. However, the primary objective of these existing protocols is to achieve reliability and energy efficiency. They do not take the spectral characteristics of the event signal to be communicated into consideration in the determination of communication parameters. In fact, different event signal characteristics may need different treatment in terms of communication requirements such as different communication bandwidth demand and diverse reliability objectives. Furthermore, they do not address the communication of high-bandwidth event signals as in the case of WASN and do not perform

¹ In this chapter, we interchangeably use audio signal and event signal.

or consider an exact reconstruction of the event signal from the delivered samples rather than averaged samples at the sink node.

On the other hand, due to the lack of the autonomous operations in these protocols, reliable sensor communication partially or fully depend on feedback messages from the sink. However, the feedback may not reach in time to provide reliable communication of event information and it may lead to waste of scarce network resources. Hence, it is imperative to develop an autonomous communication protocol for WASN that enables the sensor nodes to cooperatively and effectively communicate the event signal without the need for any coordination with the sink. This necessitates the sensor nodes to have the capability of self-organization rather than the control of a central entity such as the sink node or base station. With this self-organizing feature, the sensor nodes are expected to reliably communicate their readings with minimum energy expenditure despite any communication problems such as packet losses, link errors, and excessive delays so that *the sensor network operates in a relatively stable state*.

Similar to WASN, a major functionality of many biological organisms is the ability to autonomously maintain a relatively stable equilibrium state for the operation of vital functions [89]. This functionality is called *homeostasis*, and it is the principal quality of an organism to conserve its autonomy. Hence, this analogy between biological homeostasis and WASN gives inspiration to develop autonomous communication algorithm for WASN.

5.2 Network Model and Assumptions

We consider a network architecture in which the number of N audio sensors are deployed in an environment and they separately detect and communicate the event information to the sink node. Sensor nodes are assumed to use a fixed transmission power level. Therefore, they have a fixed transmission range, i.e., r , in which they communicate with each other. When an event occurs in the environment, the average number of S audio sensors that sense the audio signal act as source nodes to sample the event signal and generate data packets. This initiates the data transmission from sensor nodes to the sink node. The source nodes are assumed to form the average number of H clusters each of which has a cluster head. Each cluster head aggregates the source node packets received from its members and directly transmits to the sink node in one hop.

We assume that each sensor node uses a number of different frequency channels. Each sensor node selects and switches its radio front-end to one of these frequency channels in order to transmit and receive data packets. The selection of frequency channels used for the intra-cluster communication of sensor nodes is assumed to be made by cluster head. As will be introduced, HAC protocol allows the neighbor clusters, that may interfere with each other, to use different frequency channels. This aims to avoid high level of interference among the neighbor clusters. For the communication between cluster heads and the sink node, a single unique frequency channel, that is not used in intra-cluster communication of sensor nodes, is assumed to be set initially and used by the cluster heads to transmit the aggregated data packet to the sink node. Thus, the sink node also use the same frequency channel with the cluster heads in order to receive data packets from the cluster heads.

The wireless channel is assumed to be shared in fixed-duration time slots, which are, in turn, captured by sensor nodes in order to transmit. The duration of a time slot consists of two different intervals named packet transmission and acknowledgment (ACK) transmission interval. In the packet transmission interval, each sensor node transmits a packet to its destination and an ACK frame is transmitted back by the destination if the packet is successfully received. If at least two or more sensor nodes, that are in the communication range of each other, transmit using the same time slot, their packets collide and the current transmission attempt become unsuccessful.

5.3 Biological Homeostasis and Wireless Audio Sensor Networks

In this section, the basic principles of biological homeostasis are first reviewed. Then, the analogies between biological homeostasis and WASN is presented, based on which HAC protocol is introduced.

5.3.1 Biological Homeostatic System

A vital functionality of many biological organisms is the ability to maintain a stable internal state although the environmental conditions may change rapidly [89]. This functionality is called *homeostasis*, and it is the leading feature of an organism to sustain its autonomy. By means of homeostasis, the organism self-regulates its growth and development, and maintains

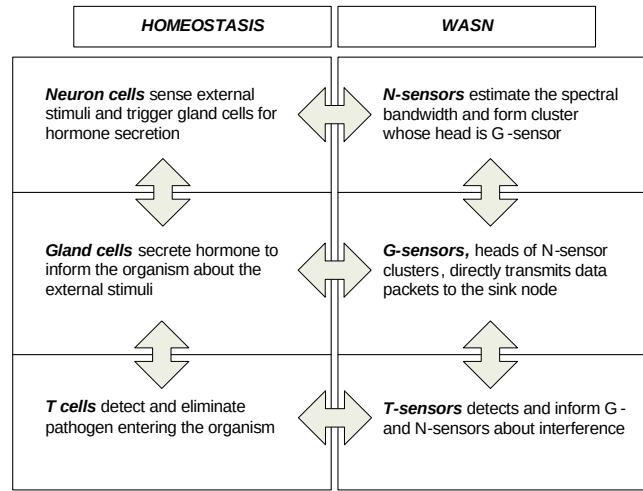


Figure 5.1: The mapping between the homeostatic system elements and sensor networks.

itself in the stable state. To maintain homeostatic stability within an organism, the nervous system, endocrine system, and immune system behave as one large, unified and complex system. The *interaction* and *communication* among all three systems are provided by the specific *receptors* on the cells [89].

A biological organism is open to various external stimuli. Neurons in the nervous system takes the stimuli, e.g., taste, smell, vision etc., via the sensory parts, and triggers an output reaction at the effectors, e.g., tissues and muscles. After the nervous system detects an input stimulus, the endocrine system produces and releases hormones through *gland cells*. Thus, the interaction between the nervous and endocrine system is the homeostatic response behavior of the organism to sustain its stable internal state. Any *malfunction* that adversely affects the operation of the organism is detected by the *immune system* of the organism. The immune system is the defense mechanism for the maintenance of homeostatic stability. This system responds to foreign substances that are generally called *pathogens*. The response is applied through *B-cells* and *T-cells* of the adaptive immune system. B-cells and T-cells are white blood cells and have the capability to bind to and eliminate pathogenic material, i.e., *antigens*. In conclusion, there is a constant interaction among the neural, endocrine, and immune systems, and the collaboration of them provides a promising model for the construction and development of self-organizing, highly-functional, and adaptable intelligent systems.

5.3.2 Homeostasis-inspired Wireless Audio Sensor Networks

Similar to a biological homeostatic system, WASN must keep itself within a stable state for successful operation. *The stable state of a WASN corresponds to a network operating condition in which the sensor network autonomously, reliably, and yet energy-efficiently communicate a wide-band event signal at the sink despite any communication problems such as packet losses, link errors, and excessive delays.* In WASN, the spectral characteristics of event signal impose a sampling frequency rate constraint on source nodes to enable successful reconstruction of the audio signal according to Nyquist Sampling Theory [90]. Spectral characteristics determine the total number of samples to be transmitted per unit time over the network and hence, the traffic load over the forward paths. When the traffic load over the network is excessively high, this leads to possible collisions and packet losses on the forward paths. Thus, it is required to estimate the spectral bandwidth of the event signal for energy efficient and reliable communication of sensor measurements. Similarly, in biological homeostasis, the properties of an external stimuli detected by neuron cells determine the response of organism to this external stimuli. When an external stimuli is detected by neuron cells, according to the properties of the external stimuli such as kind and the magnitude, neuron cells trigger the endocrine system to initiate the hormone secretion. Based on this similarity, we consider source nodes that sense and sample the event signal as neurons in neural system and call them as *N-Sensor*. Like a neuron in neural system, in WASN, the aim of an N-Sensor is to sense and sample the event signal and estimate the spectrum of the sensed event signal.

In endocrine system, gland cells are stimulated by neurons to secrete hormones so as to keep the organism in the biologically stable state. Similarly, in WASN, source nodes, i.e., N-sensors, form cluster to transmit event information to the cluster head in order to stimulate the cluster head for data transmission. Hence, we consider the cluster heads of the N-sensor clusters as gland cells and call them *G-Sensors*. In immune system, T-cells sense any malfunction in the organism and trigger the neural and endocrine systems to take an appropriate action. Similarly, in WASN, some sensor nodes must detect and notify any malfunction such as collisions and channel error and their symptoms such as packet losses and excessive delay. With this regard, we consider these sensor nodes that detect any malfunction in WASN as T-cells and call them as *T-Sensors*. The analogies and the mappings between the homeostatic system elements and sensor networks are also outlined in Fig. 5.1. Note that each sensor node

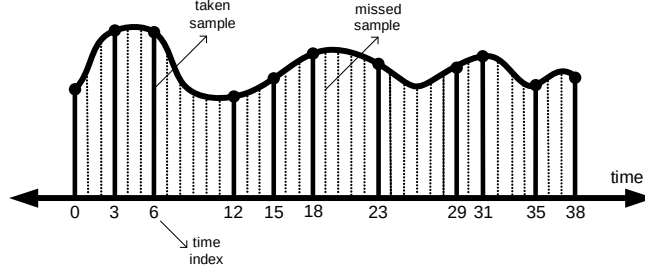


Figure 5.2: A portion of irregularly sampled event signal.

may jointly act as N-sensor, G-sensor, or T-sensor according to its current state such as source node and cluster head. Based on the analogies established in this section, in the next section, we present Homeostasis-inspired Autonomous Communication (HAC) algorithm.

5.4 Homeostasis-inspired Autonomous Communication in WASN

HAC algorithm has several unique features which allow sensor nodes to collaboratively and effectively communicate the event information based on the principles of biological homeostasis as described in Section 5.3. HAC takes the spectral characteristics of the event signal into consideration in determination of the sampling and communication parameters used by sensor nodes. Based on the spectral bandwidth of the event signal estimated by the N-Sensor, HAC enables G-sensors to determine the number of samples which must be delivered in order to enable accurate reconstruction at the sink. Furthermore, it also provides significant energy conservation by means of an efficient sampling scheme named irregular sampling [91]. In this scheme, each N-sensor takes samples from the event signal with a probability (α) called *sampling probability* while it samples the event signal with sampling frequency f . A portion of irregularly sampled event signal is illustrated in Fig. 5.2. Since the irregular sampling scheme decreases the number of samples transmitted over the network by a sensor node, it decreases excessive traffic load which may cause possible packet losses, and therefore, it provides more accurate reconstruction at the sink with significant energy conservation [91]. The overall HAC operation is composed of five distinct phases as briefly introduced below:

1. **Spectrum Estimation Phase:** N-Sensor estimates the spectrum bandwidth of the event signal, i.e., B , to determine how many samples must be taken and delivered to satisfac-

torily reconstruct the event signal at the sink. HAC algorithm uses a periodogram-based Welch Method [92] to estimate the spectral bandwidth of the event signal. The details of this procedure is given in Section 5.4.1.

2. **N-sensor Clustering Phase:** N-sensors form clusters according to their locational proximities with each other. Heads of the formed N-sensor clusters are selected as G-sensors to collect and directly transmit the data packets, received from N-sensors, to the sink as will be detailed in Section 5.4.2.
3. **Sampling Probability Determination Phase:** G-sensors, i.e., cluster heads of N-sensor clusters, determine the sampling and communication parameters of N-sensors to ensure that a sufficient number of samples can be timely delivered to the sink node. This phase is elaborated in Section 5.4.3.
4. **Channel Frequency Selection Phase:** In order to avoid excessive interference among neighbor N-sensor clusters, neighbor clusters select different operating frequency channels used for intra-cluster communication of sensor nodes. T-sensors detects frequency overlapping among neighbor clusters that use the same frequency channel and inform the clusters about that. Using this information, each cluster selects a different frequency channel in order to prevent excessive interference. This mechanism is explained in Section 5.4.4.
5. **Signal Reconstruction Phase:** The sink finally incorporates all samples of event signal, that are taken and delivered to the sink node, into a single sample set that is expected to include a sufficient number of samples that are needed for accurate reconstruction of the event signal. Detailed operation in this phase is introduced in Section 5.4.5. This phase completes the overall operation of HAC algorithm.

5.4.1 Spectrum Estimation Phase

N-Sensors that sense and sample the event signal estimate the spectral bandwidth of the event signal $f_S(t)$ and obtain the bandwidth B . For the estimation of the event signal bandwidth, the N-Sensor uses the periodogram-based Welch Method [92]. The details of the method is given in Appendix A.

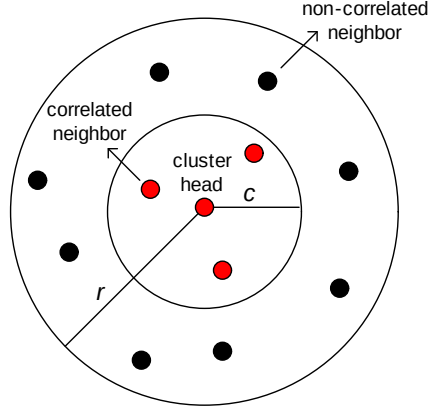


Figure 5.3: Illustration of correlated and non-correlated neighbors around a cluster head.

5.4.2 N-sensor Clustering Phase

HAC allows N-sensors that are located in close proximities of each other to form clusters by selecting a cluster head. HAC also ensures that members of clusters sense correlated event signal due to the locational proximities. Here, we characterize the correlation among N-sensors by means of the correlation radius, i.e., c . If the distance between two N-sensors is less than c , these N-sensors are assumed to sense correlated event signal [50]. Moreover, without loss of generality, we also assume that the correlation radius of sensor nodes is less than their communication radius, i.e., $c < r$. This assumption ensures that each sensor node to be able to communicate with its correlated and non-correlated neighbors within its communication radius as shown in Fig. 5.3.

Each N-sensor i decides whether or not to become a cluster head by sending *Request To Cluster* (RTC) message with a probability p at the beginning of each time slot. When N-sensor i sends a RTC message, N-sensors that are in the correlation radius of N-sensor i and receive the RTC message form a cluster whose head is N-sensor i . In order to provide quick formation of the clusters that is required for the real-time communication, it is essential to avoid possible collision of RTC messages. The optimal selection of p can provide an optimal performance in the cluster formation phase. HAC provides a nearly-optimal solution for the cluster formation phase as follows. We assume that the cluster formation is initiated with the detection of an event signal by N-sensors. We also assume that the event signal is detected in time slot 1 and N-sensor i has $n(k)$ neighbors that are neither a cluster head nor a member of a cluster at the k^{th} slot. In this case, the expected number of time slots during which N-sensor

i successfully transmits a RTC message without collision and becomes a cluster head, i.e., C , can be given as

$$C = \sum_{k=1}^{\infty} kp(1-p)^{n(k)} \left[1 - p(1-p)^{n(k)}\right]^{k-1} \quad (5.1)$$

The computation of C clearly depends on the statistical properties of $n(k)$ such as mean and variance. However, if a mean estimate for $n(k)$, i.e., $\hat{n}$, is available, an approximation for C , i.e., $\widehat{C}$, can be given as

$$\widehat{C} = \sum_{k=1}^{\infty} kp(1-p)^{\hat{n}} \left[1 - p(1-p)^{\hat{n}}\right]^{k-1} \quad (5.2)$$

In this case, $\widehat{C}$ can be computed as

$$\widehat{C} = \frac{1}{R} \quad (5.3)$$

where $R = p(1-p)^{\hat{n}}$ and $\widehat{C}$ is a strictly decreasing function of R . Thus, $\widehat{C}$ can be minimized by maximizing R . Since R is a concave function of p , it has a global maximum at the optimal p , i.e., p_{opt} , obtained by solving

$$\frac{\partial R}{\partial p} = (1-p)^{\hat{n}} - pn(1-p)^{\hat{n}-1} = 0 \quad (5.4)$$

Hence, the optimal RTC transmission probability, i.e., p_{opt} , that minimizes the average formation time of a cluster, i.e., $\widehat{C}$, can be given as

$$p_{opt} = \frac{1}{\hat{n}} \quad (5.5)$$

Surely, $\hat{n}$ has to be determined in order to find p_{opt} . However, $\hat{n}$ depends on some system parameters such as desired number of cluster heads and the average number of time slots during which overall clustering process is expected to be completed. If we assume that the overall clustering phase is completed within K slots, using the Markov inequality, an upper bound for the probability, that an N-sensor cannot become a cluster head, can be given as

$$Pr(C \geq K) \leq \frac{E[C]}{K} \quad (5.6)$$

By substituting $E[C] = \widehat{C}$ in (5.6), it can be rewritten as

$$Pr(C \geq K) \leq \frac{\widehat{C}}{K} \quad (5.7)$$

Using the upper bound in (5.7), it is expected that at least

$$H = S \left(1 - \frac{\widehat{C}}{K}\right) \quad (5.8)$$

number of N-sensors becomes cluster head at the end of an N-sensor clustering phase. If we assume that S , H , and K are known a priori by the system setting, $\widehat{C}$ can be computed as

$$\widehat{C} = K \left(1 - \frac{H}{S} \right) \quad (5.9)$$

By substituting $\widehat{C}$, computed in (5.3), (5.9) can be rewritten as

$$p(1-p)^{\widehat{n}} = \frac{S}{K(S-H)} \quad (5.10)$$

As presented above, the optimal p , i.e., $p_{opt} = \frac{1}{\widehat{n}}$, minimizes the expected number of slots in which an N-sensor becomes a cluster head, i.e., $\widehat{C}$. Thus, by setting $p = p_{opt}$, (5.10) can be written as

$$\frac{1}{\widehat{n}} \left(1 - \frac{1}{\widehat{n}} \right)^{\widehat{n}} = \frac{S}{K(S-H)} \quad (5.11)$$

HAC enables each N-sensor to determine $\widehat{n}$ by iteratively solving (5.11) that has two real roots for $0 < \frac{S}{K(S-H)} < 1$. One root of (5.11) is near 1 and the other is greater than 1. The root, that is nearly 1, i.e., $\widehat{n} \approx 1$, is not proper for our case because this makes $p_{opt} \approx 1$. However, the other root, that is greater than 1, i.e., $\widehat{n} > 1$, is a proper one. In Fig. 5.4(a), for different $\frac{S}{K(S-H)}$ values, the numerically computed $\widehat{n}$ values are shown. HAC provides a simple iterative procedure followed by all N-sensors to determine an approximate $\widehat{n}$ by iteratively testing (5.11) as follows

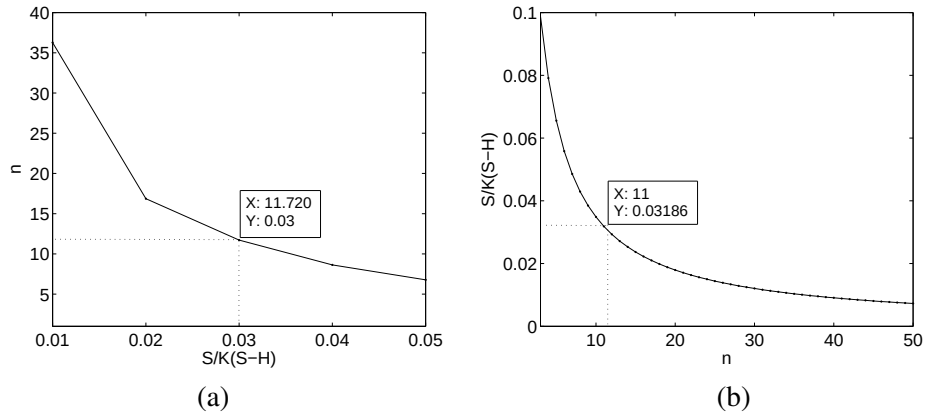


Figure 5.4: (a) $\widehat{n}$ is shown for different $\frac{S}{K(S-H)}$ values. (b) By substituting $\widehat{n} = 1, 2, \dots, b$, numerical values for the left side of (5.11) is shown.

- By substituting $\widehat{n} = 1, 2, \dots, b$, find the numerical value for the left side of (5.11) and collect them as $F_1, F_2, \dots, F_b$, respectively, where b is a sufficiently high upper bound for $\widehat{n}$ that is set a priori

- Compute the errors as $E_i = \left| F_i - \frac{S}{K(S-H)} \right|$ for $i = 1, 2, \dots, b$
- Find the minimum $E_i, i \in \{1, 2, \dots, b\}$ and set its index as $\widehat{n} = i$, where E_i is the minimum.

In Fig. 5.4(b), the above iterative procedure is illustrated to find $\widehat{n}$ for $\frac{S}{K(S-H)} = 0.03$. As observed in the data cursor, $\widehat{n}$ is found as 11, which gives a satisfactory approximation to $\frac{S}{K(S-H)} = 0.03$. The numerical solution of (5.11) is very close to the iteratively obtained solution. The numerical solution is equal to $\widehat{n} = 11.720$ while the iterative solution is $\widehat{n} = 11$. Consequently, HAC allows N-sensors to form a desired number of clusters within a desired cluster formation time. Moreover, despite the fact that cluster heads are severely energy consuming node, HAC also provides a uniformly distributed energy consumption for N-sensors by selecting cluster heads among all N-sensors with the same probability, i.e., p_{opt} .

5.4.3 Sampling Probability Determination Phase

As introduced in Section 5.4, HAC uses the irregular sampling scheme in which each N-sensor takes a sample from event signal with probability α while it samples the event signal with the sampling frequency f . Each cluster head of N-Sensor clusters acts as a G-sensor to determine α in its cluster. Using α , the probability that a specific sample of the event signal, i.e., $Sample_k$, is taken ($P[Sample_k \text{ is taken}]$) at least by one N-Sensor is expressed by

$$P[Sample_k \text{ is taken}] = \left[1 - (1 - \alpha)^M \right] \quad (5.12)$$

where M is the number of members in the cluster. If the sampling frequency determined by the cluster head, i.e., G-Sensor, is f (Hertz), then the total number of samples taken (in a second) is $f \cdot P[Sample_k \text{ is taken}]$. On the other hand, the expected number of the N-Sensors which take any sample, i.e., D , can be given as

$$D = \sum_{i=0}^M i \binom{M}{i} \alpha^i (1 - \alpha)^{M-i} \quad (5.13)$$

D is also the mean of the binomial distribution $B(M, \alpha)$, and can be expressed as $D = \alpha M$. Consequently, the average number of distinct samples delivered to the G-sensor, i.e., N_d , can be given as

$$N_d = (1 - \lambda^D) f \left[1 - (1 - \alpha)^M \right] \quad (5.14)$$

where λ is the packet loss probability observed in the cluster head, i.e., G-sensor. According to the Nyquist sampling theorem, if the event signal has bandwidth B^2 in *Hertz*, it is imperative for N-Sensors to take more than B number of distinct samples (N_d) within a second to satisfactorily reconstruct the event signal at the sink node, that is,

$$N_d > B \quad (5.15)$$

must be satisfied for the reconstruction. N_d is a function of α , f , M , and λ . Among these, λ is affected by many factors such as channel access scheme of N-sensors and signal-to-noise ratio in the sensor channels. HAC algorithm enables each G-sensor to optimize its channel access and the channel accesses of N-sensors in its cluster. For this optimization, G-sensors determine two channel access probabilities, i.e., p_a , and p_c , to provide real-time delivery for samples of event signal. p_a is the channel access probability that each N-sensor in the cluster transmits to the cluster head at the beginning of each time slot. Similarly, p_c is the channel access probability that each cluster head transmits to the sink node at the beginning of each time slot. Here, we assume that the packet losses may arise from possible collisions and channel errors. After a packet is exposed to a specific number of unsuccessful transmission attempts due to possible collisions and channel errors, it is assumed to be dropped. Possible collisions are affected by the channel access probabilities p_a and p_c while channel errors can be characterized by a packet failure probability p_f given as [93]

$$p_f = 1 - (1 - 0.5e^{\frac{-\gamma}{1.28}})^{8\kappa} \quad (5.16)$$

where γ is signal-to-noise ratio and κ is the number of samples³ in a packet. The packet losses in the network may be experienced in the links either between N-sensors and G-sensor or between G-sensors and the sink node. Thus, the packet loss probability λ given in (5.14) consists of the summation of two independent packet loss probabilities, i.e., λ_a and λ_c , that is, $\lambda = \lambda_a + \lambda_c$. λ_a is the packet loss probability observed in the link between N-sensors and the cluster head, i.e., G-sensor. Using p_a and p_f , the packet loss probability λ_a can be expressed as

$$\lambda_a = \left[1 - p_a(1 - p_f)(1 - p_a)^{M-1} - (1 - p_a)^M \right]^{z+1} \quad (5.17)$$

where z is allowable unsuccessful transmission attempts for a packet before it is dropped. Similarly, λ_c is the packet loss probability observed in the link between G-sensors and the

² Here, the spectral bandwidth of the event signal is defined as the twice of its maximum frequency component.

³ In this chapter, without loss of generality, we assume that each sample of the event signal is quantized by eight bits.

sink node and using p_c and p_f , it can be expressed as

$$\lambda_c = \left[1 - p_c(1 - p_f)(1 - p_c)^{H-1} - (1 - p_c)^H \right]^{z+1} \quad (5.18)$$

where H is the average number of cluster heads in the network as introduced in Section 5.4.2. Consequently, the overall packet loss probability λ can be expressed as $\lambda = \lambda_a + \lambda_c$. In addition to the packet loss probability, using p_a and p_c , expected number of time slots required for a successful delivery of a packet from an N-sensor to the cluster head and from the cluster head to the sink node, i.e., ξ_a and ξ_c , respectively, can be also given as

$$\xi_a = \sum_{i=0}^{\infty} i S_a (1 - S_a)^{i-1} = \frac{1}{p_a(1 - p_a)^{M-1}} \quad \text{and} \quad \xi_c = \sum_{i=0}^{\infty} i S_c (1 - S_c)^{i-1} = \frac{1}{p_c(1 - p_c)^{H-1}} \quad (5.19)$$

where S_a is the successful delivery probability of a packet transmitted from an N-sensor to the cluster head and given as $S_a = p_a(1 - p_a)^{M-1}$. Similarly, S_c is the successful delivery probability of a packet transmitted from a cluster head to the sink node and given as $S_c = p_c(1 - p_c)^{H-1}$. Hence, the expected number of time slots required for a successful delivery of a packet, i.e., ξ , can be given as the summation of ξ_a and ξ_c , i.e., $\xi = \xi_a + \xi_c$.

Here, we assume that ξ has to be less than or equal to the expected packet generation time of an N-sensor in order to provide real-time communication of event information. Each N-sensor takes sample with sampling probability α while it samples the event signal with sampling frequency f . Thus, each N-sensor takes $\alpha f \tau$ samples during a slot time, i.e., τ . Since a data packet is assumed to include κ samples from the event signal, the expected number of time slots during which a packet is generated by an N-sensor, i.e., η , can be given as

$$\eta = \frac{\kappa}{\alpha f \tau} \quad (5.20)$$

Hence, for the real-time communication of event information, it must be satisfied that

$$\xi = \xi_a + \xi_c \leq \eta \quad (5.21)$$

$$\frac{1}{p_a(1 - p_a)^{M-1}} + \frac{1}{p_c(1 - p_c)^{H-1}} \leq \frac{\kappa}{\alpha f \tau} \quad (5.22)$$

As observed in (5.22), the channel access probabilities p_a and p_c are implicitly a function of κ , τ , H , M , α , and f . Among these, κ , τ , H , and M , are assumed to be known while α and f are optimized by HAC protocol operation to satisfy the conditions given in (5.15) and (5.22). While doing this, HAC also tries to provide the minimum energy consumption. Energy consumption is mostly affected by the data traffic generated by N-sensors over the

network. More specifically, since each N-sensor takes the average number of αf samples in a second and each packet includes the same number of samples, αf determines the average reporting frequency rate of sensor nodes (*packets/sec.*) and thus, energy consumption rate over the network. Hence, in order to minimize energy consumption rate in the network, HAC aims to minimize αf . In fact, incorporating the conditions given in (5.15) and (5.22) into this objective, sampling probability determination phase of HAC can be modeled by an optimization process given as

$$\begin{aligned} & \textbf{minimize} && \alpha f \\ & \textbf{subject to} && N_d > B \quad \text{and} \quad \frac{1}{p_a(1-p_a)^{M-1}} + \frac{1}{p_c(1-p_c)^{H-1}} \leq \frac{\kappa}{\alpha f \tau} \end{aligned} \quad (5.23)$$

Using some specific relation among the parameters such as α , λ , and N_d , HAC allows G-sensors to iteratively solve the optimization problem modeled in (5.23) as follows. The successful packet delivery probabilities $S_a = p_a(1-p_a)^{M-1}$ and $S_c = p_c(1-p_c)^{H-1}$ are the concave function of p_a and p_c , respectively. Therefore, S_a and S_c have a global maximum at $p_a = \frac{1}{M}$ and $p_c = \frac{1}{H}$. Using the global maximum values of them, the condition required for the real-time communication and given in (5.22) can be rewritten as

$$\frac{1}{\frac{1}{M}\left(1 - \frac{1}{M}\right)^{M-1}} + \frac{1}{\frac{1}{H}\left(1 - \frac{1}{H}\right)^{H-1}} \leq \frac{\kappa}{\alpha f \tau} \quad (5.24)$$

By rearranging, (5.24) can be written as

$$\frac{\frac{\kappa}{\tau M H} \left(1 - \frac{1}{M}\right)^{M-1} \left(1 - \frac{1}{H}\right)^{H-1}}{\left[\left(1 - \frac{1}{M}\right)^{M-1} + \left(1 - \frac{1}{H}\right)^{H-1}\right]} \geq \alpha f \quad (5.25)$$

Clearly, (5.25) provides an upper bound for αf that is tried to be minimized by HAC protocol operation. f also has a lower bound at B since the event signal has the spectral bandwidth B according to Nyquist sampling theory. On the other hand, since p_a and p_c are selected from the interval $0 < p_a \leq \frac{1}{M}$ and $0 < p_c \leq \frac{1}{H}$, respectively, the packet loss probabilities, i.e., λ_a and λ_c are the strictly increasing function of p_a and p_c , respectively. Hence, substituting $p_a = \frac{1}{M}$ and $p_c = \frac{1}{H}$ into (5.17) and (5.18), respectively, the worst case packet loss probabilities, i.e., $\overline{\lambda}_a$ and $\overline{\lambda}_c$, can be expressed as

$$\overline{\lambda}_a = \left[1 - \frac{(1-p_f)}{M} \left(1 - \frac{1}{M}\right)^{M-1} - \left(1 - \frac{1}{M}\right)^M \right]^{z+1} \quad (5.26)$$

$$\overline{\lambda}_c = \left[1 - \frac{(1-p_f)}{H} \left(1 - \frac{1}{H}\right)^{H-1} - \left(1 - \frac{1}{H}\right)^H \right]^{z+1} \quad (5.27)$$

Combining $\bar{\lambda}_a$ and $\bar{\lambda}_c$, the overall worst case packet loss probability can be found as $\bar{\lambda} = \bar{\lambda}_a + \bar{\lambda}_c$. N_d is a strictly decreasing function of λ . Thus, using the worst case packet loss probability $\bar{\lambda}$, a worst case realization of N_d , i.e., $\bar{N}_d$, can be expressed as

$$\bar{N}_d = \left(1 - \bar{\lambda}^{M\alpha}\right) f \left(1 - (1 - \alpha)^M\right) \quad (5.28)$$

$\bar{N}_d$ is also a strictly increasing function of α for $0 < \alpha \leq 1$. In order to show this, the first derivative of $\bar{N}_d$ with respect to α , i.e., $\frac{\partial \bar{N}_d}{\partial \alpha}$, can be given as

$$\frac{\partial \bar{N}_d}{\partial \alpha} = M(1 - \alpha)^{M-1} \left(1 - \bar{\lambda}^{\alpha M}\right) - M \log(\bar{\lambda}) \left(1 - (1 - \alpha)^M \bar{\lambda}^{\alpha M}\right) \quad (5.29)$$

As observed in (5.29), $\frac{\partial \bar{N}_d}{\partial \alpha}$ is clearly positive for $0 < \alpha \leq 1$, which renders $\bar{N}_d$ strictly increasing function of α . Hence, $N_d > B$ is always satisfied if an appropriate α and f , that satisfy (5.25) and provides $\bar{N}_d > B$, can be found. Consequently, since $\bar{N}_d$ is a strictly increasing function of α , by using the upper bounds for α and f given in (5.25), a nearly optimal solution for the optimization problem given in (5.23) can be provided by HAC protocol operation as follows.

Each G-sensor initially computes the worst case packet loss probability $\bar{\lambda} = \bar{\lambda}_a + \bar{\lambda}_c$ using (5.26) and (5.27). Then, it selects the minimum sampling frequency f that satisfies (5.25) and $f > B$. Using the selected f , it iteratively searches for the minimum sampling probability α that provides $\bar{N}_d > B$ as follows. Since $\bar{N}_d$ is a strictly increasing function of α , this search can be performed by iteratively increasing α and testing whether or not $\bar{N}_d > B$ is satisfied. Here, we assume that starting with a sufficiently small value, α is increased by as $\alpha + \delta$ in each iteration, where δ is a sufficiently small step size.⁴ In the iterative search, if the current α provides $\bar{N}_d > B$, it is immediately selected as the sampling probability of N-sensors. If there is no such an α value satisfying the conditions, f is updated using a sufficiently small positive step size⁵ ζ as $f + \zeta$ and the same procedure is followed using the updated f until the minimum α that is in $0 < \alpha \leq 1$ and provides $\bar{N}_d > B$ is found. The sampling probability determination phase of HAC is also presented in Algorithm 3. After the determination of α and f , HAC also computes the minimum channel access probability p_a that satisfy the condition given in (5.22). The channel access probability of all cluster heads, i.e., p_c , is assumed to be set as $p_c = \frac{1}{H}$ by HAC protocol operation.

⁴ The initial value of α and the step size δ are set as 0.001 in the numerical analysis of this chapter.

⁵ ζ is set as $\zeta = 1$ in the numerical results.

Algorithm 3: HAC: Sampling Probability Determination Phase

```
1 foreach Each G-Sensor do
2   computes  $\bar{\lambda} = \bar{\lambda}_a + \bar{\lambda}_c$ 
3   selects the minimum  $f$  that satisfies (5.25) and  $f > B$ 
4   while  $\bar{N}_d < B$  do
5     updates  $\alpha$  as  $\alpha + \delta$ 
6   end
7   if an  $\alpha$  that provides  $\bar{N}_d > B$  can be found then
8     selects this  $\alpha$  as the sampling probability
9   else
10    updates  $f$  as  $f + \zeta$  and go to line 4
11  end
12 end
13 end
```

5.4.4 Channel Frequency Selection Phase

HAC enables each cluster of N-sensors to use different frequency channels in order not to interfere with each other. This can be achieved by means of feedback messages from T-sensors. T-sensors are the sensor node that are located at the overlapping areas of the N-sensor clusters. In Fig. 5.5, two clusters, i.e., cluster 1 and 2, are shown. They have an overlapping area in which there are four T-sensors. T-sensors in this region clearly receives N-sensor transmissions from both clusters if these clusters use the same frequency channel. Therefore, they recognize whether or not these two clusters use the same frequency channel. If they use the same channel, this imposes a high level of interference on the communication in these cluster. In order to avoid this high interference level, HAC enables T-sensors to change the operating frequency channel used in the neighbor clusters as follows.

Since T-sensors are also in the communication range of each other, they first determine two different frequency channel IDs and sends these IDs to the cluster heads. If the cluster heads accept these channels, they immediately switches to these new channels. Even if one of the cluster heads does not accept the new frequency channels, T-sensors find two new frequency channel IDs to send the cluster heads. The acceptance of the newly proposed frequency channels depends on whether or not cluster heads already have some neighbor clusters that use these channels. If such a situation is not arisen, the proposed frequency channels are immediately accepted by the cluster heads for the intra-cluster communication. Moreover,

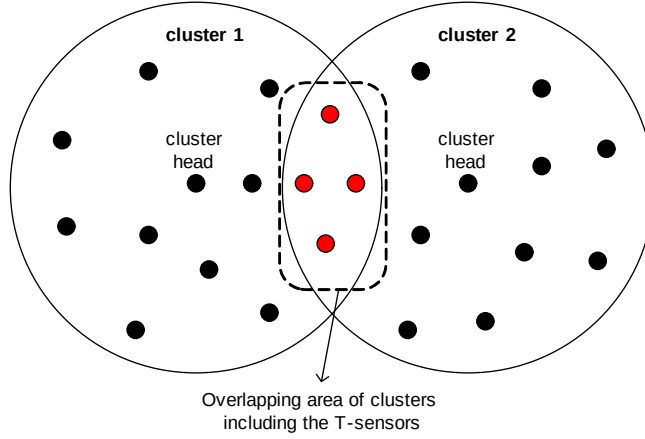


Figure 5.5: Two clusters, i.e., cluster 1 and cluster 2, with their overlapping area including a number of T-sensor.

once a cluster head switches to a proposed frequency channel, it does not change this channel anymore. This mechanism prevents a cluster head to uncertainly oscillate different frequency channels.

5.4.5 Signal Reconstruction Phase

In this phase, the sink node forms the complete sample set of the event signal $f_S(t)$ using the samples generated and transmitted by N-sensors using sampling frequency f and sampling probability α . Here, we assume that each N-sensor i also generates and transmit an indicator set I_i for its packets to inform the sink node about which samples are taken from $f_S(t)$ by it. I_i is a vector that includes 1 for the taken samples and includes 0 for the missed samples. Hence, using I_i , the sink node knows which samples are taken by N-sensor i during a sampling interval T (sec.) if the data packet including these samples and their indicator set is received. If I_i is not received with a data packet, indicator set of this packet is set as 0. Furthermore, I_i is simply transmitted as a $\lceil fT \rceil$ length bit stream because it includes a string of 0 and 1 bits. Therefore, the usage of I_i incurs a considerably low transmission overhead. In the sink node, the sample set generated and transmitted by N-sensor i , i.e., K_i , can be given as

$$K_i = \left[I_i(1)f_S(0), I_i(2)f_S\left(\frac{1}{f}\right), \dots, I_i(\lceil fT \rceil)f_S(T) \right] \quad (5.30)$$

The sink node combines the sampling sets of all N-sensors, i.e., $K_i, \forall i$, to generate a complete sample set of event signal, i.e., $\bar{K}$. Then, the sink node reconstructs the event signal $f_S(t)$ using the complete sample set $\bar{K}$. Here, the reconstruction method of the signal is not in the

Table 5.1: Evaluation Parameters

<i>Number of Sensors in the Network (N)</i>	100
<i>Number of Sensors in a Cluster (M)</i>	5-25
<i>Number of Clusters (H)</i>	4-10
<i>Event Signal Bandwidth (B)</i>	4-10 kHz
<i>Packet Size (κ)</i>	50 samples
<i>Slot Duration (τ)</i>	0.0001 (sec.)
<i>Packet Failure Probability (p_f)</i>	0.01

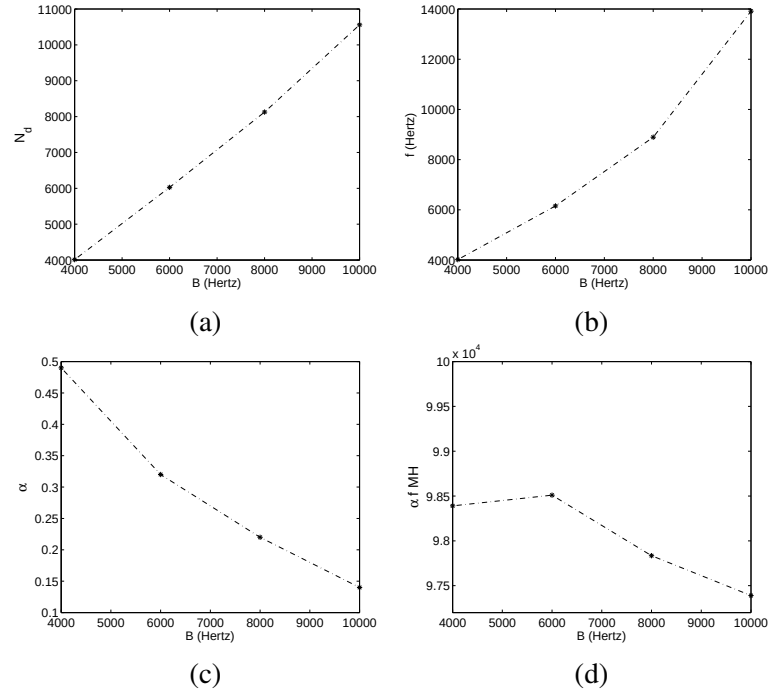


Figure 5.6: (a) The average number of distinct samples delivered within a second, i.e., N_d , is shown with the event signal bandwidth, i.e., B , values. (b) f is shown with B . (c) α is shown with B . (d) The total number of samples taken by all N -sensors during a second, i.e., $\alpha f MH$, is shown with B .

scope of this chapter. We only assume that the event signal can be reconstructed from the complete sample set of the event signal if a sufficient number of distinct samples can be taken and delivered to the sink by N -sensors.

5.5 Numerical Results

In this section, we numerically analyze the performance of HAC algorithm. We assume that the total number of N sensor nodes are deployed in a WASN environment and by means of the

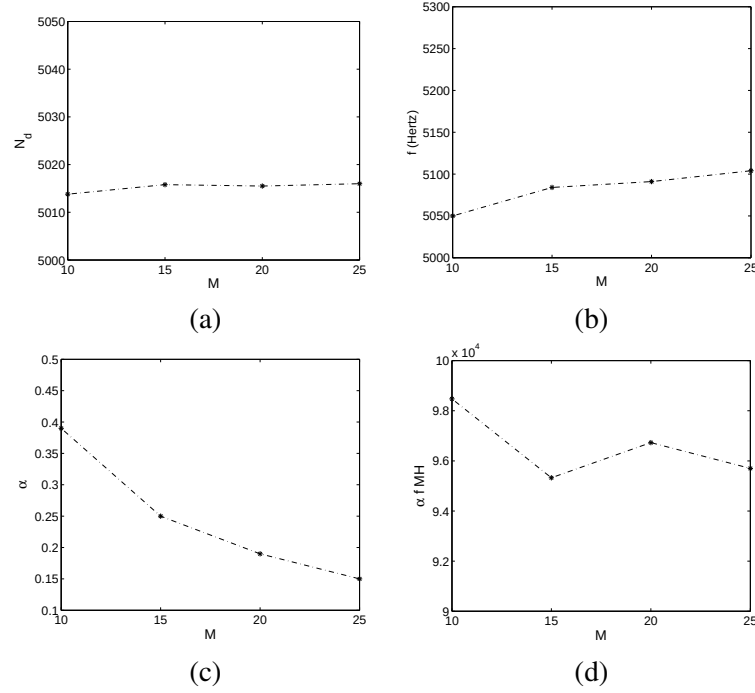


Figure 5.7: (a) The average number of distinct samples delivered within a second, i.e., N_d , is shown with the number of members in a cluster, i.e., M . (b) f is shown with M . (c) α is shown with M . (d) The total number of samples taken by all N-sensors during a second, i.e., $\alpha f M H$, is shown with M .

N-sensor clustering phase, N-sensors form H number of clusters whenever an event signal is detected in the environment. Each cluster is assumed to include M N-sensors. Moreover, the bandwidth of the event signal may change from 4 kHz to 10 kHz . The parameters used in the numerical results are given in Table 5.1. The analysis are evaluated by using MATLAB. The performance of the HAC algorithm is investigated by numerically evaluating the performance of *sampling probability determination phase*.

In Fig. 5.6, assuming that the number of $H = 5$ clusters each of which includes $M = 10$ N-sensors are formed in the N-sensor clustering phase, we change the event signal bandwidth from $B = 4 \text{ kHz}$ to $B = 10 \text{ kHz}$ in order to show the effect of B on the performance of HAC. In Fig. 5.6(a), the number of distinct samples delivered to the sink node within a second, i.e., N_d , is shown. As B increases, N_d is increased by HAC protocol operation such that N_d is slightly more than the event signal bandwidth B . This enables the exact reconstruction of event signal at the sink node while providing the delivery of minimum number of samples and thus, the minimum energy consumption of sensor nodes. Similarly, as B increases, f is increased by HAC protocol operation. For each event signal bandwidth B , HAC provides a

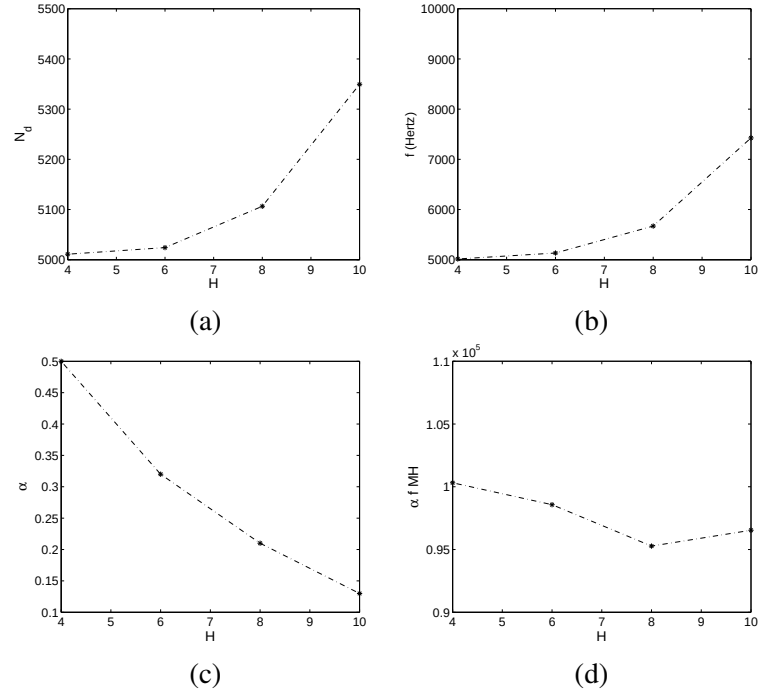


Figure 5.8: (a) The average number of distinct samples delivered within a second, i.e., N_d , is shown with the number of clusters in the network, i.e., H . (b) f is shown with H . (c) α is shown with H . (d) The total number of samples taken by all N-sensors during a second, i.e., $\alpha f M H$, is shown with H .

different f that is slightly more than B .

On the other hand, for the same case, as B increases, the sampling probability α is reduced by HAC protocol operation as shown in Fig. 5.6(c). This is because sampling frequency increases with the event signal bandwidth B . As f increases with B , HAC allows N-sensors to increase the number of distinct samples taken from the event signal by using lower sampling probabilities. This also causes a significant improvement in the energy conservation rate of the sensor nodes. In order to show the energy consumption rate of the sensor nodes, in Fig. 5.6(d), the expected number of samples taken by all N-sensors in the network within a second, i.e., $\alpha f M H$ is shown. As observed, $\alpha f M H$ slightly changes although the bandwidth B changes from $B = 4 \text{ kHz}$ to $B = 10 \text{ kHz}$. This can be achieved by decreasing the sampling probability α as seen in Fig. 5.6(c). This results shows the significant energy conservation capability of HAC.

Similarly, in Fig. 5.7, the effects of the number of members in the clusters, i.e., M , on the performance of HAC are shown. In Fig. 5.7(a), the number of distinct samples delivered in a

second, i.e., N_d , is shown with M . As M increases, the contention on the channel of the cluster head increases. However, HAC enables the network to deliver a sufficient number of samples that is slightly higher than the event signal bandwidth ($B = 5 \text{ kHz}$). The sampling frequencies (f) nearly over the bandwidth are also selected by HAC protocol operation as shown in Fig. 5.7(b). However, in contrary to N_d and f , the sampling frequency α is decreased by HAC as M increases. This stems from the increase in the diversity of the N-sensor samples such that as M increases, N-sensors can take sufficient number of samples, that is slightly more than B , using lower sampling probability. As observed in Fig. 5.7(d), by reducing α , HAC also maintain the average number of samples taken over the network during a second, i.e., $\alpha f M H$, in a nearly constant level regardless of the increase in the M . This provides a significant energy conservation for sensor nodes such that the energy consumption rate of the network remain almost the same regardless of the increase in M .

In Fig. 5.8, the effect of the average number of clusters in the network, i.e., H , on the performance of HAC is shown. N_d is shown with H in Fig. 5.8(a). While H increases from 2 to 10, which increases the number of N-sensors in the network 10 times, N_d slightly increases over 5000 that is the event signal bandwidth in *Hertz*. Similarly, in Fig. 5.8(d), the total number of samples taken by all N-sensors within a second also remain almost the same level, which gains significant energy conservation over the network. This provides a great stability that remains the data traffic in an almost the same level. This can be achieved by HAC protocol operation reducing the sampling probability α as shown in 5.8(c).

5.6 Conclusion

In this chapter, the homeostasis-inspired autonomous communication (HAC) protocol for wireless audio sensor networks (WASN) is introduced. Using the spectral properties of the event signal, the aim of HAC is to enable sensor nodes to achieve energy- and bandwidth-efficient communication of the event signal, i.e., audio signal, without any need for feedback messages from the sink node. Using the irregular sampling mechanism, HAC significantly improves the energy conservation performance of the network while it also ensures that a sufficient number samples needed for exact reconstruction according to Nyquist sampling theory is timely delivered to the sink node. Due to these unique features of HAC, it can provide an efficient communication solution for event signals with high frequency components.

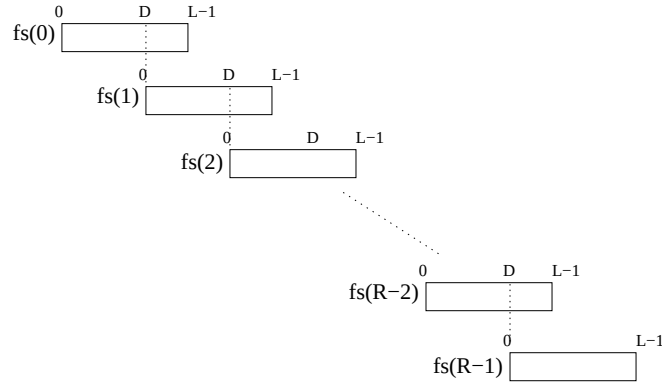


Figure 5.9: Overlapping Samples.

Moreover, HAC provides highly reliable and autonomous event-to-sink data communication based on the interactions of the network components inspired by the elements of a biological homeostatic system.

5.7 Appendix

Each N-Sensor estimates the spectral characteristics of the event signal $f_S(t)$ in the following way:

- It initially senses $f_S(t)$ and generates a sample set $f_S(0), f_S(1), \dots, f_S(R-1)$.
- It takes length L sample segments, having overlapping samples as shown in Fig. 5.7, from the sample set such that each segment has D units apart such that $D < L$. Let $f_S^1(j)$ be the first segment, then

$$f_S^1(j) = f_S(j), \quad j=0,1,\dots,L-1$$

$$f_S^2(j) = f_S(j+D), \quad j=0,1,\dots,L-1$$

$$\vdots$$

$$f_S^K(j) = f_S(j+(K-1)D), \quad j=0,1,\dots,L-1$$

where K denotes the number of segments. All samples in the segments cover the entire signal $((K-1)D + L = R)$.

- It selects a data window denoted as $W(j)$, $j = 0, 1, \dots, L - 1$.
- Using the data window $W(j)$, it forms the sequence $f_S^1(j)W(j), \dots, f_S^K(j)W(j)$. Then, it takes Fourier transform $A_1(n), \dots, A_K(n)$ of these sequences such that

$$A_k(n) = \frac{1}{L} \sum_{j=0}^{L-1} f_S^k(j)W(j)e^{-2Kijn/L}$$

where $i = \sqrt{-1}$.

- It obtains K modified periodograms such that

$$I_k(f_n) = \frac{L}{U} |A_k(n)|^2 \quad k = 1, 2, \dots, K$$

where $f_n = \frac{n}{L}$, $n = 0, \dots, \frac{L}{2}$ and $U = \frac{1}{L} \sum_{j=0}^{L-1} W(j)^2$.

- It computes the estimate of power spectral density of $f_S(t)$ as the average of these K periodograms ($I_k(f_n)$) as follows

$$\widehat{P}(f_n) = \frac{1}{K} \sum_{j=1}^K I_j(f_n). \quad (5.31)$$

- It detects the maximum frequency component f_{max} of $f_S(t)$, which satisfies $f_{max} < \frac{L}{2}$. Thus, the spectral bandwidth of $f_S(t)$ is equal to $B = 2f_{max}$.

To update the spectral information of the event signal, each N-Sensor repeats above spectrum estimation procedure in each time interval τ .

CHAPTER 6

Bio-inspired Cross-layer Communication and Coordination in Sensor and Vehicular Actor Networks

In this chapter, based on the prey model in foraging theory, we introduce the BIO-inspired Cross-layer (BIOX) communication and coordination protocol for wireless sensor and vehicular actor nodes (WSAN). BIOX was introduced in [94]. BIOX provides three different nature-inspired profitability measures called *next-hop selection*, *channel access*, and *task allocation profitability* for the sensor and actor nodes. The aim of these measures is to control the rate of gain in energy-efficiency, reliability, and stability of the network operations. Using the next-hop selection profitability measures, each sensor node selects its next hop node in an energy-efficient way. Sensor nodes also schedule their transmission by means of the channel access profitability to provide a reliable sensor-actor communication. Based on the task allocation profitability, each actor node selects and accomplishes a set of tasks associated with the event. BIOX is a fully autonomous algorithm and produces an optimal performance in prolonging network lifetime and reliable sensor-actor communication. Moreover, it also provides a stable allocation of available tasks in a way that each task is accomplished within a bounded time delay by the most appropriate actor node.

In Section 6.1, the motivation and related works are presented. In Section 6.2, we introduce the design principles of BIOX. In Section 6.3, we derive a bio-inspired cross-layer communication model for sensor-actor communication. The task allocation model of actor nodes is introduced in Section 6.4. In Section 6.5, we present the details of the BIOX protocol operation. In Section 6.6, performance evaluations are presented and the concluding remarks are given in Section 6.7.

6.1 Motivation and Related Works

Wireless sensor and actor networks (WSAN) consist of a number of communicating sensor and vehicular actor nodes¹ to perform distributed sensing and acting tasks for a wide range of applications such as tracking, surveillance, homeland security, and biological or chemical attack detection. In WSAN, sensor nodes gather information about the observed environment and transmit to vehicular actor nodes. Based on the information received from sensors, the actor nodes take decisions, and then, perform appropriate actions upon the sensed phenomenon. Sensor nodes are, in general, constrained devices in terms of battery lifetime, sensing, processing, and wireless communication capabilities. Actor nodes are relatively resource-rich nodes equipped with better processing and communication capabilities [95].

The realization of WSAN mandates addressing the challenges introduced by the coexistence of sensors and vehicular actors. Energy-efficient, timely, and reliable sensor-actor communication are the main challenges for WSAN. In particular, due to the energy scarcity in sensor batteries, sensor-actor communication must be performed with minimum and balanced energy consumption to prolong overall network lifetime. Sensor-actor communication must also provide minimum latency and packet loss rate to enable actor nodes to timely and reliably act upon the environment. Furthermore, cooperative control of actor nodes is also essential to realize an effective task allocation among the actor nodes.

In the existing literature, there are many studies on energy-efficient communication protocols for wireless sensor networks (WSN) and WSAN [32], [28], [29], [33], [34]. However, the major common drawback of these proposals is the lack of autonomy in the operations of the network nodes. This clearly imposes a tight coordination requirement on sensor communication since each sensor node must always wait for a control command from a central entity to schedule its transmission. These protocols also assume 100% reliable exchange of the feedback messages among sensor and sink or actor nodes to successfully carry out their operations. Clearly, this is not a realistic assumption, and hence, all protocol operations may easily fail in the case of an erroneous feedback from the sink or actor nodes. Furthermore, as the network size increases, the reliable and delay-aware communication may be hampered with excessive coordination and feedback message exchanges. Although these protocols aim

¹ Here, all actor nodes are assumed to be vehicular node. Hence, we interchangeably use actor node or vehicular actor node throughout the chapter to refer to the vehicular actor node.

to minimize energy consumption, they do not uniformly distribute the energy consumption over the sensor nodes. Non-uniform energy consumption results in depletion of mostly used sensor batteries, which, in turn, reduces overall network lifetime. Furthermore, none of the above algorithms considers the cross-layer interactions among MAC, routing, and transport layers, which would further improve energy-efficiency and reliability in sensor-actor communication.

A sub-optimal cross-layer solution is introduced in [96] to maximize the network lifetime in WSN. However, the network is assumed to be static and fully connected, and the cross-layer optimization procedure is accomplished over all active links by the sink node. This makes the optimization procedure too complex to timely react the dynamically occurring events with spatial and temporal uncertainty. Dynamic nature of WSN, however, renders such complex cross-layer optimization procedures infeasible. On the other hand, in order to exploit active mobility of actor nodes, a probabilistic route design mechanism is also given for mobile actor nodes. The given mechanism allows actor nodes to adaptively visit sensor locations so as to collect the sensor measurements with minimum latency and uniformly distributed energy consumption [97]. Furthermore, the mobility of actor nodes is also manipulated to mitigate congestion and severe energy consumption of the sensor nodes [98].

In addition, there is an emerging body of literature on actor coordination, recovery, and localization issues in WSN. For example, a bio-inspired coordination scheme is introduced in [99] for an effective task performance of actor nodes. A distributed actor recovery algorithm is given to recover the connectivity of the actor nodes in the case of an actor failure [100]. Furthermore, a localized service discovery algorithm is proposed for WSN [101]. In order to allow each sensor node to find its coarse-grain location, a scalable and energy-efficient training protocols are given for WSN [102]. The stability of such networked systems with a number of mobile agents is also investigated in [103], [104]. These works clearly provide quite novel insights for the realization of future WSN applications. However, in all of above works, entire communication and coordination functionalities are centrally controlled by the actor nodes. In other words, the network nodes do not have any autonomous decision-making mechanism and easily fail in the lack of a central entity that controls the entire network operations.

In nature, each biological entity of a colony has an autonomy to attain a decision-making

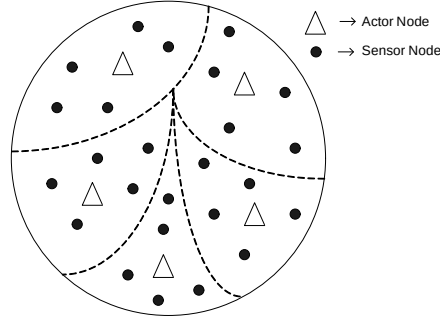


Figure 6.1: Illustration of an event area in WSN. In the environment, sensor nodes take measurements and forward the measurements towards vehicular actor nodes.

that may not hamper the overall stability of the colony despite the fact that it only accesses to a local information. For example, foraging strategy of a biological predator includes a significant autonomy in searching and handling prey. Based on a profitability measure, it trades-off searching and handling efforts to maximize its energy intake while examining its starvation risk. However, at the colony level, this mechanism also eventually leads to an equilibrium point, where each entity has the same consumption rate. Similarly, in WSN, each network node needs an autonomous decision-making mechanism to efficiently figure out its communication and task allocation strategy without hampering overall operations at the network level. Clearly, an analogy can be drawn from foraging strategies in biology toward the development of efficient and autonomous communication and coordination protocols for the WSN domain.

6.2 BIOX Protocol Design Principles

In this section, we first describe the network model and assumptions and formulate the cross-layer optimization problem. Then, we review a *prey model in foraging theory*, based on which we develop a bio-inspired cross-layer communication and coordination model for WSN.

6.2.1 Network Model and Problem Formulation

We consider a network architecture in which the number of N sensor nodes are deployed in an environment. Sensor nodes separately detect and transmit the event information to M vehicular actor nodes as shown in Fig. 6.1. Sensor nodes use a fixed transmission power level.

Therefore, they have a fixed transmission range, i.e., r , in which they communicate with each other. When an event occurs in the environment, a number of sensor nodes, which sense the event signal, act as source nodes to sample the event signal and generate data frames². According to spectral properties of the event signal, each source node sample the event signal and transmits the generated data frames with an average reporting frequency, i.e., f ($frames/s$).

Sensor nodes participating in data transmission and/or reception are called active sensor nodes. If a sensor node receives a data frame, it selects one of its neighbors toward the actor node as its next hop node and transmits. Each data path is originated from a source node, and ends with an actor node. $\mathcal{Y}$ is the set of all paths from source nodes to actor nodes. Each path also consists of a set of directed arcs. For example, (i, j) denotes a directed arc from sensor node i to sensor node j .

The wireless channel is assumed to be shared in fixed-duration time slots, which are, in turn, captured by sensor nodes in order not to interfere with each other. Using a time slot, a sensor node either transmits or receives a data frame. If at least two or more sensors, that are in the communication radius of each other, transmit or receive using the same slot, the current transmission or reception attempt becomes unsuccessful. The duration of a time slot consists of two intervals named frame transmission and acknowledgment (ACK) interval as depicted in Fig. 6.2.

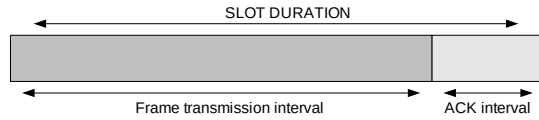


Figure 6.2: A slot duration consisting of two different intervals named frame transmission and ACK interval.

In the frame transmission interval, sensor nodes transmit the frame header and payload. The frame header includes the ID of source node that initially generates the current packet. This enables each sensor node to infer how many source node it serves to route toward the actor node. If sensor node i is used by the number of s_i source nodes to route their packets, it should achieve an average frame transmission rate, i.e., η_i ($frames/s$), that is greater than or equal to $s_i f$, i.e., $s_i f < \eta_i$, where f is the average reporting frequency rate of each source node.

² Here, we assume that each data packet is split to a number of data frames and transmitted frame by frame. Hence, the reporting frequency of the source nodes is given in ($frames/s$).

This guarantees that there is no packet loss due to possible congestion on sensor node i . In the ACK interval, each sensor node transmits an ACK frame whenever it successfully receives a data frame.

The network of the actor nodes is assumed to be fully connected. The aim of the actor-actor communication is to make data exchange and localization³ [105]. This communication is also expected to lead to a consensus among actor nodes for task allocation as introduced in Section 6.4.

6.2.1.1 Network Lifetime

Network lifetime is considered as the elapsed time until when a node runs out of its energy. Let L_i be the lifetime of sensor node i . Hence, the network lifetime can be expressed as $L_{nw} = \min_{i \in \{1, \dots, N\}} L_i$. Although lifetime maximization stands as a strict optimization problem that tackles the maximization of L_{nw} , there are several alternative ways to optimize the lifetime [106]. As will be introduced in the following sections, in order to maximize the network lifetime, BIOX aims to increase the ratio of the total residual energies in active sensors to their energy consumption. This can clearly prolong the network lifetime by simultaneously maximizing the total residual energies of active sensors and minimizing their energy consumption.

6.2.1.2 Minimization of Energy Consumption

Sensor nodes initially have the same level of energy, i.e., E , and each sensor node i periodically measures its residual energy level E_i . Sensor node i is assumed to consume $H(d_{ij})$ joule of energy to receive and transmit a bit to sensor node j . $H(d_{ij})$ is given as $H(d_{ij}) = u + v d_{ij}^\alpha$, where α is a positive constant that characterizes power loss and $\alpha \geq 2$ such that $\alpha = 2$ for free space and $\alpha = 4$ for multi-path fading channel. u is the dissipated energy in the transmitter and receiver circuitries. v is the energy needed by transmitter amplifier and d_{ij} denotes the Euclidean distance between sensor node i and sensor node j [107].

For a sensor node, transmission cost, i.e., level of dissipated energy, heavily depends on its hop distance (d_{ij}). Furthermore, overall energy consumption along a path is also affected by

³ We assume that actor nodes periodically discover their locations using one of the existing localization methods in the literature [105]. However, the details of localization procedure is beyond the scope of this chapter.

the effective hop distances (e_{ij}) that is the absolute Euclidean distance towards the destination actor by a hop as shown in Fig. 6.3. Clearly, the hop distance d_{ij} and the effective hop distance e_{ij} are the main determinants in order to discover an efficient next-hop selection strategy that provides the minimum energy consumption for sensor nodes. The total energy consumption

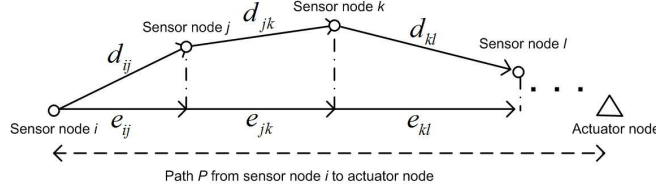


Figure 6.3: A data path from sensor node i to the actor node. Hop distances ($d_{ij}, \forall i, j$) and effective hop distances ($e_{ij}, \forall i, j$) are shown in the path.

throughout the path P in Fig. 6.3, i.e., C_P , is given as

$$C_P = \sum_{(i,j) \in P} H(d_{ij}) \quad (6.1)$$

where $(i, j) \in P$ means that the directed arc (i, j) between sensor node i and j is on the path P . Furthermore, the Euclidean distance D_P of path P can be also given by adding the effective hop distances (e_{ij}) as

$$D_P = \sum_{(i,j) \in P} e_{ij} \quad (6.2)$$

Using the log-sum inequality, C_P and D_P can be also incorporated so as to derive an upper bound for C_P as

$$\sum_{(i,j) \in P} H(d_{ij}) \log \frac{H(d_{ij})}{e_{ij}} \geq \left(\sum_{(i,j) \in P} H(d_{ij}) \right) \log \frac{C_P}{D_P} \quad (6.3)$$

By modifying (6.3), it can be also rewritten as

$$\log D_P + \frac{\sum_{(i,j) \in P} H(d_{ij}) \log \frac{H(d_{ij})}{e_{ij}}}{\sum_{(i,j) \in P} H(d_{ij})} \geq \log C_P \quad (6.4)$$

Assume that $\frac{H(d_{ij})}{e_{ij}}$ is a random variable. Then, its expected value, i.e., $\gamma_P = E \left[\frac{H(d_{ij})}{e_{ij}} \right]$, can be expressed as

$$\gamma_P = \sum_{(i,j) \in P} \sum_{j \in n_i} p_{ij} \frac{H(d_{ij})}{e_{ij}} \quad (6.5)$$

where p_{ij} is the probability that sensor node i transmits to sensor node j . n_i is the set including all node IDs in the communication range of sensor node i . By replacing $\frac{H(d_{ij})}{e_{ij}}$ with $\gamma_P =$

$E \left[\frac{H(d_{ij})}{e_{ij}} \right]$, (6.4) can be also approximated as

$$\log D_P + \frac{\log \gamma_P \sum_{(i,j) \in P} H(d_{ij})}{\sum_{(i,j) \in P} H(d_{ij})} \geq \log C_P \quad (6.6)$$

By further simplifying (6.6), an upper bound for C_P can be found as

$$\begin{aligned} \log D_P + \log \gamma_P &\geq \log C_P \\ \log D_P \gamma_P &\geq \log C_P \\ D_P \gamma_P &\geq C_P \end{aligned} \quad (6.7)$$

By using intermediate sensor nodes, source nodes are assumed to deliver their readings to a master actor node that is selected among all actor nodes. Note that the selection of the master actor node is presented in Section 6.5.1. The master actor node does not move during the data forwarding from source to actor nodes. Therefore, the Euclidean distance D_P between source and the master actor node is constant during the data forwarding and C_P can be reduced by decreasing γ_P regardless of D_P . On the other hand, the minimization of γ_P can be only achieved by finding the optimal p_{ij} values in (6.5). Thus, this clearly turns the minimization of C_P into the determination of the optimal p_{ij} values so as to provide the minimum γ_P .

In addition to the minimization of C_P , it is also imperative to provide balanced energy consumption over the network nodes in order to maximize the network lifetime. For this aim, sensor nodes having high residual energy should be mostly activated. Therefore, along the path P , if sensor node j has a high level of residual energy, each sensor node i maximizes p_{ij} in order to mostly transmit to sensor node j . This strategy clearly maximizes the expected total residual energy of active sensor nodes, i.e., β_P , that can be expressed as

$$\beta_P = \sum_{(i,j) \in P} \sum_{j \in n_i} p_{ij} E_j \quad (6.8)$$

where E_j is the residual energy of sensor node j . Briefly, the maximization of (6.8) allows the sensor nodes with high level of residual energy to be activated, which eventually provides balanced energy consumption throughout the network.

6.2.1.3 Cross-layer Optimization Problem

Based on the above definitions and notations, the cross-layer optimization problem to ensure energy-efficient and reliable sensor-actor communication can be given as follows.

- For lifetime optimization, it is imperative for each path $P \in \mathcal{Y}$
 - to minimize the energy consumption of the active sensor nodes by minimizing γ_P in (6.5) and
 - to mostly activate the sensor nodes having high residual energies by maximizing β_P in (6.8). These objectives can be achieved by easily maximizing the ratio J_P of β_P to γ_P , i.e., $J_P = \frac{\beta_P}{\gamma_P}$.
- For reliable sensor-actor communication,
 - the average frame transmission rate of sensor node i , i.e., η_i , must be greater than the average frame rate, i.e., $s_i f$, transmitted to sensor node i . This clearly entails satisfying a condition, i.e., $s_i f < \eta_i, \forall i$ in order to prevent possible congestion and packet losses on sensor node i .

Using the above requirements, the cross-layer optimization problem can be formulated as follows:

$$\textbf{maximize} \quad J_P, \forall P \in \mathcal{Y} \quad (6.9)$$

$$\textbf{subject to} \quad s_i f < \eta_i, \forall i \quad (6.10)$$

where J_P is given as

$$J_P = \frac{\beta_P}{\gamma_P} = \frac{\sum_{(i,j) \in P} \sum_{j \in n_i} p_{ij} E_j}{\sum_{(i,j) \in P} \sum_{j \in n_i} p_{ij} \frac{H(d_{ij})}{e_{ij}}} \quad (6.11)$$

The cross-layer optimization problem formulated in (6.9) and (6.10) can be solved for some reduced sub-optimal cases using well-known optimization methods [108] to obtain optimal route and link schedule of sensor nodes. All of the possible traditional solutions either need a central entity to centrally tackle with the optimization or need a network level coordination requirement among sensor nodes. However, all of these traditional solutions continuously require many network parameters and topology information that dynamically change through-

out the network. This severely imposes a latency and energy overhead on the network along with the harsh scalability and security problems.

In nature, biological networks such as immune network and insect colony have incredible capabilities to optimize many vital problems to survive in continuously changing environments. These capabilities also provide significant scalability and autonomy for biological networks. However, such a remarkable behavior of the biological networks is accomplished despite the fact that each individual can access to limited local information instead of global network information [67], [109]. These capabilities can also be modeled and adopted to characterize and analyze coordination and control of networked control systems such as WSN. In this chapter, inspired by a prey model in foraging theory, we propose an optimal solution for the cross-layer optimization problem formulated in (6.9) and (6.10). To this end, we next review the prey model in foraging theory.

6.2.2 Prey Model in Foraging Theory

Foraging theory is a field in behavioral ecology to mathematically describe the models based on which the foraging animals search for nutrients and choose which ones to consume [110], [111], [112]. The fundamental claim in foraging theory is that animals search for and obtain nutrients by maximizing their energy intake E per unit time T . This strategy is mathematically characterized by the maximization of an objective function of $\frac{E}{T}$ [112]. Clearly, maximization of such a function provides nutrient sources and additional time for foraging animals to accomplish other important activities such as reproducing, fighting, or shelter building. Foraging strategies vary from species to species, which causes animals with poor foraging strategies to be eliminated while promoting the animals with successful foraging strategies. This mechanism eventually leads the animals having significant foraging strategies to grow up. This evolutionary principle enables the foraging strategies of animals to be adopted and modeled as an optimization process that is referred to as *optimal foraging theory* [112]. The aim of the optimization is to allow a forager to maximize its energy intake based on interaction with its environment. For example, if there are different types of preys in the environment, it is essential for a forager to decide which types should be attacked to maximize overall energy intake.

One of the classical foraging models is the prey model. The prey model describes a forager

searching for different prey types in a particular environment⁴. Each prey holds a certain energy intake. The forager must search for and recognize a prey for the energy intake. Once it encounters and recognizes a prey, it decides whether to handle it [110]. Assume that there are k different types of prey in the environment. t_i is the expected time required to handle prey type i and v_i is the expected amount of energy intake obtained from handling prey i . The average rate of encounter with prey i in searching process is λ_i . p_i is the probability that prey type i is handled once it is found and recognized. Hence, the average rate of gain of a forager, i.e., J , can be defined as a ratio of total expected energy intake to expected amount of time spent for searching and handling the prey types. Hence, J can be expressed as

$$J = \frac{\sum_{i=1}^k p_i \lambda_i v_i}{\sum_{i=1}^k p_i \lambda_i t_i} \quad (6.12)$$

The prey model allows a forager to maximize J by finding the optimal values of $p_i, \forall i$. Suppose that a forager has previously handled the number of $n - 1$ prey types, i.e., prey 1, prey 2, $\dots$ prey $(n - 1)$. In this case, the rate of gain, J , can be given as $\sum_{i=1}^{n-1} p_i \lambda_i v_i / \sum_{i=1}^{n-1} p_i \lambda_i t_i$. Assume that the forager currently encounters prey type n . In this case, the rate of the gain of the forager becomes

$$J = \frac{p_n \lambda_n v_n + \sum_{i=1}^{n-1} p_i \lambda_i v_i}{p_n \lambda_n t_n + \sum_{i=1}^{n-1} p_i \lambda_i t_i} \quad (6.13)$$

In order to find the optimal p_n , J can be differentiated with respect to p_n and by considering $\frac{\partial J}{\partial p_n} = 0$, the optimal p_n that maximizes J can be found by the following zero-one-rule [110], i.e.,

$$\text{set } p_n = 0 \quad \text{if } \frac{v_n}{t_n} < \frac{\sum_{i=1}^{n-1} \lambda_i v_i}{\sum_{i=1}^{n-1} \lambda_i t_i} \quad \text{and} \quad \text{set } p_n = 1 \quad \text{if } \frac{v_n}{t_n} > \frac{\sum_{i=1}^{n-1} \lambda_i v_i}{\sum_{i=1}^{n-1} \lambda_i t_i} \quad (6.14)$$

where $\frac{v_n}{t_n}$ is called *prey selection profitability* and used to characterize the profitability of prey n that the forager obtains by handling it. Using the above zero-one-rule, each forager can autonomously decide whether to handle a prey whenever it encounters.

⁴ In the original version of the prey model, agent and task are used instead of forager and prey, respectively. Here, without loss of generality, we use forager and prey to constitute a bio-inspired analogy between prey model and WSN.

Table 6.1: Relationship between Prey Model and WSAN.

PREY MODEL	WSAN
<i>Forager</i>	<i>Sensor Node & Vehicular Actor Node</i>
<i>Prey</i>	<i>Prey for Sensor Node: Possible next hop nodes and available time slots & Prey for Vehicular Actor Node: Available tasks</i>

In the prey model, there is a natural uncertainty in geographical availability of the prey. Therefore, the forager has always a starvation risk if it cannot find a sufficient amount of prey to obtain adequate energy intake. However, as a result of natural evolution, biological foragers have acquired great capabilities to effectively tackle with the starvation risk. The aim of these capabilities is to allow each forager to balance its search effort and average energy intake. If it depletes most of its residual energy in prey search and it cannot obtain a sufficient energy intake by capturing and handling the prey, it changes its search strategy to mitigate the starvation risk by slowing down its energy consumption rate in the prey search [113]. The interaction between foragers and their prey often constitutes a *foraging game* in which forager individuals balance conflicting demands of prey and starvation [114]. The applicability of foraging theory to many engineering problems is carried out by viewing a forager as an agent and viewing prey as tasks that must be processed [115], [116], [117].

6.2.3 Inspiration from Prey Model

Clearly, the foraging model reviewed here establishes a solid basis for design of solution approaches to virtually any kind of control and optimization problem [67]. Here, in order to establish a bio-inspired coordination and communication model for WSAN, we consider sensor and actor nodes as forager. Similar to a biological forager searching for prey, each sensor node searches for possible next hop node and available time slots to forward its packets toward actor node. Each actor also searches for and performs available tasks associated with the event. Hence, for a sensor node, its possible next hop nodes and available time slots are considered as prey in the prey model. Available tasks are also considered as prey types that are searched and performed by the actor nodes. Hence, similar to the prey model, we define three profitability measures named as *the next hop selection*, *channel access*, and *task allocation profitability*. Using the next hop selection profitability, each sensor node selects its next hop

node. Based on the channel access profitability, each sensor node determines its transmission strategy. Vehicular actor nodes use the task allocation profitability to share and perform the available tasks. The analogies and relationships between WSA and the prey model are also outlined in Table 6.1. In the following sections, we apply the prey model to devise an efficient cross-layer communication and coordination model for WSA.

6.3 Bio-inspired Cross-layer Sensor-actor Communication Model

In this section, we introduce the next hop selection profitability and channel access profitability measures based on which sensor nodes select and access their next hops.

6.3.1 Next-hop Selection Profitability

In the prey model, the rate of gain, i.e., J , in (6.12) is maximized by determining the optimal p_i values via the zero-one-rule in (6.14) with the prey selection profitability v_i/t_i . Similarly, as given in the formulation of the cross-layer optimization problem in (6.9)-(6.11), J_P can be considered as a rate of gain of the next-hop selection process. J_P can be also maximized by determining the optimal p_{ij} values as follows.

Theorem 6.3.1 *In order to find the optimal p_{ij} , that maximizes J_P , a zero-one-rule can be derived via the next-hop selection profitability measure θ_{ij} given by*

$$\theta_{ij} = \frac{E_j e_{ij}}{H(d_{ij})}, \forall i, j \quad (6.15)$$

where E_j is the residual energy of sensor node j . $H(d_{ij})$ denotes the energy consumed by sensor node i to receive and transmit a bit to sensor node j . e_{ij} is the effective hop distance between sensor node i and j as shown in Fig. 6.3.

Proof. In order to find the optimal p_{ij} , J_P in (6.11) can be first rewritten as

$$J_P = \frac{p_{ij} E_j + \delta_{ij}}{p_{ij} \frac{H(d_{ij})}{e_{ij}} + \kappa_{ij}} \quad (6.16)$$

where δ_{ij} and κ_{ij} are expressed as

$$\delta_{ij} = \sum_{\substack{(z,k) \in P \\ z \neq i}} \sum_{\substack{k \in n_z \\ k \neq j}} p_{zk} E_k \quad \text{and} \quad \kappa_{ij} = \sum_{\substack{(z,k) \in P \\ z \neq i}} \sum_{\substack{k \in n_z \\ k \neq j}} p_{zk} \frac{H(d_{zk})}{e_{zk}} \quad (6.17)$$

Then, J_P can be differentiated with respect to p_{ij} as

$$\begin{aligned}\frac{\partial J_P}{\partial p_{ij}} &= \frac{E_j \left(p_{ij} \frac{H(d_{ij})}{e_{ij}} + \kappa_{ij} \right) - \frac{H(d_{ij})}{e_{ij}} (p_{ij} E_j + \delta_{ij})}{\left(p_{ij} \frac{H(d_{ij})}{e_{ij}} + \kappa_{ij} \right)^2} \\ &= \frac{E_j \kappa_{ij} - \frac{H(d_{ij})}{e_{ij}} \delta_{ij}}{\left(p_{ij} \frac{H(d_{ij})}{e_{ij}} + \kappa_{ij} \right)^2}\end{aligned}\quad (6.18)$$

By considering $\frac{\partial J_P}{\partial p_{ij}} = 0$, the optimal p_{ij} can be determined as follows.

- If $\frac{E_j e_{ij}}{H(d_{ij})} < \frac{\delta_{ij}}{\kappa_{ij}}$, then the numerator of (6.18) is negative. In this case, J_P can be maximized by choosing the lowest possible p_{ij} .
- Conversely, if $\frac{E_j e_{ij}}{H(d_{ij})} > \frac{\delta_{ij}}{\kappa_{ij}}$, then the numerator is positive. In this case, J_P can be maximized by choosing the possible highest p_{ij} .

Since $0 \leq p_{ij} \leq 1$, the lowest and highest possible p_{ij} can be given as $p_{ij} = 0$ and $p_{ij} = 1$, respectively. Consequently, the selection of p_{ij} to maximize J_P leads directly to a zero-one-rule:

$$\begin{aligned}\text{set } p_{ij} &= 0 & \text{if } \frac{E_j e_{ij}}{H(d_{ij})} < \frac{\delta_{ij}}{\kappa_{ij}} \\ \text{set } p_{ij} &= 1 & \text{if } \frac{E_j e_{ij}}{H(d_{ij})} > \frac{\delta_{ij}}{\kappa_{ij}}\end{aligned}\quad (6.19)$$

where $\frac{E_j e_{ij}}{H(d_{ij})}$ is the profitability measure, i.e., θ_{ij} , defined in (6.15). Hence, using the zero-one-rule, each sensor node i can determine its possible next hops such that sensor node j is a possible next hop of sensor node i if $p_{ij} = 1$ and it is not a possible next hop node if $p_{ij} = 0$. Since each sensor node i selects a single next hop node, it selects its next hop as one of its possible next hop nodes satisfying $\frac{E_j e_{ij}}{H(d_{ij})} > \frac{\delta_{ij}}{\kappa_{ij}}$ with the maximum next hop selection profitability $\frac{E_j e_{ij}}{H(d_{ij})}$. Such a next hop node can clearly maximize J_P by maximizing $\frac{\partial J_P}{\partial p_{ij}}$ provided that $\frac{E_j e_{ij}}{H(d_{ij})} > \frac{\delta_{ij}}{\kappa_{ij}}$. This determination will be detailed in Section 6.5. Next, the channel access profitability is introduced. ■

6.3.2 Channel Access Profitability

Here, we define the *channel access profitability* that enables each sensor node to regulate its channel access persistence to provide reliable sensor-actor communication. For a sensor

node, the regulation of its channel access persistence is affected by the data traffic that passes on it. As this data traffic increases, the channel access persistence should be increased to achieve a higher packet transmission rate, which can prevent possible congestion. However, increasing channel access persistence may result in excessive collisions and energy consumption. Therefore, channel access persistence should be carefully regulated to obtain a sufficient packet transmission rate while avoiding excessive collisions and energy consumption. The aim of channel access profitability is to allow each sensor node to obtain a sufficient level of packet transmission rate by regulating its channel access persistence.

In the prey model, forager is assumed to encounter prey types with an encounter rate (λ_i) as introduced in (6.12). Similarly, we assume that sensor node i encounters with an available time slot to successfully transmit a data frame with an encounter rate λ . When sensor node i encounters with an available slot to transmit, this means that there is no other transmission in the communication radius of sensor node i and its neighbors. Suppose that sensor node i transmits its data frames at the beginning of each time slot with a fixed probability p_i . Let us also assume that T_k is the time elapsed between successful transmissions of $(k - 1)^{th}$ and k^{th} frames. The beginning of a new frame transmission period can be called a *renewal point* or *event* [118]. T_k values are drawn from the same probability distribution and are assumed to be independent of one another.

The sequence of T_k can be modeled as inter-arrival times of a homogeneous Poisson process with the exponential density function $f_T(T) = \lambda e^{-\lambda T}$, where λ is the encounter rate of sensor node i with an available slot. Using the sequence of measured $T_k, k \in \mathcal{Z}^+$, it is possible for a sensor node to obtain an unbiased and efficient maximum-likelihood (ML) estimate of λ [119]. Assume that a total of B data frames are delivered with the interarrival times $T_1, T_2, \dots, T_B$. Likelihood function for this case is as follows:

$$f_{T_1, T_2, \dots, T_B; \lambda}(T_1, T_2, \dots, T_B; \lambda) = \prod_{i=1}^B \lambda e^{-\lambda T_i} \quad (6.20)$$

To find the ML estimate for λ , the derivative of the log-likelihood function, i.e., the logarithm of (6.20), can be taken and equated to 0. This yields

$$\frac{\partial}{\partial \lambda} \left[B \ln \lambda - \lambda \sum_{i=1}^B T_i \right] = \frac{B}{\lambda} - \sum_{i=1}^B T_i = 0. \quad (6.21)$$

Solving (6.21) for λ , the maximum likelihood estimate of λ , i.e., $\hat{\lambda}$ can be obtained as

$$\hat{\lambda} = \frac{1}{\frac{1}{B} \sum_{i=1}^B T_i} \quad (6.22)$$

Briefly, the estimate is equal to the reciprocal of the average inter-arrival time measures, i.e., $T_1 \dots T_B$. Since the duration of a time slot is assumed to be on the scale of μ -second, $\widehat{\lambda}$ will be sufficiently large and the homogeneous Poisson distribution can be approximated as a normal distribution given as $\mathcal{N}(\widehat{\lambda}, \widehat{\lambda})$. Hence, the random variable that characterizes the inter-arrival times reduces a normally distributed random variable, i.e., $\mathcal{N}(\widehat{\lambda}, \widehat{\lambda})$.

Let us now assume that sensor node i serves s_i source nodes to route their packets towards the actor node and each source node uses the average reporting frequency rate f (*frames/s*). Sensor node i receives the average number of fs_i data frames in a second. Therefore, in order to avoid a possible congestion, sensor node i should achieve a sufficiently high frame transmission rate, i.e., η_i , so as to provide $fs_i < \eta_i$. In that sense, the channel access profitability of sensor node i , i.e., ϑ_i , can be characterized by the probability that there is no congestion on sensor node i and formulated as follows:

$$\begin{aligned}\vartheta_i &= P(fs_i < \eta_i) \\ &= 1 - \Phi_i(fs_i)\end{aligned}\tag{6.23}$$

where η_i denotes the successful frame transmission rate of sensor node i and η_i is a random variable characterized by the normal distribution $\mathcal{N}(\widehat{\lambda}, \widehat{\lambda})$. $\Phi_i(\cdot)$ is the cumulative distribution function of $\mathcal{N}(\widehat{\lambda}, \widehat{\lambda})$. Maximization of the channel access profitability, i.e., ϑ_i , also favors the subject of the cross-layer optimization problem given in (6.10). Using the channel access profitability, BIOX allows each sensor node to keep its channel access profitability above a feasible maximum value. Hence, this enables each sensor node to separately regulate its transmission strategy in order to prevent possible congestion. The details of this regulation are introduced in Section 6.5.

6.4 Bio-inspired Task Allocation for Actor Nodes

In WSAN, the main goal of actor task allocation is to establish an organizational framework, based on which vehicular actor nodes efficiently share the available tasks associated with the event. For this goal, by following the original prey model given in Section 6.2.2, we introduce a task allocation model. We first introduce a *rate of gain for the task allocation*. Then, we derive a *task allocation profitability* measure that enables each actor to decide which tasks to perform in order to maximize the rate of the gain of the task allocation. Then, we investigate

the stability of the given task allocation mechanism.

6.4.1 Task Allocation Profitability

We assume that initially, M vehicular actor nodes randomly roam in the environment. Among these M actors, an actor node is selected as the master actor node. The details of this selection is presented in Section 6.5.1. Based on the event information received from the source nodes, the master actor node is assumed to estimate a 4-tuple for each of available tasks⁵, i.e., $F_i = \{b_i, m_i, x_i, y_i\}$ to characterize the properties of task i . b_i denotes the prioritized time⁶ for task i and given as $b_i = o_i T_i(t)$, where o_i is the constant priority of task i and $T_i(t)$ is the elapsed time at time t since task i first appears in the environment. As soon as task i is performed at time t , $T_i(t)$ is set to zero. m_i is the encounter rate of the master actor node with the task i . x_i and y_i are the coordinates of the estimated location of task i .

The master actor node periodically estimates or updates the 4-tuples for each task appearing in the environment, i.e., $F_i, \forall i$, and broadcasts them to other actor nodes. Using the received 4-tuples, each actor node decides which task to perform first and sends a request to the master to allocate the decided task. Once the request message is acknowledged by the master, the task is performed by the actor node. In order to perform a task, the vehicular actor node must also come to the close proximity of the task location. Hence, based on the task location, the task allocation time, and its capability in task performance, each actor node estimates its task completion time, i.e., t_i . Briefly, t_i is a time duration in which the actor node allocates task i and goes to its close proximity, and performs it.

Above mentioned concepts and parameters can be visualized by a realistic task allocation example. Consider an agriculture application in which a set of vehicular actor nodes are employed to keep the humidity level of a specific environment in a desired level in order to allow a special kind of plant to grow efficiently. The humidity level of the environment is continuously sensed by sensor nodes and dry areas of the environment and their dryness levels are continuously reported to the master actor node by sensor nodes. These dry areas must be allocated and made sufficiently wet by the vehicular actor nodes. Each of these dry

⁵ Here, we assume that each actor node is able to separately estimate a 4-tuple for each of available tasks whenever it is selected as a master actor node. However, the estimation of the 4-tuples is out of the scope of this chapter.

⁶ The measure of the prioritized time has been extensively used in the existing literature in order to investigate the stability of networked control systems [103].

areas can be viewed as a task to make wet. Each task is allocated and performed by an actor node. Then, the task is assumed to be accomplished until it appears again in the environment. Here, the level of dryness can be considered as priority for a task. Besides the capabilities of an actor node, the task completion time of the actor also depends on the actor and the dry area location.

Let us assume that K is the set of task IDs that are performed by actor node A and actor node A allocates task i with the probability a_i . Similar to rate of the gain in the prey model given in (6.12), for actor node A , the rate of gain in task performance, i.e., J_A , can be expressed as

$$J_A = \frac{\sum_{i \in K} a_i m_i b_i}{\sum_{i \in K} a_i m_i t_i} \quad (6.24)$$

The probability of performing each task, i.e., $a_i \forall i$, is the decision variable for the actor node. Thus, the goal of the actor node is to choose $a_i, \forall i$ that maximizes J_A .

Theorem 6.4.1 *The optimal $a_i, \forall i$ to maximize J_A can be found by means of a zero-one rule with the task allocation profitability, i.e., ψ_i , given by*

$$\psi_i = \frac{b_i}{t_i}, \forall i \quad (6.25)$$

Proof. To choose the optimal a_i , J_A is differentiated with respect to a_i , i.e.,

$$\begin{aligned} \frac{\partial J_A}{\partial a_i} &= \frac{m_i b_i (c_i + a_i m_i t_i) - m_i t_i (a_i m_i b_i + k_i)}{(c_i + a_i m_i t_i)^2} \\ &= \frac{m_i b_i c_i - m_i t_i k_i}{(c_i + a_i m_i t_i)^2} \end{aligned} \quad (6.26)$$

where k_i and c_i are given as

$$k_i = \sum_{\substack{j \in K \\ j \neq i}} a_j m_j b_j \quad \text{and} \quad c_i = \sum_{\substack{j \in K \\ j \neq i}} a_j m_j t_j \quad (6.27)$$

By considering $\frac{\partial J_A}{\partial a_i} = 0$, it can be possible to select optimal a_i as follows.

- $\frac{\partial J_A}{\partial a_i}$ is negative if $\frac{b_i}{t_i} < \frac{k_i}{c_i}$. In this case, J_A is maximized by selecting the lowest a_i .
- $\frac{\partial J_A}{\partial a_i}$ is positive if $\frac{b_i}{t_i} > \frac{k_i}{c_i}$. In this case, the highest a_i maximizes J_A .

Due to $0 \leq a_i \leq 1$, the lowest and the highest a_i are 0 and 1, respectively. Hence, J_A can be maximized by choosing a_i via a zero-one rule:

$$\begin{aligned} \text{set } a_i = 0 & \quad \text{if } \frac{b_i}{t_i} < \frac{k_i}{c_i} \\ \text{set } a_i = 1 & \quad \text{if } \frac{b_i}{t_i} > \frac{k_i}{c_i} \end{aligned} \quad (6.28)$$

where $\frac{b_i}{t_i}$ is the task allocation profitability measure, i.e., $\psi_i = \frac{b_i}{t_i}$, defined in (6.25). The zero-one rule is employed by each actor node to decide which tasks to perform as follows:

- It first ranks all profitability values as $\psi_1 > \psi_2 > \dots > \psi_K$. This sorting also changes the ID of the available tasks in a way that the task having the highest task allocation profitability becomes task 1 with ψ_1 and so on.
- Then, it performs the tasks when it encounters starting with the most profitable one until the task j that satisfies

$$\frac{\sum_{i=1}^j m_i b_i}{\sum_{i=1}^j m_i t_i} > \frac{b_{j+1}}{t_{j+1}} \quad (6.29)$$

The task allocation profitability is an efficient local metric to allow each actor node to select and perform available tasks in the environment. In fact, the zero-one rule clearly provides an optimal performance for actor nodes to maximize the rate of gain of the task performance, i.e., J_A . However, it is essential for the task allocation process to determine whether it provides a stable allocation of the tasks, where any task is accomplished within a bounded time delay. ■

6.4.2 Stability of BIOX Task Allocation

In this section, we investigate the stability of the bio-inspired task allocation mechanism introduced above. The stability of such a task allocation mechanism aims to ensure the boundedness of time delay, i.e., $T_i(t)$, during which task i is accomplished by an actor node. Here, the boundedness of $T_i(t)$, $\forall i$ is deduced to prove the stability of the given task allocation mechanism by the following theorem.

Lemma 6.4.2 *Task allocation algorithm in BIOX guarantees that any task i is definitely allocated and performed within a delay bound given as*

$$\lim_{t \rightarrow \infty} T_i(t) < \frac{t_{\min} o_{\max} t_i}{o_{\max} t_i - o_i t_{\min}} + t_{\max} \quad (6.30)$$

where o_{max} is the maximum of the priority values, i.e., $o_i < o_{max}, \forall i$ and t_{min} and t_{max} are the minimum and maximum of the task completion time values, i.e., $t_{min} < t_i < t_{max}, \forall i$.

Proof. Assume that at time t , task i has not been performed during τ units of time and all of other tasks are performed at least once in this duration. Hence, at time t task selection profitability of task i can be given as

$$\psi_i = \frac{T_i(t)o_i}{t_i} = \frac{\tau o_i}{t_i} \quad (6.31)$$

Suppose also that task k has the maximum possible profitability, i.e., ψ_k , at time t . Since task k has been performed at least once before the time t , the possible maximum value for $T_k(t)$ can be given as $T_k(t) = \tau - t_{min}$. Task k may also have the maximum priority and the minimum task performing time as $o_k = o_{max}$ and $t_k = t_{min}$, respectively. Hence, substituting $T_k(t) = \tau - t_{min}$, $o_k = o_{max}$, and $t_k = t_{min}$, the maximum possible ψ_k can be given as

$$\psi_k = \frac{(\tau - t_{min})o_{max}}{t_{min}} \quad (6.32)$$

At time t , in order to guarantee that task i is performed, $\psi_i > \psi_k$ has to be satisfied, which yields

$$\tau < \frac{t_{min}o_{max}t_i}{o_{max}t_i - o_it_{min}} \quad (6.33)$$

Task i is definitely allocated within the bound in (6.33). After the allocation, it is also performed within a time duration that is bounded by t_{max} . Hence, task i is definitely allocated and performed within a delay bound, i.e.,

$$\lim_{t \rightarrow \infty} T_i(t) < \frac{t_{min}o_{max}t_i}{o_{max}t_i - o_it_{min}} + t_{max} \quad (6.34)$$

■ Consequently, BIOX provides an optimal and stable task allocation for actor nodes. Next, we introduce the BIOX protocol operation.

6.5 Bio-inspired Cross-layer (BIOX) Protocol Operation

The operation of BIOX includes four main phases. Whenever an event occurs in WSN environment, these four phases are consecutively accomplished by the network nodes for the efficient communication and coordination of sensor and actor nodes. In the first phase, an actor node is selected as master actor to collect the event information and estimate the 4-tuples for all available tasks, which is explained in the next subsection. In the second and third

phase, the next-hop selection and channel access profitability measures are computed and employed by sensor nodes to provide energy-efficient routing and autonomous congestion and rate control for sensor-actor communication. Finally, task allocation profitability measures are computed by the master actor and all available tasks are shared and performed by the vehicular actor nodes. Next, these four phases of BIOX are presented.

6.5.1 Event-driven Master Actor Selection and Positioning

When an event occurs, each source node generates a *source announcement packet*, that includes its ID, in order to announce that it is a source node. The source announcement packet is then transmitted by the source node to all of sensor and actor nodes in its communication radius. Then, this strategy is followed by every sensor node that receives the source announcement packet. This allows the diffusion of the source announcement packet along the network. During this diffusion, the actor node, that first receives the source announcement packet, becomes the master actor and it broadcasts a *master actor announcement packet* to all actor nodes. Then, it waits for ACK messages from other actors to ensure that it becomes master actor node. Here, we assume that each actor node has a unique ID, i.e., $i \in \{1, 2, 3, \dots, M\}$. Hence, if more than one actor nodes, that receive the *source announcement packet* at the same time, broadcast *master actor announcement packet*, the actor node having maximum ID is assumed to be selected as the master actor node. After the determination, the master actor node broadcasts master actor announcement packet to all sensor nodes to announce that it becomes the master actor node. Then, all source node start to forward their data packets towards the master actor node.

The aim of the master actor node is to collect event data from all source nodes. Thus, its position is critical for an energy-efficient sensor communication. As already given in (6.7), total energy consumption along the data path P , i.e., C_P , depends on the Euclidean distance D_P between the source and actor node by $D_P \gamma_P \geq C_P$. Clearly, the total cost C_P can be decreased by reducing the Euclidean distance, i.e., D_P . Thus, the position of the master actor node must be closest to all source node positions in Euclidean sense to provide minimum energy consumption for sensor nodes. Therefore, the position of the master actor node, i.e., φ , must be selected by minimizing

$$\sum_{i \in S} |\varphi - \rho_i| \quad (6.35)$$

where ρ_i is the position of source node i and S denotes the set including the IDs of all source nodes. Since all source positions and its position are assumed to be known by the master actor node, the master actor node first computes its new position to efficiently collect the source data packets by minimizing (6.35). Then, the master actor moves to this new position in order to minimize the energy consumption of sensor nodes.

6.5.2 Next-hop Selection

BIOX enables each sensor node to select its next hop node by means of the next hop selection profitability, i.e., θ_{ij} , given in (6.15). For the computation of $\theta_{ij}, \forall j$, sensor node i needs to find $H(d_{ij})$, e_{ij} , and $E_j, \forall j$ for all of its neighbors towards the master actor node. Here, we assume that each sensor node foreknows its own location as well as the location of its neighbors by means of an existing localization technique⁷ [105]. Hence, it can compute all Euclidean distance to its neighbors and corresponding effective distances, i.e., $d_{ij}, e_{ij}, \forall j$, and also computes $H(d_{ij}), \forall j$ as given in Section 6.2.1.2. Each sensor node periodically measures and broadcasts its residual energy to all of its neighbors. Therefore, each sensor node i also knows the residual energy level of its neighbors, i.e., $E_j, \forall j$.

Substituting $H(d_{ij})$, e_{ij} , and E_j into (6.15), each sensor node i determines $\theta_{ij}, \forall j$ and establishes its next hop selection set U_i including all IDs of its possible next hops. Then, using U_i , sensor node i selects its next hop node as a sensor node providing the maximum next hop selection profitability, that is, the next hop of sensor node i is sensor node j such that $\{j \mid \forall k \in U_i : \theta_{ik} < \theta_{ij}\}$.

6.5.3 Channel Access and Rate Control

After the selection of its next hop, each sensor node computes its channel access profitability, i.e., ϑ_i in (6.23), to select its transmission probability p_i . For the computation of ϑ_i , sensor node i periodically estimates the arrival rate of its successful transmissions using the maximum likelihood estimation mechanism given in Section 6.3.2. In one second, sensor node i records time intervals between successful packet transmissions, i.e., $T_k, \forall k$. At the end of each second, it computes the estimation of the arrival rate of the successful transmissions,

⁷ The details of localization procedure is beyond the scope of this chapter.

i.e., $\widehat{\lambda}$, using (6.22). Based on the cumulative distribution function of the normal distribution $\mathcal{N}(\widehat{\lambda}, \widehat{\lambda})$, i.e., $\Phi(\cdot)$, sensor node i computes its channel access profitability as $\vartheta_i = 1 - \Phi(f s_i)$. Here, f is the average reporting frequency rate of source nodes and s_i is the number of source nodes that sensor node i serves to route their packets towards the master actor node.

Each sensor node i regulates its transmission probability in an interval given as $p_{min} < p_i < p_{max}$, where p_{min} and p_{max} are the minimum and maximum of the transmission probability, respectively. Initially, each sensor node sets its transmission probability as p_{min} and starts to transmit to its next hop. For a successful transmission of sensor node i , there must be no other transmission in the first, second, and third collision domains as shown in Fig. 6.4. Assume that there are g active sensor nodes that transmit with the transmission probability p in the first, second, and third collision domain. Therefore, the probability of a successful transmission of sensor node i , i.e., R_i , can be given as

$$R_i = p(1 - p)^{g-1} \quad (6.36)$$

To find the optimal p that makes R_i maximum,

$$\frac{\partial R_i}{\partial p} = (1 - p)^{g-1} - (g - 1)p(1 - p)^{g-2} = 0 \quad (6.37)$$

This yields the optimal p , i.e., $\bar{p}$, as $\bar{p} = 1/g$. In order to compute $\bar{p}$, each sensor node i first discovers the number of active sensor nodes in its collision domain, i.e., g . To this end, each active sensor node sends a message to all of its neighbors to notice that it is an active sensor node. This enables each sensor node to discover the number of active sensor nodes, g , in its collision domains. Then, it sets p_{max} as $p_{max} = \bar{p} = 1/g$. Furthermore, p_{min} is set to a sufficiently low value within the interval $[0, 1]$. In the simulation experiments, p_{min} is selected as $p_{min} = 0.05$.

BIOX enables each sensor node to periodically update its transmission probability starting with p_{min} . The aim of this update is to keep the channel access profitability ϑ_i above a predefined threshold value⁸ called as $\widetilde{\vartheta}$. This implicitly provides a congestion control mechanism since the channel access profitability of a sensor node is the probability that there is no congestion on this sensor node. The transmission probability update strategy of BIOX is given as

⁸ This threshold called as $\widetilde{\vartheta}$ is set as 0.9 in the simulation experiments of this chapter.

- If sensor node i finds ϑ_i as $\vartheta_i < \widetilde{\vartheta}$ and $p_i < p_{max}$, it updates p_i as $p_i = p_i + \zeta$, where ζ is a small positive constant.
- If sensor node i finds ϑ_i as $\vartheta_i \geq \widetilde{\vartheta}$, it does not change the current value of p_i .

Based on the updated transmission probability, for the access to the channel, each sensor node generates a uniformly distributed random number h from the interval $[0, 1]$ at the beginning of each time slot. If $p_i > h$, sensor node i transmits to its next hop at the beginning of current slot. Otherwise, sensor node i does not make any transmission during current time slot.

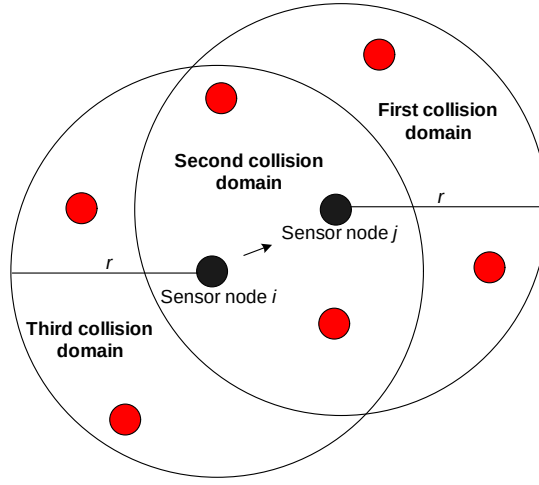


Figure 6.4: Illustration of collision domains for the link between sensor nodes i and j .

For a successful delivery from sensor node i to sensor node j , there must be no other transmission in the first, second, and third collision domains as shown in Fig. 6.4. After a successful delivery, ACK frame, that sensor node j transmits to sensor node i , can still collide with the possible ACK frames that are transmitted by the sensor nodes in the first collision domain. For example, assume that in the first, second, and third collision domains there is no other transmission except sensor node i . Assume also that in the first collision domain, at least one sensor node receives a frame from the outside of the collision domains. In this case, sensor node i successfully delivers a frame to sensor node j . However, if the sensor node in the first collision domain also succeeds to receive a data frame from outside of the collision domains, ACK frame of sensor node j and this sensor node can collide and current transmission attempts become unsuccessful.

In order to mitigate the probability of ACK collision, BIOX allows each sensor node j to send

multiple ACKs in the ACK transmission interval for each of successfully received frames as follows.

- During the frame transmission interval, sensor node i transmits a data frame to sensor node j with probability p_i . Sensor node i also listens the channel whether or not the transmitted data frame is collided.
- If the transmitted frame is collided, it is retransmitted with the probability p_i .
- If sensor node i successfully delivers a packet to sensor j within the frame transmission interval of a slot, sensor node j immediately sends an ACK frame for this frame in the ACK transmission interval of the slot. ACK frame includes the IDs of all frames previously received within a fixed time interval.
- This mitigates the probability of ACK collision in a way that sensor node i can eventually receive an ACK frame including ID of successfully delivered data frames even if sensor node i did not previously receive an ACK frame belonging to the data frame that is successfully delivered.

Eventually, using next-hop selection and channel access profitability, BIOX allows sensor nodes to autonomously estimate and adapt their communication parameters so as to provide energy-efficient, reliable, and autonomous sensor-actor communication in WSN.

6.6 Performance Evaluation

In this section, we present the performance evaluation results of BIOX protocol. For the performance experiments, we develop a simulation environment using MATLAB. We model the event with a point source located in the sensor field which generates an event signal to be sensed and sampled by the source nodes. The parameters used in the performance analysis are given in Table 6.2.

6.6.1 Energy Conservation

To show the BIOX performance on energy conservation, we investigate how the next hop selection profitability can prolong the network lifetime. When an event occurs, 80 source

Table 6.2: Simulation Parameters

<i>Number of sensor nodes (N)</i>	100
<i>Environment size</i>	$100 \times 100m - 200 \times 200m$
<i>Transmission range of sensor nodes (r)</i>	20m
<i>Queue Size of Sensor Nodes</i>	15KBytes
<i>Number of Actor Nodes (M)</i>	2-8
<i>Number of Source Nodes</i>	80
<i>Frame Length</i>	30 Bytes
<i>ACK Frame</i>	5 Bytes
<i>Slot Duration</i>	0.001 s

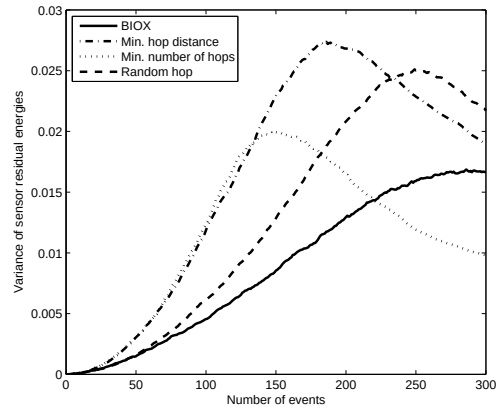


Figure 6.5: Variance of sensor residual energies with increasing number of events.

nodes closest to the event source act as source node and generate data traffic with reporting frequency rate of 40 (*frames/s*). Events are consecutively generated in the different locations of the environment and each event continues during 1 *minute*.

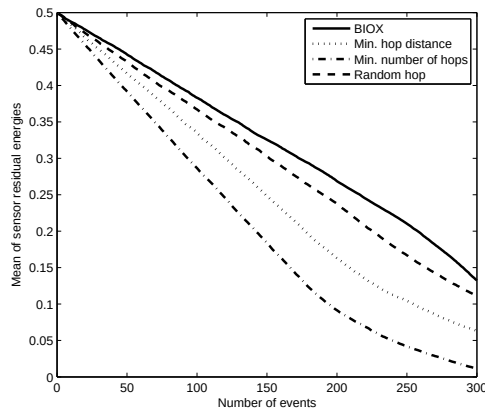


Figure 6.6: Mean of sensor residual energies with increasing number of events.

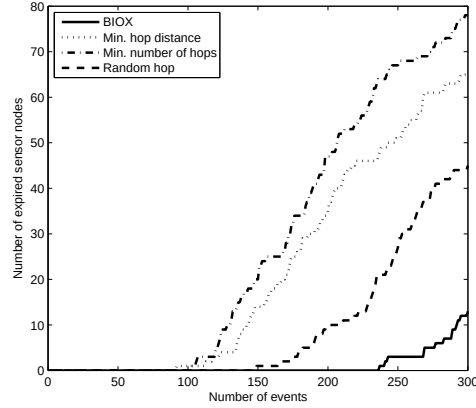


Figure 6.7: The number of expired nodes with increasing number of events.

In Fig. 6.5, the variance of sensor residual energies is shown according to increasing number of events and we compare the change in the variance with three fundamental next hop selection approaches. These approaches provide minimum hop distance, minimum number of hops, and random hop with the same uniform probability. BIOX maintains the variance of sensor residual energies at the minimum level and significantly outperforms the three fundamental approaches. This manifests that BIOX enables sensor nodes to uniformly distribute their energy consumption to provide balanced energy consumption throughout the network. This can be easily seen in Fig. 6.6 by observing the mean of sensor residual energies. BIOX clearly minimizes total energy consumption throughout the network. In Fig. 6.7, the number of expired nodes is shown with the increasing number of consecutive events. BIOX outperforms minimum hop distance and minimum number of hops-based approaches by nearly 150%, and the random hop-based approach by 75% improvement in network lifetime. Hence, this reveals that BIOX significantly maximizes the network lifetime.

6.6.2 Timely and Reliable Sensor-actor Communication

In this section, we observe the packet delivery performance of BIOX by means of packet delivery ratio, i.e., ratio of the number of packets delivered to the actors to the total number of transmitted packets by source nodes. In Fig. 6.8, the packet delivery ratio is shown with time for the different number of source nodes (n). While n increases, the packet delivery ratio decreases. However, BIOX enables each sensor node to achieve to prevent possible congestion

and to eventually deliver more than 80% of total packets generated by source nodes.

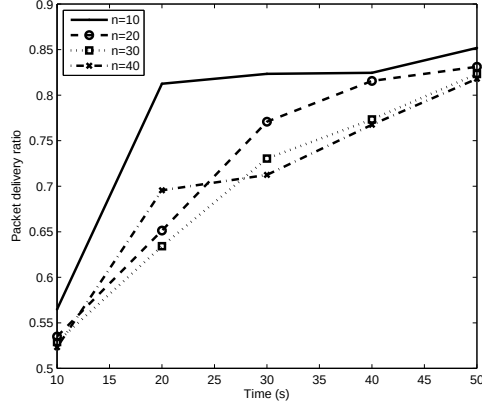
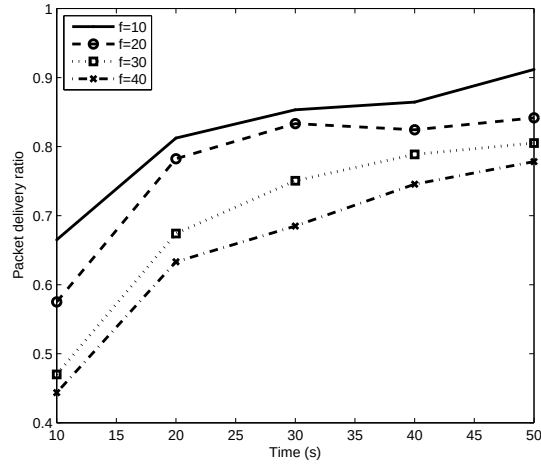
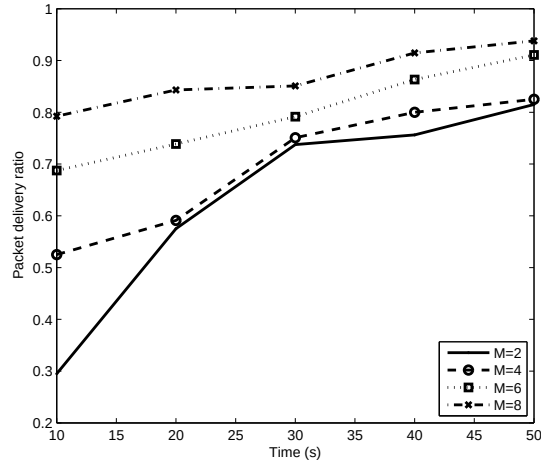


Figure 6.8: Packet delivery ratio with time for the different number of source nodes.

In Fig. 6.9(a), the packet delivery ratio is shown with time for the different number of actor nodes. As the number of actor nodes increases, the number of hops over the network decreases. This naturally decreases the collision probability over the network and BIOX protocol can converge to more than 90% of packet delivery ratio as the number actor nodes increases. However, even if two actor nodes are used, BIOX also provides high packet delivery ratio.

In Fig. 6.9(b), the packet delivery ratio is evaluated for the different reporting frequency rate of source nodes (f). As the reporting frequency rate increases, average collision probability over the network increases, which clearly decreases the packet delivery ratio. However, BIOX provides high packet delivery ratio near 70 % of packet delivery ratio even if $f = 40$ is used by the source nodes.

In Fig. 6.10, we compare the packet delivery performance of BIOX with event-to-sink reliable transport protocol (ESRT) [28]. ESRT aims to provide a desired event reliability that is defined as the minimum number of data packets required for reliable event detection. In the simulation experiments of ESRT, the desired event reliability are set as 4000 packets per 10s. To evaluate the performance, the packet delivery ratio is obtained dividing the number of packets delivered within 10s by the event reliability (4000 packets). Furthermore, 81 source nodes are used in the $200 \times 200m$ sensor field. Based on the feedback messages from the sink node, ESRT operates to provide the desired event reliability by updating the report-



(a)

(b)

Figure 6.9: (a) Packet delivery ratio with time for the different number of actor nodes. (b) Packet delivery ratio with time for the different reporting frequency of source nodes.

ing frequency rate of source nodes. Using the same simulation setting with ESRT, in Fig. 6.10, we compare the packet delivery performance of BIOX and ESRT according to a reporting frequency update trace of ESRT (10, 8.79, 6.52, 5.73, 5.33, 5.12 *packets/s*). BIOX provides approximately the same packet delivery performance of ESRT although BIOX does not need any feedback message from the actor nodes for the packet delivery performance, which reveals that BIOX is an autonomous protocol with high and reliable packet delivery performance.

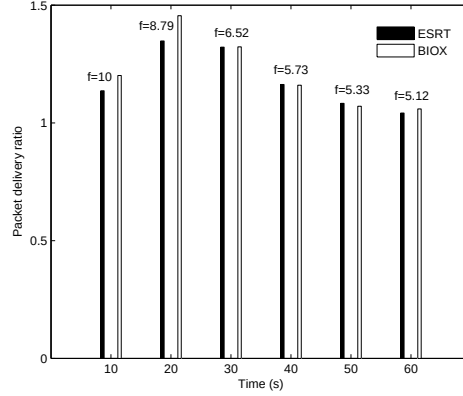


Figure 6.10: Performance comparison of BIOX and ESRT in terms of packet delivery ratio.

6.6.3 Efficient Task Allocation

Here, we observe the BIOX task allocation performance by comparing the task performance delay of any task and theoretical upper bound of the task performance delay given in (6.30). We assume that in a WSAN environment, the number of 20 tasks consecutively appear and are accomplished by 5 actor nodes. Every task has different priority value and each actor node has different task completion time according to its task performing capability. We assume that the minimum and maximum of the priority values are $o_{min} = 1.5$ and $o_{max} = 3$, respectively. Similarly, the lower and upper bound for the task performance time of actor nodes are set as $t_{min} = 1.5$ and $t_{max} = 3$, respectively, where the task performance time is given in a unit of time. We assume that the appear rate of the tasks follows a uniform distribution. In Table 6.3, the average task performance delay, i.e., the average time elapsed until a task is accomplished, is compared with the theoretical delay bound in (6.30) for the different values of t_{min} , t_{max} , o_{min} , and o_{max} . It is shown that the given theoretical delay bound can be justified by the simulation experiments. The delay and theoretical upper bound increase with t_{min} however, they can be reduced by decreasing t_{max} . Furthermore, the task performance delay can be also decreased by reducing o_{min} . Conversely, decreasing o_{max} slightly affects the task performance delay, which stems from the fact that the difference between o_{min} and o_{max} becomes too low such that almost all tasks have the same priority values.

Table 6.3: Comparison of average task performance delay and its theoretical bound given in (6.30)

Minimum task completion time ($t_{\min}$)	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5
<i>Average delay for task performance ($T_i(t)$)</i>	2.9317	2.9126	3.0862	3.1294	3.2322	3.3243	3.3207	3.4426	3.4293	3.5261	3.5462
<i>Theoretical upper bound for $T_i(t)$ in (6.30)</i>	5.9464	6.4328	6.7390	7.0952	7.7072	8.0385	8.6083	8.7181	9.3503	10.2582	10.9938
Maximum task completion time ($t_{\max}$)	3.0	2.9	2.8	2.7	2.6	2.5	2.4	2.3	2.2	2.1	2.0
<i>Average delay for task performance ($T_i(t)$)</i>	2.9227	2.9142	2.8992	2.8614	2.8208	2.8501	2.7405	2.7112	2.6917	2.6965	2.5206
<i>Theoretical upper bound for $T_i(t)$ in (6.30)</i>	5.9454	6.1074	5.9455	5.9495	5.9114	5.6980	5.9030	5.8371	6.0052	5.8482	5.8910
Minimum task priority ($\phi_{\min}$)	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5
<i>Average delay for task performance ($T_i(t)$)</i>	2.9166	2.9231	2.8585	2.8689	2.7713	2.7854	2.7157	2.7263	2.6776	2.6569	2.6813
<i>Theoretical upper bound for $T_i(t)$ in (6.30)</i>	5.9833	6.1608	6.0345	6.0987	6.2792	6.3411	6.5253	6.7544	6.5550	6.7811	6.9354
Maximum task priority ($\phi_{\max}$)	3.0	2.9	2.8	2.7	2.6	2.5	2.4	2.3	2.2	2.1	2.0
<i>Average delay for task performance ($T_i(t)$)</i>	2.8779	2.9213	2.9439	2.8217	2.8937	2.8476	2.7902	2.8056	2.7763	2.7954	2.8193
<i>Theoretical upper bound for $T_i(t)$ in (6.30)</i>	5.9515	5.9881	6.1458	6.3452	6.0813	6.1194	6.2540	6.2275	6.4212	6.5920	6.5088

6.7 Conclusion

In this chapter, we introduce BIO-inspired Cross-layer (BIOX) Communication and Coordination Protocol for WSN. BIOX is a unified algorithm that incorporates MAC, routing and transport layer functionalities to enable each sensor node to separately select and access its next hop node using the nature-inspired next-hop selection and channel access profitability measures. BIOX also uses the nature-inspired task allocation profitability measure for efficient allocation of the available tasks associated with the event. BIOX provides a sub-optimal performance for the network lifetime, timely and reliable sensor-actor communication, and task allocation. Due to its fully-autonomous operations, BIOX can steadily keep the network in a highly reliable and energy-efficient state. This renders BIOX a robust and energy-efficient protocol for the realization of future WSN applications.

CHAPTER 7

Biological Foraging-inspired Communication in Intermittently-Connected Mobile Cognitive Radio Ad Hoc Networks

In this chapter, based on a *prey model in foraging theory*, Biological Foraging-inspired Communication (BFC) algorithm is introduced for Intermittently-connected Mobile Cognitive Radio ad hoc Networks (IMCRN). BFC was presented in [120]. BFC is based on two profitability measures called *relay selection profitability* (RSP) and *channel selection profitability* (CSP). The aim of RSP is to allow each node to quantify its rate of gain in copying a message to a relay while examining its transmission effort. Using RSP, each node decides which relays should be used in order to provide minimum energy consumption without sacrificing end-to-end delay performance. Based on CSP, each node decides and switches to a licensed channel to maximize spectrum utilization while keeping the interference in a minimum level. BFC maximizes overall spectrum utilization in a way that any idle channel is always allocated by an IMCRN node within a delay bound. BFC provides a great autonomy for IMCRN by adaptively regulating and tuning its communication strategy in the relay and channel selection. This also provides an optimization procedure to determine optimal relay and channel selection rules.

The motivation and the related works are discussed in Section 7.1. In Section 7.2, the network model and assumptions are given. In Section 7.3, a prey model in foraging theory is briefed, based on which BFC algorithm is introduced in Section 7.4. The performance evaluation of BFC is presented in Section 7.5. Finally, concluding remarks are given in Section 7.6.

7.1 Motivation and Related Works

Rapid development in communication technologies has enabled a large set of communicating devices that may be interconnected by an existing infrastructure or ad hoc manner. The infrastructured connection technologies such as cellular access and WiFi do not exist everywhere and are expensive for limited bandwidth. Therefore, the ad hoc connection, where each node only communicates with its neighbors to eventually deliver its message to a destination node, is indispensable for many wireless networks. However, such an ad hoc connection brings many crucial challenges into networking functionalities, i.e., the absence of permanently available end-to-end paths and intermittent connectivity. Networks with these challenges are mainly categorized as Intermittently Connected Networks (ICN) [121]. The examples of ICN include terrestrial mobile networks, military and tactical networks, satellite networks, and wireless sensor & actor networks.

Until recently, ICN have mostly used the licensed portions of wireless spectrum. However, underutilization of the spectrum recently stimulates unlicensed communication in ICN via cognitive radio (CR) technology. Intermittently-connected Mobile Cognitive Radio ad hoc Networks (IMCRN) can be viewed as a promising example of ICN with CR technology. IMCRN allows intermittently-connected mobile unlicensed nodes to exploit temporarily available contacts and idle licensed channels for end-to-end message delivery. However, the realization of IMCRN also induces serious research challenges that must be addressed. In particular, due to different node mobility and spectrum availability patterns, IMCRN is frequently divided into unpredictable partitions. These partitions are essentially intermittently-connected and deficient in complete end-to-end paths. Hence, the realization of IMCRN mandates efficiently addressing intermittent and spectrum-aware connection of network nodes.

There is an emerging body of literature on routing protocols in cognitive radio ad hoc networks [22]. All of these protocols are also based on the assumption that at least one end-to-end path always exists between any two nodes in the network. However, this assumption is invalid for IMCRN that is usually lacking in complete end-to-end paths. Hence, *spectrum-aware flooding* (SAF) is more relevant for IMCRN. SAF permits a message to be first copied to a set of relay nodes using available channels. Then, one of these relay nodes delivers the message to the destination provided that it encounters. In SAF, if the message is tried to be copied to all relays that do not have the message, the end-to-end message delay can be clearly

minimized. However, such a forwarding strategy is energy-inefficient and may cause a severe interference to primary system. The energy consumption and high interference can be clearly mitigated by deciding the appropriate relay nodes and channels. Hence, it is imperative to decide which relay nodes and licensed channels should be used for an efficient communication in IMCRN.

In nature, each biological entity in a colony has an autonomous decision-making mechanism to decide whether to perform some specific actions needed for vital functionalities. Despite the fact that each entity only accesses to a local information, this decision-making does not also hamper the overall stability of the biological system. For example, foraging strategy of a biological predator includes a significant decision-making in searching and handling prey types. Using the profitability measures of prey types, each predator optimizes the trade off between its prey searching efforts and its energy intake obtained by handling the prey. This optimization enables the predator to decide which prey types should be handled to maximize its energy intake while examining its starvation risk. Moreover, at the colony level, this mechanism also eventually leads to an equilibrium point, where each predator has the same prey consumption rate [112], [110].

Similar to a biological forager, in IMCRN, each node needs an autonomous decision-making mechanism to efficiently determine relay nodes and channels used in message delivery. To select a relay node, it optimizes the trade off between consumed energy and gained utility in copying the message to the relay. Furthermore, to determine whether to use a licensed channel, it reconciles spectrum utilization and interference level imposed on primary system. Hence, an analogy can be drawn from biological foraging toward developing efficient and autonomous communication algorithms for the IMCRN domain.

The earliest routing algorithm devised for intermittently-connected ad hoc networks is a flooding-based routing algorithm called Epidemic routing [122]. It follows a simple strategy that a source node copies its message to the relay nodes that are temporarily contacted with itself and do not have the message. Then, the relay nodes with the message also follow the same strategy and one of these relays contacts and delivers the message to the destination node. Epidemic routing guarantees the minimum end-to-end message latency if the network resource is assumed to be infinite. However, in reality, it is inherently prone to quick depletion of network resources such as communication bandwidth, buffer space, and battery of nodes.

In order to alleviate the resource wastefulness of epidemic routing, many routing protocols have been also proposed for ICN. In [123], the number of hops to deliver the message is restricted for reducing the communication overhead. Source node first copies the message to a number of its neighbors. Then, each of these neighbors only delivers the message when it contacts with the destination. Hence, the number of hops between source and destination are restricted to 2 and this strategy is called as 2-hop relaying. In addition to number of hops, a spray and wait strategy is suggested to bound the total number of message copies distributed to the relay nodes [124]. Furthermore, a mobility-assisted routing is also presented for ICN in which special mobile nodes called message ferries is employed to provide end-to-end connection [125].

Although these works efficiently address the intermittent connectivity, none of them concurrently considers secondary access to licensed spectrum with CR technology. On the other hand, there is also a rising frame of literature on the routing protocols in cognitive radio ad hoc networks (CRAHN) [22]. These routing protocols mainly focus on the joint design of spectrum management and route selection in CRAHN. For example, in [35], traditional geographical routing and spectrum allocation are combined to reduce end-to-end delay without interfering primary system. In [36], the spectrum allocation and next hop selection are also considered as two distinct phases of route set up to minimize end-to-end message latency while explicitly protecting primary user activities. Furthermore, for CR networks, spectrum management and route selection are integrated at the network layer to simultaneously select route and spectrum band for conflict-free channel usage [37]. A data-adaptive routing is also introduced. For a given amount of data, the given algorithm sets up the fastest routes by considering capacity, disruption probability, and propagation delay of existing links [38].

All of above mentioned works and many others [126] clearly address the joint design of spectrum management and route selection in CRAHN and CR networks. However, these works are also based on the assumption that at least one path can be always found between a source and destination node in the network. In particular, they do not consider the cases in which the network is intermittently-connected and frequently lacking in complete end-to-end paths due to different node mobility and spectrum availability patterns. In this chapter, BFC protocol is introduced in order to address unlicensed communication of intermittently-connected mobile nodes in IMCRN. BFC enables each node to adaptively determine its relay and channel selection strategy. It employs an autonomous decision-making mechanism that

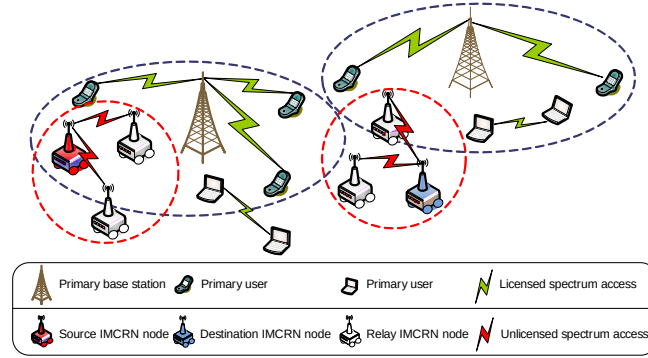


Figure 7.1: An IMCRN scenario is illustrated. A source node intends to deliver a message to a destination node by using the relay nodes in an environment in which a number of primary users and primary base stations coexist and operate on the licensed spectrum bands.

does not need any a priori information on node mobility and spectrum availability patterns. The decision-making mechanism eventually leads to an optimization procedure to determine optimal relay and channel selection rules.

7.2 Network Model and Assumptions

We consider a network with N mobile unlicensed nodes that move in an environment according to some stochastic mobility models as depicted in Fig. 7.1. We also assume that entire spectrum is divided into number of M non-overlapping orthogonal channels having different bandwidth. The access to each licensed channel is regulated by fixed duration time slots. Slot timing is assumed to be broadcast by the primary system. Before transmitting its message, each transmitter node, which is a node with the message, first selects a relay node and a frequency channel to copy the message. Note that the details of the relay and channel selection are given in Section 7.4. After the relay and channel selection, the transmitter node negotiates and handshakes with its relay node and declares the selected channel frequency to the relay. The communication needed for this coordination is assumed to be accomplished by using a common control channel (CCC) which is also a licensed channel.

After the transmitter selects a frequency channel and handshakes with its relay, relay node tunes to the selected frequency to receive the message. Then, the transmitter starts to allocate the selected channel. In the channel allocation, the transmitter node splits a slot duration into four distinct intervals, i.e., primary sensing, secondary sensing, frame transmission, and ACK

intervals as illustrated in Fig. 7.2(a). In the primary sensing interval, it senses the selected frequency channel¹ in order to deduce whether a primary user transmits. If the channel is not used by any primary user, in the secondary sensing interval, the transmitter senses the channel to infer if an IMCRN node, i.e., a secondary user, has previously allocated and currently uses the channel. If the channel is not used by an IMCRN node, the transmitter node allocates the channel and starts to transmit the frames via the frame transmission interval. If the channel is used by a secondary user, this means that the channel has been previously allocated by a transmitter IMCRN node. In this case, the transmitter node determines another channel among the available channels and it follows the same steps in order to allocate the newly selected channel. Notice that the details of the channel selection are given in Section 7.4. Here, we assume that a message is divided to a number of frames and hence, the delivery of a message to a relay node lasts for more than one slot duration.

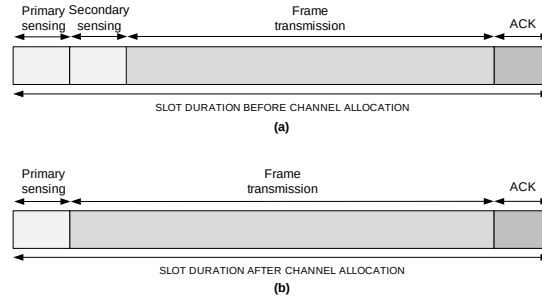


Figure 7.2: (a) In channel allocation phase, the slot duration consists of the primary sensing, secondary sensing, frame transmission, and ACK intervals. (b) After the channel allocation, the slot duration includes the primary sensing, frame transmission, and ACK intervals.

After the allocation of a channel, each transmitter node delivers the message to its relay by splitting the slot duration into three intervals called primary sensing, frame transmission, and ACK intervals as seen in Fig. 7.2(b). Here, the secondary sensing interval in the channel allocation phase is used for the frame transmission interval as shown in Fig. 7.2. This clearly permits other IMCRN nodes in the channel allocation phase to figure out that the channel has been previously allocated and used. Furthermore, in the ACK interval, the transmitter node waits for an ACK frame from its relay if it can successfully transmit a frame.

Besides the instantaneous channel sensing in some parts of a slot duration as explained above, we also assume that each IMCRN node also periodically senses a subset of M licensed chan-

¹ Note that we assume that channel sensing is accomplished by an energy-based channel sensing mechanism [77].

nels. More specifically, each node is assumed to randomly select a number of channels among M available licensed channels and periodically senses to investigate how long each of the selected channels is unused. As introduced in Section 7.4, this information is used by the nodes to select the licensed channels to transmit.

Whenever any two nodes are within their communication radius, they are assumed to meet with each other and they are called as contacted. In order to announce its existence, each node periodically broadcasts a beacon message to its contacts using CCC. Whenever CCC is not used by the primary system, each node is assumed to receive the beacon messages from their contacts that are transiently in its communication radius. This allows each node to keep its temporarily changing contact list whenever CCC is available. BFC also authorizes the network nodes to compute a utility metric to quantify the expected performance of a relay node in message delivery, which is defined as follows.

Definition 7.2.1 (Utility for Message Delivery) *The utility v_{jd} is a metric that implicitly specifies the message delivery performance of node j if it tries to deliver a message to node d . v_{jd} consists of two parts given as $v_{jd} = v_{jd}^t + v_{jd}^s$. The first part v_{jd}^t characterizes the transient utility while the second part v_{jd}^s characterizes the steady-state utility. By initially setting $v_{jd}^t = 0$ and $v_{jd}^s = 0$, in each time interval I , each node j computes $v_{jd}, \forall d$ using the following utility update strategy:*

- *If node j contacts with node d at the time interval I , it updates the transient and steady-state components of v_{jd} as $v_{jd}^t = v_{jd}^t + \xi_1$ and $v_{jd}^s = v_{jd}^s + \xi_2$, where ξ_1 and ξ_2 are positive constants.*
- *If node j does not contact with node d , it decreases v_{jd} by only updating the transient part as $v_{jd}^t = v_{jd}^t - \xi_1$. Here, the lower-bound of v_{jd}^t is assumed to be zero such that if the update, $v_{jd}^t = v_{jd}^t - \xi_1$, makes v_{jd}^t less than zero, v_{jd}^t is immediately set to zero and the update is skipped as long as $v_{jd}^t = 0$.*

As will be introduced in the following sections, the given utility metric is used by IMCRN nodes to determine their message copying strategies. Next, we brief a prey model in foraging theory, based on which we develop a biological foraging-inspired communication (BFC) algorithm for IMCRN.

7.3 A Prey Model in Foraging Theory

Foraging theory mathematically describes how the foraging animals search for prey types and decide which prey types to be handled [112][110][111]. Animals are assumed to seek and handle a prey by maximizing a ratio given as $\frac{E}{T}$. E denotes the energy intake obtained by searching and handling the prey while T is the time needed for searching and handling the prey. The maximization of $\frac{E}{T}$ clearly provides nutrient sources and additional time for foraging animals to perform other activities such as reproducing, fighting, or shelter building. Animals with poor foraging strategies are eliminated while animals with successful foraging strategies are promoted. This mechanism eventually allows the animals having significant foraging strategies to grow up. Based on this evolutionary principle, the foraging strategies of animals are adopted and modeled as an optimization process that is referred to as *optimal foraging theory* [112]. For example, this optimization permits a forager to decide the most appropriate prey types so as to maximize overall energy intake within the minimum time interval.

Among the foraging models, one of the most common models is the prey model. The prey model considers that each prey holds a certain energy intake. Suppose that there are k different types of prey in the environment. t_i is the expected time required to seek and handle prey type i and v_i is the expected amount of energy intake obtained by handling prey i . The average rate of encounter with prey i in searching process is λ_i . p_i is the probability that prey type i is handled once it is found and recognized. Hence, the average rate of gain of a forager, i.e., J , can be defined as a ratio of expected energy intake to expected time spent for searching and handling the prey types and expressed as

$$J = \frac{\sum_{i=1}^k p_i \lambda_i v_i}{\sum_{i=1}^k p_i \lambda_i t_i} \quad (7.1)$$

p_i is a decision variable for a forager and J can be maximized by finding the optimal values of $p_i, \forall i$. Assume that the forager has previously handled the number of $n-1$ prey types, i.e., prey 1, prey 2, ..., prey $(n-1)$. The rate of gain, J , can be given as $J = \sum_{i=1}^{n-1} p_i \lambda_i v_i / \sum_{i=1}^{n-1} p_i \lambda_i t_i$.

Let assume that the forager currently encounters with prey type n . In this case, J becomes

$$J = \frac{p_n \lambda_n v_n + \sum_{i=1}^{n-1} p_i \lambda_i v_i}{p_n \lambda_n t_n + \sum_{i=1}^{n-1} p_i \lambda_i t_i} \quad (7.2)$$

For the optimal p_n , J can be differentiated with respect to p_n and by considering $\frac{\partial J}{\partial p_n} = 0$, the optimal p_n to maximize J can be found by the following zero-one-rule [110]:

$$\begin{aligned} \text{set } p_n = 0 \quad & \text{if } \frac{v_n}{t_n} < \frac{\sum_{i=1}^{n-1} \lambda_i v_i}{\sum_{i=1}^{n-1} \lambda_i t_i} \\ \text{set } p_n = 1 \quad & \text{if } \frac{v_n}{t_n} > \frac{\sum_{i=1}^{n-1} \lambda_i v_i}{\sum_{i=1}^{n-1} \lambda_i t_i} \end{aligned} \quad (7.3)$$

where $\frac{v_n}{t_n}$ is called *prey selection profitability* and used to characterize the profitability of prey n . Based on the above zero-one-rule, each forager can autonomously decide prey types to handle.

In the prey model, there exists a natural uncertainty in geographical availability of the prey types. Therefore, the forager is subject to a starvation risk if it cannot obtain a sufficient energy intake. However, biological foragers have great capabilities to effectively tackle with the starvation risk. These capabilities enable a forager to balance its search effort and energy intake. If it consumes most of its residual energy in prey search, it cannot harvest an adequate energy intake by capturing and handling the prey. In this case, it changes its search strategy to alleviate the starvation risk by slowing down its energy depletion rate in the prey search [113]. Such interactions of foragers and prey types form a *foraging game* in which forager individuals balance conflicting demands of prey and starvation risk [114]. The foraging model reviewed here establishes a solid basis for design of solution approaches to virtually any kind of control and optimization problem [115][116][117][67].

7.4 Biological Foraging-inspired Communication in IMCRN

In this section, we first present the analogies between the prey model and IMCRN. Then, based on the given analogies, we introduce biological foraging-inspired communication (BFC) algorithm.

Table 7.1: Relationship between Prey Model and IMCRN.

PREY MODEL	IMCRN
<i>Forager</i>	<i>IMCRN node</i>
<i>Prey types</i>	<i>Relays & Licensed channels</i>
<i>Forager determines which prey types should be handled to maximize its energy intake with minimum search effort</i>	<i>IMCRN node determines which relay nodes should be used to provide minimum energy consumption without sacrificing end-to-end delay & IMCRN node determines which licensed channels should be used to maximize spectrum utilization with minimum interference to primary system</i>

7.4.1 Inspiration from Prey Model

Here, in order to create a biological foraging-inspired communication model for IMCRN, we consider the network nodes as foragers. Similar to a biological forager searching for prey, each node with message searches for possible relay nodes to copy its message. Hence, possible relay nodes of a node are considered as prey types in the prey model. Similar to the prey model, we define a profitability measure named as *relay selection profitability* (RSP). Using RSP, each node having message selects its relay nodes in order to provide a sufficient level of end-to-end latency while examining its transmission effort.

We also consider available licensed channels as prey types and define another profitability measure called *channel selection profitability* (CSP). Based on CSP, each node with message selects and uses licensed channels so as to maximize spectrum utilization with minimum interference to primary system. The established analogies and relationships, introduced above, are outlined in Table 7.1. Next, we derive RSP and CSP measures by using the given analogies.

7.4.2 Relay Selection Profitability

In the prey model, predator i needs the time duration t_j to search for and handle prey type j . t_j also implicitly corresponds to an amount of energy that predator i consumes in searching and handling the prey type j . If predator i handles the prey type j , it also gains an expected energy intake, i.e., v_j , as introduced in Section 7.3. Similar to a forager, IMCRN node i with message needs an amount of energy e_{ij} to copy the message to relay node j . If node i copies the message to node j , it also gains the utility v_{jd} to deliver the message to destination node

d . Hence, assuming that node i copies its message to each node j with the probability p_{ij} , the rate of gain of node i can be expressed by the ratio, $\mathcal{G}_i$, given as

$$\mathcal{G}_i = \frac{\sum_{j \in S_i} p_{ij} v_{jd}}{\sum_{j \in S_i} p_{ij} e_{ij}} \quad (7.4)$$

where the numerator of $\mathcal{G}_i$ denotes the expected utility gained by copying the message to the selected relay nodes while the denominator is the expected energy consumed by copying the message to the relays. S_i is the contact set that includes IDs of nodes in the communication radius of node i . p_{ij} is a decision variable for node i in selection of its relay nodes. By finding optimal $p_{ij}, \forall j \in S_i$, $\mathcal{G}_i$ can be maximized. This also enables node i to copy the message to the most appropriate relay nodes with high utility values by consuming minimum energy. These relay nodes with high utility can also minimize the end-to-end message delay. In order to show this, let assume that node i copies the message to relay node j and relay node j encounters with destination node d with probability α in each time interval I . The expected increase of v_{jd} in each time interval I , i.e., u , can be given as

$$\begin{aligned} u &= \alpha(\xi_1 + \xi_2) - (1 - \alpha)\xi_1 \\ &= \alpha(2\xi_1 + \xi_2) - \xi_1 \end{aligned} \quad (7.5)$$

u is a strictly increasing function of α , that is, v_{jd} increases with α . On the other hand, using α , the expected number of time intervals I in which relay node j meets with destination node d , i.e., γ , can be also found as

$$\begin{aligned} \gamma &= \sum_{i=1}^{\infty} i\alpha(1 - \alpha)^{i-1} \\ &= \frac{1}{\alpha} \end{aligned} \quad (7.6)$$

As observed in (7.6), γ is a strictly decreasing function of α . Consequently, combining the results obtained in (7.5) and (7.6), it can be easily concluded that by maximizing v_{jd} , the expected delay of node j to encounter with destination node d can be minimized. Thus, the aim of node i is to maximize $\mathcal{G}_i$ by choosing optimal p_{ij} values. To this end, (7.4) can be rewritten as

$$\mathcal{G}_i = \frac{p_{ij} v_{jd} + \eta_i}{p_{ij} e_{ij} + \kappa_i} \quad (7.7)$$

where η_i and κ_i can expressed as

$$\eta_i = \sum_{\substack{z \in S_i \\ z \neq j}} p_{iz} v_{zd} \quad \text{and} \quad \kappa_i = \sum_{\substack{z \in S_i \\ z \neq j}} p_{iz} e_{ij} \quad (7.8)$$

For the optimal p_{ij} , (7.7) can be differentiated with respect to p_{ij} , i.e.,

$$\begin{aligned}\frac{\partial \mathcal{G}_i}{\partial p_{ij}} &= \frac{v_{jd}(\kappa_i + p_{ij}e_{ij}) - e_{ij}(\eta_i + p_{ij}v_{jd})}{(\kappa_i + p_{ij}e_{ij})^2} \\ &= \frac{v_{jd}\kappa_i - e_{ij}\eta_i}{(\kappa_i + p_{ij}e_{ij})^2}.\end{aligned}\quad (7.9)$$

Based on the differentiation in (7.9), p_{ij} can be found in order to maximize $\mathcal{G}_i$ by the following way:

- If $\frac{v_{jd}}{e_{ij}} < \frac{\eta_i}{\kappa_i}$ is satisfied, $\frac{\partial \mathcal{G}_i}{\partial p_{ij}}$ is negative and $\mathcal{G}_i$ can be maximized by choosing the lowest possible p_{ij} , that is, $p_{ij} = 0$.
- Conversely, if $\frac{v_{jd}}{e_{ij}} > \frac{\eta_i}{\kappa_i}$ is satisfied, $\frac{\partial \mathcal{G}_i}{\partial p_{ij}}$ is positive and $\mathcal{G}_i$ can be maximized by choosing the highest possible p_{ij} , that is, $p_{ij} = 1$.

Consequently, the selection of p_{ij} to maximize $\mathcal{G}_i$ leads to a zero-one-rule:

$$\begin{aligned}\text{set } p_{ij} &= 0 & \text{if } \frac{v_{jd}}{e_{ij}} < \frac{\eta_i}{\kappa_i} \\ \text{set } p_{ij} &= 1 & \text{if } \frac{v_{jd}}{e_{ij}} > \frac{\eta_i}{\kappa_i}\end{aligned}\quad (7.10)$$

where $\frac{v_{jd}}{e_{ij}}$ is the *relay selection profitability* and characterizes the profitability of node j that node i gains if it uses node j as a relay to deliver the message to destination node d .

7.4.3 Channel Selection Profitability

Here, we derive the CSP measure to permit IMCRN nodes to decide which licensed channels should be used. The aim of CSP is to maximize spectrum utilization with minimum interference to primary system. Let assume that there are M licensed channels with different bandwidth values and ω_c denotes the bandwidth of channel c , where $c \in \mathcal{K} = \{1, 2, \dots, M\}$. Each IMCRN node is also assumed to periodically sense a set of M licensed channels. A_i denotes the set including IDs of licensed channels that are periodically sensed by node i . Suppose that channel c is periodically sensed by node i in each slot and channel c is idle during the time interval ρ_c called *channel idle duration*. Here, we use the product of channel bandwidth ω_c and the channel idle duration ρ_c , i.e., $\mathcal{E}_c = \omega_c \rho_c$, as a metric to examine the channel idleness. Furthermore, failures in the sensing of primary users are assumed to cause the collisions among the transmissions of primary users and IMCRN nodes. Such collisions

are also assumed to impose an interference level on the primary system. As the primary user activities become heavy, the likelihood of the collisions and thus, interference level increases. Let n_{ic} denotes the interference level incurred by IMCRN node i if it uses the licensed channel c . Hence, total rate of gain of node i in licensed channel selection, i.e., C_i , is given as the ratio of expected $\mathcal{E}_c$ to expected n_{ic} , i.e.,

$$C_i = \frac{\sum_{c \in A_i} a_{ic} \mathcal{E}_c}{\sum_{c \in A_i} a_{ic} n_{ic}} = \frac{\sum_{c \in A_i} a_{ic} \omega_c \rho_c}{\sum_{c \in A_i} a_{ic} n_{ic}} \quad (7.11)$$

where a_{ic} is the probability that node i uses channel c . a_{ic} is also a decision variable for node i . The maximization of C_i clearly allows node i to use mostly idle licensed channels with high bandwidth while minimizing expected interference level imposed by node i on primary system. Thus, the aim of node i is to maximize C_i by choosing optimal $a_{ic}, \forall c$. For this aim, (7.11) can be rewritten as follows

$$C_i = \frac{a_{ic} \omega_c \rho_c + \zeta_i}{a_{ic} n_{ic} + \lambda_i} \quad (7.12)$$

where ζ_i and λ_i are

$$\zeta_i = \sum_{\substack{z \in \mathcal{K} \\ z \neq c}} a_{iz} \omega_z \rho_z \quad \text{and} \quad \lambda_i = \sum_{\substack{z \in \mathcal{K} \\ z \neq c}} a_{iz} n_{iz} \quad (7.13)$$

(7.12) can be also differentiated with respect to a_{ic} in order to find optimal a_{ic} , i.e.,

$$\begin{aligned} \frac{\partial C_i}{\partial a_{ic}} &= \frac{\omega_c \rho_c (a_{ic} n_{ic} + \lambda_i) - n_{ic} (a_{ic} \omega_c \rho_c + \zeta_i)}{(a_{ic} n_{ic} + \lambda_i)^2} \\ &= \frac{\omega_c \rho_c \lambda_i - n_{ic} \zeta_i}{(a_{ic} n_{ic} + \lambda_i)^2}. \end{aligned} \quad (7.14)$$

Based on the differentiation, a_{ic} can be determined in order to maximize C_i by the following way:

- If $\frac{\omega_c \rho_c}{n_{ic}} < \frac{\zeta_i}{\lambda_i}$ is satisfied, $\frac{\partial C_i}{\partial a_{ic}}$ is negative and C_i can be maximized by choosing the lowest possible a_{ic} , that is, $a_{ic} = 0$.
- Conversely, if $\frac{\omega_c \rho_c}{n_{ic}} > \frac{\zeta_i}{\lambda_i}$ is satisfied, $\frac{\partial C_i}{\partial a_{ic}}$ is positive and C_i can be maximized by choosing the highest possible a_{ic} , that is, $a_{ic} = 1$.

Consequently, the selection of a_{ic} leads to a zero-one-rule:

$$\begin{aligned} \text{set } a_{ic} = 0 & \quad \text{if } \frac{\omega_c \rho_c}{n_{ic}} < \frac{\zeta_i}{\lambda_i} \\ \text{set } a_{ic} = 1 & \quad \text{if } \frac{\omega_c \rho_c}{n_{ic}} > \frac{\zeta_i}{\lambda_i} \end{aligned} \quad (7.15)$$

where $\frac{\omega_c \rho_c}{n_{ic}}$ is the *channel selection profitability* and characterizes the profitability of channel c that node i gains if it uses channel c . Next, using the results summarized in (7.10) and (7.15), we introduce the relay selection and channel selection algorithms in BFC.

7.4.4 Relay Selection Algorithm in Biological Foraging-inspired Communication

BFC selects the relay nodes within only three hops. Source node first copies the message to a set of relay nodes that it encounters and these relay nodes of source also follow the same strategy with the source node. Then, one of the relay nodes having the message immediately delivers the message to the destination node d provided that it meets. BFC allows each node to copy the message using a predefined energy budget called as E . After the depletion of its energy budget E , nodes stop copying the message. The relay selection is accomplished via the zero-one-rule given in (7.10). Let assume that node i with message encounters with node j and receives the utility value v_{jd} from node j . In order to decide whether it copies the message to node j , node i first computes the energy needed for delivering the message to node j . Here, we assume that each node i discovers the locations of its and their contacts using one of the existing localization methods² in the literature [105]. By using the location information, it also computes the Euclidean distances to their neighbors. Let l_{ij} be the Euclidean distance between node i and j . Using l_{ij} , node i computes energy e_{ij} required to deliver the message to node j as $e_{ij} = o + r l_{ij}^s$, where s is a positive constant that characterizes the power loss in wireless transmission, o is the dissipated energy in the transmitter circuitries. r is the energy needed by transmitter amplifier [107]. Hence, as node i meets with relay nodes and discovers v_{jd} and e_{ij} , $\forall j \in S_i$, it decides whether to copy the message to relay node j by directly using the zero-one-rule given in (7.10) as follows. It first sorts all RSP values of its current contacts, i.e., $\frac{v_{jd}}{e_{ij}}$, $\forall j \in S_i$, in ascending order and copies the message to node j such that $\{j \mid \forall k \in S_i : \frac{v_{kd}}{e_{ik}} > \frac{v_{jd}}{e_{ij}}\}$. Then, according to the ascending order of RSP values, it checks which contacts of it satisfy the zero-one-rule in (7.10). Whenever a contact satisfies the zero-one-rule, node i copies the message this contact and updates η_i and κ_i in (7.8). Then, it continues to check whether or not its remaining contacts satisfy the zero-one-rule by considering the updated η_i and κ_i . Furthermore, node i also updates and checks its residual energy budget whenever it copies the message to a relay node. If it depletes all of its energy budget E , it immediately stops copying the message.

² The details of localization procedure is beyond the scope of this chapter.

7.4.5 Channel Selection Algorithm in Biological Foraging-inspired Communication

In BFC, the channel selection follows the zero-one-rule given in (7.15). Each node i periodically senses a set of all available licensed channels and it determines and updates the idle duration of all sensed channels, i.e., $\rho_c, \forall c \in A_i$. Bandwidth of each channel, i.e., $\omega_c, \forall c$, is assumed to be known by all IMCRN nodes. For each sensed channel, node i is also assumed to estimate possible interference level that it may impose on a channel if it uses this channel.³ Hence, using ω_c, ρ_c , and n_{ic} , node i computes all CSP values and selects a channel c such that $\{c \mid \forall k \in A_i : \frac{\omega_k \rho_k}{n_{ik}} < \frac{\omega_c \rho_c}{n_{ic}}\}$. Such a selection strategy does not conflict with the zero-one-rule in (7.15). Inherently, it usually satisfies $\frac{\omega_c \rho_c}{n_{ic}} > \frac{\xi_i}{\lambda_i}$ as given in (7.15).

Lemma 7.4.1 *Channel selection algorithm of BFC enables each idle licensed channel c to be definitely allocated by an IMCRN node within a time duration, i.e., τ , such that τ always has an upper-bound, i.e., τ_u , that can be expressed as*

$$\tau_u = \frac{\omega_{\max} n_{ic} m}{\omega_{\max} n_{ic} - \omega_c n_{\min}} \quad (7.16)$$

where $n_{\min}$ and $n_{\max}$ are the minimum and maximum of the interference, i.e., $n_{\min} \leq n_{ic} \leq n_{\max}, \forall i, c$. $\omega_{\max}$ is the maximum of all channel bandwidth values, i.e., $\omega_c \leq \omega_{\max}, \forall c$. Furthermore, m is the time duration required for successfully transmitting the message to a relay and m is called as message duration.

Proof. Assume that at a time instant, channel c has been unused during τ and all of other channels are used at least once within τ . In this time instant, the channel selection profitability of channel c , i.e., ψ_{ic} , can be computed by any node i as

$$\psi_{ic} = \frac{\omega_c \rho_c}{n_{ic}} = \frac{\omega_c \tau}{n_{ic}} \quad (7.17)$$

where n_{ic} is the interference level imposed on channel c by node i . Also, suppose that channel k has the maximum CSP at this time instant. Since channel k has been used at least once within τ , the possible maximum value for ρ_k can be given as $\rho_k = \tau - m$. Channel k may also have the maximum bandwidth and any node j may incur the minimum interference to primary system if it uses channel k . In this case, ω_k and n_{jk} can be considered as $\omega_k = \omega_{\max}$ and $n_{jk} = n_{\min}$, respectively. Hence, by setting $\rho_k = \tau - m$, $\omega_k = \omega_{\max}$, and $n_{jk} = n_{\min}$, node j

³ Note that the estimation of the interference level is out of the scope of this chapter.

computes the CSP of channel k , i.e., ψ_{jk} , as follows

$$\psi_{jk} = \frac{\omega_{max}(\tau - m)}{n_{min}} \quad (7.18)$$

In order to ensure that channel c is used before channel k ,

$$\begin{aligned} \psi_{ic} &> \psi_{jk} \\ \frac{\omega_c \tau}{n_{ic}} &> \frac{\omega_{max}(\tau - m)}{n_{min}} \end{aligned} \quad (7.19)$$

has to be satisfied provided that

$$\tau < \tau_u = \frac{\omega_{max} n_{ic} m}{\omega_{max} n_{ic} - \omega_c n_{min}} \quad (7.20)$$

■

In IMCRN, successful channel utilization of a transmitter depends on three different duration.

1. Contact duration between the transmitter and its receiver, i.e., ϕ_1 .
2. Idle duration of the channel, i.e., ϕ_2 .
3. Idle duration of CCC, i.e., ϕ_3 .

More specifically, the successful utilization relies on how long nonzero ϕ_1 , ϕ_2 , and ϕ_3 coexist at the same time. Let consider that ϕ_1 , ϕ_2 , and ϕ_3 coexist during a time interval, i.e., θ , which is called *feasible channel utilization interval*. For a given θ , BFC allows transmitter nodes to utilize any vacant channel by a *successful channel utilization probability* derived by the following theorem.

Theorem 7.4.2 *Assuming that n_{ic} and ω_c , $\forall i, c$ are characterized by the normal random variables $\mathbf{n} \sim \mathcal{N}(\mu_n, \sigma_n^2)$ and $\omega \sim \mathcal{N}(\mu_\omega, \sigma_\omega^2)$, respectively, the successful channel utilization probability of BFC, i.e., β , can be approximated as*

$$\beta \cong \int_{-\infty}^{\vartheta} \Phi(s) ds \quad (7.21)$$

where $\Phi(s)$ is the standard normal density function and ϑ is given as

$$\vartheta = \frac{\frac{(\theta - m)(\omega_{max} \mu_n - n_{min} \mu_\omega) - \omega_{max} \mu_n m}{\sigma_n \omega_{max} m}}{\sqrt{1 + \left(\frac{(\theta - m) \sqrt{\omega_{max}^2 \sigma_n^2 - n_{min}^2 \sigma_\omega^2}}{\sigma_n \omega_{max} m} \right)^2}} \quad (7.22)$$

Proof. In order to ensure that any idle licensed channel can be always successfully used by an IMCRN node, the feasible channel utilization interval θ must be higher than the sum of the delay bound τ_u and the message duration m , i.e.,

$$\begin{aligned}\tau_u + m &< \theta \\ \frac{\omega_{max}n_{ic}m}{\omega_{max}n_{ic} - \omega_c n_{min}} + m &< \theta\end{aligned}\quad (7.23)$$

Based on $\mathbf{n} \sim \mathcal{N}(\mu_n, \sigma_n^2)$ and $\omega \sim \mathcal{N}(\mu_\omega, \sigma_\omega^2)$, (7.23) can be modified as

$$\frac{\omega_{max}\mathbf{n}m}{\omega_{max}\mathbf{n} - \omega n_{min}} < \theta - m \quad (7.24)$$

Since ω_{max} and n_{min} are assumed to be constant and the linear operations over normal random variables also produce normal random variables, (7.24) can be also expressed as

$$\frac{x}{y} < \theta - m \quad (7.25)$$

where $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ and

$$\begin{aligned}\mu_x &= \omega_{max}\mu_n m, \quad \sigma_x^2 = (\sigma_n \omega_{max} m)^2 \\ \mu_y &= \omega_{max}\mu_n - n_{min}\mu_\omega, \quad \sigma_y^2 = \omega_{max}^2 \sigma_n^2 - n_{min}^2 \sigma_\omega^2\end{aligned}\quad (7.26)$$

By standardizing the normal random variables x and y , (7.25) can be also written as

$$\frac{z_x + \frac{\mu_x}{\sigma_x}}{z_y + \frac{\mu_y}{\sigma_y}} < \frac{\sigma_y(\theta - m)}{\sigma_x} \quad (7.27)$$

where z_x and z_y are the standard normal random variables, i.e., $z_x \sim \mathcal{N}(0, 1)$ and $z_y \sim \mathcal{N}(0, 1)$. Consequently, if (7.27) is satisfied, it can be ensured that any idle licensed channel can be successfully used by an IMCRN node. The probability that (7.27) is satisfied, i.e., β , can be also approximated as [127]

$$\beta = Pr\left(\frac{z_x + \frac{\mu_x}{\sigma_x}}{z_y + \frac{\mu_y}{\sigma_y}} < \frac{\sigma_y(\theta - m)}{\sigma_x}\right) \cong \int_{-\infty}^{\vartheta} \Phi(s) ds \quad (7.28)$$

where ϑ can be given as

$$\vartheta = \frac{\frac{\mu_y(\theta - m) - \mu_x}{\sigma_x}}{\sqrt{1 + \left(\frac{\sigma_y(\theta - m)}{\sigma_x}\right)^2}} \quad (7.29)$$

By substituting μ_x, σ_x, μ_y and σ_y into (7.29), ϑ can be also rewritten as given in (7.22). The successful channel utilization probability, β , is used to examine the performance of BFC algorithm in Section 7.5.2. ■

7.5 Performance Evaluation

In this section, we present the performance evaluation of BFC algorithm. The simulation environment is created using MATLAB. We consider a generic mobility and spectrum availability model. The model assumes that any two nodes are contacted and able to communicate with each other using an available licensed channel with the probability p_x . The probability p_x is also assumed to be uniformly distributed over the interval $(0, p_u]$. p_u implicitly characterizes the dynamics of node contact statistics and heaviness of primary user activities. In fact, many popular mobility models such as random direction, random waypoint, and many synthetic mobility models such as community-based mobility follow the similar models [128], [129], [130]. Throughout this section, end-to-end message delay is given in number of time intervals, I , used in the utility update as introduced in Section 7.2 and node energy values are scaled to the interval $[0, 1]$. Next, we first present the results on the relay selection performance of BFC then, the channel selection performance is given.

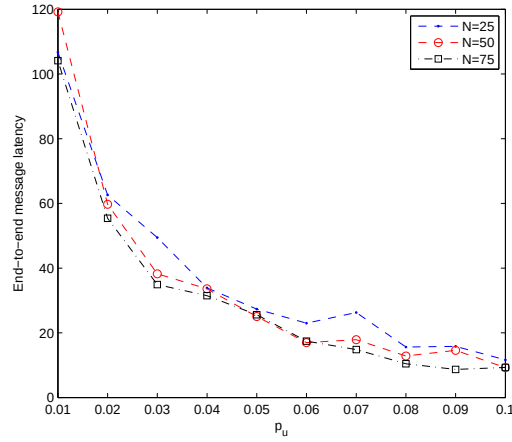


Figure 7.3: End-to-end delay is shown with varying upper bounds of encounter probability, i.e., p_u , for different number of nodes.

7.5.1 Relay Selection Performance

In Fig. 7.3, end-to-end delay provided by BFC is shown with varying p_u values according to different number of nodes in the network, i.e., N . As p_u increases, the average of p_x selected from $0 < p_x \leq p_u$ increases, which reduces the end-to-end message delay. However, the

message delay slightly decreases with N . This is observed more apparently as p_u is relatively small, i.e., less than 0.05. Hence, it can be easily concluded that BFC can successfully utilize all contact opportunities regardless of the number of nodes in the network because it provides nearly the same end-to-end performance even if N changes from 25 to 75.

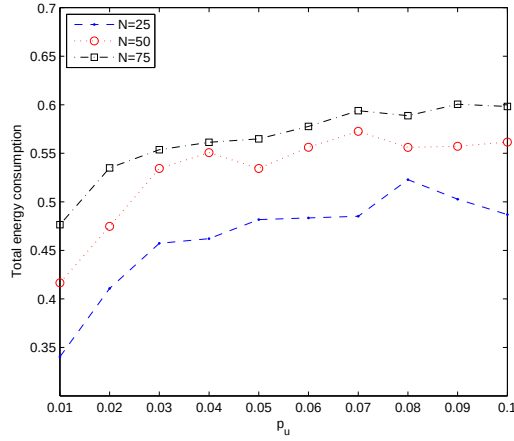


Figure 7.4: Total energy consumption of nodes in message delivery is shown with varying upper bounds of encounter probability, i.e., p_u , for different number of nodes.

In Fig. 7.4, total energy consumption of network nodes in message delivery is shown for the same setting in Fig. 7.3. For a fixed N , the total energy consumption hardly increases while end-to-end latency significantly reduces as seen in Fig. 7.3. This reveals that BFC can always select the most appropriate relay nodes that can minimize the energy consumption. For example, for $N = 25$, end-to-end latency can be decreased from nearly 100 to 10. On the other hand, total energy consumption only increases for nearly 0.47 to 0.55. This reveals that BFC can always follow the changes in the network and regulate its strategy to provide an acceptable latency with minimum energy consumption.

In Fig. 7.5, for the same scenario evaluated in Fig. 7.3 and 7.4, the number of relay nodes that are employed for the message delivery is shown with varying p_u values for the different values of N . The number of relay nodes increases with p_u . However, as observed in Fig. 7.4, total energy consumption slightly changes although the number of employed relays increases nearly 350%. Hence, it can be easily inferred that BFC can always decide the relay nodes that minimize total energy consumption. Furthermore, as p_u increases, BFC can adaptively track the changes in p_u and decrease the end-to-end latency by increasing the number of relays although it keeps the energy consumption in almost the same level.

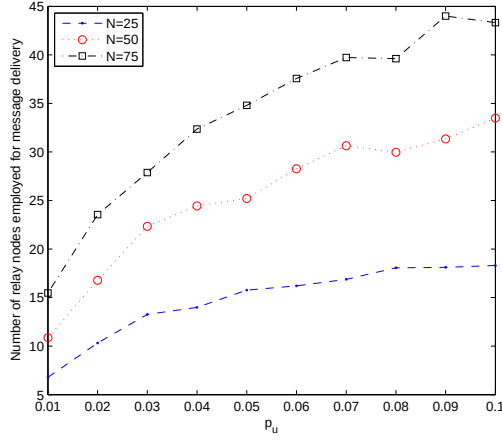


Figure 7.5: Number of relay nodes that selected by BFC is shown with p_u according to different N values.

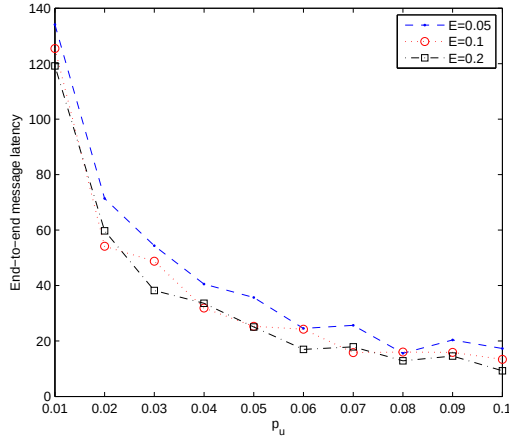


Figure 7.6: End-to-end delay is shown with varying p_u values for different values of energy budget E .

In Fig. 7.6, we observe the effect of energy budget E on the end-to-end latency by varying p_u for different values of E . BFC reduces the end-to-end latency regardless of E as p_u increases. Furthermore, for the same setting, total energy consumption is also shown in Fig. 7.7. Total energy consumption can be also reduced by using a low level of energy budget without sacrificing end-to-end delay performance. Hence, it is essential for BFC to determine an optimal energy budget level that is not redundant to provide a desired end-to-end latency. Apparently, as a result of the increase in total energy consumption, the number of relays also increases with p_u as shown in Fig. 7.8. The number of redundant relays can be also alleviated by determining an optimal energy budget of nodes. Note that this determination is remained

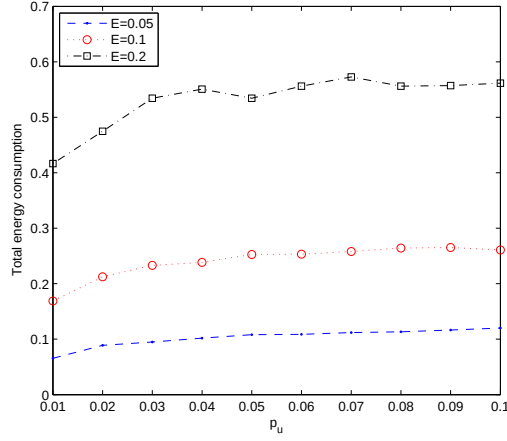


Figure 7.7: Total energy consumption with varying p_u for different E values.

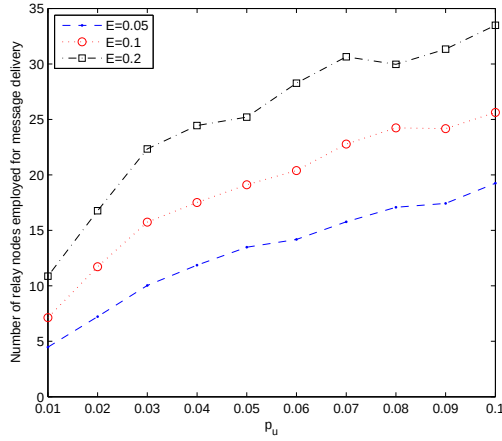


Figure 7.8: Number of relay nodes with varying p_u for different values of E .

a future work of this chapter.

7.5.2 Channel Selection Performance

In this section, the channel selection performance of BFC is presented in order to understand how BFC enables IMCRN nodes to efficiently utilize the available spectrum channels. We give the performance results on the upper bound of channel allocation time, i.e., τ_u , given in (7.16). Interference level $\mathbf{n}$ and channel bandwidths ω are assumed to be normal random variables, i.e., $\mathbf{n} \sim \mathcal{N}(\mu_n, \sigma_n^2)$ and $\omega \sim \mathcal{N}(\mu_\omega, \sigma_\omega^2)$. Since 99.7% of values drawn from a

normal distribution are within three standard deviations, n_{min} and ω_{max} can be given as $n_{min} \approx \mu_n - 3\sigma_n$ and $\omega_{max} \approx \mu_\omega + 3\sigma_\omega$. Note that throughout the evaluations, all of the parameters μ_n , σ_n , μ_ω , and σ_ω are scaled to the interval (0,1].

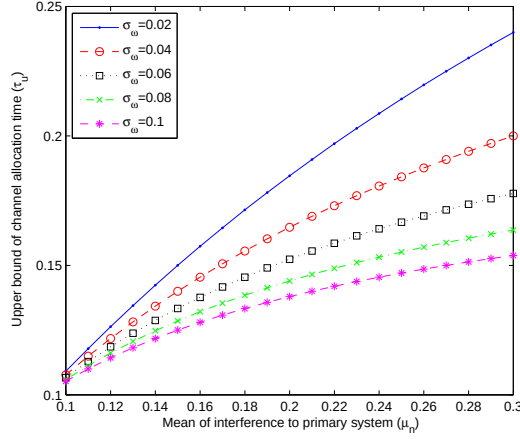


Figure 7.9: Upper bound of channel allocation time, τ_u , is shown with varying μ_n values for different values of σ_ω .

In Fig. 7.9, the upper bound τ_u of the channel allocation time is shown with varying values of μ_n for different σ_ω values. The upper bound τ_u increases with μ_n . This helps the nodes that impose a higher level of interference on some channels to access these channels later than the other nodes. Hence, BFC mitigates interference along the network by delaying the accesses of the nodes that cause high interference. Furthermore, τ_u can be reduced as the standard deviation σ_ω of the channel bandwidth increases. This is because diversity of the channel bandwidths can be improved with σ_ω and as σ_ω increases, the zero-one-rule in (7.15) selects various channels more rapidly. Thus, BFC improves the performance as the diversity of these channels increases.

In Fig. 7.10, τ_u is shown with the standard deviation σ_n of interference for varying values of σ_ω . τ_u can be decreased by increasing σ_n . In fact, as σ_n increases, number of nodes, that can easily find various licensed channel opportunities without interfering the primary system, increases. This reveal that BFC has a potential to be able to convert the various network conditions into a performance improvement.

In Fig. 7.11, τ_u is presented with varying mean μ_ω of channel bandwidth for different σ_n values. τ_u increases with μ_ω . For a fixed σ_ω , as μ_ω increases, channel bandwidth values

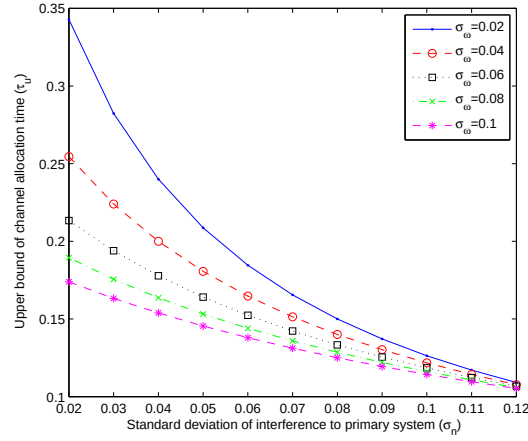


Figure 7.10: Upper bound of channel allocation time τ_u is shown with varying σ_n values for different σ_w values.

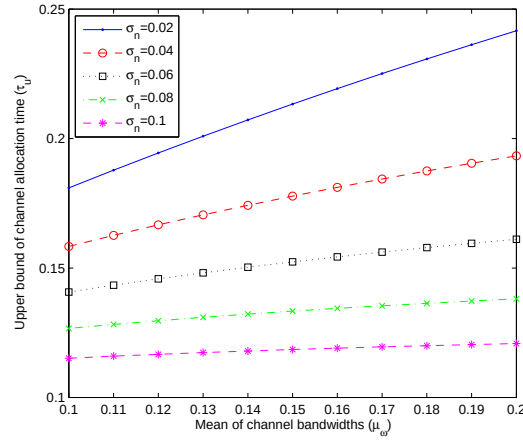


Figure 7.11: Upper bound of channel allocation time τ_u is shown with varying μ_w values for different σ_n values.

becomes closer to each other. This slows down the relative increment of CSP values with time and the channels are selected more slowly. Similar to Fig. 7.9 and 7.10, the increase in σ_n reduces τ_u since nodes can find the channels with low level of interference as σ_n increases. Contrary to Fig. 7.11, in Fig. 7.12, increase in σ_w reduces the upper bound of the channel allocation delay. This stems from the fact that more diverse channel bandwidth values allows the nodes to easily select alternative channels with high level of bandwidth without severely interfering the primary system. Hence, it can be inferred that BFC can provide sufficiently small upper bound for the delay in the channel allocation process. As the network condition and parameters change, BFC can adaptively regulate its strategy in channel selection so as to

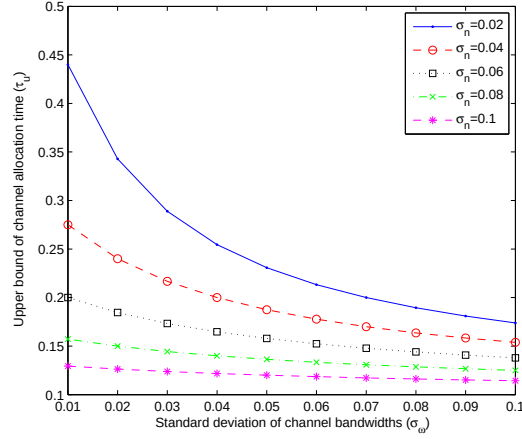


Figure 7.12: Upper bound of channel allocation time τ_u is shown with varying σ_ω values for different σ_n values.

maximize spectrum utilization while minimizing interference to primary system.

Table 7.2: The successful channel utilization probability, β , according to varying m and σ_n

Message duration (m)	0.03	0.04	0.05	0.06	0.07	0.08	0.09
$\sigma_n = 0.1$	0.99999	0.999906	0.998877	0.98954	0.94388	0.829704	0.664625
$\sigma_n = 0.15$	0.999041	0.997196	0.98983	0.963556	0.895365	0.777109	0.634034
$\sigma_n = 0.2$	0.99626	0.991719	0.978061	0.941121	0.864124	0.748379	0.618475

In Table 7.2, the successfully channel utilization probability, i.e., β , is presented with varying message duration values, i.e., m , that are changed from $m = 0.03$ to $m = 0.09$ while the feasible channel utilization interval is $\theta = 0.1$. β is very close to 1 for $m = 0.03 - 0.06$. This means that for these message duration values, BFC provides almost 100% spectrum utilization. As m is further increased up to $m = 0.09$, β is still relatively high although m is very close to the feasible channel utilization interval, i.e., $\theta = 0.1$.

Similar to Table 7.2, in Table 7.3, it can be observed that BFC provides nearly %100 spectrum utilization when the message duration is selected appropriately. Note that BFC also produces almost %80 spectrum utilization although the message duration is very close to the feasible channel utilization interval such that $m = 0.08$ while $\theta = 1$.

Finally, in Table 7.4, β is presented for varying values of m with different θ values. Here, the

Table 7.3: The successful channel utilization probability, β , according to varying m and σ_ω

Message duration (m)	0.03	0.04	0.05	0.06	0.07	0.08	0.09
$\sigma_\omega = 0.1$	0.998334	0.995626	0.986009	0.955435	0.883087	0.76515	0.627292
$\sigma_\omega = 0.15$	0.998772	0.99658	0.988282	0.960167	0.890131	0.771947	0.631107
$\sigma_\omega = 0.2$	0.999041	0.997196	0.98983	0.963556	0.895365	0.777109	0.634034

results are taken by setting m always less than θ . Therefore, some of the boxes in the table remains empty. As observed, BFC provides a significantly high spectrum utilization even if the message duration is selected in close proximity of the feasible channel utilization interval.

Table 7.4: The successful channel utilization probability, β , for varying m and θ

Message duration (m)	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$\theta = 0.9$	0.995168	0.979435	0.928433	0.820497	0.682914	0.57217	0.512942
$\theta = 0.8$	0.994649	0.974788	0.907276	0.774672	0.631597	0.537278	—
$\theta = 0.7$	0.993734	0.965855	0.869008	0.706673	0.569292	—	—
$\theta = 0.6$	0.991879	0.946159	0.797996	0.612853	—	—	—
$\theta = 0.5$	0.987222	0.895997	0.674295	—	—	—	—

7.6 Conclusion

In this chapter, based on a prey model in foraging theory, biological foraging-inspired communication (BFC) protocol is introduced for intermittently-connected mobile cognitive radio ad hoc networks (IMCRN). Using a foraging-inspired relay selection profitability measure, BFC allows each node with message to decide whether to copy the message to a relay node by optimizing the trade off between its transmission effort and utility in message delivery. This mechanism eventually controls total number of relay nodes used in message delivery in order to provide a sufficient level of message delay. Using a channel selection profitability measure, BFC also provides nearly %100 spectrum utilization while it minimizes the interference level to primary system. BFC is a fully-autonomous protocol. Without need for any a priori information or regulation, it enables network nodes to adaptively regulate their communication strategies according to dynamically changing network environment.

CHAPTER 8

On Channel Capacity and Error Compensation in Molecular Communication

In this chapter, we introduce an information theoretical approach for molecular communication and propose a closed form expression for molecular communication capacity between two nanomachines and propose an adaptive error compensation technique for molecular communication by significantly extending our preliminary work in [131]. The extended version of this work was also introduced in [132]. Using the principles of mass action kinetics, we first model the molecular delivery between two nanomachines called Transmitter Nanomachine (TN) and Receiver Nanomachine (RN). Then, based on the molecular delivery model, we derive the closed form expression for capacity of the channel between TN and RN. In this chapter, we also propose an adaptive Molecular Error Compensation (MEC) scheme for the molecular communication between TN and RN. We first define an interval for selection of the most appropriate molecular bit transmission probability providing higher molecular communication capacity with minimum error. Then, using this interval, we introduce a selection strategy to enable TN to select the most appropriate molecular bit transmission probability with respect to some environmental factors such as temperature, binding rate, distance between nanomachines. MEC allows TN and RN to collaboratively select the most appropriate molecular bit transmission probability providing high molecular communication capacity. Finally, using the capacity expression and the error compensation scheme, we investigate how the conditions such as temperature of environment, concentration of emitted molecules, distance between nanomachines and duration of molecule emission affect the molecular communication capacity and molecular bit transmission probability that provides higher molecular communication capacity. We further discuss under which conditions the molecular communication can be feasible with high capacity.

In Section 8.1, the motivation and the related works are introduced. In Section 8.2, we introduce a molecular communication model. In Section 8.3, we introduce a molecule delivery approach for the molecular communication between two nanomachines. In Section 8.4, based on the molecule delivery scheme we introduce an information theoretical approach for the molecular communication between two nanomachines. In Section 8.5, we propose an adaptive error compensation scheme. In Section 8.6, we provide the numerical results and we give concluding results in Section 8.7.

8.1 Motivation and Related Works

Molecular Communication is a new interdisciplinary research area including the nanotechnology, biotechnology, and communication technology [39]. In nature, molecular communication is one of the most important biological functions in living organisms to enable biological phenomena to communicate with each other. For example, in an insect colony, insects communicate with each other by means of pheromone molecules. When an insect emits the pheromone molecules, some of them bind the receptors of some insects in the colony and these insects convert the bound pheromone molecules to biologically meaningful information. This enables the insects in the colony to communicate with each other. Similar to insects, almost all of the biological systems in nature perform intra-cellular communication through vesicle transport, inter-cellular communication through neurotransmitters, and inter-organ communication through hormones [39].

As in nature, molecular communication is also indispensable to enable nano-scale machines to communicate with each other. Nanotechnology is one of the most important promising technology which enables nano-scale machines called as nanomachines. Nanomachines are molecular scale objects that are capable of performing simple tasks such as actuation and sensing [39]. Nanomachines are categorized into two types. While one type mimics the existing machines, other type mimics nature made nanomachines such as molecular motors and receptors [9]. In the biological systems, communication among the cells forming the biological system is essential to enable the cells to effectively accomplish their tasks. For example, in natural immune system, the white blood cells called as B-cells and T-cells communicate with each other to eliminate the pathogen entering the body. Similar to biological systems, communication among nanomachines is essential for effective sensing and action.

Due to size and capabilities of nanomachines, the traditional wireless communication with electromagnetic waves cannot be possible to communicate nanomachines that constitute of just several moles of atoms or molecules [39]. Instead, the molecular communication is a viable communication paradigm, which enables nanomachines to communicate with each other using molecules as information carrier [39]. Therefore, a molecular channel is envisioned as a communication channel for the molecular communication between two nanomachines. For this channel, it is essential to find out molecular delivery capacity between two nanomachines to understand how to enable molecular communication with high molecule delivery capacity. The molecule delivery capacity may be affected by some parameters specific to the nanomachines and physical properties of the environment such as diffusion coefficient and temperature. Therefore, it is imperative to find out capacity of the molecular channel and to understand how it varies with the properties of the nanomachines and environment.

There exist several research efforts about the molecular communication in the literature. In [39], research challenges in molecular communication is manifested. In [133], the concept of molecular communication is introduced and first attempt for design of molecular communication system is performed. In [134], a molecular motor communication system for molecular communication is introduced. In [135], a molecular communication system which will enable future health care applications is investigated. In [136], based on intercellular calcium signaling networks, the design of a molecular communication system is introduced. In [137], an autonomous molecular propagation system is proposed to transport information molecules using DNA hybridization and biomolecular linear motors. The existing studies on the molecular communication include feasibility of the molecular communication and design schemes for molecular communication system. However, none of these studies investigate the capacity of a molecular channel to understand possible conditions in which the molecular communication can be feasible and high molecular communication capacity can be achieved.

8.2 Molecular Communication Model

In nature, molecular communication between biological entities is based on the ligand-receptor binding mechanism. According to ligand-receptor binding mechanism, ligand molecules are emitted by one biological phenomenon then, the emitted ligand molecules diffuse in the environment and bind the receptors of another biological phenomenon. This binding enables

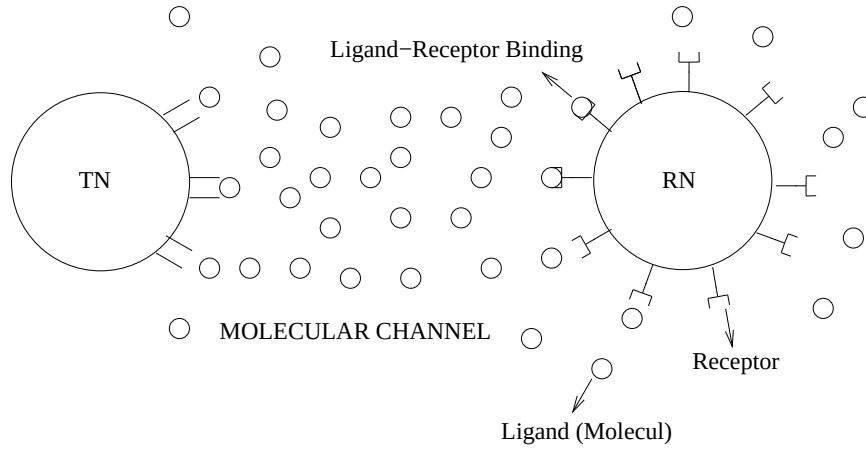


Figure 8.1: Molecular Communication Model.

the biological phenomenon to receive the bound molecules by means of the diffusion on cell membrane. The received ligand molecules allow the biological phenomenon to understand the biological information. For example, in biological endocrine system, gland cells emit hormones to inter-cellular environment then, hormone molecules diffuse and are received by corresponding cells. According to the type of emitted hormone, the corresponding cells convert the hormone molecule to biologically meaningful information. This natural mechanism provides the molecular communication for almost all biological phenomena.

In this chapter, we adopt this natural ligand-receptor binding mechanism to enable the molecular communication between nanomachines analogous to the biological mechanisms and called Transmitter Nanomachine (TN) and Receiver Nanomachine (RN) as shown in Fig. 8.1. In the literature, artificial ligand-receptor binding schemes have been previously introduced [138], [139]. In this chapter, we assume an artificial ligand-receptor binding model developed in [138]. We assume that TN is a nano-scale machine or a biological entity and it can emit one kind of molecule called A . We also assume that TN emits molecules A with a time-varying concentration of $L(t)$ according to the following emission pattern [139] which is similar to alternating square pulse, i.e.,

$$R(t) = \begin{cases} L_{ex}, & \text{for } jt_H \leq t \leq (j+1)t_H \\ 0, & \text{otherwise} \end{cases} \quad (8.1)$$

where $j = (0, 1, \dots)$, t_H is the duration of a pulse and L_{ex} is concentration of molecules A emitted by TN. Furthermore, we assume that RN is a nano-scale machine and it has N receptors called R on its surface. The receptors enable RN to receive the molecules which bind their

surface.

In traditional digital communication, information sequences are transmitted via two bits, logic 1 and 0. If a transmitter detects a voltage level which is greater than a prescribed voltage level in the channel, it decides that transmitter transmitted logic 1. If the voltage level in the channel is less than the prescribed level, the receiver decides that the transmitter transmitted logic 0. Using this traditional idea, we propose a similar molecular communication scheme. According to this scheme, during time interval t_H , TN either emits molecules A corresponding to logic 1 in digital communication or it transmits no molecule corresponding to logic 0 in digital communication. If a TN intends to transmit molecules A , we assume that during the time interval t_H , it emits molecules A to its surrounding environment with a specific concentration L_{ex} . Similar to logic 1 and logic 0 in traditional digital communication, we denote the case that TN transmits molecules A with A and we denote the case that TN transmits no molecule with 0. Hence, for the molecular communication model, we have two molecular communication bits called A and 0.

At RN side, these bits are inferred via concentration of molecules A such that if an RN receives a concentration of molecules A greater than a prescribed concentration S ($\mu\text{mol/liter}$), the RN decides that the TN transmitted molecular bit A during the time interval t_H . Conversely, if the RN receives molecules A with a concentration less than S , the RN decides that the TN transmitted molecular bit 0.

In traditional digital communication, noise level in the channel leads to channel errors such that when a transmitter intends to transmit logic 0, the receiver may detect logic 1, or for logic 1, the receiver may detect logic 0 due to the noise in the channel. Similarly, in the molecular communication, it may be possible to observe erroneous molecular communication bits at the RN side. During the molecular communication, the molecules A are emitted by TN and the emitted molecules continuously diffuse to surrounding environment including the RN such that molecules A always exist and diffuse in the environment. Therefore, due to the emitted molecules A which diffuse in the surrounding environment, it is possible for RN to receive molecular bit A although TN transmits molecular bit 0. Furthermore, due to delay in diffusion of molecules A to RN it is also possible for RN to receive molecular bit 0 although TN transmits molecular bit A . Moreover, erroneous molecular bits can arise due to some additional factors which affect the molecular diffusion between TN and RN, such as

temperature of the environment, concentration of emitted molecules A , distance between TN and RN, duration of molecule emission, binding and release rates and number of receptors on RN.

Consequently, similar to traditional digital communication channel, the molecular communication channel between TN and RN has a molecule delivery capacity which is defined as maximum number of non-erroneous molecular bits which can be delivered within a specific time duration.

Next, we introduce a molecule delivery model for the molecular communication between TN and RN according to the molecular communication approach introduced here.

8.3 Molecule Delivery

For the molecular communication between TN and RN, it is important to understand how molecules A can be delivered to RN by means of the binding between molecules A and receptors R on the RN. In this section, assuming the ligand-receptor binding model in [138], we introduce a molecule delivery model for the molecular communication between TN and RN.

According to the ligand-receptor binding reaction kinetics, when molecules A , emitted by TN, encounter with receptors R on RN, molecules A bind the receptors R . These bound molecules A and receptors R constitute complexes C (bound receptors) according to the following chemical reaction,



where k_1 ($\mu\text{mol}/\text{liter}/\text{sec.}$) is the rate of binding reaction. Similar to the binding reaction, it is possible to release molecules A from receptors R according to the following chemical reaction,



where k_{-1} ($1/\text{sec.}$) is rate of release reaction.

As given in (8.1), TN emits molecules A via a square pulse with amplitude L_{ex} during t_H

(*sec.*). In this duration, concentration of bound receptors $C(t)$ ($\mu\text{mol/liter}$) can be given [138] as follows

$$C(t) = C_\infty(1 - e^{-t(k_{-1} + k_1 L_{ex})}) \quad (8.4)$$

where k_1 and k_{-1} are the binding and release rates, respectively, L_{ex} ($\mu\text{mol/liter}$) is concentration of molecules A which is emitted by TN. C_∞ is steady-state level of bound receptors and can be given [138] as follows

$$C_\infty = \frac{k_1 L_{ex} N}{k_{-1} + k_1 L_{ex}} \quad (8.5)$$

where N ($\mu\text{mol/liter}$) is the concentration of receptors (R) on RN.

During the pulse duration t_H , $C(t)$ rises exponentially according to (8.4). At time t_0 when the pulse duration ends, $C(t)$ starts to decay [138] according to

$$C(t) = C_{t_0} e^{[-k_{-1}(t-t_0)]} \text{ for } t > t_0 \quad (8.6)$$

The rates of molecule/receptor interaction, k_1 and k_{-1} , may depend on molecular diffusion from TN to RN. More specifically, while the binding rate k_1 heavily depends on the molecular diffusion parameters from TN to RN such as diffusion coefficient, temperature of environment, distance between TN and RN [140], the release rate k_{-1} depends on some environmental factors such as interaction range and temperature [141]. Here, we only assume that binding rate $(k_1)^1$ is inversely proportional with distance (α) between TN and RN such that $k_1 \propto 1/\alpha$ and it is directly proportional with temperature of environment (T) such that $k_1 \propto 2T$. For the release rate k_{-1} , we use the model given in [141] as follows

$$k_{-1} = k_{-1}^0 e^{\alpha f / k_B T} \quad (8.7)$$

where k_{-1}^0 is the zero-force release rate, α is the distance between TN and RN, k_B and T are the Boltzmann constant and absolute temperature, respectively. f is the applied force per bound. f is related with the energy of the emitted molecules, the distance between TN and

¹ Here, we do not predict k_1 according to the diffusion parameters of the environment. In fact, binding rate k_1 can be captured with analytical expressions [142]. However, this is out of scope of this chapter.

RN, and the environmental factors [143]. Here, we consider f as positive constant throughout this chapter. k_{-1}^0 can be predicted by fitting the experimental measurements [141] and it is related with the capability of molecule capturing of RN receptors. Therefore, we assume that k_{-1}^0 is a variable which depends only on the properties of RN receptors.

Based on the models introduced in Section 8.2 and 8.3, we next develop an information theoretical approach for the capacity of the molecular channel between TN and RN. According to total concentration of complex molecules ($C(t)$) expressed in (8.4), (8.5) and (8.6), we derive probability of erroneous molecular bits which cannot be successfully delivered to RN. Then, we model the molecular communication channel similar to binary symmetric channel and we derive a capacity expression of the molecular channel between TN and RN.

8.4 An Information Theoretical Approach for Molecular Communication

As introduced in Section 8.2, for the molecular communication between TN and RN, two molecular bits are available. Every time when TN transmits a molecular bit, concentration of delivered molecules determines the success of the transmission. If TN transmits molecular bit A , at least S number of molecules² A must be delivered to RN within time interval t_H for a successful delivery of a molecular bit A . If TN transmits molecular bit 0, number of molecules A delivered within t_H must be less than S for a successful delivery of molecular bit 0. Therefore, it is imperative to find the number of delivered molecules in each transmission interval t_H to determine the success of the molecular bit transmission from TN to RN.

Here, using (8.4), (8.5), (8.6) and (8.7), the closed form expressions for expected value of number of delivered molecules A during t_H , i.e., N_A , can be given by

$$N_A = \int_0^{t_H} C(t) dt \quad (8.8)$$

$$N_A = \int_0^{t_H} \frac{k_1 L_{ex} N}{k_{-1} + k_1 L_{ex}} (1 - e^{-(k_{-1} + k_1 L_{ex})t}) dt \quad (8.9)$$

Since the molecular diffusion continues after every t_H interval, the previous molecular bits can

² Since concentration of molecules ($\mu\text{mol/liter}$) can be converted to number of molecules by multiplying Avogadro constant (6.02×10^{23}), we interchangeably use the number of molecules for the concentration of molecules.

be received in the current interval by RN. Therefore, the number of delivered molecules A in a given interval also depends on molecular bits transmitted in the previous intervals. Here, we assume that the last molecular bit only affects the current molecular transmission since the number of delivered molecules exponentially decay after t_H seconds according to (8.6). If we assume that TN emits A molecules with probability P_A in each time interval t_H and it emits molecular 0 bit with probability $(1 - P_A)$. Hence, the effect of the last emitted molecular bit on the current molecular bit transmission can be considered as expected number of delivered molecules coming from the previous interval, i.e., N_p . Thus, using (8.6) and (8.9), N_p can be given as follows

$$N_p = \int_0^{t_H} P_A N_A e^{(-k_{-1}t)} dt \quad (8.10)$$

$$N_p = \int_0^{t_H} \left(P_A \int_0^{t_H} \frac{k_1 L_{ex} N}{k_{-1} + k_1 L_{ex}} (1 - e^{-t(k_{-1} + k_1 L_{ex})}) dt \right) e^{(-k_{-1}t)} dt \quad (8.11)$$

Combining (8.9) and (8.11), for the case that TN emits A during t_H , expected value of total number of delivered molecules A , i.e., $E[N_{TA}]$, can be given as follows

$$E[N_{TA}] = N_A + N_p \quad (8.12)$$

At the RN side, if RN receives S number of molecules A , it infers that TN emitted the molecular bit A during t_H . Thus, using the well-known Markov inequality, we obtain a maximum bound for the probability p_1 that TN achieves to deliver molecular bit A as follows

$$p_1(N_{TA} \geq S) \leq \frac{E[N_{TA}]}{S} \quad (8.13)$$

Hence, TN achieves to deliver molecular bit A with maximum probability $p_1 = \frac{E[N_{TA}]}{S}$ and RN receives molecular bit 0 instead of the molecular bit A such that TN does not succeed to deliver A with probability $(1 - p_1)$.

For the transmission of molecular bit 0 during t_H , the number of delivered molecules A only depends on lastly emitted molecular bit since TN transmits no molecules during the transmission of molecular bit 0. Therefore, following (8.11), the expected value of total number of

delivered molecules A within t_H for the transmission of molecular bit 0, i.e., $E[N_{T0}]$, is given by

$$E[N_{T0}] = N_p \quad (8.14)$$

For the successful delivery of a molecular bit 0, TN must deliver a number of molecules A that is less than S and ($N_{T0} \leq S$) to RN. Using the Markov inequality, the maximum bound for probability p_2 that TN achieves to deliver molecular bit 0 is given by

$$p_2(N_{T0} \leq S) \leq \frac{S}{E[N_{T0}]} \quad (8.15)$$

Hence, for the transmission of molecular bit 0, TN achieves to deliver molecular bit 0 with maximum probability $p_2 = \frac{S}{E[N_{T0}]}$ and it does not achieve to deliver molecular bit 0, instead, it incorrectly delivers molecular bit A with probability $(1 - p_2)$. Here, it is critical to select an appropriate S to maximize p_1 and p_2 . As seen in (13), for $E[N_{TA}] \geq S$, $p_1 \geq 1$ is obtained although it is not possible for p_1 to take a value greater than 1. This implies that $p_1 \approx 1$ can be obtained by appropriately selecting S . For example, if we assume that N_{TA} is a random variable with the normal distribution $N(E[N_{TA}], \sigma_{TA}^2)$, $p_1 \approx 1$ can be obtained by selecting S as $0 < S < E[N_{TA}] - 3\sigma_{TA}$. This is because in any normal distribution, % 99.7 of the observations fall within 3 standard deviations of the mean. Similarly, if we assume that N_{T0} has the normal distribution $N(E[N_{T0}], \sigma_{T0}^2)$, $p_2 \approx 1$ can be succeed selecting S as $S > E[N_{T0}] + 3\sigma_{T0}$. Consequently, p_1 and p_2 can be maximized by selecting S from the interval $0 < E[N_{T0}] + 3\sigma_{T0} < S < E[N_{TA}] - 3\sigma_{TA}$.

According to the transmission probabilities p_1 and p_2 , we can model a channel similar to the symmetric channel. If we consider that TN emits molecular bit X and RN receives molecular bit Y , then the transition matrix of the molecular channel can be given as follows

$$P(Y/X) = \begin{pmatrix} p_1 P_A & (1 - p_2)(1 - P_A) \\ (1 - p_1)P_A & p_2(1 - P_A) \end{pmatrix}$$

Based on the transition matrix $P(Y/X)$, we can give the mutual information $I(X; Y)$ between X and Y which states the number of distinguishable molecular bits, i.e., M as follows

$$M = \left(H(p_1 P_A + (1 - p_2)(1 - P_A), (1 - p_1)P_A + p_2(1 - P_A)) \right) - \left(P_A H(p_1, 1 - p_1) + (1 - P_A) H(p_2, 1 - p_2) \right) \quad (8.16)$$

$$M = - \left[P_A \frac{E[N_{TA}]}{S} + (1 - P_A) \left(1 - \frac{S}{E[N_{T0}]} \right) \right] \log \left[P_A \frac{E[N_{TA}]}{S} + (1 - P_A) \left(1 - \frac{S}{E[N_{T0}]} \right) \right] - \left[P_A \left(1 - \frac{E[N_{TA}]}{S} \right) + (1 - P_A) \frac{S}{E[N_{T0}]} \right] \log \left[P_A \left(1 - \frac{E[N_{TA}]}{S} \right) + (1 - P_A) \frac{S}{E[N_{T0}]} \right] - P_A \left[\frac{E[N_{TA}]}{S} \log \left(\frac{E[N_{TA}]}{S} \right) - \left(1 - \frac{E[N_{TA}]}{S} \right) \log \left(1 - \frac{E[N_{TA}]}{S} \right) \right] - (1 - P_A) \left[\frac{S}{E[N_{T0}]} \log \left(\frac{S}{E[N_{T0}]} \right) - \left(1 - \frac{S}{E[N_{T0}]} \right) \log \left(1 - \frac{S}{E[N_{T0}]} \right) \right] \quad (8.17)$$

where $H(\cdot)$ denotes the entropy. Using (8.17), the capacity of the molecular channel between TN and RN i.e., C_M , can be expressed as

$$C_M = \max(M) \quad (8.18)$$

Traditionally, in a communication channel it is necessary to design codes that enable minimum error rate and maximum capacity. Similarly, in a molecular communication channel, it is necessary to find a molecular bit transmission probability that can maximize molecular communication capacity. Next, we introduce an adaptive error compensation scheme in the molecular communication, which enable TN to select most appropriate molecular bit transmission probability that can maximize molecular communication capacity.

8.5 Adaptive Error Compensation in the Molecular Communication

In the traditional digital communication, erroneous bits are frequently observed due to noise in the channel. To compensate bit errors, various kinds of channel coding schemes have been proposed in the literature. Generally, the aim of these techniques is to reduce bit error probability of communication between transmitter and receiver. For this aim, transmitter organizes transmitting communication bits to generate fixed-length codewords such that these codewords enable the receiver to detect and correct the erroneous communication bits. However, detection and correction of erroneous communication bits necessitate efficient processors, algorithms, and circuits with high computational power at the receiver side.

As in traditional digital communication, two molecular communication bits are available for the communication between TN and RN. However, existing channel coding techniques are not appropriate for the molecular communication since they necessitate high computational power, which may not be realizable for nanomachines with limited computational and storage capabilities. Therefore, the molecular communication needs proactive error compensation schemes, which do not necessitate any computational processing to compensate possible errors on the molecular channel. These error compensation schemes should proactively prevent the possible errors on the molecular communication channel by adapting some molecular communication parameters according to changing environmental factors such as temperature, binding rate, and distance between nanomachines.

In the molecular communication, we assume that some communication parameters such as concentration of emitted molecules A (L_{ex}), duration of emission pulse (t_H), and concentration of receptors on RN (N) are specific to the TN and RN and related with the design issues of the nanomachines. Therefore, they cannot be changed by neither nanomachines nor environmental factors. However, other parameters such as temperature of the environment (T), applied force per bound (f), distance between TN and RN (α), binding rate (k_1), release rate (k_{-1}), and zero-force release rate (k_{-1}^0) only depend on some environmental factors such as diffusion coefficients of the environment and deployment strategies of the nanomachines. Here, we assume that the probability of molecular bit A emission (P_A) can only be changed by TN. Therefore, a proactive error compensation scheme exploits regulation of P_A such that the regulation can proactively compensate the possible channel errors, which stem from some environmental factors and some properties of the nanomachines. For example, in an environment generating high binding rate (k_1), transmission of molecular bit 0 can be erroneous since high amount of molecules A can be delivered to RN during the transmission of molecular bit 0. However, selection of the most appropriate P_A decreasing the number of delivered molecules can enable TN to compensate such kind of errors.

Theoretically, it is possible to optimize (8.17) to find a molecular transmission probability (P_A), which minimizes the errors on molecular communication channel and maximizes molecular communication capacity. This can enable TN to encode transmitted molecular bit such that TN can minimize the errors. However, this kind of optimization process is computationally impossible for TN. In stead of some optimization process with high computational burden, it is possible to find some simple methods that enable TN to decide which P_A is the

most appropriate in which environmental conditions.

According to the molecular communication model introduced in Section 8.2, to successfully deliver molecular bit A TN must deliver at least S number of molecules A to RN. Therefore, the condition

$$E[N_{TA}] = N_A + N_p > S \quad (8.19)$$

must hold for successful delivery of molecular bit A . Substituting N_p given in (8.10), we rewrite (8.19) as

$$N_A + P_A N_A \int_0^{t_H} e^{(-k_{-1}t)} dt > S \quad (8.20)$$

Using (8.20), a lower bound for P_A , i.e., LB , can be given as

$$P_A > \frac{S - N_A}{N_A \int_0^{t_H} e^{(-k_{-1}t)} dt} = LB \quad (8.21)$$

where $N_A \int_0^{t_H} e^{(-k_{-1}t)} dt$ states concentration of molecules A that are received by RN within an exponential decaying phase after TN transmits molecular bit A as introduced in (8.11). We denote $N_A \int_0^{t_H} e^{(-k_{-1}t)} dt$ with N_{ex} and rewrite (8.21) as

$$LB = \frac{S - N_A}{N_{ex}} \quad (8.22)$$

To successfully deliver molecular bit 0, TN must deliver a number of molecules A less than S to RN. Therefore, the condition that must be met for successful delivery of molecular bit 0 is expressed as

$$E[N_{T0}] = N_p \leq S \quad (8.23)$$

where $E[N_{T0}]$ and N_p denote the expected number of delivered molecules coming from the previous interval as introduced in (8.10). Using (8.10), (8.23) can be rewritten as

$$P_A N_A \int_0^{t_H} e^{(-k_{-1}t)} dt \leq S \quad (8.24)$$

Similar to lower bound given in (8.22), using (8.24), an upper bound for P_A , i.e., UB , can be given as follows

$$P_A \leq \frac{S}{N_A \int_0^{t_H} e^{(-k_{-1}t)} dt} = UB \quad (8.25)$$

$$UB = \frac{S}{N_{ex}} \quad (8.26)$$

Combining the lower and upper bounds given in (8.22) and (8.26), respectively, an interval for selection of the most appropriate P_A that minimizes the channel errors in the molecular communication can be stated as

$$\frac{S - N_A}{N_{ex}} < P_A \leq \frac{S}{N_{ex}} \quad (8.27)$$

$$LB < P_A \leq UB \quad (8.28)$$

N_{ex} includes an integral operation that is impossible for TN to practically compute due to its very limited computational power. Since N_{ex} states concentration of molecules A that are received by RN within an exponential decaying phase after TN transmits molecular bit A , RN can obtain N_{ex} by computing concentration of molecules A within an exponential decaying phase after TN transmits a molecular bit A . Here, we assume that similar to the molecular communication from TN to RN, molecular communication from RN to TN can be possible³. We also assume that RN computes the concentration within an exponential decaying phase after TN transmits a molecular bit A and it communicates this concentration to RN before initiating the molecular communication. Thus, to determine an appropriate P_A providing satisfactory molecular communication capacity, TN evaluates the lower and upper bounds given in (8.27). However, TN needs a selection strategy to select the most appropriate molecular bit transmission probability (P_A) from the interval given in (8.28).

³ Here, we also note that TN and RN have the same molecule delivery and reception capability. However, we do not assume full duplex molecular communication between TN and RN such that TN and RN cannot simultaneously deliver or receive molecular bits. Hence, we assume a half duplex molecular communication between TN and RN

8.5.1 A Selection Strategy for Molecular Bit Transmission Probability

For the molecular communication, it is critical to select the most appropriate molecular bit transmission probability (P_A) providing high molecular communication capacity. Using the interval given in (8.27), we investigate how the variation of P_A affects the error rate in the molecular communication such that we derive a selection strategy for P_A , which allows TN to minimize error rate and to maximize molecular communication capacity.

While P_A increases, the number of delivered molecules increases. Therefore, higher P_A decreases the errors in transmission of molecular bit A . Conversely, while P_A increases, errors increase in transmission of molecular bit 0 . However, P_A less than UB enables non-erroneous molecular bit 0 transmission as introduced in derivation of UB . Therefore, P_A should be selected as high as possible for non-erroneous molecular bit A transmission while it should be selected less than UB for non-erroneous molecular bit 0 . Hence, P_A should be selected as a value that is the closest to UB ($P_A \cong UB$). This selection strategy can minimize error rate and maximize molecular communication capacity. According to this P_A selection strategy, UB is more important than LB because $P_A \cong UB$ can provide higher molecular communication capacity. Therefore, in the numerical analysis in Section 8.6 we evaluate only UB for the selection of P_A .

8.5.2 An Adaptive Molecular Error Compensation Scheme

In the molecular communication, error rate is heavily affected by some environmental factors such as binding rate (k_1), temperature (T), distance (α) between nanomachines. Therefore, it is imperative to compensate the errors due to the changing environmental factors to achieve higher molecular communication capacity. For this compensation, it is essential to regulate molecular bit transmission probability (P_A) with respect to the changing environmental factors. However, the regulation of P_A to compensate the errors needs some coordination between TN and RN. This coordination enables an adaptive error compensation scheme that is periodically conducted by TN and RN to compensate possible channel errors. Here, we introduce an adaptive Molecular Error Compensation (MEC) scheme for the molecular communication between TN and RN as outlined below:

1. Initially, TN sets the molecular bit transmission probability P_A to an initial value de-

noted by $\overline{P_A}$ and initiates the molecular communication.

2. When error rate increases, RN detects the increasing error rate⁴.
3. RN emits a molecular bit stream denoted by BS_1 to terminate current molecular communication between TN and RN and to initiate the error compensation scheme. BS_1 is a special stream⁵ such that when it is received by TN, TN can immediately terminate the current molecular communication and infer the initiation of the error compensation scheme. Therefore, by means of BS_1 , RN can initiate the error compensation scheme in any time when it detects increasing error rate in the molecular channel.
4. Once TN receives BS_1 , it immediately emits the molecular bit stream BS_2 ⁶ like A000A000 to enable RN to compute N_{ex} and UB .
5. Using N_{ex} and UB , RN selects P_A as a value closest to UB ($P_A \cong UB$).
6. RN informs TN about the selected molecular bit transmission probability. Here, we do not assume that RN communicates the actual value of the selected P_A , which is possibly a floating point number, to TN. We only assume that RN emits specific molecular bit patterns corresponding to the different level of molecular bit transmission probability such that according to the emitted molecular bit pattern, TN infers the selected molecular bit transmission probability.
7. TN sets P_A as the selected molecular bit transmission probability. After the setting of the molecular bit transmission probability, TN emits the molecular bit stream BS_3 ⁷ to again initiate the molecular communication. Then, TN again initiates the molecular communication according to the updated molecular bit transmission probability, which minimizes the error rate and maximizes the molecular communication capacity.

Using the adaptive Molecular Error Compensation (MEC) scheme given above, TN and RN can collaboratively select P_A to minimize the error rate and maximize molecular communication capacity. We also give MEC in the pseudo-code given in Algorithm 4.

⁴ Note that while it may be possible to develop some error detection mechanisms for molecular communication, it is beyond the scope of this chapter

⁵ BS_1 is a fixed molecular bit stream, which may be determined in the design stage of the molecular communication system. For example, A0A may be selected as BS_1 .

⁶ Since N_{ex} is the number of molecules delivered within an exponential decaying phase after TN emits A, BS_2 is appropriate to enable RN to compute N_{ex} . Furthermore, other molecular bit streams that can enable this computation can be selected for this computation.

⁷ Similar to BS_1 , BS_3 is also a fixed molecular bit stream, which determined in the design. For example, 0A0 may be selected as BS_3

Algorithm 4: MEC

```
1  TN sets  $P_A$  as  $\overline{P_A}$ 
2  TN initiates the molecular communication
3  foreach  $P_A$  do
4      RN detects the increasing error rate
5      RN emits  $BS_1$ 
6      TN terminates the molecular communication
7      TN emits  $BS_2$ 
8      RN computes  $N_{ex}$  and  $UB$ 
9      RN selects  $P_A$ , ( $P_A \cong UB$ )
10     RN informs TN about the selected  $P_A$ 
11     TN updates  $P_A$  as the selected  $P_A$ 
12     TN emits  $BS_3$ 
13     TN again initiates the molecular communication
14 end
```

8.6 Numerical Analysis

In this section, we first present the numerical analysis performed over the mutual information expression given in (8.17) to show how the molecular communication capacity varies according to the some environmental parameters and some other parameters specific to the nanomachines TN and RN. Then, we give the numerical analysis for the performance of MEC over the selection of the most appropriate P_A that provides high molecular communication capacity with minimum error rate. The aim of this analysis is to determine appropriate configuration of molecular communication parameters, which can achieve high molecular communication capacity, according to changing environmental factors. We perform the numerical analysis using Matlab. We assume that TN and RN are randomly positioned in an environment, which may have different diffusion coefficients such that it allows TN to achieve different binding rates (k_1). Due to the principles of mass action kinetics, we also assume that k_1 varies with temperature of environment (T) and distance (α) between TN and RN such that k_1 is directly proportional with $2T$ ($k_1 \propto 2T$) and inversely proportional with α ($k_1 \propto 1/\alpha$), respectively. Moreover, we assume that k_{-1}^0 depends only on the properties of RN receptors and cannot be

changed. The simulation parameters of the analysis are given in Table 8.1.

Table 8.1: Simulation Parameters

Binding rate (k_1)	0.1-0.5 ($\mu\text{mol/liter/s}$)
Zero-force release rate (k_{-1}^0)	0.08 (s^{-1})
Temperature (T)	300-1000 K
Distance between TN and RN (α)	$5^{-10} - 4 \times 10^{-9} m$
Applied force per bound (f)	$10^{-12} (J/m)$
Concentration of molecules A (L_{ex})	1-8 ($\mu\text{mol/liter/s}$)
Duration of the pulses (t_H)	0.5-1 s
Number of receptors R (N)	0.001-0.01 ($\mu\text{mol/liter}$)
S ($\mu\text{mol/liter/s}$)	$10^{-6} - 4 \times 10^{-6}$

8.6.1 Effect of Environmental Factors on Molecular Communication Capacity

8.6.1.1 Binding Rate:

For the first analysis, we investigate the effect of binding rate (k_1) on capacity of the molecular channel. In Fig. 8.2, mutual information (M) given in (8.17) is shown with varying molecular bit transmission probability (P_A) for different k_1 . For higher $k_1 = 0.2 - 0.4 \mu\text{mol/liter/s}$, TN delivers higher number of molecules A that is greater than S . Therefore, for higher k_1 , transmission of molecular bit 0 can be erroneous and molecular communication capacity decreases. However, for a smaller $k_1 = 0.1 \mu\text{mol/liter/s}$ that enables TN to deliver sufficient molecules for molecular bit A and 0, M can be maximized selecting an appropriate P_A . TN can deliver greater than S number of molecules in transmission of molecular bit A and deliver less than S number of molecules in transmission of molecular bit 0. Therefore, to achieve high molecular communication capacity, it is necessary to select appropriate S with respect to binding rate (k_1). For an environment imposing higher k_1 , S should be selected as a sufficiently high value that can hinder the delivery of erroneous molecular bit 0. For an environment imposing smaller k_1 , a smaller S should be used such that it can hinder the delivery of erroneous molecular bit A .

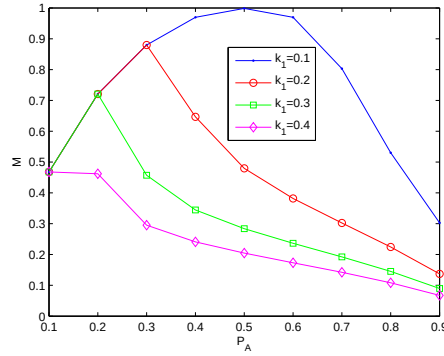


Figure 8.2: M with varying P_A for different k_1 .

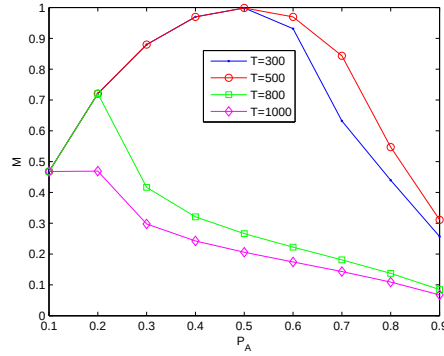


Figure 8.3: M with varying P_A for different T .

8.6.1.2 Temperature:

Temperature of the environment (T) is another important parameter since it heavily affects the binding rate (k_1) and molecular communication capacity. However, T has similar effects with k_1 on the molecular communication capacity. In Fig. 8.3, M is shown with varying P_A for different T . For $T = 300 - 500$ K, higher molecular communication capacity can be achieved. However, the capacity decreases while T is further increased from 500 K to 1000 K ($T = 500 - 1000$ K). This stems from binding rate (k_1) that increases with the higher temperature (T) such that the increasing k_1 results in delivery of excessively high number of molecules and erroneous molecular bit 0 and the capacity decreases. Hence, for an environment having higher T , S should be selected as a sufficiently high value to prevent erroneous molecular bit 0 and to maximize molecular communication capacity.

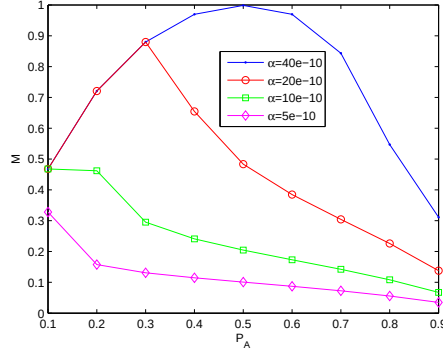


Figure 8.4: M with varying P_A for different α .

8.6.1.3 Distance between TN and RN:

As the traditional wireless communication, distance (α) between TN and RN heavily affects molecular communication capacity since k_1 depends on α . While α decreases, k_1 increases. To evaluate the effect of α on molecular communication, in Fig. 8.4, M is shown with varying P_A for different α . For smaller $\alpha = 5 \times 10^{-10} - 20 \times 10^{-10} \text{ m}$, k_1 increases and excessive number of molecules that is greater than S is delivered to RN. This results in erroneous molecular bit 0 and decreases the capacity. However, for a sufficiently high $\alpha = 40 \times 10^{-10} \text{ m}$ providing appropriate k_1 , TN delivers a number of molecules less than S in transmission of molecular bit 0 and delivers greater than S number of molecules in transmission of molecular bit A. This can maximize the molecular communication capacity. Therefore, S should be selected with respect to distance between TN and RN. While the distance increases, S should be decreased to prevent erroneous molecular bit A. While the distance decreases, S should be increased to prevent erroneous molecular bit 0.

8.6.2 Effect of the Parameters Specific to TN and RN on Molecular Communication

In this section, we present the numerical results for effect of some parameters specific to TN and RN. S is one of the most important molecular communication parameter specific to TN and RN. In Fig. 8.5(a), M is shown with varying P_A for different S . For smaller $S = 1 \times 10^{-6} - 2 \times 10^{-6} \mu\text{mol/liter/s}$, it is most likely that TN can deliver a number of molecules A that is greater than S . Therefore, molecular communication capacity decreases while S decreases. However, using a sufficiently high $S = 3 \times 10^{-6} - 4 \times 10^{-6} \mu\text{mol/liter/s}$, which

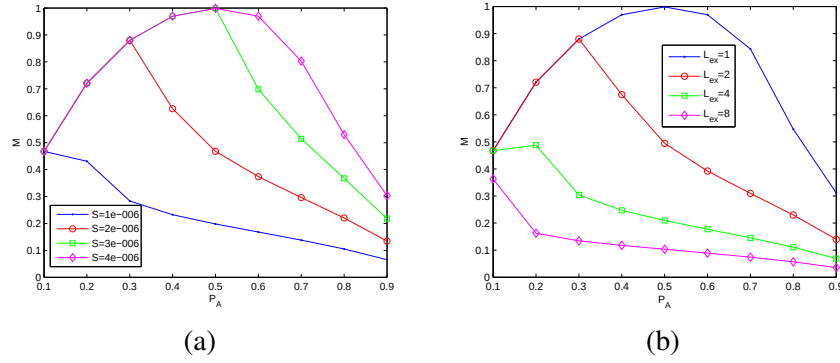


Figure 8.5: (a) M for different S . (b) M for different L_{ex} .

enables non-erroneous molecular communication bits, maximum molecular communication capacity can be achieved. Hence, S should be selected according to some environmental factors such as binding rate (k_1), temperature (T), and distance between TN and RN (α). For an environment imposing high k_1 and T , S should be used as a smaller value to prevent erroneous molecular bit 0.

Concentration of emitted molecules (L_{ex}) is also one of the most important parameters specific to TN and RN, which affects the number of delivered molecules in each transmission of molecular bits. In Fig. 8.5(b), M is shown with varying P_A for different L_{ex} . For $L_{ex} = 1 \mu\text{mol/liter/s}$, L_{ex} is sufficiently high such that TN can achieve to deliver the needed concentration to RN for molecular bits A and 0 . Therefore, high molecular communication capacity can be achieved and they can be maximized using appropriate P_A . However, for $L_{ex} = 2 - 8 \mu\text{mol/liter/s}$, TN delivers excessive concentration to RN for molecular bits 0 , which is greater than S and erroneous bits arise and the molecular communication capacity decreases. Therefore, to achieve higher molecular communication capacity, L_{ex} must be selected as an appropriate value according to the environmental factors such as binding rate, temperature and distance between TN and RN.

Duration of emission pulse (t_H) is critical for performance of the molecular communication. While t_H increases, number of delivered molecules increases. Therefore, erroneous molecular bits 0 arise while t_H increases. In Fig. 8.6(a), M is shown with varying P_A for different t_H . For $t_H = 0.5 \text{ s}$, high molecular communication capacity can be achieved. However, the capacity decreases at higher values of P_A while t_H increases. As t_H and P_A increase, number of delivered molecules increases such that TN cannot deliver the concentration smaller than

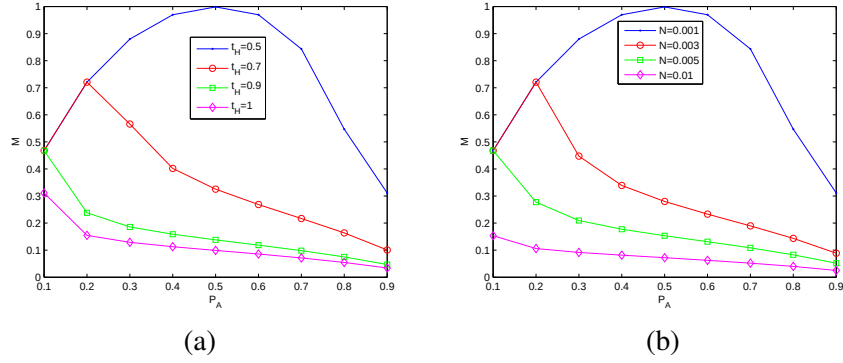


Figure 8.6: (a) M for different t_H . (b) M for different N .

S in transmission of molecular bit 0. Therefore, erroneous molecular bit 0 arises at higher P_A while t_H increases. Hence, appropriate t_H is needed to achieve higher molecular communication capacity.

Concentration of receptors (N) on RN also affects the number of delivered molecules in the molecular communication. While N increases, number of delivered molecules increases. In Fig. 8.6(b), M is shown with varying P_A for different N . For $N = 0.003 - 0.01 \mu\text{mol/liter}$, TN delivers excessive number of molecules in transmission of molecular bit 0 and erroneous molecular bits arise and the molecular communication capacity decreases. For $N = 0.001 \mu\text{mol/liter}$, the concentration of receptors on RN is sufficient to enable TN to deliver sufficient molecular concentration for molecular bit A and 0. Therefore, selecting appropriate N , the molecular communication capacity can be maximized. When the environment allows delivery of smaller number of molecules, N should be selected as a higher value to enable TN to deliver sufficient number of molecules for non-erroneous molecular bit A .

8.6.3 Adaptive Molecular Error Compensation Scheme

In this section, we present the numerical analysis on performance of adaptive Molecular Error Compensation (MEC) scheme. For the performance of MEC, we show the mutual information given in (8.17) with and without MEC scheme according to the varying environmental factors with higher error rate. For the case with MEC, we allow MEC scheme to select molecular transmission probability (P_A) to compensate possible molecular errors and to achieve high molecular communication capacity according to varying environmental factors. For the case

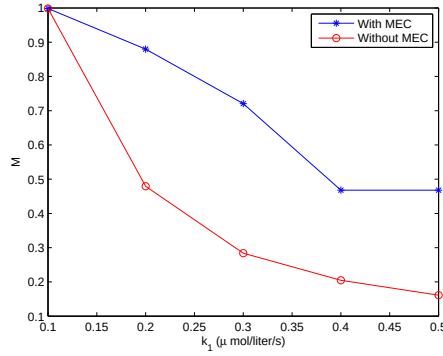


Figure 8.7: M for varying k_1 with and without MEC.

without MEC, we statically set P_A as $P_A = 0.5$ regardless of changing environmental factors. We show the effect of MEC on the molecular communication capacity in terms of variation of three environmental factors, binding rate (k_1), temperature (T), and distance between TN and RN (α). Throughout the analysis, MEC sets molecular transmission probability (P_A) as $P_A = \lfloor 10UB \rfloor / 10$, where $\lfloor \cdot \rfloor$ denotes the floor function. This selection strategy provides an appropriate P_A for MEC such that $P_A \cong UB$.

8.6.3.1 Binding Rate:

In Fig. 8.7, mutual information (M) given in (8.17) is shown for varying binding rate (k_1) with and without MEC. As k_1 increases ($k_1 = 0.1 - 0.5 \mu\text{mol/liter/s}$), molecular communication capacity decreases due to erroneous molecular bit 0. However, MEC can compensate some possible errors selecting the most appropriate P_A . MEC provides % 100 capacity improvement with respect to the case that statically selects $P_A = 0.5$ without MEC.

8.6.3.2 Temperature:

In Fig. 8.8, M is shown for varying temperature (T) with and without MEC. While T increases, the molecular communication capacity decreases due to increasing error rate in transmission of molecular bit 0. For $T = 300 - 500 \text{ K}$, almost the same capacity can be achieved with and without MEC because $\lfloor 10UB \rfloor / 10 \cong 0.5$. However, as T further increases, MEC significantly overcomes the static selection strategy in which P_A is set as $P_A = 0.5$ selecting the most appropriate P_A and MEC can compensate possible errors and achieve high molecular

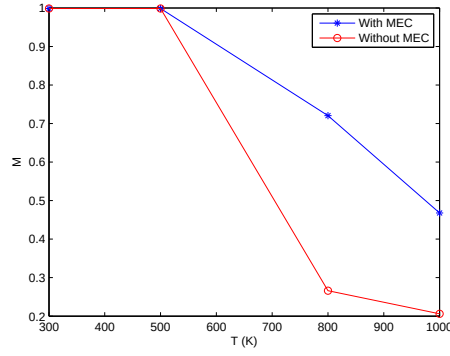


Figure 8.8: M for varying T with and without MEC.

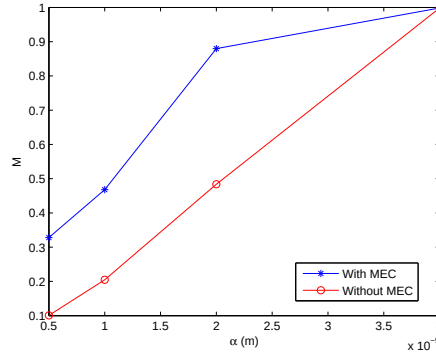


Figure 8.9: M for varying α with and without MEC.

communication capacity.

8.6.3.3 Distance between TN and RN:

In Fig. 8.9, M is shown for varying distance between TN and RN (α). As α decreases, the molecular communication capacity decreases due to increasing error rate in transmission of molecular bit A . However, MEC can compensate these errors and achieve high molecular communication capacity selecting the most appropriate P_A . MEC provides more than % 100 capacity improvement with respect to the static selection strategy using $P_A = 0.5$.

8.7 Conclusion

In this chapter, we derive a closed form expression for capacity of the channel between TN and RN. Furthermore, we propose an adaptive Molecular Error Compensation (MEC) scheme for the molecular communication between TN and RN. MEC allows TN and RN to collaboratively select the most appropriate appropriate molecular bit transmission probability to maximize molecular communication capacity with respect to environmental factors such as temperature, binding rate, distance between nanomachines. Using the capacity expression, we investigate how the conditions such as temperature of environment, concentration of emitted molecules, distance between nanomachines and duration of molecule emission, binding rate, concentration of receptors affect the molecular communication capacity. Numerical analysis reveals that MEC provides more than % 100 capacity improvement in the molecular communication selecting the most appropriate molecular transmission probability that proactively compensate the possible errors in the molecular channel. Numerical analysis also shows that the molecular communication with high capacity is only possible by arranging the molecular communication parameters such that cross-relation between the parameters should be carefully considered to compensate their negative effects over each other. Furthermore, a possible design scheme for the molecular communication should consider the environmental factors to provide high molecular communication capacity. The design scheme should select the parameters specific to TN and RN according to the environmental factors.

CHAPTER 9

Deterministic Capacity of Information Flow in Molecular Nanonetworks

In this chapter, based on stochastic models of molecular reactions, a realistic channel model is first introduced for molecular communication. Then, in order to reveal the unique characteristics of molecular information flow in nanonetworks, we define three different molecular channels as follows:

- *Point-to-point molecular channel* is a molecular communication channel between a Transmitter Nanomachine (TN) and Receiver Nanomachine (RN).
- *Molecular broadcast channel* is a molecular communication channel in which a single TN simultaneously transmits molecular information to a number of RNs.
- *Molecular multiple-access channel* is a molecular communication channel in which multiple TNs transmit molecular information to a single RN.

For each of these channels, we derive a deterministic capacity expression and investigate the deterministic capacity of molecular information flow for the realization of efficient communication and networking techniques in molecular nanonetworks. This work in this chapter was also presented in [144].

In Section 9.1, the motivation and related works are given. In Section 9.2, we present a realistic channel model for molecular communication. In Section 9.3, we first derive deterministic capacity expressions for molecular point-to-point, broadcast, and multiple-access channels. Then, we investigate information flow in a molecular nanonetwork. In Section 9.4, we give the numerical results on the capacity of the molecular channels and information flow. Finally,

we discuss concluding remarks in Section 9.5.

9.1 Motivation and Related Works

Nanotechnology envisages the practical realization of very low-end nanomachines that have tiny components to accomplish a simple specific task such as communication, computation, and sensing in a scale ranging from 1 to 100 nanometers. For example, the programmable molecular nanomachines that are formed by precisely organized molecule and atom configurations are envisioned to accomplish specific tasks. These machines are also designed with the significant abilities to physically configure themselves based on the task requirements or sensing information. They are also considered as part of the potential solution approaches for pollution, scarce food resources, and cellular repair [6].

At the same time, the existing nanoscale biological entities such as functional living cell components, bacterium, and viruses are also viewed as nanomachines. The realization of such nanomachines has also been worked by a large set of biochemical methods for many years. For example, biochemists already tackle with the production of functional cell components such as ribosome that can be used for the synthesis of beneficial proteins. Moreover, bacteria are also genetically programmed to generate principle vital components ranging from human growth hormone to beneficial enzymes for chess production. Moreover, by mixing viral proteins with a special DNA in a test tube, it might be possible to generate a kind of T₄ phage viruses that stick to a bacterium and inject viral DNA to program the bacterium for the generation of excessive viral DNA and infections [6].

Doubtlessly, network of communicating nanomachines can also be programmed to share nanoscale information over a nanonetwork so as to fulfill more complex tasks such as collaborative drug delivery, health monitoring, and biological or chemical attack detection [12], [13]. In nature, biological systems such as immune system and insect colony already allow their biological entities to cooperatively share information to achieve specific tasks. For example, in natural immune system, the white blood cells called B-cells and T-cells communicate with each other via molecular communication to cooperatively sense and eliminate the hazardous pathogen entering the body. The molecular communication among white blood cells enables the biological immune network that is an excellent defense mechanism of organism. As in

nature, nanomachines may also be interconnected via molecular communication to form a molecular nanonetwork for a specific task [13].

In the literature, there exist several research efforts toward addressing the unique challenges of molecular communications. In [12], an extensive survey on molecular nanonetworking paradigm is given. In [133], the concept of molecular communication is introduced and the first attempt for design of molecular communication system is performed. An autonomous molecular propagation system is proposed in [39] that uses DNA hybridization and biomolecular linear motors to transport information molecules. In [40], a physical channel model for molecular communication is introduced by incorporating three different phases of molecular communication called as molecular emission, diffusion, and reception into a single channel model. In [145], achievable information rate is investigated in the molecular communication channel that is modeled as a timing channel. The time delay experienced in delivery of molecules is considered as a random variable for transferring molecular information. In our previous work [132], we introduced an information theoretical approach for molecular communication, derived a closed-form expression for the capacity of molecular communication channel between two nanomachines and proposed an adaptive error compensation technique for molecular communication. However, none of these works investigates the deterministic capacity of information flow in a nanonetwork and explores the feasibility of networking techniques for the realization of future molecular nanonetworks.

In [146] and [147], we model the point-to-point and multiple-access molecular channels and discuss their achievable molecular communication capacities. The channel models are formed by using the concept of ligand-receptor binding process with deterministic mass-action kinetic mechanisms. However, due to the random nature of the biochemical reactions in molecular communication system, random fluctuations and chemical noise are inevitable for molecular reaction rates [148]. Therefore, it is essential to devise realistic molecular channel models that completely characterize the stochastic nature of physical molecular channel.

9.2 A Realistic Channel Model for Molecular Communication

In nature, molecular communication among biological entities is based on molecular emission and reception. Molecules are emitted by a biological entity and the emitted molecules freely

diffuse in the medium. Then, some of them collide with another entity that has specific receptors for the molecular reception. If the collision point of a molecule overlaps with a receptor site, the collided molecule and receptor react with each other, and a set of such reactions triggers the perception of a biologically meaningful information to fire an action potential [149]. Hence, biological entities communicate with each other by emitting and perceiving molecular information. Based on this biological phenomenon, in this chapter, we devise a realistic channel model for molecular communication as follows.

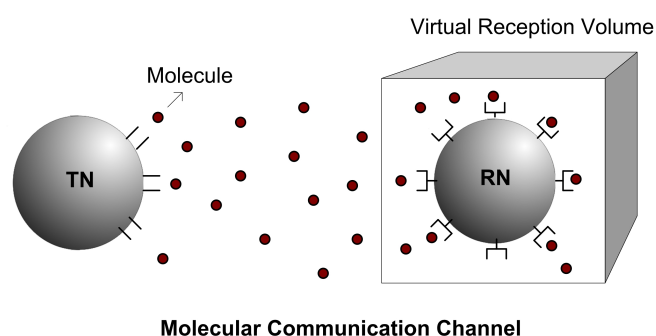


Figure 9.1: Molecular communication channel in which TN communicates with RN by emitting molecules. All reactions among the emitted molecules and receptors on RN occur in a virtual reception volume that surrounds the RN.

We assume that a TN and RN are located in an aqueous medium as shown in Fig. 9.1. The molecules emitted by TN diffuse in the medium and some of them collide and react with the receptors on the surface of RN. Each reaction is an instantaneous event that is triggered with a collision between a molecule and receptor and results in an infinitesimally small electrochemical pulse to allow RN to infer the delivery of a molecular information. A reaction between a molecule and receptor occurs if the collision point of the molecule overlaps with the location of a receptor.

Here, we also define a *virtual reception volume* (VRV), center of which is the location of RN, and all of the reactions among the molecules and receptors take place in VRV as in Fig. 9.1. VRV is selected as a sufficiently small volume that surrounds RN to exactly follow the stochastic dynamics of the reactions among molecules and receptors. Size of VRV, i.e., v_r is assumed to be constant and greater than the volume of RN. The molecules that collide and react with the receptors are not grasped and permanently enclosed by RN. After the reaction, the molecule continues its free diffusion in the environment.

We assume that TN emits the molecules L with an initial concentration¹ x to the medium at time t_0 . Note that x is the input to the molecular channel derived below. The emitted molecules start to diffuse in the medium and at time t , the concentration of molecules in VRV, i.e., x_t can be given as [13]

$$x_t = \frac{x}{(4\pi Dt)^{3/2}} e^{-\frac{d^2}{4Dt}} \quad (9.1)$$

where D is the translational diffusion coefficient for molecules L and d is the distance between TN and RN. Hence, at time t , the number of molecules that might be available and chemically react with the receptors in VRV is x_t .

There are mainly two approaches to mathematically characterize the temporal evolution of a spatially homogeneous chemically reacting systems [150]. The first approach called deterministic approach identifies the time evolution as a predictable process by a set of coupled ordinary differential equations also called reaction-rate equations. The second is the stochastic approach that considers the time evolution as a random-walk process which is characterized by a single differential (or difference) equation called as the master equation. The deterministic approach considers the time evolution of a chemically reacting system as continuous and deterministic. However, the evolution is not a continuous process since number of molecules due to chemical reactions changes by discrete amounts. Furthermore, the evolution is evidently a non-deterministic process such that it is impossible to predict the exact molecular population levels during a time interval unless the precise positions and velocities of all molecules are considered [148]. Here, in order to devise a realistic channel model in molecular communication, we follow a stochastic approach introduced for describing the stochastic simulation of coupled chemical reactions [148]. The number of reactions occurring in VRV during a time interval is actually a random variable since it is not possible to track locations and positions of all molecules. Here, our aim is to describe the evolution of the number of reactions within a time interval. For this aim, it can be possible to define a probability, i.e., $a(x_t)dt$, that one reaction occurs between a molecule and receptor in the next infinitesimally small time interval $[t, t + dt)$. $a(x_t)$ is commonly referred to as propensity function in the stochastic analysis of chemically reacting systems and also provides the logical basis for stochastic chemical kinetics in the stochastic simulation algorithms [151]. $a(x_t)$ has

¹ Since concentration of molecules ($\mu\text{mol/liter}$) can be converted to number of molecules by multiplying Avogadro constant (6.02×10^{23}), we interchangeably use the number of molecules for the concentration of molecules.

also a special mathematical form [152] as

$$a(x_t) = \eta\gamma \quad (9.2)$$

where η is the specific probability rate constant for the molecular reaction among the molecules and receptors in VRV, and γ is the number of distinct combinations of molecule and receptor pairs available in VRV. ηdt also gives the probability that a randomly chosen pair of molecule and receptor reacts during the next dt . ηdt can be computed² using the average velocity of molecules, the size of VRV, and the volume of the molecules and RN [148]. Note that γ can be directly obtained by inspecting the reaction $L+R \rightarrow \text{electrochemical pulse} + L+R$. Clearly, since there are rx_t distinct combinations of molecules and receptors for such a reaction, γ can be set as $\gamma = rx_t$, where r is the concentration of receptors R on RN. Substituting η , r , and x_t given in (9.1) into (9.2), $a(x_t)$ can be also written as

$$a(x_t) = \left[\frac{\eta r e^{-\frac{d^2}{4Dt}}}{(4\pi Dt)^{3/2}} \right] x \quad (9.3)$$

Using $a(x_t)$, we establish the stochastic dynamics of reactions in VRV as follows. Let $y(t, \tau)$ be the number of reactions that occur within the time interval $[t, t + \tau]$. $y(t, \tau)$ is clearly a random variable and its computation for an arbitrary $\tau > 0$ is nontrivial and necessitates the solution of stochastic chemical master equation that does not always have a solution. However, it can be possible to obtain an easier and excellent approximation for $y(t, \tau)$ if the following conditions can be satisfied [151]:

1. τ should be small enough such that during the interval $[t, t + \tau]$, the changes in the system state will be very slight, that is,

$$a(x_t) \cong a(x_{t+t'}) \quad \forall t' \in [t, t + \tau] \quad (9.4)$$

If (9.4) is satisfied, the rate of reactions does not significantly change in the interval $[t, t + \tau]$. With this condition, all reaction events within the interval $[t, t + \tau]$ can be clearly independent with each other. Hence, $y(t, \tau)$ is simply a Poisson random variable with the rate $\lambda = a(x_t)\tau$ [151].

2. τ should be large enough such that the expected number of reactions in the interval $[t, t + \tau]$ should be much larger than 1, that is,

$$a(x_t)\tau \gg 1 \quad (9.5)$$

² The computation of η is beyond the scope of this chapter. In the numerical analysis of this chapter, we select η as a physically feasible value and only observe its effect on the molecular communication capacity.

If (9.5) is satisfied, the Poisson random variable $y(t + \tau)$ can be approximated by the corresponding normal random variable $N(\lambda, \lambda)$, that is, $N(a(x_t)\tau, a(x_t)\tau)$ [151]. In order to satisfy (9.5), x must be greater than a lower bound, i.e., x_l , since $a(x_t)$ is a strictly increasing function of x as seen in (9.3). Thus, $x \geq x_l$ must be satisfied in order to obtain a coherent normal distribution approximation for y . The normal random variable y given as $N(a(x_t)\tau, a(x_t)\tau)$ also can be simplified as a standard normal random variable based on

$$N(\mu, \sigma^2) = \mu + \sigma N(0, 1). \quad (9.6)$$

By setting $\mu = a(x_t)\tau$ and $\sigma = [a(x_t)\tau]^{\frac{1}{2}}$ and substituting into (9.6), $y(t, \tau)$ can be expressed as

$$y(t, \tau) = a(x_t)\tau + [a(x_t)\tau]^{\frac{1}{2}}z \quad (9.7)$$

where z is the standard random variable $N(0, 1)$. Using the derivation of x_t and $a(x_t)$ given above and substituting them into (9.7), $y(t, \tau)$ can be rewritten as

$$y(t, \tau) = \left[\frac{\eta r \tau e^{-\frac{r^2}{4Dt}}}{(4\pi Dt)^{3/2}} \right] x + \left[\frac{\eta r \tau e^{-\frac{r^2}{4Dt}}}{(4\pi Dt)^{3/2}} x \right]^{\frac{1}{2}} z \quad (9.8)$$

Here, in order to simplify the capacity analysis of the molecular channels in the following section, we also normalize the concentration of emitted molecules (x) over the interval $[0, 1]$ by using an upper bound for x . This is similar with the transmission power constraint in traditional communication channel. Hence, by normalizing x over the interval $[0, 1]$, (9.8) can be given as

$$y(t, \tau) = \left[\frac{\eta r \tau e^{-\frac{r^2}{4Dt}} x_u}{(4\pi Dt)^{3/2}} \right] \frac{x}{x_u} + \left[\frac{\eta r \tau e^{-\frac{r^2}{4Dt}} x_u}{(4\pi Dt)^{3/2}} \frac{x}{x_u} \right]^{\frac{1}{2}} z \quad (9.9)$$

where x_u is the upper bound for x . (9.9) can be also simplified as

$$y(t, \tau) = h(t, \tau)\bar{x} + [h(t, \tau)\bar{x}]^{\frac{1}{2}}z \quad (9.10)$$

where $\bar{x}$ denotes the normalized x , i.e., $\bar{x} = \frac{x}{x_u}$, and $h(t, \tau)$ is defined as

$$h(t, \tau) = \frac{\eta r \tau e^{-\frac{r^2}{4Dt}} x_u}{(4\pi Dt)^{3/2}} \quad (9.11)$$

Clearly, (9.10) forms a channel model for molecular communication and is similar with a Gaussian channel model in which $\bar{x}$ is the channel input, $y(t, \tau)$ is the channel output, $h(t, \tau)$ is the channel gain and $[h(t, \tau)\bar{x}]^{\frac{1}{2}}z$ is the noise term that includes a white noise term z and an input-dependent term $[h(t, \tau)\bar{x}]^{\frac{1}{2}}$. Since the channel gain $h(t, \tau)$ is a function of time, the molecular communication channel has time-varying characteristics. However, based on the

assumption that $h(t, \tau)$ slightly changes within each $[t, t + \tau]$, the characteristics of the molecular communication channel can be investigated throughout each consecutive interval of τ . Hence, in the following sections, capacity and numerical analyses of the molecular channel given in (9.10) are performed within the consecutive intervals of τ . Note that, hereafter for ease of illustration, we denote the normalized channel input as x instead of $\bar{x}$. Next, using the channel model derived in this section, we investigate the achievable molecular capacities for molecular communication channels.

9.3 Deterministic Analysis of Molecular Nanonetworks

In this section, based on the deterministic analysis of wireless networks [153] and the molecular channel model given in (9.10), we first introduce a deterministic capacity analysis and derive a capacity expression for point-to-point, broadcast, and multiple-access molecular channels. Then, based on these analyses and capacity expressions, we investigate information flow in a molecular nanonetwork and discuss the feasibility of networking techniques that enable the predicted capacity levels in molecular nanonetworks.

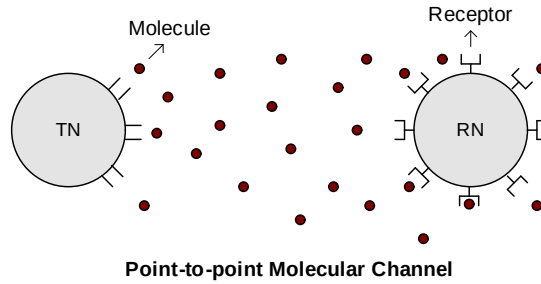


Figure 9.2: Point-to-point molecular channel in which a single TN communicates with a single RN by emitting molecules.

9.3.1 Point-to-point Molecular Channel

In the point-to-point molecular channel, a TN and RN exist in the system and TN emits molecules with concentration x to its surroundings in a single puff as a point source at time $t = t_0$. Some of the emitted molecules diffuse toward RN and collide with the receptors on its surface. Each of the collided molecules is assumed to cause molecular information delivery to RN. During a sufficiently long τ , the concentration of the delivered molecules, i.e., $y(t, \tau)$,

can be given as in (9.10). Using the binary expansion of x and z , (9.10) can be rewritten as follows [153]

$$y(t, \tau) = 2^{\log h(t, \tau)} \sum_{i=1}^{\infty} x(i)2^{-i} + 2^{\frac{1}{2} \log h(t, \tau)x} \sum_{i=-\infty}^{\infty} z(i)2^{-i} \quad (9.12)$$

where $\log(\cdot)$ is used with base 2. In order to simplify the effect of the noise, setting the peak power of the white noise term z as 1, (9.12) can be approximated as

$$y(t, \tau) \approx 2^{n(t, \tau)} \sum_{i=1}^{\infty} x(i)2^{-i} + 2^{k(t, \tau)} \sum_{i=1}^{\infty} z(i)2^{-i} \quad (9.13)$$

where $\lceil \log h(t, \tau) \rceil = n(t, \tau)$ and $\lceil \frac{1}{2} \log h(t, \tau)x \rceil = k(t, \tau)$. By further simplifying, (9.13) can be also expressed as

$$y(t, \tau) \approx 2^{n(t, \tau)} \sum_{i=1}^{n(t, \tau)-k(t, \tau)} x(i)2^{-i} + 2^{k(t, \tau)} \sum_{i=1}^{\infty} [x(i + n(t, \tau) - k(t, \tau)) + z(i)]2^{-i} \quad (9.14)$$

Ignoring the carry bit from the second term $\sum_{i=1}^{\infty} [x(i + n(t, \tau) - k(t, \tau)) + z(i)]2^{-i}$, (9.14) approximately represents the point-to-point molecular channel given in (9.12). As seen in (9.14), RN receives $[n(t, \tau) - k(t, \tau)]$ most significant bits of the input x without any noise while the rest are not seen at all due to the noise [153]. Here, the concept of bit is not the same with the traditional digital communication since we do not consider any digital coding mechanism for the transmission of molecular information. If $[n(t, \tau) - k(t, \tau)]$ most significant bits of the channel input are received by RN in the interval $[t, t + \tau]$, this means the reception of a number of molecules which is equal to the decimal equivalent of $[n(t, \tau) - k(t, \tau)]$ most significant bits in the binary version of the channel input x emitted by TN.

The molecular information rate achieved in the interval $[t, t + \tau]$ in the deterministic point-to-point molecular channel, i.e., $R_p(t, \tau)$ (*bits*/ τ *sec.*), can be given as $R_p(t, \tau) = n(t, \tau) - k(t, \tau)$. However, since $k(t, \tau)$ is a function of the channel input x as $\lceil \frac{1}{2} \log h(t, \tau)x \rceil = k(t, \tau)$, $R_p(t, \tau)$ is also a function of x such that $R_p(t, \tau)$ decreases while x increases. This stems from the fact that in the channel model given in (9.10), the effect of the noise z increases with the factor $[h(t, \tau)x]^{\frac{1}{2}}$ and the increasing noise effect reduces the molecular information rate. However, as introduced in Section 9.2, x has also a lower bound, i.e., x_l , in order to obtain a coherent approximation for the channel output $y(t, \tau)$. Therefore, the maximum molecular information rate of the point-to-point molecular channel in the interval $[t, t + \tau]$, i.e., $R_p(t, \tau)$, can be obtained by selecting the channel input x as its lower bound x_l and can be given as

$$R_p(t, \tau) = n(t, \tau) - k(t, \tau) = \lceil \log h(t, \tau) \rceil - \left\lceil \frac{1}{2} \log \left(h(t, \tau) \frac{x_l}{x_u} \right) \right\rceil \quad (9.15)$$

Due to the extremely long propagation delay of the emitted molecules, the molecular information rate $R_p(t, \tau)$ is a time-varying function. Here, in order to define an ultimate average capacity for the point-to-point molecular channel, starting from $t_0 = 0$, we simply average consecutive $R_p(t, \tau)$ values computed during a sufficiently high number of consecutive intervals of τ . For example, during T consecutive intervals of τ , we have T molecular information rates as $R_p(t_0, \tau)$, $R_p(t_0 + \tau, \tau)$, $R_p(t_0 + 2\tau, \tau)$, $\dots$, $R_p(t_0 + (T-1)\tau, \tau)$ and the capacity achieved during these intervals can be obtained by averaging these rates. Hence, the average capacity of the point-to-point molecular channel, i.e., Π , can be given as

$$\Pi = \frac{1}{T} \sum_{i=0}^{T-1} R_p(t_0 + i\tau, \tau) \quad (9.16)$$

In the point-to-point molecular channel, the molecular communication rate between TN and RN can be clearly increased with the channel gain $h(t, \tau)$ on account of $\lceil \log h(t, \tau) \rceil = n(t, \tau)$. However, the channel gain $h(t, \tau)$ is mostly affected by the physical system parameters such as diffusion coefficient and distance between nanomachines as introduced in Section 9.2. Therefore, the regulation of such physical parameters to improve the gain may not be always possible since these parameters are initially set and cannot be changed throughout. Similar to the traditional wireless communication, by increasing the upper bound of the channel input (x_u), the channel gain $h(t, \tau)$ and the capacity of point-to-point molecular channel can be increased. However, x_u may be also determined by the physical molecular emission capabilities of nanomachines. Hence, it might not be possible to improve x_u without the regulation of physical system parameters. On the other hand, since TN needs an energy resource to emit each of molecules L , there may be a rough trade off between energy consumption and the achievable capacity level. Thus, similar to traditional wireless communication, in point-to-point molecular channel, communication rate is clearly prone to extremely low level of nanomachine energy resource.

9.3.2 Molecular Broadcast Channel

In the molecular broadcast channel, single TN communicates with multiple RNs as shown in Fig. 9.3. Based on the point-to-point molecular channel devised in Section 9.3.1, it is straightforward to find the achievable molecular communication rate in each RN for molecular broadcast channel. If we consider the case in which a single TN communicates with two RNs called as RN₁ and RN₂, following the channel model given in (9.10), the channel output in

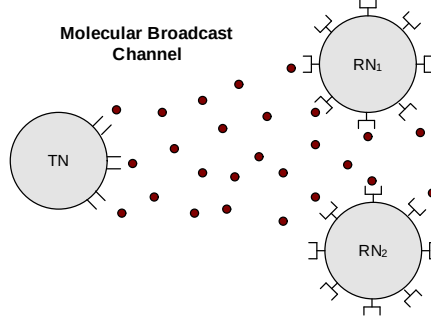


Figure 9.3: Molecular broadcast channel with one TN simultaneously communicating with two RNs.

RN₁ and RN₂, i.e., $y_1(t, \tau)$ and $y_2(t, \tau)$, respectively, can be given as

$$y_1(t, \tau) = h_1(t, \tau)x + [h_1(t, \tau)x]^{\frac{1}{2}}z \quad \text{and} \quad y_2(t, \tau) = h_2(t, \tau)x + [h_2(t, \tau)x]^{\frac{1}{2}}z \quad (9.17)$$

Using the binary expansion of x and z , the channel outputs $y_1(t, \tau)$ and $y_2(t, \tau)$ can be expressed as

$$y_j(t, \tau) = 2^{\log h_j(t, \tau)} \sum_{i=1}^{\infty} x(i)2^{-i} + 2^{\frac{1}{2} \log h_j(t, \tau)} \sum_{i=1}^{\infty} z(i)2^{-i}, \quad \text{for } j = 1, 2. \quad (9.18)$$

In order to give a deterministic model for molecular broadcast channel, (9.18) can be simplified as

$$y_j(t, \tau) \approx 2^{n_j(t, \tau)} \sum_{i=1}^{n_j(t, \tau) - k_j(t, \tau)} x(i)2^{-i} + 2^{k_j(t, \tau)} \sum_{i=1}^{\infty} [x(i + n_j(t, \tau) - k_j(t, \tau)) + z(i)]2^{-i} \quad \text{for } j = 1, 2. \quad (9.19)$$

where $\lceil \log h_j(t, \tau) \rceil = n_j(t, \tau)$ and $\lceil \frac{1}{2} \log h_j(t, \tau) \rceil = k_j(t, \tau)$ for $j = 1, 2$. Based on the deterministic model given in (9.19), the molecular communication rate between TN and RN_j in the time interval $[t, t + \tau]$, i.e., $R_b^j(t, \tau)$ for $j = 1, 2$, can be given as

$$R_b^j(t, \tau) = n_j(t, \tau) - k_j(t, \tau) = \lceil \log h_j(t, \tau) \rceil - \left\lceil \frac{1}{2} \log \left(h_j(t, \tau) \frac{x_l}{x_u} \right) \right\rceil \quad \text{for } j = 1, 2. \quad (9.20)$$

Here, similar to the point-to-point molecular channel, by averaging the molecular communication rates obtained during T consecutive intervals of τ , average capacity achieved by each RN_j, i.e., Λ_j can be given as

$$\Lambda_j = \frac{1}{T} \sum_{i=0}^{T-1} R_b^j(t_0 + i\tau, \tau) \quad \text{for } j = 1, 2. \quad (9.21)$$

Due to free diffusion of the emitted molecules toward all directions, unicast and multicast routing among nanomachines may not be feasible in molecular communication. Therefore, the

molecular broadcast channel is indispensable for molecular nanonetworks to provide an efficient flooding-based routing scheme among nanomachines. In fact, due to the free diffusion of emitted molecules, each molecular transmission can ultimately reach all nanomachines in the system. This imposes a high level of molecular interference on the nanonetwork. Therefore, it is essential for an efficient routing mechanism to choose a set of broadcaster nanomachines to efficiently route molecular information from a source to a destination nanomachine by imposing the minimum level of molecular interference on the other nanomachines.

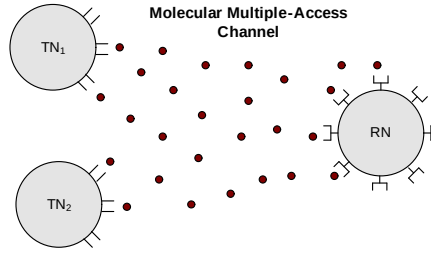


Figure 9.4: Molecular multiple-access channel with two TNs simultaneously communicating with a single RN.

9.3.3 Molecular Multiple-access Channel

In the molecular multiple-access channel, multiple TNs communicate with a single RN. Therefore, a single RN simultaneously receives multiple molecular transmissions from all TNs. Here, we consider a molecular multiple-access channel in which two TNs (TN₁ and TN₂) communicate with a single RN as shown in Fig. 9.4. The molecular communication channel model given in (9.10) also has a useful characteristic based on which molecular multiple-access channel can be also modeled using (9.10). This follows the fact that the superposition of Poisson random variables is also a Poisson random variable, rate of which is sum of the rates of all Poisson random variables. Let us assume that two TNs called TN₁ and TN₂ transmit to a single RN in a molecular multiple-access channel and each TN_{*i*} has the channel gain $h_i(t, \tau)$ and channel input x_i for $i = 1, 2$. Let us also assume that x_i has a lower and upper bound given as x_l^i and x_u^i , respectively, for $i = 1, 2$. Hence, the channel output $y(t, \tau)$ of the molecular multiple-access channel can be given as

$$y(t, \tau) = h_1(t, \tau)x_1 + h_2(t, \tau)x_2 + [h_1(t, \tau)x_1 + h_2(t, \tau)x_2]^{\frac{1}{2}}z \quad (9.22)$$

(9.22) clearly leads to a realistic channel model for molecular multiple-access channel. Using the binary expansion of x_1 , x_2 , and z , (9.22) can be also expressed as

$$y(t, \tau) = 2^{\log h_1(t, \tau)} \sum_{i=1}^{\infty} x_1(i) 2^{-i} + 2^{\log h_2(t, \tau)} \sum_{i=1}^{\infty} x_2(i) 2^{-i} + 2^{\frac{1}{2} \log [h_1(t, \tau) x_1 + h_2(t, \tau) x_2]} \sum_{i=1}^{\infty} z(i) 2^{-i} \quad (9.23)$$

Simplifying (9.23), the deterministic model of the molecular multiple-access channel can be given as

$$\begin{aligned} y(t, \tau) \approx & 2^{n_1(t, \tau)} \sum_{i=1}^{n_1(t, \tau) - n_2(t, \tau)} x_1(i) 2^{-i} + 2^{n_2(t, \tau)} \sum_{i=1}^{n_2(t, \tau) - k(t, \tau)} [x_1(i + n_1(t, \tau) - n_2(t, \tau)) + x_2(i)] 2^{-i} \\ & + 2^{k(t, \tau)} \sum_{i=1}^{\infty} [x_1(i + n_1(t, \tau) - k(t, \tau)) + x_2(i + n_2(t, \tau) - k(t, \tau)) + z(i)] 2^{-i} \end{aligned} \quad (9.24)$$

where $n_j(t, \tau) = \lceil \log h_j(t, \tau) \rceil$ for $j = 1, 2$ and $k(t, \tau) = \lceil \frac{1}{2} \log (h_1(t, \tau) x_1 + h_2(t, \tau) x_2) \rceil$. As seen in (9.24), $[n_1(t, \tau) - n_2(t, \tau)]$ most significant bits of x_1 can be received by RN without any noise and $[n_2(t, \tau) - k(t, \tau)]$ bits of x_1 and x_2 interfere with each other. The other bits of x_1 and x_2 cannot be seen by RN since they interfere with the noise [153]. Similar to traditional wireless multiple-access channel, the molecular multiple-access channel is also prone to the interference among TNs that simultaneously transmit to the same channel. In the traditional wireless multiple-access channel, the interference can be relieved by means of efficient multiple-access mechanisms such as code division multiple access (CDMA). However, in the molecular multiple-access channel, a coding mechanism that can separate the molecular transmissions in order not to interfere with each other may not be practical due to extremely low propagation speed of molecules and lack of the synchronization among TNs. Instead, it may be possible to enable each TN to use different molecules as carriers that are distinguished by RN when they are perceived. This can clearly prevent the high interference level in the molecular multiple-access channel. Hence, if each TN uses a different molecule for communication carrier the molecular communication rate achieved by TN_1 and TN_2 in the interval $[t, t + \tau]$, i.e., $R_m^1(t, \tau)$ and $R_m^2(t, \tau)$, respectively, can be given as

$$R_m^j(t, \tau) = n_j(t, \tau) - k(t, \tau) = \lceil \log h_j(t, \tau) \rceil - \left\lceil \frac{1}{2} \log \left(h_1(t, \tau) \frac{x_l^1}{x_u^1} + h_2(t, \tau) \frac{x_l^2}{x_u^2} \right) \right\rceil \quad \text{for } j = 1, 2 \quad (9.25)$$

If such a multiple-access mechanism is not used, communication rate in the molecular multiple-access channel excessively decreases due to high interference level between x_1 and x_2 . More specifically, in this case, TN_1 achieves $[n_1(t, \tau) - n_2(t, \tau)]$ bits of molecular communication rate while TN_2 may not use any capacity in the molecular multiple-access channel due to

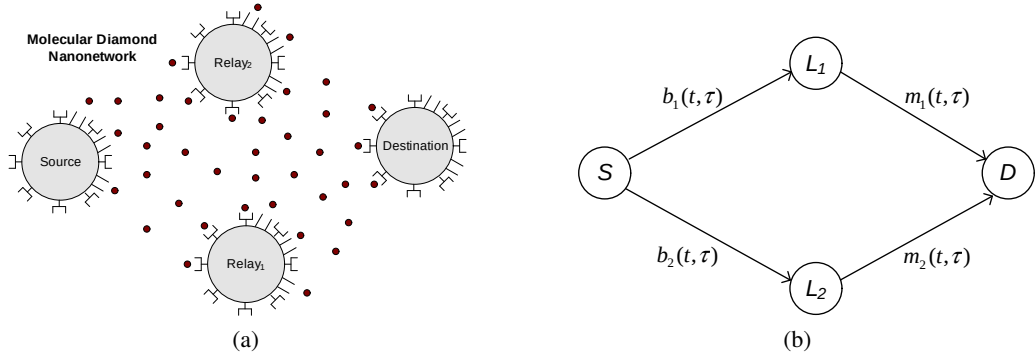


Figure 9.5: (a) Molecular diamond nanonetwork. A source nanomachine communicates with a destination nanomachine with the help of two relay nanomachines. (b) Schematic diagram of diamond-like molecular nanonetwork.

the excessive interference level. In order to obtain the average capacity, i.e., Γ_j achieved by each TN in the molecular multiple-access channel, molecular communication rates $R_m^j(t, \tau)$ measured along T consecutive intervals of τ are averaged as follows

$$\Gamma_j = \frac{1}{T} \sum_{i=0}^{T-1} R_m^j(t_0 + i\tau, \tau) \quad \text{for } j = 1, 2. \quad (9.26)$$

Next, using the capacity expressions derived for molecular point-to-point, broadcast, and multiple-access channel, we investigate the deterministic capacity of information flow in a molecular nanonetwork.

9.3.4 Information Flow in a Molecular Nanonetwork

Here, we investigate deterministic capacity of information flow between a source nanomachine S and a destination nanomachine D in a diamond-like molecular nanonetwork as shown in Fig. 9.5(a). We assume that S cannot directly communicate with D and two relay nanomachines called as L_1 and L_2 help the molecular communication between S and D . We also assume that all nanomachines in the network have the capability of molecule emission and reception to transmit and receive the molecular information as a relay. The network consists of a molecular broadcast channel among S , L_1 , and L_2 and a molecular multiple-access channel among L_1 , L_2 , and D . These channels are schematically shown in Fig. 9.5(b) and achievable molecular communication rates within the interval $[t, t + \tau]$ in these channels can be computed using the deterministic capacity expressions given in (9.20) and (9.25) in Section 9.3.2 and 9.3.3. $b_i(t, \tau)$ denotes the achievable molecular communication rate between S and L_i in the

interval $[t, t + \tau]$ for $i = 1, 2$ and $m_i(t, \tau)$ denotes the achievable rate between L_i and D in the interval $[t, t + \tau]$ for $i = 1, 2$.

Here, we aim to find the capacity of maximum information flow between S and D using the maximum-flow minimum cut theorem [154]. We first focus on the time interval $[t, t + \tau]$ to find the achievable molecular communication rate in this interval. Then, we average the rates obtained along a sufficiently high number of consecutive intervals of τ to find the average capacity of the information flow between S and D . As observed in Fig. 9.5(b), there are four cut sets, i.e., ξ_i , $i = 1, \dots, 4$, in the diamond-like molecular nanonetwork [153]. The first cut set, i.e., $\xi_1 = \{(S, L_1), (S, L_2)\}$ forms a broadcast channel among S , L_1 , and L_2 . Based on the deterministic capacity analysis of molecular broadcast channel given in Section 9.3.2, the achievable molecular communication rate within the interval $[t, t + \tau]$ in the first cut set can be given as $\max(b_1(t, \tau), b_2(t, \tau))$. The second cut set, i.e., $\xi_2 = \{(L_1, D), (L_2, D)\}$ forms a molecular multiple-access channel among L_1 , L_2 , and D . Using the deterministic capacity of molecular multiple-access channel given in Section 9.3.3, molecular communication rate achieved during the interval $[t, t + \tau]$ in this cut set can be given as $\max(m_1(t, \tau), m_2(t, \tau))$. The third cut set, i.e., $\xi_3 = \{(S, L_1), (L_2, D)\}$ consists of two molecular communication links and the molecular rate in this set can be given as $(b_1(t, \tau) + m_2(t, \tau))$. The last cut set, i.e., $\xi_4 = \{(S, L_2), (L_1, D)\}$, also consists of two molecular communication links and the molecular information rate in this set can be given as $(b_2(t, \tau) + m_1(t, \tau))$. According to the maximum-flow minimum cut theorem [154], capacity of information flow between S and D during the interval $[t, t + \tau]$, i.e., $R_d(t, \tau)$ is equal to the minimum cut set capacity in the molecular nanonetwork shown in Fig. 9.5 and can be given as [153]

$$R_d(t, \tau) = \min \left[\max(b_1(t, \tau), b_2(t, \tau)), \max(m_1(t, \tau), m_2(t, \tau)), (b_1(t, \tau) + m_2(t, \tau)), (b_2(t, \tau) + m_1(t, \tau)) \right] \quad (9.27)$$

The average capacity of information flow, i.e., Φ , can be obtained by averaging $R_d(t, \tau)$ values observed in T number of consecutive intervals of τ as

$$\Phi = \frac{1}{T} \sum_{i=0}^{T-1} R_d(t_0 + i\tau, \tau) \quad (9.28)$$

The capacity analysis of information flow between S and D reveals that in the molecular nanonetwork, routing phenomenon among nanomachines that do not directly communicate with each other may be feasible with a quantifiable molecular communication rate. Moreover, in order to provide a satisfactorily high molecular communication rate, it may be essential to

develop a multiple-access mechanism that allows nanomachines to use different molecules as communication carrier that can be distinguished by RN. This stems from the fact that as introduced in Section 9.3.3, if TNs in the molecular multiple-access channel emit the same kind of molecules to the channel, the capacity of molecular multiple-access channel given as $\max(m_1(t), m_2(t))$ is extremely low and this may result in an extremely low Φ .

9.4 Numerical Analysis

In this section, we first present the numerical analysis on the molecular point-to-point, broadcast and multiple-access channels. Then, we present the numerical results on the information flow between a source and destination nanomachines in the diamond-like molecular nanonetwork. The aim of this analysis is to find the numerical values of molecular channel and information flow capacities and gain some insight for the feasibility of networking techniques in molecular nanonetworks. We also aim to observe the variations in these capacities with respect to the molecular communication parameters such as diffusion coefficient, internode distance, and receptor concentration r on RN. Since the molecular communication channel has a time-varying gain $h(t, \tau)$ as introduced in Section 9.2, we take the numerical results in consecutive time steps. In each time step, we compute the molecular information rate for the time interval of τ . Therefore, all capacity values are given in $\text{bits}/\tau\text{sec}$. In addition to the time-dependent molecular information rate, we also observe the average capacity values of molecular channel given in Section 9.3. We perform the numerical analysis using Matlab. The simulation parameters of the analysis are given in Table 9.1.

Table 9.1: Simulation Parameters

<i>Diffusion coefficient (D)</i>	$1e2 - 1e8 (\mu\text{m}^2/\text{sec.})$
<i>Upper bound for channel input (x_u)</i>	$1e16 - 1e24$
<i>Lower bound for channel input (x_l)</i>	$1e - 15$
<i>Distance between nanomachines (d)</i>	$1e2 - 5e4 (\mu\text{m})$
<i>Receptor concentration (r)</i>	$100 - 12500$
<i>Reaction prob. in VRV (η)</i>	0.01
<i>Time duration (τ)</i>	10 sec.
<i>Observation interval (T)</i>	$500 - 500000$
<i>Size of VRV (v_r)</i>	$1e - 18 (m^3)$

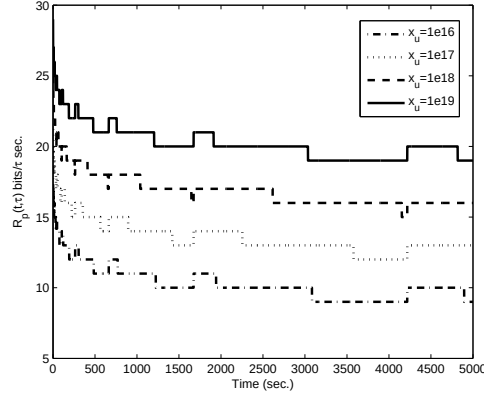


Figure 9.6: Information rate of point-to-point molecular channel for different upper bounds of channel input with time.

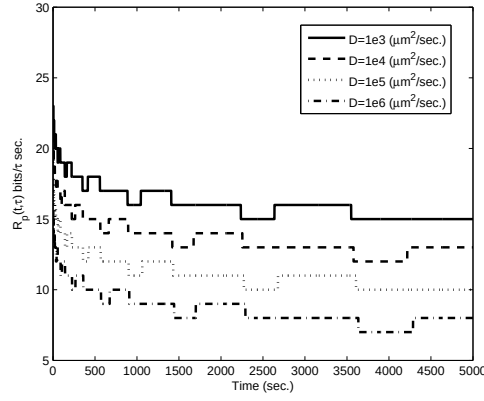


Figure 9.7: Information rate of point-to-point molecular channel for the different diffusion coefficients of the medium with time.

9.4.1 Point-to-point Molecular Channel

We first observe the effect of channel input upper bound (x_u) on molecular information rate given in (9.15). In Fig. 9.6, the molecular information rate of point-to-point molecular channel in the interval $[t, t + \tau]$, $R_p(t, \tau)$, is shown for different x_u with time. In the traditional wireless communication channel, as the upper bound of input power increases, signal-to-noise ratio increases and the communication rate can be improved. Similarly, in the point-to-point molecular channel, as x_u increases, $R_p(t, \tau)$ also increases. However, as seen in Fig. 9.6, while x_u is changed from $1e16$ to $1e19$, that is, x_u is increased 1000 times, the molecular information rate can be approximately doubled. Hence, due to very low energy supply of nanomachines [13], there is a clear trade off between the achievable capacity and energy conservation of

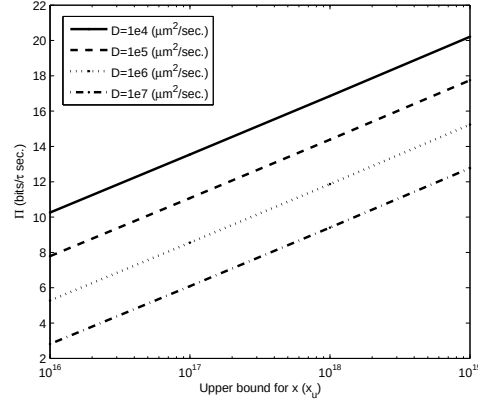


Figure 9.8: Average capacity of point-to-point molecular channel according to varying x_u with different D values.

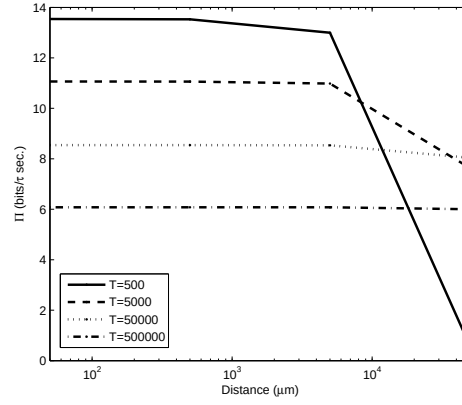


Figure 9.9: Average capacity of point-to-point molecular channel according to varying internode distance (d) with different observation intervals (T).

nanomachines. Therefore, the achievable molecular rate in point-to-point molecular channel is prone to very limited power resource of nanomachines. In Fig. 9.7, $R_p(t, \tau)$ is shown for different diffusion coefficients D ($\mu m^2/sec.$). The diffusion coefficient is a major parameter for the molecular information rate since it determines the average number of molecules that react with the receptors in VRV during the time interval $[t, t + \tau]$. As the diffusion coefficient increases, $R_p(t, \tau)$ also decreases. When the molecules are initially emitted by TN, the molecules diffuse in the medium and immediately reach to close proximity of RN and react with the receptors in VRV. After a course of time, the number of molecules in VRV starts to decrease since the molecules continue to diffuse toward all directions to which they do not reach yet. This decreases the molecule concentration in VRV and also decreases $R_p(t, \tau)$ with

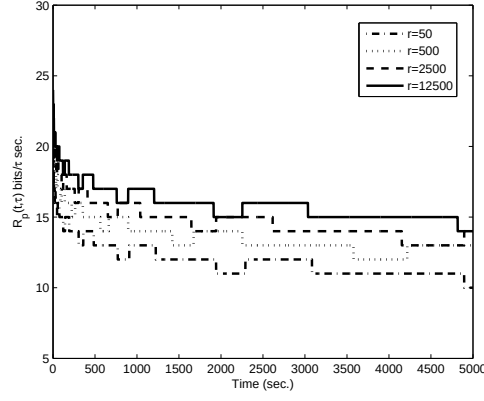


Figure 9.10: Information rate of point-to-point molecular channel for the different receptor concentration with time.

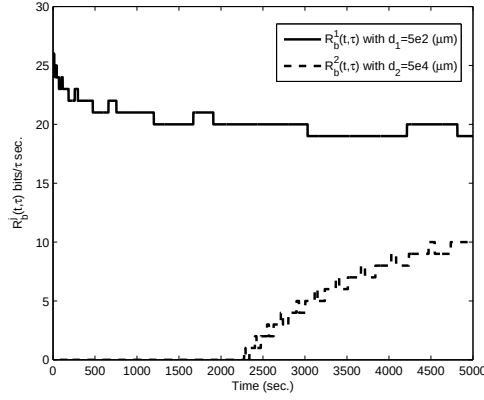


Figure 9.11: Information rate achieved by two distinct RNs located to different distances from a single TN in a molecular broadcast channel.

time. Hence, as the diffusion coefficient increases, the number of molecules in VRV decreases more quickly and the molecular communication rate decreases as shown in Fig. 9.7.

The average capacity Π is also shown in Fig. 9.8 according to the different diffusion coefficient levels with increasing x_u . Fig. 9.8 clearly follows the results observed in Fig. 9.6 and 9.7 since Π is obtained by simply averaging $R_p(t, \tau)$ values observed in T consecutive intervals of τ . Π increases with x_u , which is also shown in Fig. 9.6. Moreover, as the diffusion coefficient D increases, Π decreases. This result can also be seen in Fig. 9.7. In Fig. 9.9, Π is shown according to different distances between TN and RN with varying observation intervals T . For the lowest level of observation interval, i.e., $T = 500$, the molecules cannot completely reach to the longest distance, i.e., $d = 5e4 \mu m$ at the last of the $T = 500$ intervals. This results

in an extreme reduction in Π . However, as T increases, the molecules easily reach to the long distances and provide high level of molecular communication capacity. On the other hand, as T increases, due to the continuous diffusion of the molecules, the capacity decreases with time and consequently, average capacity becomes low for higher T values.

In Fig. 9.10, we observe the effect of receptor concentration (r) on the information rate of point-to-point molecular channel ($R_p(t, \tau)$). Similar to the diffusion coefficient, r also determines the reaction rate between the molecules and receptors in VRV such that $R_p(t, \tau)$ increases with r . However, due to continuous diffusion of molecule, the molecule concentration in VRV decreases with time and rate also decreases regardless of the receptor concentration as in Fig. 9.10.

9.4.2 Molecular Broadcast Channel

Here, we observe the molecular information rate in a molecular broadcast channel. Due to the fact that molecular communication is based on the free diffusion of molecules toward all directions, the molecular broadcast channel has similar characteristics with the point-to-point molecular channel. Especially, the effects of x_u , D , and r are the same as the point-to-point molecular channel. However, if RNs have different channel gains due to their different distances to TN, molecular broadcast channel may provide varying rate levels in two RNs, i.e., RN₁ and RN₂. In Fig. 9.11, we show the molecular information rate ($R_b^j(t, \tau)$) within the interval $[t, t + \tau]$ in the molecular broadcast channel with a single TN and RN _{j} for $j = 1, 2$. We assume that the distances from TN to RN₁ and RN₂ are different and given as $d_1 = 5e2 \mu m$ and $d_2 = 5e4 \mu m$, respectively. Due to $d_1 < d_2$, the emitted molecules reach to RN₂ later than RN₁. As observed in Fig. 9.11, until $t = 2300 \text{ sec.}$, RN₂ does not receive any molecule in its VRV although RN₁ already experiences many reactions. Hence, internode distances are the main determinant for the molecular information rate experienced by RN _{j} in a molecular broadcast channel.

9.4.3 Molecular Multiple-access Channel

Similar to the traditional wireless multiple-access channel, the most critical barrier toward the realization of molecular multiple-access communication is the interference among the trans-

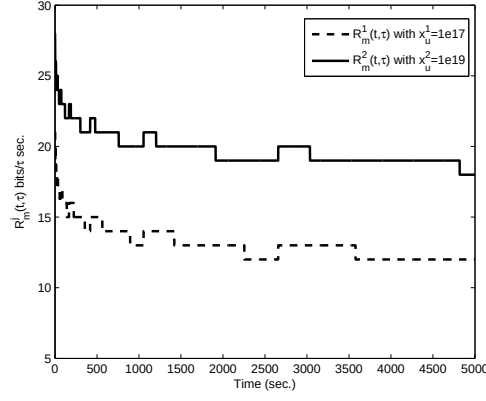


Figure 9.12: Information rate achieved by two distinct TNs using different level of upper bound for channel input to access a single RN.

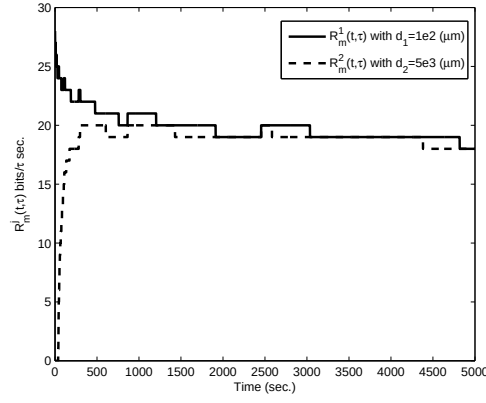


Figure 9.13: Information rate achieved by two TNs located in different distances from RN in a molecular multiple-access channel.

mitter nanomachines that simultaneously access to the same receiver nanomachine. However, unlike the traditional multiple access mechanisms such as code division multiple access (CDMA) and time division multiple access (TDMA), in the molecular multiple-access channel, the interference among the transmitter nanomachines can be prevented by enabling each TN to use different kind of molecules that can be uniquely distinguished by RN. In Fig. 9.12, the molecular information rates, $R_m^j(t, \tau)$, achieved by TN_1 and TN_2 with $x_u^1 = 1e17$ and $x_u^2 = 1e19$, respectively, are illustrated.

Clearly, TN_2 reaches a rate higher than TN_1 since the average number of reactions in the VRV of TN_2 is much larger than the reactions in VRV of TN_1 . In Fig. 9.13, $R_m^j(t, \tau)$ achieved by each of two distinct TNs located in different distances from a single RN is shown for

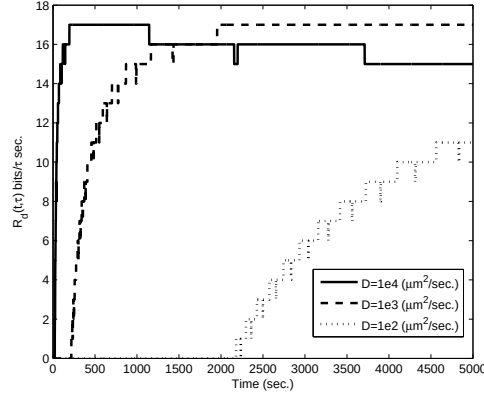


Figure 9.14: Information flow capacity in diamond-like molecular nanonetwork for different diffusion coefficients.

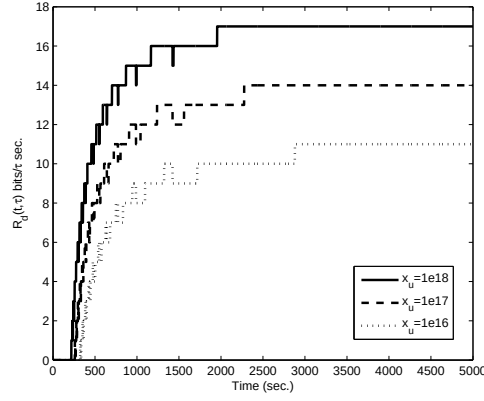


Figure 9.15: Information flow capacity in diamond-like molecular nanonetwork for different upper bound levels of channel inputs.

$j = 1, 2$. Similar to the molecular broadcast channel, in molecular multiple-access channel, the transmission distance extremely affects the molecular information rate. The distances from RN to TN_1 and TN_2 , i.e., d_1 and d_2 , respectively, are given as $d_1 = 1e2 \mu m$ and $d_2 = 5e3 \mu m$. Clearly, as $d_1 < d_2$, the molecules emitted by TN_1 reach to RN more quickly than the other molecules emitted by TN_2 . Therefore, TN_1 quickly experiences higher molecular communication rate and as the emitted molecules diffuse. This decreases the average number of molecules in VRV and the molecular communication rate decreases. Conversely, since the distance between TN_2 and RN is higher, the molecules emitted by TN_2 reach to VRV of RN later than the other molecules emitted by TN_2 . Thus, TN_2 communicates with RN after its molecules reach to VRV of RN and the rate abruptly increases. After a course of time, TN_1

and TN_2 start to achieve the same information rate level.

9.4.4 Diamond-like Molecular Nanonetwork

The diamond-like molecular nanonetwork consists of a molecular broadcast and multiple-access channel. Therefore, the capacity of information flow between a source and destination nanomachines is affected by the characteristics of these channels. Here, we observe the effects of D and x_u on the information flow rate in the diamond-like molecular nanonetwork. In Fig. 9.14, the information flow rate is shown according to different D . Contrary to the point-to-point molecular channel, as D increases, the information flow rate also increases. Moreover, due to the low level of diffusion coefficient that results in high propagation delay, the rate becomes 0 during a time interval. However, as soon as the molecules reach to the VRV of the destination nanomachine, it abruptly increases. In Fig. 9.15, we observe the effect of x_u on the information flow capacity. As observed, similar to the point-to-point molecular channel, as x_u increases, the information flow rate increases in the diamond-like molecular nanonetworks.

9.5 Conclusion

In this chapter, using the stochastic models in chemically reacting systems, we first devise a realistic channel model for molecular communication. Then, we define and analyze the molecular point-to-point, broadcast, and multiple-access channels. We also investigate the deterministic capacity of information flow in a molecular nanonetwork. Our models and investigations clearly reveal that molecular communication links among nanomachines can be established with sufficiently high communication rate in nanonetworks. Unlike the traditional Gaussian communication channel, in molecular communication channel, achievable capacity is affected by the both of the lower and upper bound of the channel input. This arises from the input-dependent noise term. In the point-to-point and broadcast molecular communication channel, communication rate can be improved by increasing the upper bound of the channel input. However, there is a rough trade off between communication rate and energy consumption of nanomachines. In the molecular multiple-access channel, in order to obtain sufficiently high molecular communication rate, transmitter nanomachines must either schedule their transmissions or use different kind of molecules that are distinguished by the

receiver nanomachines. Due to the very low-end computation and synchronization capabilities of nanomachines, it may not be possible for transmitter nanomachines to schedule their transmissions. Instead, the transmitter nanomachines may use different kinds of molecules to obtain sufficient level of molecular communication rate. In a molecular nanonetwork, it may be possible to communicate the network nodes that cannot directly communicate with each other with the helps of relay nodes with a satisfactorily high communication rate. This may allow a possible realization of a routing scheme in a molecular nanonetwork, which needs to be further investigated.

CHAPTER 10

Mobile Ad Hoc Nanonetworks with Collision-based Molecular Communication

The aim of this chapter is to introduce the concept of mobile ad hoc nanonetworks (MAMNET) and provide an analytical framework in order to show feasibility of MAMNET. The communication of mobile nanomachines is enabled by three main phases, namely, collision, adhesion and transmission. In collision phase, nanomachines randomly collide with each other. In adhesion phase, nanomachines stuck to each other. Finally, the information is transmitted in the transmission phase. For the transmission of information, we introduce a new communication paradigm, called *molecular neuro-spike communication*, that is inspired by electrochemical communication among biological neuron cells. After the adhesion phase, molecular neuro-spike communication scheme enables the communication of adherent nanomachines. MAMNET was also presented in [155]

The motivation and the related works are introduced in Section 10.1. In Section 10.2, we give an overview of MAMNET, and underline the design issues and assumptions. Then, we model the collision and adhesion of nanomachines in Section 10.3. In Section 10.4, we introduce molecular neuro-spike communication of adherent nanomachines and an analytical capacity expression for molecular neuro-spike channel. In Section 10.5, we derive the distribution of communication latency and find average throughput in MAMNET. We present the numerical results in Section 10.6 and conclude in Section 10.7.

10.1 Motivation and Related Works

Rapid growth in nanotechnology provides favorable development in miniaturization and fabrication of nanomachines with simple sensing, computation, communication, and action capabilities. In literature, three different ways are proposed for the development of nanomachines, namely, bottom-up, top-down and bio-hybrid approaches [12]. In bottom-up approach, the molecules or atoms are assembled to form nanomachines. In top-down approach, the design of nanomachines is realized by downsizing current microelectronic devices. In bio-hybrid approach, biological entities can either be genetically modified to develop nanomachines or used as the building blocks of nanomachines. In nature, there are already many molecular-scale phenomena consisting of an arranged set of molecules, which are able to perform very simple tasks. These phenomena can be envisaged as nature-made nanomachines. For example, a chloroplast in a plant cell stands as a nanomachine including arrays of molecules that act as tuned optical antennas for absorption and transformation of solar energy. Mitochondrion can be envisioned as a nanomachine used for controlled combustion of organic molecules to generate adenosine triphosphate (ATP) for fulfilling energy needs of cellular activities. A flagellar motor attached on the membrane of many bacterial cells is a highly structured combination of proteins for providing cellular movement [9]. Nature-made nanomachines can be exploited to learn and understand the principles governing the operation of nanomachines and their interactions [9].

The limited capabilities of nanomachines also limit the applications that can be achieved by a single nanomachine. However, the networks of communicating nanomachines, i.e., *nanonetworks*, are expected to enable very large set of new applications in various research fields such as genetic engineering, health monitoring, and military surveillance systems [12]. For the realization of frontier nanonetworks, it is imperative to develop new and efficient communication and networking techniques that aim to overcome unique challenges and requirements of nanonetworks which can be briefed as follows:

- Scale of the nanomachines is on the order of micrometers, therefore, classical transceiver circuitries cannot be mounted into nanomachines.
- Current encoding and decoding techniques are not feasible due to very limited processing capability of nanomachines.

- For in-vivo application scenarios, nanomachines need to be biocompatible in order not to be rejected by the organism.
- Mobility of nanomachines is governed by the physical rules in nano domain [156].
- Communication or noise signal characteristics cannot be easily anticipated due to severely unreliable nature of the communication medium.

In literature, different communication techniques, namely, acoustic, electromagnetic, molecular, and nanomechanical, are proposed for the communication of nanomachines [13]. Traditional electromagnetic and acoustic communication technologies cannot be directly used in the communication between nanomachines because of the size restrictions. Nanomechanical communication requires a strict physical contact between transmitter and receiver in order to enable information transmission, which makes it inconvenient for many application scenarios. On the other hand, molecular communication, which is already used by biological entities, is a promising approach for the communication of nanomachines.

Existing communication paradigms between cells may be adopted for the realization of nanonetworks [39]. In [136], a design of a molecular communication system based on intercellular calcium signaling is introduced. Furthermore, [157] presents an information theoretical model of calcium signaling channel, which is a good interdisciplinary approach combining biology and information theory. Using this model, the capacity of calcium signaling channel is defined and simulations are given to indicate the optimal design of biological systems. The simulation experimental results in [157] provide an important understanding of the design principles of calcium signaling communication. In [134] and [158], molecular communication systems using biological molecular motors and vesicles as communication carriers are introduced. Molecular communication channel is modeled as a binary symmetric channel and mutual information and capacity expressions are derived for that channel in [132]. Information theoretical capacity analysis of diffusion-based molecular communication in nanonetworks is performed in [159]. In [160], modulation techniques for communication via diffusion in nanonetworks are introduced. A flagellated bacteria and catalytic nanomotors based molecular nanonetworks is proposed in [161]. In [162], a computational model for mobile nanomachines using molecular communication is introduced. In [40], the first realistic and very comprehensive physical channel model of diffusion-based molecular communication has been developed.

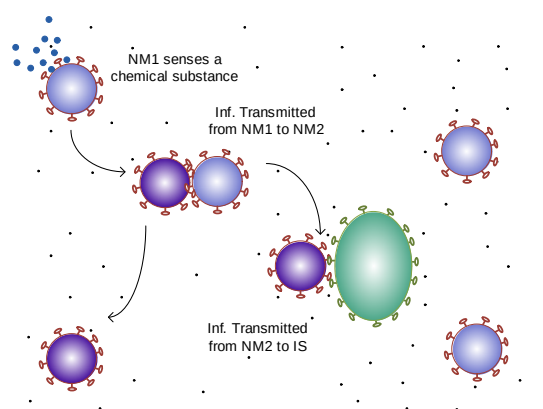


Figure 10.1: Information flow in MAMNET.

In literature, generally, nanomachines are considered as immobile nodes. However, mobile nanomachines may be indispensable for many nanonetwork applications. A possible example that necessitates the mobile nanomachines is a nanonetwork designed for coordinated cancer cell detection by identifying cancer cells and informing a central controller to take an appropriate action. Clearly, these applications necessitate the realization of mobile ad hoc molecular nanonetworks (MAMNET). In MAMNET, nanomachines collect some environmental information such as a chemical state, or the existence of a certain concentration level and deliver the information collected by mobile nanomachines to mobile *infostations* as shown in Fig. 10.1. The infostations are central control units that make decisions according to the collected information, or gateways that connect MAMNET to a micro-device.

However, the mobility of nanomachines also incurs a new set of crucial challenges that must be addressed for the realization of MAMNET. One of the main challenges introduced by the mobility is the intermittent connectivity, i.e., nanomachines can communicate only when they are in physical contact. Therefore, this intermittent connectivity clearly imposes a high level of latency on nanoscale communication. To the best of our knowledge, the feasibility and performance of an ad hoc nanonetwork that is composed of mobile nanomachines communicating through electrochemical means have not yet been investigated.

10.2 Mobile Ad Hoc Molecular Nanonetwork Communication Model

MAMNET is composed of two kinds of mobile nanonodes, namely, nanomachines and infostations as shown in Fig. 10.1. Nanomachines and infostations are assumed as genetically modified cells with additional capabilities. Nanomachines are assumed to sense the environment and gather some environmental information to deliver infostations. Based on the collected information coming from nanomachines, infostations are assumed to make decision for an appropriate action.

In order for a nanomachine to transmit collected information to an infostation, the nanomachine and the infostation should first collide, then adhere to each other. Since the system volume containing MAMNET is much larger than the size of a single nanonode, the meeting probability of a nanomachine and an infostation is very low. Thus, intermediary nanomachines can be used as relay nodes that help the communication between the nanomachines and the infostations as follows.

Fig. 10.1 describes the information flow in MAMNET. Nanomachine 1 is assumed to be the source of the information that is to be communicated to the infostation. Nanomachine 1 transmits the information to every nanomachine, i.e., relay node, with which it collides and adheres. The nanomachines that acquire the information follow the same strategy and this clearly increases the probability of information delivery. Note that this communication strategy is clearly similar to the spreading of epidemic disease.

As will be detailed in Section 10.3, nanomachines diffuse in the environment and randomly collide with each other. The collided nanonodes adhere to each other via the surface mounting molecules called, ligands and receptors. This phenomenon is commonly known as the ligand-receptor binding process. According to this process, ligand molecules on the surface of one biological entity, i.e., nanonode, bind to the receptors of an other nanonode. Once a certain number of bonds is established between the nanonodes, they can adhere to each other.

The performance of MAMNET directly depends on the interaction of nanonodes that are triggered by collision and adhesion of nanomachines and infostation. Therefore, it is important to investigate the collision rate, i.e., the number of collisions occur per unit time, and the adhesion probability of collided nanomachines and infostations.

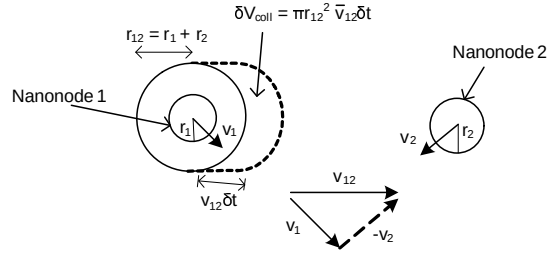


Figure 10.2: Collision of two nanomachines.

10.3 Collision and Adhesion of Nanomachines

In nature, mobile biological cells collide with each other and cellular adhesion occurs between the collided cells [143]. The cell adhesion is accomplished by the ligand-receptor binding process on the surface of the cells. The adhesion has a critical role in intracellular and inter-cellular signaling, i.e., communication, that governs basic cellular activities and coordinates cell actions [163].

In this chapter, we adopt the natural cellular communication for the communication among nanomachines. Similar to living cells, nanomachines should be in physical contact with each other in order to communicate. The physical contact is established through the collision and adhesion of nanonodes that are modeled as follows.

10.3.1 Collision of Nanomachines

As in traditional mobile ad hoc networks (MANET), in MAMNET stochastic nature of nanomachine collision rate is governed by some mobility models such as Brownian motion and random waypoint. Since nanomachines do not have any capability of controlling their movement, they are assumed to freely diffuse in the aqueous medium. Hence, in this chapter, it is assumed that the mobility of nanomachines is governed by the dynamics of Brownian motion that explains the movement of small particles that freely diffuse in an aqueous medium. However, in some realistic scenarios, nanomachines can be deployed in flows which introduces a drift for the motion of nanomachines. Therefore, in these scenarios, Brownian motion with drift might be used to model the collision of nanomachines, and such analysis is left for a future work.

Nanomachines with radius r are assumed to be contained in a volume V such that $V \gg r$. To find the collision rate of nanomachines, we first obtain the probability that the nanomachines collide within the next infinitesimally small time interval δt , given that the first nanomachine's center is located at the position (x_1, y_1, z_1) . Two nanomachines collide in the next δt , only if the second nanomachine is in the volume covered by the first nanomachine with respect to the second nanomachine. This collision volume δV_{coll} is shown in Fig. 10.2 and expressed as

$$\delta V_{coll} = \pi r_{12}^2 v_{12} \delta t \quad (10.1)$$

where $r_{12} = r_1 + r_2$ and v_{12} is the relative velocity of the first nanomachine with respect to the second nanomachine. By using the relative velocity, the second nanomachine is considered as stationary while the first one is considered moving with velocity v_{12} rather than v_1 . Thus, probability that the second nanomachine is in volume δV_{coll} , i. e., two nanomachines collide, is expressed by

$$p_{x_1, y_1, z_1} = \iiint_{\delta V_{coll}} f(x, y, z) dx dy dz \quad (10.2)$$

where $f(x, y, z)$ is the probability density function (pdf) of spatial node distribution in the volume V . For small values of r , the points $f(x, y, z)$ in V can be approximated by $f(x_1, y_1, z_1)$. This gives the probability of a collision within the next infinitesimally small time interval, given that the first node's starting position is (x_1, y_1, z_1) , i.e.,

$$p_{x_1, y_1, z_1} \approx \pi r_{12}^2 v_{12} \delta t f(x_1, y_1, z_1) \quad (10.3)$$

The first node's starting position can be changed. Therefore, by integrating p_{x_1, y_1, z_1} over all starting positions of the first node, the probability of collision within the next infinitesimally small time interval δt can be obtained as

$$\begin{aligned} p &= \iiint_V p_{x_1, y_1, z_1} f(x_1, y_1, z_1) dx_1 dy_1 dz_1 \\ &\approx \pi r_{12}^2 v_{12} \delta t \iiint_V f^2(x_1, y_1, z_1) dx_1 dy_1 dz_1 \end{aligned} \quad (10.4)$$

(10.4) is specific to the relative speed v_{12} . Averaging the relative velocity over the velocity distributions of the nanonodes, the collision rate for nanodes can be approximated as

$$R_c \approx \pi r_{12}^2 E[v_{nn}] \iiint_V f^2(x_1, y_1, z_1) dx_1 dy_1 dz_1 \quad (10.5)$$

where $E[v_{nn}]$ is the average relative speed of the nanomachines. In order to evaluate the collision rate, the spatial node distribution $f(x, y, z)$, and average relative speed $E[v_{nn}]$ for the

nanomachines moving according to Brownian motion are required. In [216], it is shown that the steady-state node distribution in Brownian motion is uniform. Thus, $f(x, y, z) = 1/V$ and R_c can be approximated as

$$\begin{aligned} R_c &\approx \pi r_{12}^2 E[v_{nn}] \iiint_V 1/V^2 dx_1 dy_1 dz_1 \\ &\approx \frac{\pi r_{12}^2 E[v_{nn}]}{V} \end{aligned} \quad (10.6)$$

Since the size of nanomachines are identical, r_{12} can be given as $r_{12} = 2r$, where r is the radius of a nanomachine. Then, from (10.6), the collision rate for nanomachines can be expressed as

$$R_c \approx \frac{4\pi r^2 E[v_{nn}]}{V} \quad (10.7)$$

The collision rate of an infostation and a nanomachine follows a similar argument except that infostations could have a different radius compared to nanomachines. For nanomachine-infostation r_{12} is set as $r_{12} = r + r_i$, where r_i is the radius of an infostation. Then, similar to the collision rate of nanomachines, the collision rate of an infostation and a nanomachine can be written as

$$R_{ic} \approx \frac{\pi(r + r_i)^2 E[v_{ni}]}{V} \quad (10.8)$$

where $E[v_{ni}]$ is the average relative speed of the nanomachine and infostation.

The collision rates R_c and R_{ic} are essential to understand the interaction rate of nanomachines and infostations. However, the collision rates are not the ultimate parameters that affect the MAMNET performance since the collided nanomachines should also adhere with each other to enable molecular neuro-spike communication.

10.3.2 Adhesion of nanomachines

After the collision between nanomachines, the collided nanomachines adhere with each other via the binding of the surface molecules called ligands and receptors, as shown in Fig. 10.3. In nature, cell adhesion is involved in a variety of processes such as migration, invasion, embryogenesis, wound healing and cell-to-cell communication [163]. In literature, there exists several research efforts on modeling the adhesion between cells [164], [143]. Here, the adhesion model developed in [164] is adopted. Accordingly, nanomachines are assumed to have ligands and receptors which mediate adhesion.

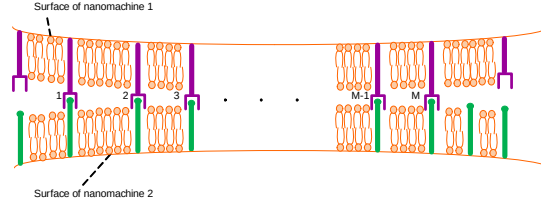


Figure 10.3: Nanomachine 1 and Nanomachine 2 adhere with each other by M ligand receptor pairs.

Adhesion is considered as a random event and the state of the system is considered as a probability vector $[p_0, p_1, \dots, p_n, \dots, p_{A_c m_{min}}]$ where m_{min} is the minimum of surface densities of receptors and ligands and A_c is the area of contact [164]. In other words, adhesion could be mediated by any number of bonds ranging from 0 to $A_c m_{min}$. For forming any number of bonds, there is a defined likelihood given by p_n . Here, nanomachines are assumed to adhere when at least c bonds are formed between them. Hence, our aim is to derive the probability that nanomachines adhere via at least c bonds. In [164], a closed-form expression is found for p_n under a set of certain assumptions for two cells, such as either ligands or receptors outnumbers the other one. When two nanomachines are just brought together ($t = 0$), there is no bond, hence,

$$p_n(0) = \begin{cases} 1 & , n = 0 \\ 0 & , n \neq 0 \end{cases} \quad (10.9)$$

With the contact of nanomachines, bonds start to form according to a single step reversible reaction. The chemical reaction of v_r receptors (designated M_r) binding to v_l ligands (designated M_l) to form v_b bonds (designated M_b) can be expressed as



where k_f^0 and k_r^0 are the forward and reverse rate constants, respectively.

In the stochastic model of ligand-receptor binding process, the state of the system, described by the single step reversible chemical reaction in (10.10), is represented by the probability vector $[p_0, p_1, \dots, p_n, \dots, p_{A_c m_{min}}]$. In [165], a stochastic model of ligand-receptor binding process is discussed and master equations are derived in order to describe the state of the system. Actually, these master equations can be used to describe the rate of change in probability $p_n(t)$. In the current literature, two simplified versions of master equations are discussed in order to obtain closed-form expressions. In [164], a simplified version is discussed under the

condition that either ligands or receptors excessively outnumber the other one. Under that condition and with $v_r = v_l = v_b = 1$, i.e., a bond is formed by the binding of a ligand to a receptor, the master equations can be simplified as [164]

$$\begin{aligned} \frac{dp_n}{dt} = & \left[A_c m_{min} - (n-1) \right] m_{max} k_f^0 p_{n-1} - \\ & - \left[(A_c m_{min} - n) m_{max} k_f^0 + n k_r^0 \right] p_n + (n+1) k_r^0 p_{n+1} \end{aligned} \quad (10.11)$$

Assuming either ligands or receptors excessively outnumber the other one, $p_n(t)$ is found to be in the form of binomial distribution and can be given as [164]

$$p_n(t) = \binom{A_c m_{min}}{n} [p(t)]^n [1 - p(t)]^{A_c m_{min} - n} \quad (10.12)$$

where $p(t)$ is the probability of forming one bond given by

$$p(t) = \frac{1 - e^{-kt}}{1 + (m_{max} K_a^0)^{-1}} \quad (10.13)$$

where $K_a^0 = k_f^0/k_r^0$ is the equilibrium association constant and $k = m_{max} k_f^0 + k_r^0$ is the overall rate of reaction.

After the collision event between two nanomachines, they are assumed to stay in contact with each other during an average contact duration τ_c . τ_c is affected by the physical properties of the environment, relative velocities and physical surface properties of the nanomachines. The effect of τ_c on performance is discussed in Section 10.6. We also assume that the collided nanomachines adhere with each other if at least c bonds are formed during a contact duration τ_c . Hence, the probability that the collided nanomachines adhere with each other, i.e., R_a , can be given as

$$R_a = 1 - \sum_{i=0}^{c-1} p_i(\tau_c) \quad (10.14)$$

Here, we note that in [164], using single cell pair of a chinese hamster ovary cell interacting with a red blood cell, a comprehensive experimental data set on biological adhesion process is presented and analyzed. In [164], it is also shown that the probabilistic cell-to-cell adhesion model that is adopted for our model satisfactorily follows and is validated by this experimental data set. Furthermore, we do not present a different adhesion rate for the nanomachine-infostation interaction. This is because the physical properties used in the derivation of adhesion rates are the same for both nanomachines and infostations. The adhesion probability

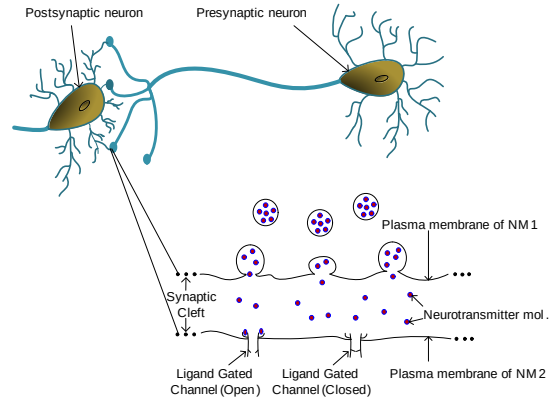


Figure 10.4: Neuro-spike transmission between nanomachines.

derived in this section is very important to understand the interaction of nanomachines. The communication between nanomachines can only be possible after successful adhesion.

10.4 Molecular Neuro-Spike Communication

In this section, molecular neuro-spike communication is explained in detail. First, the basics of molecular neuro-spike communication are introduced. Then, its channel capacity and error probability are analytically investigated. Finally, successful information transmission probability in the molecular neuro-spike communication channel is derived to be used for the performance evaluation of MAMNET.

10.4.1 Molecular Neuro-Spike Communication

When a nanomachine that has an information collide and adhere to another nanomachine, it transmits its information to this nanomachine by means of molecular neuro-spike communication. After the adhesion of nanomachines, there exists a small gap, synapse, between the transmitter nanomachine (TN) and receiver nanomachine (RN). This synapse is the communication media for nanomachines. Since TN and RN cannot remain adhered for a long time, a fast and reliable communication paradigm is needed to enable the information transmission between nanomachines. Among the existing inter-cellular communication paradigms, the communication between neuron cells is the fastest one. Therefore, in our model, we adopt the principles of neural communication to enable the information transmission among

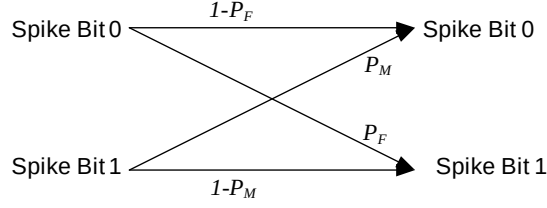


Figure 10.5: Channel model of synaptic transmission. Transmission probabilities are shown for the binary channel.

nanomachines.

In human body, there already exist at least two types of natural synapses between cells, namely, the neuronal synapse and the immunological synapse. These specialized contacts directly transfer highly controlled secretory signals between the adjacent cells [166]. The neuronal synapse is formed between the neurons and it retains the connectivity of the neuron cells throughout the life, whereas the immunological synapse is formed by the instant contacts of immune cells. Although nervous system and immune system have totally different roles in human life, in both systems cells need to communicate, and this is accomplished in the synapse. Therefore, synapses play a critical role in the cell-to-cell communication.

In nervous system, neural signals propagate in the form of electrochemical waves, which are basically action potentials that propagate along axons and transmitted to other neurons. Action potential is basically an electrical pulse which has approximately 80 mV amplitude. At synapse, action potentials or *spikes* are electrochemically transmitted from one neuron to another. The signal is transmitted by means of chemical messengers called neurotransmitters. Neurotransmitters are contained in vesicles. The coming pulse or action potential, releases the neurotransmitters on the pre-synaptic neuron, and then, the neurotransmitters bind to the receptors on the post-synaptic neuron. Binding of a neurotransmitter to a receptor opens a channel which lies between inside and outside of the post-synaptic neuron. These channels allow the flow of ions to the neuron. The net movement of ions causes the membrane potential of the post-synaptic neuron to change rapidly. In this way, spikes or action potentials are transmitted to the next neuron.

In this work, we adopt the principles of neural communication to develop molecular neuro-spike communication model. The communication between nanomachines is realized with spike transmission as shown in Fig. 10.4. The communication process based on molecular

neuro-spike communication involves the following steps:

1. *Encoding*: Similar to the traditional digital communication, we define 2 bit levels, i.e., spike bit 0 and spike bit 1 corresponding to logic 0 and 1. The information is encoded on the concentration of the released neurotransmitter molecules. For spike bit 1, neurotransmitters are released to the channel, whereas for spike bit 0 no releasing occurs.
2. *Transmission*: Actually, this process initiates the electrochemical signaling. Whenever a TN wants to send a spike bit 1, it simply activates the release of vesicles that contain neurotransmitters.
3. *Signal Propagation*: The released neurotransmitters propagate in the synaptic channel formed between the adherent nanomachines. The aim of the neurotransmitters is to bind to the receptors on the RN.
4. *Reception*: The neurotransmitters released into the synaptic channel reach the RN and bind to the receptors on the RN membrane as in Fig. 10.4. The binding of neurotransmitters to receptors opens ligand gated channels that let the flow of ions into or out of the RN. The flow of ions changes the membrane voltage.
5. *Decoding*: RN nanomachine monitors the plasma membrane voltage for certain time periods. If RN observes a rapid change in the membrane voltage it decides the received bit as spike bit 1, otherwise the decision is spike bit 0.

In an ideal channel, the spike bit sent by a TN should be perfectly received by the RN. However, in practice there exist two kinds of errors as shown in Fig. 10.5, namely, false alarm and miss detection errors. The false alarm and miss detection probabilities, i.e., P_F and P_M , respectively, can be defined as

$$P_F = \text{Prob}[Y = 1|X = 0]$$

$$P_M = \text{Prob}[Y = 0|X = 1] \quad (10.15)$$

where X and Y are the input and output symbols, respectively. In order to derive these probabilities, a close look at neuronal synapse would be helpful. In [167], a cortical synapse is modeled as a binary channel and the information-theoretical capacity is derived for two different coding paradigms, i.e., signal estimation and signal detection. Although this synaptic

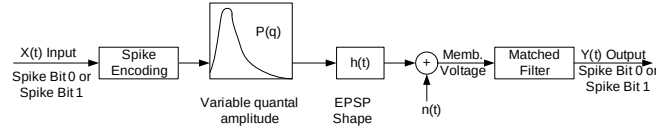


Figure 10.6: Block diagram of the channel model of synaptic transmission [167]

transmission model ignores certain aspects like paired-pulse facilitation, vesicle depletion, calcium buffering, it represents a simplified and compact picture of synaptic transmission. Therefore, we adopt the synaptic transmission model in [167] and accordingly derive the channel capacity expression.

In Fig. 10.6, the block diagram of the channel model of synaptic transmission is shown. The input to the channel, $X(t)$, is a spike bit. If the input is spike bit 1, vesicles containing neurotransmitters are released by the TN to the synaptic cleft. For spike bit 0, no vesicles are released.

Neurotransmitters binding to the receptors on the RN membrane open channels that let the flow of ions. The flow of ions create an Excitatory Postsynaptic Potential (EPSP) on the RN. In Fig. 10.6, EPSP profile or shape of RN is modeled by a filter with impulse response $h(t)$ [167]. In Fig. 10.6, the random variable q , which has probability density $P(q)$, models the variability in the size of EPSP. The number of neurotransmitter molecules released by the TN, the number of available postsynaptic receptors on the RN and several other factors affect this variability. In Fig. 10.6, there exists also an additive noise, $n(t)$, coupled to the response of RN. The noise $n(t)$ accounts for the other noise sources in the membrane of the RN such as thermal noise and channel noise [168].

10.4.2 Information Theoretical Capacity of Molecular Neuro-Spike Communication

In molecular neuro-spike communication, at the receiver side, the aim of RN is to optimally detect the presence or absence of a single spike through the knowledge of membrane voltage $V_m(t)$. In [167], it is shown that the optimal decision rule for spike detection is to compare the correlation r between $V_m(t)$ and $h(t)$ to a threshold θ . Thus, the decision rule can be written as

$$\begin{aligned} r &\geq \theta \Rightarrow Y = 1, \\ r &< \theta \Rightarrow Y = 0 \end{aligned} \tag{10.16}$$

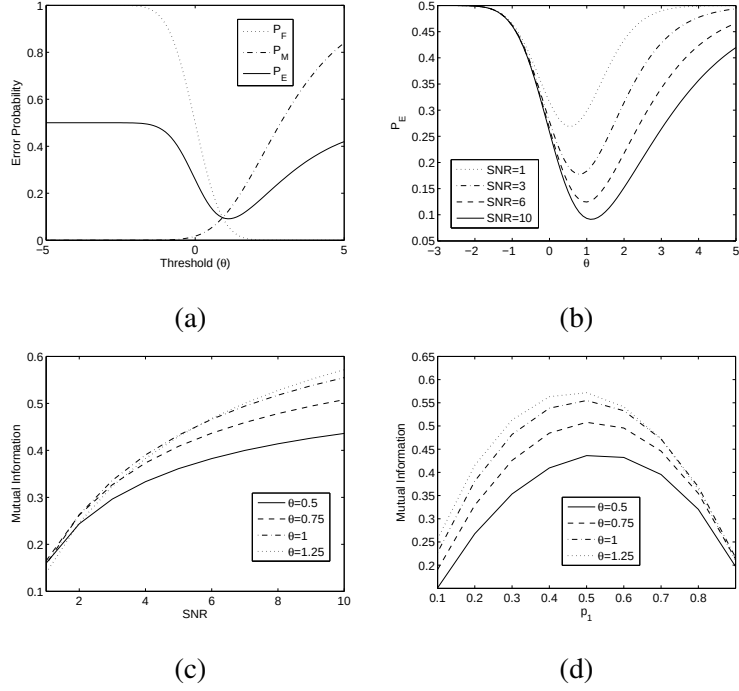


Figure 10.7: (a) False alarm, miss and channel error probabilities in molecular neuro-spike channel for a fixed value of $SNR = 10$ with varying θ , ($p_1 = 0.5$). (b) Channel error probability in molecular neuro-spike channel with varying θ for different values of SNR , ($p_1 = 0.5$). (c) Mutual information with varying SNR for different values of θ , ($p_1 = 0.5$). (d) Mutual information with varying p_1 for different values of θ , ($SNR = 10$).

With this decision rule, the false alarm and miss detection probabilities are written as

$$\begin{aligned} P_F &= Prob[r \geq \theta | X = 0] \\ P_M &= Prob[r < \theta | X = 1] \end{aligned} \quad (10.17)$$

In [167], the probabilities of miss detection and false alarm are derived for both stochastic and deterministic vesicle release processes. In our model, we use the probabilities derived for deterministic vesicle release, i.e.,

$$P_F = \frac{1}{2}[1 - Erf(\theta)] \quad (10.18)$$

$$P_M = \frac{1}{2} \left[1 + \int_0^\infty Erf(\theta - q\sqrt{SNR})P(q) dq \right] \quad (10.19)$$

where $Erf(x)$ is the error function and SNR is the signal-to-noise ratio on the post-synaptic potential.

Based on the false alarm and miss detection probabilities, the molecular neuro-spike channel

can now be modeled as a binary symmetric channel. TN emits spike bit 1 with probability p_1 and spike bit 0 with probability $(1 - p_1)$. Then, the probability of error for one spike bit can be written as

$$P_E = p_1 P_M + (1 - p_1) P_F \quad (10.20)$$

The value of P_E depends both on the threshold θ and SNR . For small values of θ , P_M is low whereas P_F is high. On the other hand, for large values of θ , P_F is low whereas P_M is high. Therefore, P_E should be evaluated with varying θ in order to find an optimum value for θ . In Fig. 10.7(a), with decreasing SNR , the probability of error increases. Furthermore, SNR also has an impact on the optimum value of threshold. As can be seen in Fig. 10.7(b), the optimum value of θ becomes closer to 0 as the value of SNR decreases.

The performance of the molecular neuro-spike communication can be further quantified by the mutual information. The transition matrix of the molecular neuro-spike channel can be written as follows

$$P(Y|X) = \begin{pmatrix} p_1(1 - P_M) & (1 - p_1)P_F \\ p_1 P_M & (1 - p_1)(1 - P_F) \end{pmatrix} \quad (10.21)$$

Using the transition matrix in (10.21), the mutual information $I(X; Y)$ between X and Y can be derived as follows

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= H(p_1 P_M + (1 - p_1)(1 - P_F)) \\ &\quad - (p_1 H(P_M) + (1 - p_1) H(P_F)) \end{aligned} \quad (10.22)$$

where $H(z)$ denotes the binary entropy function $H(z) = -z \log_2(z) - (1 - z) \log_2(1 - z)$. Mutual information corresponds to how much one can guess about the input with the knowledge of the output. θ and SNR have a direct effect on the value of mutual information. For low SNR values, mutual information can be very low. However, as SNR increases the mutual information increases as well, as shown in Fig. 10.7(c). The optimum value of θ changes with SNR .

In a synaptic channel, spikes can be transmitted by 3 ms differences which is composed of 0.5 ms action potential pulse and 2.5 ms of recovery time. Thus, a theoretical bandwidth of 333 bps can be reached. This corresponds to an ideal synaptic channel with $I(X; Y) = 1$, i.e.,

every bit transmitted contains exactly 1 bit of information. However, in practice there exist channel errors and this rate cannot be achieved.

The capacity of the molecular-neuro spike channel can be obtained by maximizing the mutual information over all input distributions, i.e.,

$$C = \max_{p_X}(I(X; Y)) \quad (10.23)$$

As can be seen in Fig. 10.7(d), the mutual information can be at most 0.57 for the channel under consideration. This corresponds to a bandwidth of approximately $333 \times 0.57 \approx 190$ bps, which is a considerably high bandwidth.

Moreover, neuro-spike communication model introduced above is realistic and consistent with some experimental work in terms of observed information rate. For example, in [169], similar to our model, neuronal spikes are considered as an information carrier and the carried information is quantified by bits. Using experimental data obtained by a motion sensitive neuron in the fly visual system, an information rate up to 90 bps is observed. This rate is clearly consistent with the rate region of neuro-spike communication. In fact, in Section 10.6, information message is encoded with 8 bits. Thus, the information capacity of neuro-spike communication is beyond the rate needed to successfully transmit the message.

By analyzing single spike bit error probability and channel capacity expressions, it is shown that molecular neuro-spike channel promises a reliable and fast communication. Therefore, molecular neuro-spike communication stands as a promising solution for the communication of adherent nanomachines in MAMNET. In evaluating the performance of MAMNET, information transmission probability is needed. Information transmission probability can be described as the successful transmission of all the information from TN to RN.

The information collected by the TN is encoded to spikes and sent to the RN. The probability of successful information transmission is analogous to the successful transmission of all spike bits in the message. Hence, successful information transmission probability is

$$R_t = (1 - P_E)^n \quad (10.24)$$

where n is the number of spike bits contained in the message. The probability in (10.24) is used in the analytical model of MAMNET.

10.5 Mathematical Model of Mobile Ad Hoc Molecular Nanonetwork

In this section, a closed-form expression for average message delivery delay is derived for MAMNET. Furthermore, average throughput and incurred traffic rate expressions are obtained. The propagation of a single message is modeled using the principles of epidemic disease spreading modeled by a store-carry-forward scheme.

When a nanomachine having an information encounters with another nanomachine that does not have a copy of the information, it forwards the information to this nanomachine. Such a transmission strategy is clearly similar to the Epidemic routing schemes [122] in partially-connected mobile ad hoc networks where each mobile node with the message copies the message to its neighbors that do not have the message. Epidemic routing strategies can provide the minimum message delay, if each node is assumed to have infinite buffer size. However, this strategy is clearly resource-inefficient since it causes unnecessary diffusion of the message even after the message is delivered to destination node. In the existing literature, there are many works that tackle with reducing number of transmissions while minimizing the message delay [170] in order to provide resource-efficient message delivery in partially-connected mobile ad hoc networks. In MAMNET, the delivery of the message is also prone to a similar resource inefficiency that must be addressed by optimizing number of neuro-spike transmissions in order to provide minimum message delay. However, this is considered as a future work of this chapter.

In mobile ad hoc networks with intermittent connectivity, many routing protocols inspired by epidemic disease spreading have been previously developed [171], [172], [173]. Similar to these works, inspiring by epidemic disease spreading nanomachines can be in three different states, i.e., *infected*, *susceptible* and *recovered*. The message which should be transmitted to an infostation, is analogous to the agent of a disease. An infected nanomachine is the one that has a copy of the message. A nanomachine is said to be suspicious when it does not have a copy of the message, but could potentially acquire a copy of the message from the infected ones. A nanomachine is recovered after it has offloaded the message to the infostation.

Our model is developed upon the Markov model devised for the basic epidemic disease spreading [174]. In Fig. 10.8, S , I and R are the susceptible, infected and recovered states, respectively. $S(t)$, $I(t)$ and $R(t)$ are the numbers of nanomachines in the susceptible, infected

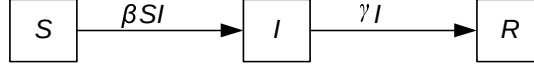


Figure 10.8: S-I-R Model. Markov chain model of an epidemic disease spreading.

and recovered states, respectively. The model is based on three assumptions, which can be adopted for our model as follows.

- i. An infected nanomachine makes contacts, that are ended with the successful transmission of message, with $\beta(N - 1)$ others per unit time where N represents the total number of nanomachines in MAMNET. In MAMNET, β represents the rate of information spreading among nanomachines. Information spreading involves collision, adhesion and molecular neuro-spike transmission phases. Therefore, the collision rate of nanomachines (R_c), the adhesion probability of collided nanomachines (R_a) and the information transmission probability (R_t) are the factors constituting the information spreading capability, hence, $\beta = R_c \times R_a \times R_t$. Since $S/(N - 1)$ proportion of these contacts are with the suspicious ones, the number of infections per unit time can be derived as $(\beta(N - 1))(S/(N - 1))I = \beta SI$.
- ii. γ represents the rate at which infected nanomachines become non-infected (recovered) and therefore, it contains infostation-nanomachine collision, adhesion and transmission rates. Therefore, γ can be given as $\gamma = R_{ic} \times R_a \times R_t$. If there are more than one infostation, the parameter γ should be multiplied by the number of infostations. This is because, the delivery of message to any infostation means recovery.
- iii. The total number of individuals in the system is constant.

The message delivery delay T_d is defined as the time elapsed from an information message is first generated by a nanomachine to the time when this message is first offloaded to an infostation. The initial conditions for the system can be defined as follows. At time $t = 0$, only one nanomachine is infected, i.e., $I(0) = 1$, $S(0) = N - 1$. Until offloading, the entire number of nanomachines N will be contained in either I state or S state, i.e., $S + I = N$, $R(t) = 0$ for $t < T_d$. At the time of offloading, $R(T_d) = 1$.

In order to find message delay distribution, it should be required that the transient solution of

the Markov Chain given in Fig. 10.8 and modeled as

$$\frac{dI}{dt} = \beta S I = \beta N I - \beta I^2 \quad (10.25)$$

For $I(0) = 1$, the solution of (10.25) is given in [171] as

$$I(t) = \frac{N}{1 + (N - 1)e^{-\beta N t}} \quad (10.26)$$

Using (10.26), the cumulative distribution function (CDF) of the message delay can be found. The CDF describes the probability that the message is delivered to an infostation by time t and denoted by $F(t) = \Pr(T_d < t)$. Here, the argument about the derivation of the CDF, in [171], is adopted. In [171], nodes are assumed to transmit information successfully as long as they are in the transmission range of each other. However, in MAMNET, delivery of the information message is also affected by adhesion and information transmission rates. These rates, specific to MAMNET, distinguish our model from the existing work devised for traditional mobile ad hoc networks. This difference, is captured in the definition of β and γ . The differential equation for $F(t)$ can be written as [171]

$$\frac{dF}{dt} = \lim_{\epsilon \rightarrow 0} \frac{F(t + \epsilon) - F(t)}{\epsilon} \quad (10.27)$$

where ϵ is an arbitrary positive small number close to 0. Note that, $F(t) = 1 - \Pr(T_d > t)$ and $F(t + \epsilon) = 1 - \Pr(T_d > t + \epsilon)$. In order to find $F(t)$ using (10.27), the probability $\Pr(T \geq t + \epsilon)$ is needed. Assuming $\Pr(\text{event in } [0, t))$ is independent of $\Pr(\text{event in } [t, t + \epsilon))$,

$$\begin{aligned} & \Pr(T_d > t + \epsilon \mid T_d > t) \\ &= 1 - \Pr(t < T_d < t + \epsilon \mid T_d > t) \\ &= 1 - \epsilon \gamma I(t) \end{aligned} \quad (10.28)$$

From (10.28), $\Pr(T_d > t + \epsilon)$ can be calculated as

$$\begin{aligned} \Pr(T_d > t + \epsilon) &= \Pr(T_d > t) \Pr(T_d > t + \epsilon \mid T_d > t) \\ &= \Pr(T_d > t) (1 - \epsilon \gamma I(t)) \end{aligned} \quad (10.29)$$

This can be used to derive the differential equation for $F(t)$, i.e.,

$$\begin{aligned} \frac{dF}{dt} &= \lim_{\epsilon \rightarrow 0} \frac{[1 - \Pr(T_d > t + \epsilon)] - [1 - \Pr(T_d > t)]}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{-1}{\epsilon} \Pr(T_d > t) (1 - \epsilon \gamma I(t) - 1) \\ &= \gamma I(t) \Pr(T_d > t) \\ &= \gamma \frac{N}{1 + e^{-\beta N t} (N - 1)} [1 - F(t)] \end{aligned} \quad (10.30)$$

For the solution of (10.30), an initial condition is required. Note that, $F(0)$ represents the probability that the nanomachine which generates the message is adherent with infostation such that it can directly transmit the message to the infostation. Therefore, $F(0)$ can be given by

$$F(0) = (4\pi(r_i^3 - r^3)) / 3V \quad (10.31)$$

Using $F(0)$, $F(t)$ can be derived as

$$F(t) = 1 - K \left(\frac{N-1}{N-1 + e^{\beta N t}} \right)^{\frac{\gamma}{\beta}} \quad (10.32)$$

where $K = \left[\frac{N-1}{N} \right]^{\frac{\gamma}{\beta}} [1 - F(0)]$. Hence, the probability that a message is delivered to an infostation for a given time can be obtained. Conversely, the average time needed to deliver the information to an infostation can also be calculated. Average delivery delay can be found using CDF of message delivery delay, i.e.,

$$E[T_d] = \int_0^{\infty} 1 - F(t) dt = \int_0^{\infty} K \left(\frac{N-1}{N-1 + e^{\beta N t}} \right)^{\frac{\gamma}{\beta}} dt \quad (10.33)$$

Assuming $\gamma/\beta \in N^+$, $E[T_d]$ is analytically expressed as

$$E[T_d] = \begin{cases} \frac{F(0) \ln N}{\gamma(N-1)} & , \gamma/\beta = 1 \\ \frac{F(0)(1-N)}{\beta N^2} \left[\sum_{k=2}^{\gamma/\beta} \frac{\beta}{\beta + \gamma - k\beta} \left(\frac{N}{N-1} \right)^k - \left(\frac{N}{N-1} \right)^{1+\gamma/\beta} \log N \right] & , \gamma/\beta \geq 2 \end{cases} \quad (10.34)$$

Note that, the analytical result in (10.34) are only valid for $\gamma/\beta \in N^+$. On the other hand, by numerically evaluating the integral in (10.33), $E[T_d]$ can be calculated for any γ and β . The numerical results for average delivery delay are presented in Section 10.6.

Here, we define two new metrics specific to MAMNET to evaluate the performance. Average throughput of MAMNET can be defined as the average number of bits transmitted per second to an infostation assuming that there is a single message propagating in the network. Using the average delivery delay, average throughput of MAMNET can also be calculated. Considering from an infostation point of view, a message encoded by n spike bits is transmitted from a nanomachine source to an infostation in an average time of $E[T_d]$. Hence, the average throughput of MAMNET can be described as

$$T_{avg} = n/E[T_d] \quad (10.35)$$

Table 10.1: Simulation Parameters

<i>Radius of a nanomachine (r)</i>	7.5 (μm)
<i>Radius of an infostation (r_i)</i>	15 (μm)
<i>Average relative speed of nanomachines ($E[v_{nn}]$)</i>	1 ($\mu\text{m}/\text{sec}$)
<i>Average relative speed of nanomachines and infostation ($E[v_{ni}]$)</i>	1 ($\mu\text{m}/\text{sec}$)
<i>Volume (V)</i>	1000000 (μm^3)
<i>Area of contact (A_c)</i>	3 (μm^2)
<i>Surface density of receptors (m_r)</i>	100 ($1/\mu\text{m}^2$)
<i>Surface density of ligands (m_l)</i>	75 ($1/\mu\text{m}^2$)
<i>Forward binding rate (k_f^0)</i>	1.32×10^{-7} ($\mu\text{m}^2/\text{sec}$)
<i>Reverse binding rate (k_r^0)</i>	0.36 ($1/\text{sec}$)
<i>Contact time τ_c</i>	1 (sec)
<i>Minimum number of required bonds for proper adhesion (c)</i>	40
<i>Prior probability of sending spike bit 1 (p_1)</i>	0.5
<i>Threshold for decision of spike bit (θ)</i>	1.1
<i>Signal to noise ratio on RN (SNR)</i>	10
<i>Length of message (n)</i>	8 (bits)
<i>Number of nanomachines (N)</i>	20

On the other hand, the incurred traffic rate in the network is defined as the average number of bits transmitted in the whole system per second assuming that there is a single message propagating in the network. In the whole system, until the message arrives into an infostation, an average number of $I(E[T_d])$ nanomachines get infected, i.e., a message containing n spike bits is transmitted between nanomachines $(I(E[T_d]) - 1)$ times, which further means $(I(E[T_d]) - 1) \times n$ bits are transmitted. Furthermore, for n spike bits transmitted to infostation, the incurred traffic rate (T_{itr}) can be written as

$$T_{itr} = I(E[T_d])n/E[T_d] \quad (10.36)$$

Consequently, all aspects of MAMNET are covered by this model. Hence, the expressions derived for $E[T_d]$, T_{avg} and T_{itr} can be used to investigate the feasibility and evaluate performance of MAMNET. Using the model derived above, next, the performance of MAMNET is evaluated.

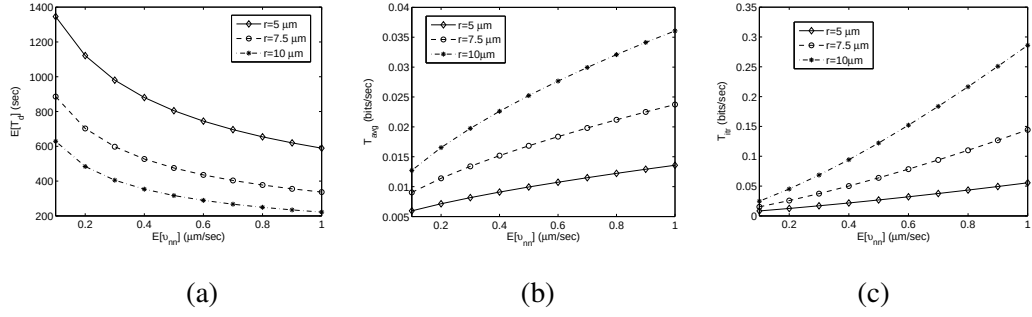


Figure 10.9: (a) Average message delay (b) Average throughput (c) Incurred traffic rate with varying average relative speed of nanomachines for different values of nanomachine radius.

10.6 MAMNET Performance Analysis

In this section, we evaluate the performance of MAMNET by using average message delay, average throughput and incurred traffic rate expressions derived in previous section. We use MATLAB to obtain analytical results. The aim is to investigate delay and throughput performance and gain insight on the feasibility of MAMNET. The parameters used in the analysis are given in Table 1.

10.6.1 Speed and Size of Nanomachines

We first observe the effect of relative speed and size of nanomachines on average delay, average throughput and incurred traffic rate. Nanomachine size is on the order of typical mammalian cells, i.e. having a diameter of 5-10 μm . Since infostations have more complex roles than nanomachines, they are assumed to have larger dimensions than nanomachines, e.g., around 10-20 μm .

In Fig. 10.9(a), Fig. 10.9(b) and Fig. 10.9(c), average message delay, average throughput and incurred traffic rate are shown for different r with varying $E[v_{nm}]$, respectively. For $r = 7.5 \mu m$, average message delivery delay ranges from a few hundreds to a thousand seconds. The volume containing nanomachines is 565 times larger than a nanomachine volume. Furthermore, the average relative speed of nanomachines is very low. Hence, the collision rate in MAMNET is low, e.g., with the values specified in Table 10.1, on average, only 14 collisions occur between nanomachines in 1000 seconds. Therefore, it is reasonable to end up with message delays on the order of a few thousand seconds. The incurred traffic rate follows

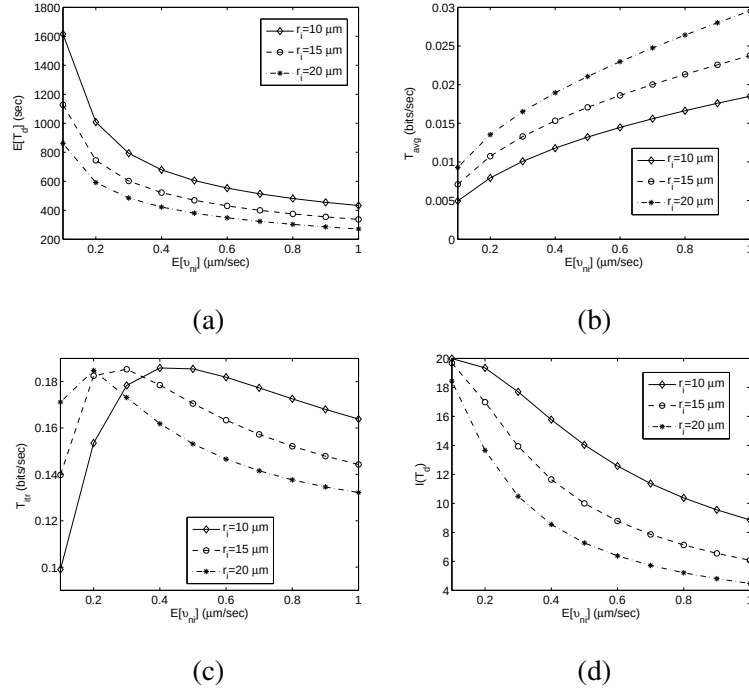


Figure 10.10: For varying average relative speed of nanomachine and infostation for different values of infostation radius, the plot of (a) Average message delay. (b) Average throughput. (c) Incurred traffic rate. (d) Average number of infected nanomachines at the instant of message delivery.

a similar argument. Considering the effect of low collision rate and the information message is encoded by 8 spike bits, it is also reasonable to encounter with a incurred traffic rate on the order of few spike bits per 10 seconds.

In Fig. 10.9(a), average message delivery delay decreases with increasing relative speed of nanomachines. On the other hand, in Fig. 10.9(b) and Fig. 10.9(c), average throughput and incurred traffic rate increase with $E[v_{nm}]$. This is because, by increasing the relative speed of nanomachines the collision rate is amplified and this causes more nanomachines to get infected. Therefore, the number of information transfers in MAMNET increases, which yields a higher throughput. Moreover, the message is delivered to an infostation more quickly.

Fig. 10.9(a), (b) and (c) also show the effect of r on $E[T_d]$, T_{avg} and T_{itr} , respectively. The increase in r also causes collision rate to increase and this yields a higher throughput and lower message delivery delay. Actually, r is analogous to the transmission range of mobile nodes in traditional mobile ad hoc networks. The larger the transmission range, the higher the probability of meeting of nodes. However, the transmission range is limited in traditional

mobile ad hoc networks because of limited battery lifetime of network nodes. Similarly, it seems reasonable to make larger nanomachines to reduce the message delay while providing resource-efficiency. However, one has to consider that these nanomachines can be used in applications where the size of the nanomachines must be extremely small to function in an extremely small volume.

10.6.2 Speed and Size of Infostations

$E[T_d]$, T_{avg} and T_{itr} are shown for different infostation radius values with varying relative speed in Fig. 10.10(a), Fig. 10.10(b) and Fig. 10.10(c), respectively. Increasing r_i and $E[v_{ni}]$ improves the collision rate R_{ic} of nanomachines and infostation. Therefore, increasing interactions between nanomachines and infostation yields a better delay performance and higher average throughput.

However, the incurred traffic rate performance does not have such a direct relation. This is because, incurred traffic rate is defined as the average number of spike bits transmitted in the system until the offloading of a message. For $E[v_{ni}] = 0.1$, the collision rate of nanomachines and infostations are extremely low, which creates a high message delay. Until the delivery of the message, almost all of the nanomachines get infected. Doubling $E[v_{ni}]$ causes the message delivery delay to reduce significantly. However, the number of infected nanomachines does not change at the same pace as seen in Fig. 10.10(d). Thus, the throughput is significantly enhanced because of the significant decrease in average message delivery delay. Further increasing of $E[v_{ni}]$ causes message delivery delay to decrease, and also the number of infected nanomachines decreases as well. Therefore, the incurred traffic rate performance decreases. This creates a critical point in T_{itr} performance as in Fig. 10.10(c). While $E[v_{ni}]$ increases, T_{itr} can be improved. However, T_{itr} starts to decrease after a specific $E[v_{ni}]$.

10.6.3 Contact Time

In Fig. 10.11(a), average message delay and incurred traffic rate are shown with varying contact time. A long contact time yields a high adhesion probability and more of the collisions result with successful adhesion. Therefore, $E[T_d]$ decreases and T_{itr} increases with increasing τ_c . In Fig. 10.11(a), after a certain value of τ_c , the performance of MAMNET does not

change. The reason can be observed more clearly in Fig. 10.11(b). The contact time of nanonodes merely affects the adhesion probability. Actually, at $\tau_c = 1.1$ second, the adhesion probability becomes 1, and the adhesion rate does not change. For $\tau_c = 0.75$ second, adhesion probability (R_a) is 0.3, i.e., approximately one third of the collisions resulted in successful adhesion. On the other hand, for $\tau_c = 1.1$, $R_a = 1$, i.e., all the collisions resulted in successful adhesion. Thus, $E[T_d]$ and T_{itr} become constant after $\tau_c = 1.1$ second. Consequently, $\tau_c = 1.1$ can be also considered as a critical point at which average message delay performance no longer changes.

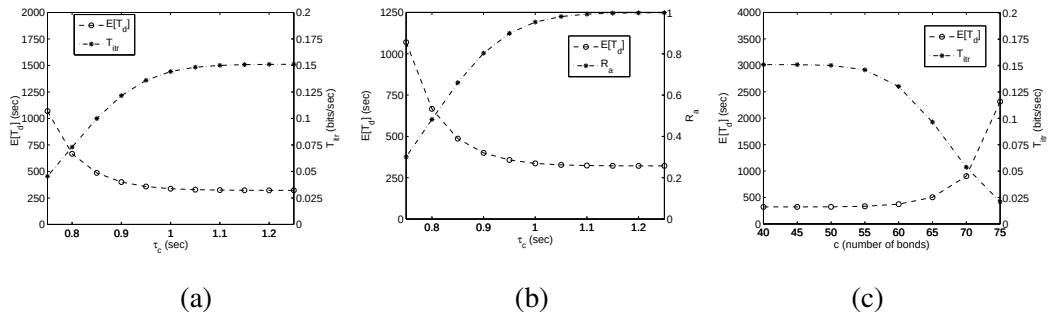


Figure 10.11: (a) Average message delay and incurred traffic rate with varying contact time. (b) Adhesion probability and average message delay with varying contact time. (c) Average message delay and incurred traffic rate with varying number of bonds.

10.6.4 Number of Bonds Required for Adhesion

Another critical parameter defined in Section 10.3 was the number of bonds required for successful adhesion. The effect of number of bonds required for successful adhesion on MAMNET performance is analyzed in Fig. 10.11(c). With the increase in the number of bonds required for successful adhesion, the adhesion of nanomachines becomes more difficult. Therefore, the adhesion probability decreases, which implies a reduction in the interaction of nanomachines. Hence, average message delivery delay increases and incurred traffic rate decreases with increasing number of bonds required for successful adhesion.

10.6.5 Neuro-Spike Communication Parameters

In Fig. 10.12(a), average message delivery delay is shown with varying θ for different SNR . The effect of SNR on delay performance is obvious. With decreasing SNR , the transmission

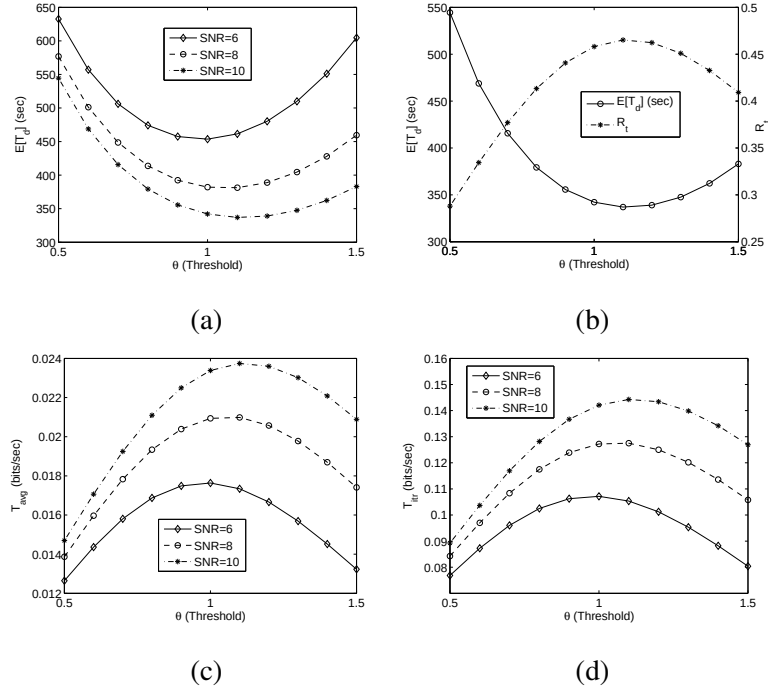


Figure 10.12: (a) Average message delay (b) Average message delay and transmission probability (c) Average throughput (d) Incurred traffic rate with varying threshold value for different values of SNR.

probability of information decreases and this yields higher delays. The effect of the threshold used for spike detection is more complicated. In Section 10.4, it is pointed out that the choice of the threshold for detection of spikes is very important for molecular neuro-spike communication. There exists a critical value for θ to achieve a high channel performance. In Fig. 10.12(a), the average message delivery delay decreases until a certain value of θ . After that point, $E[T_d]$ starts to increase. This is because of the effect of θ on transmission probability. In Fig. 10.12(b), the effects of the threshold for spike detection on transmission probability and average message delivery delay are shown. It can be seen that, there exists a critical value of θ which makes the transmission probability maximum.

Similar arguments follow for the average throughput and incurred traffic rate. The threshold value that makes the transmission probability maximum also maximizes the throughput. In Fig. 10.12(c) and (d), a reverse hook shape exists for the average throughput and incurred traffic rate with varying threshold. From Fig. 10.12(a), Fig. 10.12(c) and Fig. 10.12(d), we deduce that the critical value of threshold slightly changes with different SNR values. Therefore, a threshold that goes well with all SNR values should be chosen to achieve low message delay and high incurred traffic rate.

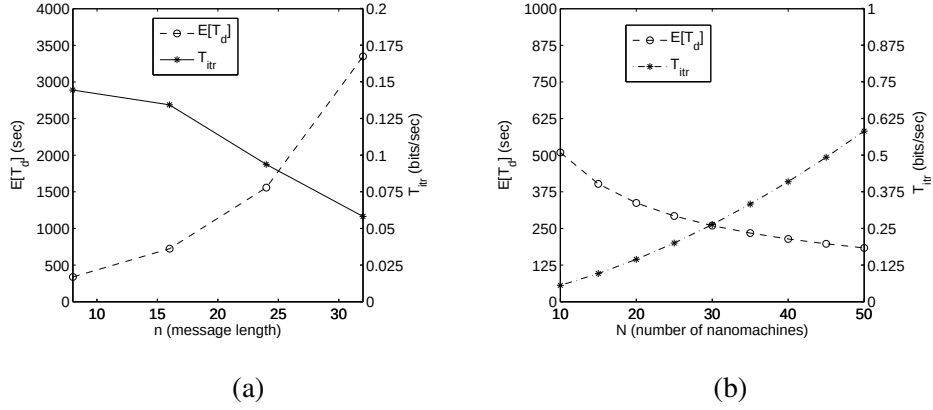


Figure 10.13: Average message delay and incurred traffic rate with varying (a) Number of spike bits contained in the message. (b) Number of nanomachines.

10.6.6 Number of Spike Bits in a Message

The number of spike bits, n , used to encode a message is a critical parameter. Because of the limited capabilities of nanomachines, we do not expect them to generate a large amount of information. Nanomachines sensing the environment, generates information about the observed phenomena. The information generated by nanomachines are on the order of a few dozen of spike bits. The predefined message strings, and their length are determined before the deployment, and hence, nanomachines are accordingly programmed. The length of the messages depends on the diversity of the phenomena that is desired to be observed in the deployment environment, e.g., with 8 bits 256 different phenomena can be represented.

In Fig. 10.13(a), MAMNET performance is observed to decrease with increasing n . This is due to the decreased transmission probability of a message. Note that, the successful transmission probability of a message given in (10.24) exponentially decreases with increasing message length. This causes fewer nanomachines to get infected, and yields higher message delivery delay. On the other hand, having more number of spike bits transmitted in a single message, one can expect a higher incurred traffic rate. However, decrease in successful message transmission probability overcomes the increase in the number of spike bits contained in a message. Therefore, incurred traffic rate decreases with increasing number of spike bits contained in a message.

10.6.7 Number of Nanomachines

The number of nanomachines in MAMNET is also a critical parameter for the performance MAMNET. In Fig. 10.13(b), average message delay and incurred traffic rate are shown for varying number of nanomachines. The increasing number of nanomachines increases the connectivity in MAMNET. Therefore, the message can be delivered to an infostation in a lower time. Furthermore, the incurred traffic rate increases with increasing number of nanomachines.

10.7 Conclusion

In this chapter, using the stochastic models for collision and adhesion of nanomachines we first derive the rates for the interaction of nanomachines. Then, using the principles of neural communication we provide a new communication paradigm for the realization of mobile ad hoc molecular nanonetworks. We also investigate the performance of molecular neuro-spike channel model by examining single spike bit error probability and mutual information. Then, we model the flow of a single message in MAMNET by using the principles of infectious disease spreading. We evaluate the performance of MAMNET based on average message delivery delay and incurred traffic rate expressions. Our models and numerical analysis clearly show that a mobile ad hoc nanonetwork can be realized with sufficiently low message delivery delay and high incurred traffic rate.

CHAPTER 11

Nanoscale Communication with Molecular Arrays in Nanonetworks

In this chapter, we introduce the Molecular ARray-based COmmunication (MARCO) scheme in which the transmission order of different molecules is used to convey molecular information without any need for time synchronization between transmitter nanomachine (TN) and receiver nanomachine (RN). Unlike the previous channel approaches, MARCO also follows an efficient nature-inspired approach in which information is carried by the arrays of different molecules as in the natural genetic information carrying mechanism, i.e., DNA arrays consisting of different base pairs. We first theoretically model MARCO channel behavior. Then, we comprehensively derive channel characteristics such as inter-symbol interference and error probabilities based on the principles of Brownian motion. Using these probabilities, we finally obtain a communication rate expression for the MARCO channel. Numerical results and performance comparisons reveal that MARCO provides significantly higher communication rate, i.e., on the scale of 100 *Kbps*, than the previously proposed molecular communication schemes without any need for time synchronization. MARCO was also introduced in [175].

In Section 11.1, the motivation and the related works are discussed. In Section 11.2, we present the MARCO scheme and derive the inter-symbol interference and error probabilities based on which a rate expression for the MARCO channel is given. In Section 11.3, we give the numerical results on the MARCO performance. We also present the performance comparisons of MARCO with the previously proposed molecular communication schemes in Section 11.4. Finally, we discuss concluding remarks in Section 11.5.

11.1 Motivation and Related Works

Nanotechnology envisages the practical realization of very low-end nanomachines that have tiny components to accomplish a simple specific task such as communication, computation, and sensing in a scale ranging from 1 to 100 nanometers. For example, molecular machines, a particular type of nanomachines, consist of organized molecular components changing their relative positions to achieve a specific task as a response to an external stimulus. Nanomachines are considered as a part of the potential solution approaches to many crucial problems such as pollution, scarce food resources, and cellular repair [6]. However, the design and construction of nanomachines is one of the most attractive challenges in nanotechnology [7].

The construction of nanomachines is currently pursued by the top-down approach that progressively manipulates more and more tiny nanomachine components by photolithography and related techniques. However, as Moore's law races towards 50 *nm* feature sizes on chips, it is not yet clear whether smaller nanomachines can be manufactured by using a top-down approach. Moreover, the top-down approach is subject to a drastic cost limitation as the size of the nanomachine components comes closer to nanometer scale [7], [8].

On the other hand, the bottom-up approach follows a more attractive strategy in the construction of frontier nanomachines. Rather than combining individual nanomachine components as in the top-down approach, the bottom-up approach entails the self-assembly of nanoscale building blocks with controlled size, shape, surface structure, and chemical functionality. In fact, by considering nature-made nanomachines constructed by an incredible self-assembly process, it can be easily inferred that the bottom-up approach has many promising advantages over the top-down approach [8]. For example, a chloroplast in a plant cell stands as a nanomachine including arrays of molecules that act as tuned optical antennas for absorption and transformation of solar energy. Mitochondrion can be envisioned as a nanomachine used for controlled combustion of organic molecules to generate adenosine triphosphate (ATP) for fulfilling energy needs of cellular activities. A flagellar motor attached on the membrane of many bacterial cells is a highly structured combination of proteins for providing cellular movement [9].

Biologically-inspired design or genetic adaptation from Nature is currently adopted to construct new nanomachines and nanomaterials through a large set of methods used in biochem-

istry and supramolecular chemistry [7], [10]. For example, biochemists already tackle with the production of functional cell components such as ribosome that can be used for the synthesis of beneficial proteins. Inspired by the flagella rotary motor in bacteria, a hybrid bio-nanochemomechanical rotary device has been assembled from ATP synthase and a nickel propeller [11]. Moreover, bacteria are also genetically programmed to generate principle vital components ranging from human growth hormone to beneficial enzymes for chess production [6].

While a single nanomachine can perform a very simple specific task, a network of communicating nanomachines can also be engineered to share nanoscale information over a nanonetwork so as to fulfill more complex tasks such as collaborative drug delivery, health monitoring, and biological or chemical attack detection [12], [13]. In fact, such nature-made nanonetworks already exist and are indispensable for nature to cooperatively share nanoscale information in order to accomplish a specific task. For example, the quorum sensing mechanism allows bacteria to communicate with each other by producing, emitting, and receiving hormone-like messenger molecules called autoinducer. This natural process permits bacteria to synchronize all colony activities and to change the colony state as a response to an external stimulus [14]. Similar to bacteria, in natural immune system, the white blood cells called B-cells and T-cells communicate with each other using molecules to cooperatively sense and eliminate the hazardous pathogen [15]. The molecular communication among white blood cells forms the biological immune network that is an excellent defense mechanism of organism.

As in nature, nanomachines can also be artificially interconnected via molecular communication to form a nanonetwork for a specific task [13]. For example, two artificial cell-cell communication systems for yeast are developed in [16]. In the first system, the engineered yeasts, considered as sender cells, synthesize and emit the plant hormone cytokinin and the emitted cytokinin molecules diffuse into the environment and activate a target signaling pathway in nearby engineered yeasts, considered as receiver cells. In the second system, like quorum sensing in bacteria, the molecular communication between sender and receiver cells provides population density-dependent gene expression process for the synthesis of a functional biochemical material. These artificial cell-cell communication systems are expected to provide new biotechnology applications such as biomaterial fabrication and tissue engineering. Moreover, in [17], using nitric oxide signaling elements, an artificial cell-cell com-

munication system in mammalian cells is introduced in order to govern a population of cells with the synthetic genetic circuits. This system may also serve as a functional tool for many applications such as gene therapy and artificial gene regulatory networks.

On the other hand, there also exist several research efforts toward addressing the unique challenges of molecular communications by the viewpoint of communication theory. In [39], the concept of molecular communication is introduced and the first attempt for design of molecular communication system is performed. An extensive survey on nanonetworking with molecular communication is presented in [12]. A physical channel model for molecular communication is introduced by incorporating three different phases of molecular communication, i.e., molecular emission, diffusion, and reception, into a single channel model in [40]. Achievable information rate is investigated for the molecular communication channel that is modeled as a timing channel [41]. Transmission time of the emitted molecules are considered as the input to the timing channel and time delay experienced in delivery of molecules is also considered as a random channel noise to communicate molecular information.

The molecular noise in diffusion-based molecular communication systems is investigated and modeled in [176]. The effect of noise on the molecular communication performance is investigated in [177] by considering various molecular transmission, propagation, encoding, and decoding mechanisms. In [178], an information theoretical analysis for a molecular relay channel is proposed to reveal the exact dynamics of the relay channel and find out its communication capacity. In-sequence delivery of different molecules in a molecular communication channel is investigated in [179] and different transmission strategies are proposed to provide in-sequence molecule delivery. However, this work does not explore the capabilities of in-sequence molecular delivery, and does not provide a comprehensive analysis of the molecular sequence-based communication by considering the inter-symbol interference and error probability of the channel.

In [132], we introduced an information theoretical approach for molecular communication, derived a closed-form expression for the capacity of molecular communication channel and proposed an adaptive error compensation technique for molecular communication. We provided the deterministic capacity of information flow in a molecular nanonetwork [144], and explored the necessity of networking techniques for the realization of future nanonetworks. We also modeled the point-to-point and multiple-access molecular channels and discuss their

achievable molecular communication rates [146]. However, we notice that we have not considered molecular array-based communication in any of our previous work.

All of these existing and previously developed molecular channel models between a transmitter nanomachine (TN) and a receiver nanomachine (RN) have been modeled as either concentration channel or timing channel. However, these channel models require exact time synchronization between TN and RN, which is beyond the state-of-the-art capabilities of very low-end nanomachines. Furthermore, they provide a relatively low communication bandwidth, which may hamper some frontier applications such as molecular processor with molecular inter/intra-chip communication. Therefore, it is imperative for the realization of future nanonetworks to develop synchronization-free and practical molecular communication schemes providing relatively high communication bandwidth.

11.2 Molecular Array-Based Communication

Molecular communication enables nanomachines to exchange information by emitting molecules to their surrounding environment [13], [39]. Molecules emitted by a TN diffuse in the medium. Then, they are received by an RN via chemical or electrochemical reactions between the molecules and the receptors of RN. These reactions allow RN to receive the molecular information transmitted by TN. The information can be encoded by using the concentration levels of transmitted molecules if the molecular communication channel is considered as a concentration channel. On the other hand, the information can be also encoded using transmission time of the molecules if the channel is a timing channel.

In Molecular ARray-based COmmunication (MARCO), information is encoded *by the transmission order of the molecules* as shown in Fig. 11.1. TN selects a molecule at a time among n distinct types of molecules and emits to the medium. Inter-emitting time of the molecules is assumed to be set to t_e . Thus, if the first molecule is emitted at time t , the second is emitted at time $t + t_e$, and, so on. In the conventional transmission systems, whose alphabet includes two symbols, i.e., $\{0, 1\}$, if a codeword is a sequence of k symbols, there are at most 2^k different codewords. In general, if the alphabet contains l different symbols, the resulting number of possible codewords is l^k . As in conventional transmission channels, let us investigate the binary case, that is, $l = 2$ and there are only two different molecules, namely a and b to transmit

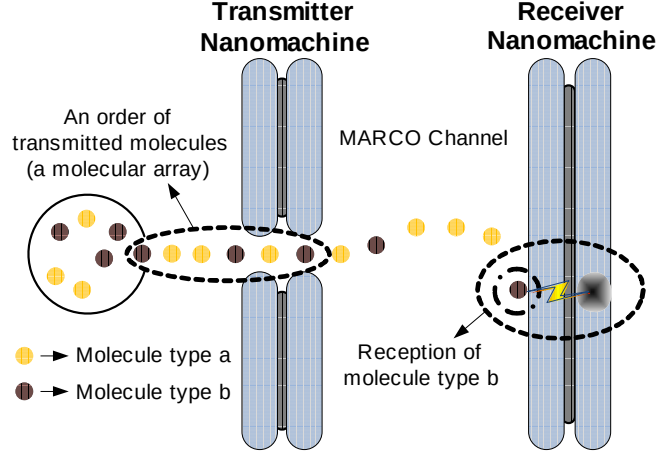


Figure 11.1: Illustration of molecular order-based communication. Here, two distinct molecules, i.e., molecule type a and b , are used to transmit information.

information. With no loss of generality, let us associate the transmission order $a - b$ with bit 0 and $b - a$ with bit 1. Hence, *the information is encoded with the order of the transmitted molecules*.

We assume that the movement of molecules is governed by a Brownian motion [180]. Then, the delay t experienced by any molecule to reach RN located at distance d obeys the following probability density function:

$$f(t) = \frac{d}{\sqrt{4\pi Dt^3}} e^{-\frac{d^2}{4Dt}}, \quad t > 0 \quad (11.1)$$

where D is the diffusion coefficient of the molecules. $f(t)$ only depends on the ratio $\frac{d}{\sqrt{D}}$. If $\frac{d}{\sqrt{D}}$ is set to $\lambda = \frac{d}{\sqrt{D}}$, (11.1) can be also rewritten as follows:

$$f(t) = \frac{\lambda}{\sqrt{4\pi t^3}} e^{-\frac{\lambda^2}{4t}}, \quad t > 0 \quad (11.2)$$

The cumulative distribution function associated with the density function $f(t)$, i.e., $F(t)$, can be given as

$$F(t) = 1 - \text{Erf}\left(\frac{d}{2\sqrt{Dt}}\right), \quad t > 0 \quad (11.3)$$

Let us now consider the scenario in which TN sends a symbol 0 using the molecular order $a - b$ to RN and RN correctly receives this symbol. The time diagram of this scenario is

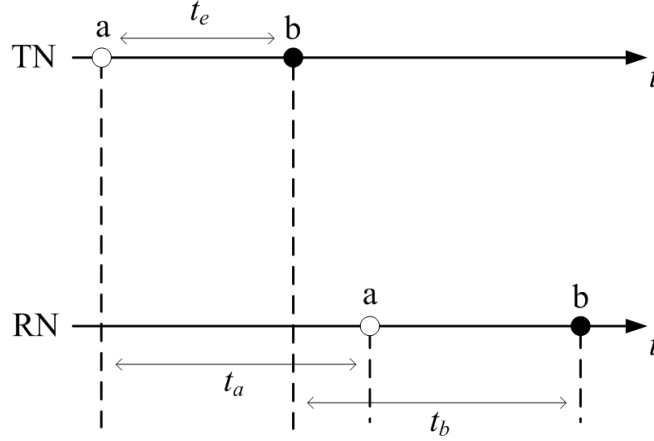


Figure 11.2: Transmission and correct reception of two molecules in order $a - b$ which corresponds to the symbol 0.

presented in Fig. 11.2. The probability of such event can be characterized as

$$Pr(0|0) = Pr(t_b > t_a - t_e) = \begin{cases} 1, & t_a < t_e \\ I(t_e, \lambda), & t_a > t_e \end{cases} \quad (11.4)$$

where $Pr(0|0)$ denotes the probability that symbol 0 is transmitted by TN and correctly received by RN. t_a and t_b are the random time delays experienced by molecule a and b , respectively. Note that since t_a and t_b have the same probability distribution given in (11.2), the transmission of symbol 0 with the order $a - b$ is statistically equal to the transmission of symbol 1 with the order $b - a$, that is, $Pr(0|0) = Pr(1|1)$. In (11.4), $I(t_e, \lambda)$ denotes the probability function that depends on the inter-emission time t_e and the ratio λ . In fact, (11.4) can be rewritten in a more compact form as

$$Pr(0|0) = 1 \times Pr(t_a < t_e) + I(t_e, \lambda) \times (1 - Pr(t_a < t_e)) \quad (11.5)$$

Using the probability distribution of t_a and t_b in (11.1) and (11.2), respectively, the function $I(t_e, \lambda)$ can be derived as

$$\begin{aligned} I(t_e, \lambda) &= \int_{t_e}^{\infty} \int_{t_a - t_e}^{\infty} f(t_a, t_b) dt_a dt_b \\ &= \int_{t_e}^{\infty} \int_{t_a - t_e}^{\infty} f(t_a) f(t_b) dt_a dt_b \\ &= \int_{t_e}^{\infty} f(t_a) \text{Erf}\left(\frac{\lambda}{2\sqrt{t_a - t_e}}\right) dt_a \end{aligned} \quad (11.6)$$

where $f(t_a, t_b)$ denotes the joint density function of the random delays t_a and t_b . Since these delays are assumed to be independent, $f(t_a, t_b) = f(t_a)f(t_b)$. Furthermore, $Pr(t_a < t_e)$ in

(11.5), i.e., the probability that t_a does not exceed the inter-emission time t_e , can be derived as

$$\begin{aligned}
 Pr(t_a < t_e) &= \int_0^{t_e} f(t_a) dt_a \\
 &= \int_0^{t_e} f(t) dt \\
 &= \text{Erfc}\left(\frac{\lambda}{2\sqrt{t_e}}\right)
 \end{aligned} \tag{11.7}$$

Consequently, substituting (11.6) and (11.7) into (11.5), $Pr(0|0)$ can be rewritten in a more explicit form, i.e.,

$$Pr(0|0) = \text{Erfc}\left(\frac{\lambda}{2\sqrt{t_e}}\right) + \left[1 - \text{Erfc}\left(\frac{\lambda}{2\sqrt{t_e}}\right)\right] \int_{t_e}^{\infty} f(t_a) \text{Erfc}\left(\frac{\lambda}{2\sqrt{t_a - t_e}}\right) dt_a \tag{11.8}$$

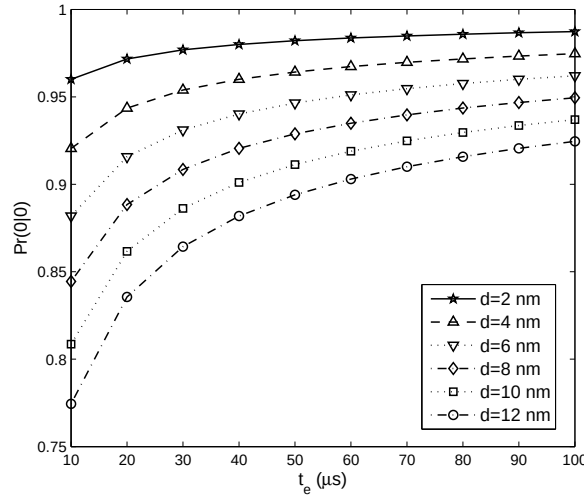


Figure 11.3: Probability of correct symbol reception, i.e., $Pr(0|0)$, with respect to the inter-emission time t_e for the different distances between TN and RN.

The successful transmission probability $Pr(0|0)$ in (11.8) is plotted in Fig. 11.3 as a function of the inter-emission time t_e for different values of the inter-node distance d between TN and RN. The results reveal an intuitively expected behavior such that, as d decreases and t_e increases, the probability of successful reception of a symbol increases. In other words, a larger d requires a larger t_e to yield a high probability of correct reception. At the same time, the numerical evaluation of $Pr(0|0)$ manifests that for sufficiently large d , e.g., in the order of ten times larger (like 120 nm), the probability of correct reception becomes almost

independent of the inter-emission time, and flattens around 0.5. This is not surprising at all, since when the distance is so large, the inter-emission time becomes irrelevant and any molecule may arrive first.

In addition to the case where a single symbol is out-of-order, two consecutive (binary) symbols may also mutually interfere with each other, which causes erroneous transmission. This phenomenon is called as inter-symbol interference (ISI). If we assume that a transmitted symbol merely interferes with either previously or next transmitted symbols, ISI may occur between the following symbol triplets:

- If 0 is currently transmitted, it may interfere with the previously or next transmitted symbols in the symbol triplets 000, 001, 101, and 100.
- If 1 is currently transmitted, it may interfere with the previously or next transmitted symbols in the symbol triplets 010, 011, 111, and 110.

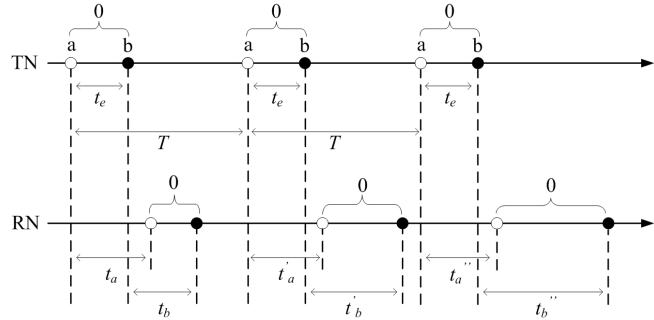


Figure 11.4: Time diagram for the transmission and correct reception of three consecutive symbols 000 using the molecular order $a - b, a - b, a - b$.

Now, considering the triplet 000, we derive the ISI probability, i.e., P_{ISI} , that the currently transmitted symbol 0 interferes with the previously or next transmitted symbol 0. The relevant delay variables needed for this derivation can be defined as follows.

11.2.1 The Relevant Delay Variables in MARCO Channel

In the transmission of symbol triplet 000, t_a and t_b are the random variables denoting the delay of molecules a and b in the first symbol 0, respectively, as shown in Fig. 11.4. The maximum of delay experienced by the molecules a and b can also be statistically characterized by the

random variable y , i.e., $y = \max(t_a, t_b + t_e)$. Note that since t_a and t_b are drawn according to the same density and cumulative distribution functions, i.e., $f(t)$ and $F(t)$ given in (11.2) and (11.3), respectively, t_a and t_b are identically distributed and equivalent to the delay of any other molecule of the same or different symbols. Therefore, the density and cumulative distribution functions of the random variable y can also be derived using $f(t)$ and $F(t)$. Due to the fact that t_a and $t_e + t_b$ are mutually independent, the cumulative distribution function of y , i.e., $F_y(t)$, can be expressed as

$$F_y(t) = F_{t_a}(t)F_{t_e+t_b}(t), \quad t \geq t_e \quad (11.9)$$

where $F_{t_e+t_b}(t)$ can be obtained by time-shifting $F_{t_b}(t)$ by an inter-emission time t_e , i.e., $F_{t_e+t_b}(t) = F_{t_b}(t - t_e)$. Because t_a and t_b are statistically equivalent, $F_{t_a}(t)$ and $F_{t_b}(t)$ are the cumulative distribution functions given in (11.3), that is,

$$F_{t_a}(t) = F_{t_b}(t) = F(t) = 1 - \text{Erf}\left(\frac{d}{2\sqrt{Dt}}\right) \quad (11.10)$$

Accordingly, (11.9) can be written as

$$\begin{aligned} F_y(t) &= F_{t_a}(t)F_{t_e+t_b}(t) \\ &= F_{t_a}(t)F_{t_b}(t - t_e) \\ &= \left[1 - \text{Erf}\left(\frac{d}{2\sqrt{Dt}}\right)\right] \left[1 - \text{Erf}\left(\frac{d}{2\sqrt{D(t - t_e)}}\right)\right] \end{aligned} \quad (11.11)$$

By taking the derivative of (11.9), the probability density function of y , i.e., $f_y(t)$, can be also given as

$$f_y(t) = f(t)F(t - t_e) + F(t)f(t - t_e), \quad t \geq t_e \quad (11.12)$$

$F_y(t)$ is plotted in Fig. 11.5 for $d = 4 \text{ nm}$ and for different values of the inter-emission time. Note that the inter-symbol time should correspond to a value of the argument t such that the distribution function is sufficiently large, for instance 0.9. In this case, the inter-symbol time is nothing else but the percentile-90 of the corresponding density function. For instance, for an inter-emission time of $20 \mu\text{s}$, the value of t for which the distribution function is around 0.9 is approximately $40 \mu\text{s}$. This would be the inter-symbol time or simply the slot duration. The corresponding transmission rate would be approximately equal to 25 Kbps .

In addition to the first symbol 0, the delay of molecules a and b in the second symbol 0 can be denoted as t'_a and t'_b as shown in Fig. 11.4. Moreover, the minimum and maximum of the delay experienced by the molecules a and b in the second symbol 0 can be also characterized

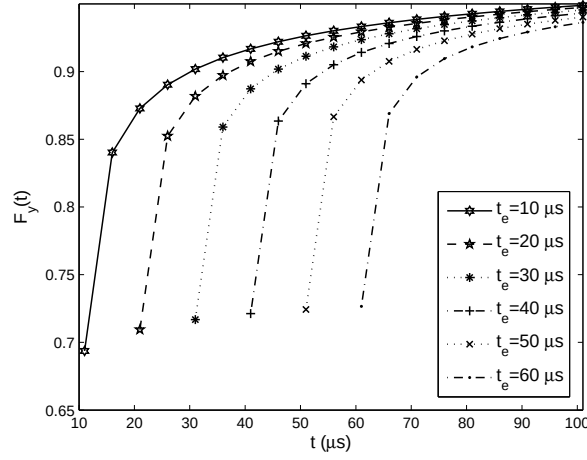


Figure 11.5: Distribution function for the evaluation of the inter-symbol time ($d = 4 \text{ nm}$).

by the random variable z' and y' , respectively, i.e., $z' = \min(t'_a + T, t'_b + t_e + T)$ and $y' = \max(t'_a + T, t'_b + t_e + T)$. Note that the second 0 is transmitted after an inter-symbol time T , which is clearly observed in Fig. 11.4. Here, the cumulative distribution function of $z' = \min(t'_a + T, t'_b + t_e + T)$, i.e., $F_{z'}(t)$, can be derived as

$$F_{z'}(t) = F(t - T) + F(t - t_e - T) - F(t - T)F(t - t_e - T), \quad t \geq T. \quad (11.13)$$

By taking the derivative of $F_{z'}(t)$ in (11.13), the density function of $z' = \min(t'_a + T, t'_b + t_e + T)$, i.e., $f_{z'}(t)$, can be also found as

$$f_{z'}(t) = f(t - T)[1 - F(t - t_e - T)] + f(t - t_e - T)[1 - F(t - T)], \quad t \geq T \quad (11.14)$$

Similarly, the time delays experienced by the molecules a and b in the third symbol 0 are denoted as t''_a and t''_b , respectively as seen in Fig. 11.4. z'' are also the random variable that characterizes the minimum of the delay experienced by the molecules a and b in the third symbol 0, i.e., $z'' = \min(t''_a + 2T, t''_b + t_e + 2T)$. Here, note also that the third 0 is transmitted after two of inter-symbol times ($2T$) as shown in Fig. 11.4. The density function of z'' , i.e., $f_{z''}(t)$, can be also obtained by replacing T in $f_{z'}(t)$ given in (11.14) with $2T$, i.e.,

$$f_{z''}(t) = f(t - 2T)[1 - F(t - t_e - 2T)] + f(t - t_e - 2T)[1 - F(t - 2T)], \quad t \geq 2T \quad (11.15)$$

Next, we derive the ISI probability in the MARCO channel using the delay variables and their related distribution functions derived above.

11.2.2 Probability of ISI in MARCO Channel

By using the delay variables defined and characterized above, the probability of ISI, i.e., P_{ISI} , can be formulated as

$$\begin{aligned} P_{ISI} &= 1 - P_{NoISI} = 1 - \Pr((z' > y) \cap (y' < z)) \\ &= 1 - \Pr(y' < z'' | z' > y) \Pr(z' > y) \end{aligned} \quad (11.16)$$

where P_{NoISI} denotes the probability that a transmitted symbol does not interfere with any other symbol. This expression highlights the fact that y' and z' are not independent random variables, because they are evaluated from the same pair (t'_a, t'_b) as introduced above and observed in Fig. 11.4. Then, the conditioned probability $\Pr(y' < z'' | z' > y)$ in (11.16) can be derived as

$$\begin{aligned} \Pr(y' < z'' | z' > y) &= \Pr(t'_a + T < z'' | t'_b + t_e + T > y) \Pr(t'_a + T > t'_b + t_e + T) + \\ &+ \Pr(t'_b + t_e + T < z'' | t'_a + T > y) \Pr(t'_b + t_e + T > t'_a + T) \end{aligned} \quad (11.17)$$

Here, the two possible cases are considered, i.e., either $y' = t'_a + T$ and $z' = t'_b + t_e + T$ or vice versa, that is, $y' = t'_b + t_e + T$ and $z' = t'_a + T$. Because of the independence of t'_a and t'_b , (11.17) can be rewritten as

$$\begin{aligned} \Pr(y' < z'' | z' > y) &= \Pr(t'_a < z'' - T) \Pr(t'_a > t'_b + t_e) + \\ &+ \Pr(t'_b < z'' - t_e - T) \Pr(t'_a < t'_b + t_e) \end{aligned} \quad (11.18)$$

However, we know that $\Pr(t'_a < t'_b + t_e) = \Pr(0|0)$ and obviously $\Pr(t'_a > t'_b + t_e) = 1 - \Pr(0|0)$, where $\Pr(0|0)$ is already known as given in (11.8). Accordingly, (11.18) can be rewritten as

$$\Pr(y' < z'' | z' > y) = \Pr(t'_a < z'' - T) [1 - \Pr(0|0)] + \Pr(t'_b < z'' - t_e - T) \Pr(0|0) \quad (11.19)$$

In order to derive $\Pr(t'_a < z'' - T)$ in (11.19), note that

$$\Pr(t'_a < z'' - T) = \Pr(t'_a - z'' < -T) \quad (11.20)$$

Let us now define $r = -z''$. Using $f_{z''}(t)$ derived in (11.15), the density function of the random variable r , i.e., $f_r(t)$, can be easily written as

$$f_r(t) = f(-t - 2T) [1 - F(-t - t_e - 2T)] + f(-t - t_e - 2T) [1 - F(-t - 2T)], \quad t \leq -2T \quad (11.21)$$

Since t_a and r are independent, the density function of $t_a + r$, i.e., $f_{t_a+r}(t)$, can be obtained by

$$f_{t_a+r}(t) = \int_{\bar{u}}^{\infty} f(u)f_r(t-u)du \quad (11.22)$$

where $\bar{u} = \max(0, t + 2T)$. Hence, using (11.22), $Pr(t'_a < z'' - T)$ can be computed as

$$\begin{aligned} Pr(t'_a < z'' - T) &= Pr(t'_a + r < -T) \\ F_{t_a+r}(-T) &= \int_{-\infty}^{-T} f_{t_a+r}(t)dt \end{aligned} \quad (11.23)$$

Similarly, $Pr(t'_a < z'' - t_e - T)$ in (11.19) can be also computed as

$$Pr(t'_a < z'' - t_e - T) = \int_{-\infty}^{-t_e-T} f_{t_a+r}(t)dt \quad (11.24)$$

Here, for the ease of illustration, let us introduce

$$q(\alpha) = \int_{-\infty}^{-\alpha} f_{t_a+r}(t)dt \quad (11.25)$$

Thus, using (11.25), $Pr(y' < z''|z' > y)$ can be written as

$$Pr(y' < z''|z' > y) = q(T)(1 - Pr(0|0)) + q(t_e + T)Pr(0|0) \quad (11.26)$$

Furthermore, substituting (11.26) into (11.16), the probability of inter-symbol interference, i.e., P_{ISI} , can be expressed as

$$P_{ISI} = 1 - \left[q(T)[1 - Pr(0|0)] + q(t_e + T)Pr(0|0) \right] Pr(z' > y) \quad (11.27)$$

In (11.27), $Pr(0|0)$, $q(T)$, and $q(t_e + T)$ are known as they are derived above. However, $Pr(z' > y)$ should be derived in order to obtain an exact expression for P_{ISI} . To this end, defining another random variable s such that $s = -y$, $Pr(z' < y)$ can be expressed as

$$Pr(z' < y) = Pr(z' + (-y) < 0) = Pr(z' + s < 0) \quad (11.28)$$

Using $f_y(t)$ in (11.12), the density function of s , i.e., $f_s(t)$, can be expressed as

$$f_s(t) = f_y(-t) = f(-t)F(-t - t_e) + F(-t)f(-t - t_e), \quad t \leq -t_e. \quad (11.29)$$

Since z' and s are two independent random variables, the density function $f_{z'+s}(t)$ can be obtained by

$$\begin{aligned} f_{z'+s}(t) &= \int_{-\infty}^{\infty} f_{z'}(u)f_s(t-u)du \\ &= \int_T^{\infty} f_{z'}(u)f_s(t-u)du \\ &= \int_0^{\infty} f_{z'}(u)f_s(t-u)du \end{aligned} \quad (11.30)$$

where $\bar{o} = \max(t + t_e, T)$ and $f_{z'}(t)$ is the density function of random variable $z' = \min(t'_a + T, t'_b + t_e + T)$ and given in (11.14). Hence, using $f_{z'+s}(t)$, $Pr(z' < y)$ can be evaluated as

$$Pr(z' < y) = Pr(z' + s < 0) = \int_{-\infty}^0 f_{z'+s}(t) dt \quad (11.31)$$

Again, for the ease of illustration, let us introduce another variable $\tilde{p}$ as

$$\tilde{p} = Pr(z' < y) = \int_{-\infty}^0 f_{z'+s}(t) dt \quad (11.32)$$

Due to $Pr(z' > y) = 1 - Pr(z' < y)$, by using $\tilde{p} = Pr(z' < y)$, $Pr(z' > y)$ can be expressed as

$$Pr(z' > y) = 1 - Pr(z' < y) = 1 - \tilde{p} \quad (11.33)$$

Finally, by substituting $Pr(z' > y) = 1 - \tilde{p}$ into (11.27), a definite expression for P_{ISI} can be given as

$$P_{ISI} = 1 - \left[q(T) [1 - Pr(0|0)] + q(t_e + T) Pr(0|0) \right] (1 - \tilde{p}) \quad (11.34)$$

Substituting $Pr(0|0)$, $q(t)$, $q(t_e + T)$, and $\tilde{p}$ into (11.34), P_{ISI} can be also rewritten as

$$\begin{aligned} P_{ISI} = & 1 - \left[\left(\int_{-\infty}^{-T} \int_{\bar{u}}^{\infty} f(u) f_r(t-u) du dt \right) \left[1 - \left(\text{Erfc}\left(\frac{\lambda}{2\sqrt{t_e}}\right) + \text{Erf}\left(\frac{\lambda}{2\sqrt{t_e}}\right) \int_{t_e}^{\infty} f(t_a) \text{Erf}\left(\frac{\lambda}{2\sqrt{t_a-t_e}}\right) \right) \right] + \right. \\ & + \left(\int_{-\infty}^{-(t_e+T)} \int_{\bar{u}}^{\infty} f(u) f_r(t-u) du dt \right) \left(\text{Erfc}\left(\frac{\lambda}{2\sqrt{t_e}}\right) + \text{Erf}\left(\frac{\lambda}{2\sqrt{t_e}}\right) \int_{t_e}^{\infty} f(t_a) \text{Erf}\left(\frac{\lambda}{2\sqrt{t_a-t_e}}\right) \right) \Big] \times \\ & \times \left(1 - \int_{-\infty}^0 \int_{\bar{o}}^{\infty} f_{z'}(u) f_s(t-u) du dt \right) \end{aligned} \quad (11.35)$$

where $f_{z'}(t)$, $f_r(t)$, and $f_s(t)$ are given in (11.14), (11.21), and (11.29), respectively. Furthermore, $\bar{u}$ and $\bar{o}$ are given as $\bar{u} = \max(0, t + 2T)$ and $\bar{o} = \max(t + t_e, T)$. Next, we also derive the error probability based on which a communication rate for the MARCO channel is developed.

11.2.3 Derivation of Error Probability in MARCO Channel

As observed in (11.34), P_{ISI} is statistically dependent on the successful transmission probability $Pr(0|0)$ ¹. Hence, due to this statistical dependence, it cannot be possible to obtain the error probability in the MARCO channel by simply factorizing P_{ISI} and $(1 - Pr(0|0))$. In fact, the probability of error, i.e., P_e , can be formulated as

$$P_e = 1 - P_{NoISI|(0|0)} Pr(0|0) \quad (11.36)$$

¹ However, this is not the case for very small t_e , and therefore, $q(T) \cong q(t_e + T)$ and $Pr(0|0)$ disappears from the P_{ISI} expression.

where $P_{NoISI|(0|0)}$ is the probability that a symbol does not interfere given the fact that it is correctly transmitted. Then, as in Fig. 11.4, $P_{NoISI|(0|0)}$ can be derived as

$$\begin{aligned} P_{NoISI|(0|0)} &= Pr((z' > y) \cap (y' < z''))|(0|0)) \\ &= Pr(y' < z''|z' > y, (0|0))Pr(z' > y|(0|0)) \end{aligned} \quad (11.37)$$

where $Pr(y' < z''|z' > y, (0|0))$ in (11.37) can be expressed as

$$\begin{aligned} Pr(y' < z''|z' > y, (0|0)) &= Pr(t'_b + t_e + T < z''|t'_a + T > y) \\ &= Pr(t'_b < z'' - t_e - T) \\ &= q(t_e + T) \end{aligned} \quad (11.38)$$

where we note that the result $Pr(t'_b < z'' - t_e - T) = q(t_e + T)$ has already been derived in (11.24)-(11.25). Moreover, in order to complete the derivation of $P_{NoISI|(0|0)}$, $Pr(z' > y|(0|0))$ in (11.37) can be also given as

$$Pr(z' > y|(0|0)) = Pr(t'_a + T > y) = Pr(t'_a - y > -T) \quad (11.39)$$

As in (11.28), let us again define the random variable $s = -y$ and rewrite (11.39) as

$$Pr(z' > y|(0|0)) = Pr(t'_a + (-y) > -T) = Pr(t'_a + s > -T) \quad (11.40)$$

Note that the density function of s , i.e., $f_s(t)$, is already derived in (11.29). Since t'_a and s are assumed to be statistically independent, the density function of $t'_a + s$, i.e., $f_{t'_a+s}(t)$, is found as

$$\begin{aligned} f_{t'_a+s}(t) &= \int_{-\infty}^{+\infty} f(u)f_s(t-u)du \\ &= \int_0^{+\infty} f(u)f_s(t-u)du \\ &= \int_{\bar{h}}^{+\infty} f(u)f_s(t-u)du \end{aligned} \quad (11.41)$$

where $\bar{h} = \max(0, t + t_e)$ and we notice that $f_{t_a+s}(t) = f_{t'_a+s}(t)$ since t_a and t'_a are the same delay variable in statistical sense. Using $f_{t_a+s}(t)$, $Pr(z' > y|(0|0))$ in (11.37) can be developed as follows

$$\begin{aligned} Pr(z' > y|(0|0)) &= Pr(t'_a + s > -T) \\ &= 1 - F_{t'_a+s}(-T) \\ &= 1 - \int_{-\infty}^{-T} f_{t'_a+s}(t)dt \end{aligned} \quad (11.42)$$

Here, let us introduce another variable, i.e., $\tilde{q}(\beta)$, for the ease of illustration such that

$$\tilde{q}(\beta) = \int_{-\infty}^{-\beta} f_{t_a+s}(t) dt \quad (11.43)$$

Hence, using $\tilde{q}(\beta)$, $Pr(z' > y|0|0)$ can be rewritten as

$$Pr(z' > y|0|0) = (1 - \tilde{q}(T)) \quad (11.44)$$

Consequently, substituting (11.38) and (11.44) into (11.37), $P_{NoIS\,I|0|0}$ can be written as

$$\begin{aligned} P_{NoIS\,I|0|0} &= Pr(y' < z''|z' > y, (0|0))Pr(z' > y|0|0) \\ &= q(t_e + T)(1 - \tilde{q}(T)) \end{aligned} \quad (11.45)$$

Finally, substituting $P_{NoIS\,I|0|0}$ given in (11.45) into (11.36), the error probability in the MARCO channel, i.e., P_e , can be expressed as

$$\begin{aligned} P_e &= 1 - P_{NoIS\,I|0|0}Pr(0|0) \\ &= 1 - q(t_e + T)(1 - \tilde{q}(T))Pr(0|0) \end{aligned} \quad (11.46)$$

where $q(\cdot)$, $\tilde{q}(\cdot)$, and $Pr(0|0)$ are expressed in (11.25), (11.43), and (11.8), respectively. Next, using the error probability P_e given in (11.46), we explore a communication rate for MARCO.

11.2.4 Rate of Molecular Array-Based Communication

For the MARCO channel, it is essential to find the communication rate to comprehensively evaluate the performance. Here, we derive a simple communication rate expression of MARCO for the case of two distinct molecules in the symbol set. In the channel, TN transmits either 0 or 1 and RN either correctly receives the information bit with probability $(1 - P_e)$ or incorrectly receives the information bit with the probability P_e . Therefore, a communication rate of the MARCO channel in bits per transmission, i.e., C , can be characterized as $C = (1 - P_e)$. Hence, by substituting the error probability P_e given in (11.46), the communication rate, i.e., $C = (1 - P_e)$, can be expressed as

$$C = q(t_e + T)(1 - \tilde{q}(T))Pr(0|0) \quad (11.47)$$

Substituting $q(t_e + T)$, $\tilde{q}(T)$, and $Pr(0|0)$ into (11.47), C can be also rewritten as

$$\begin{aligned} C &= \left[\int_{-\infty}^{-(t_e+T)} \int_{\bar{u}}^{\infty} f(u)f_r(t-u)dudt \right] \times \left[1 - \int_{-\infty}^{-T} \int_{\bar{h}}^{+\infty} f(u)f_s(t-u)dudt \right] \times \\ &\times \left[\text{Erfc}\left(\frac{\lambda}{2\sqrt{t_e}}\right) + \text{Erf}\left(\frac{\lambda}{2\sqrt{t_e}}\right) \int_{t_e}^{\infty} f(t_a)\text{Erf}\left(\frac{\lambda}{2\sqrt{t_a-t_e}}\right) \right] \end{aligned} \quad (11.48)$$

where $f_r(t)$ and $f_s(t)$ are given in (11.21) and (11.29), respectively. Furthermore, $\bar{u}$ and $\bar{h}$ are given as $\bar{u} = \max(0, t + 2T)$ and $\bar{h} = \max(0, t + t_e)$, respectively. The maximization of C provides the maximum number of symbols that can be successfully transmitted per inter-symbol time T . Thus, the rate of MARCO in bits per second (bps) can be given as $\frac{C}{T}(bps)$.

11.3 Numerical Results

In this section, we present the numerical analysis for the performance of MARCO by using the ISI probability P_{ISI} , error probability P_e , and the communication rate C given in (11.34), (11.46) and (11.48), respectively. The performance of MARCO is quantified by the numerical evaluation of these expressions. We first show the effect of inter-emission time and inter-symbol time on the MARCO performance. Then, we show how inter-emission time and inter-node distance between TN and RN affect the performance. Finally, we also present the numerical results to demonstrate the effect of diffusion coefficient on the MARCO performance.

The aim of this analysis is to numerically investigate the characteristics of the MARCO channel and gain some insight for the realization of future nanonetworks. We also aim at examining how these channel characteristics vary with the channel parameters such as inter-symbol time, inter-emission time, and internode distance. In the analysis, the inter-emission time and inter-symbol time are changed within the intervals $t_e = 2 - 8 \mu s$ and $T = 10 - 25 \mu s$, respectively. The inter-node distance is selected from the interval $d = 1 - 10 nm$ and the diffusion coefficient is changed within the interval $D = 25 - 100 \mu m^2/s$. The numerical computation of P_{ISI} , P_e , and C are performed using *Mathematica*.

11.3.1 Effects of Inter-Emission Time and Inter-Symbol Time

In Fig. 11.6, ISI probability, P_{ISI} , is shown with respect to the changing inter-emission time t_e and inter-symbol time T . Note that since the inter-emission time t_e must be less than the inter-symbol time T , t_e is increased at most up to the selected T values. As observed in Fig. 11.6, P_{ISI} can be reduced by increasing T , which can be easily verified by visualizing consecutively transmitted symbols. More specifically, as the inter-symbol time increases, it is apparent that the likelihood of inter-symbol interference decreases. Moreover, while t_e increases up to some

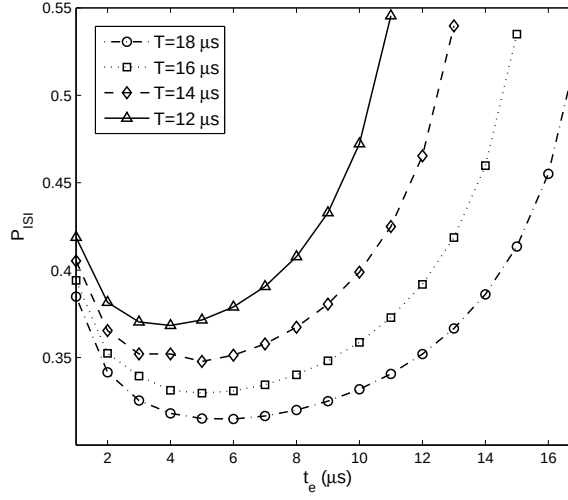


Figure 11.6: The ISI probability values are shown for varying inter-emission time t_e and inter-symbol time T values.

critical values, P_{ISI} first decreases. This result stems from the statistical dependence between P_{ISI} and the successful transmission probability $Pr(0|0)$. While t_e increases, $Pr(0|0)$ can be improved as previously shown in Fig. 11.3 and this reduces P_{ISI} . However, as t_e is further increased, P_{ISI} starts to increase because the difference between t_e and T becomes too small. That is, after some critical values of t_e , as the difference between T and t_e decreases, almost all erroneous transmissions, coming from the probability $(1 - Pr(0|0))$, results in an ISI error and P_{ISI} increases. Hence, we can conclude that, as the difference between T and t_e decreases below some critical values, P_{ISI} cannot be further reduced by increasing t_e and starts to increase with t_e .

In Fig. 11.7, the error probability P_e given in (11.46) is shown with t_e for different T values. Similar to P_{ISI} , P_e also decreases by increasing t_e and T except for the setting in which t_e and T values are close to each other. Note that P_e and P_{ISI} follow almost the same characteristics since P_e is directly a function of P_{ISI} . However, P_e values are slightly larger than P_{ISI} as observed in Fig. 11.6 and 11.7. This increase results from the cases in which there is no ISI error while the symbol cannot be correctly received due to out of order of a single symbol with the probability $(1 - Pr(0|0))$.

Using the same setting in Fig. 11.6 and 11.7, the communication rate of the MARCO channel, i.e., C/T in *bps*, is shown in Fig. 11.8 for varying t_e and T values. In fact, C is given as $C = (1 - P_e)$, where P_e is the error probability in (11.46). Therefore, the communication

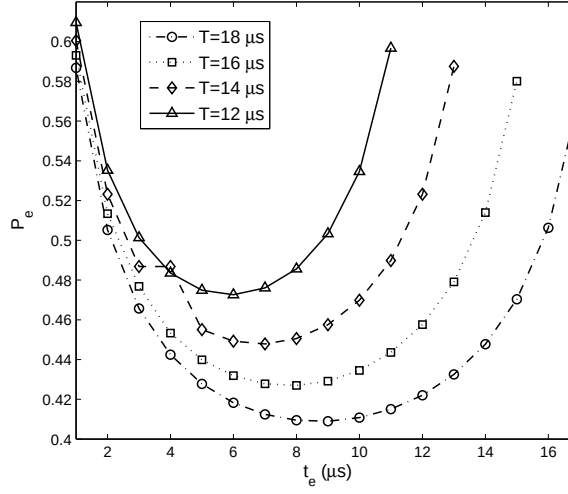


Figure 11.7: The error probability P_e values are shown with different inter-emission time t_e inter-symbol time T values.

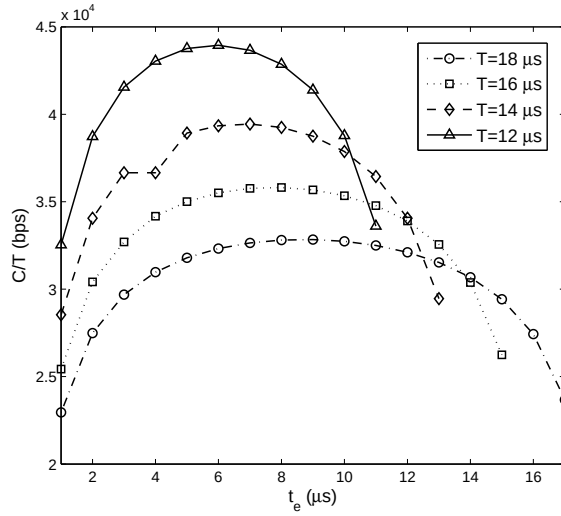


Figure 11.8: MARCO channel capacity $\frac{C}{T}$ is shown with inter-emission time t_e according to the different inter-symbol time T values.

rate C/T is a function of P_e and follows the same characteristics with $(1 - P_e)$ for a specific T . That is, C/T has a concave shape while P_e has a convex shape. This result can be easily verified by observing Fig. 11.7 and 11.8. On the other hand, T values in the denominator of C/T suppress the effect of this increase, and T completely controls the rate in a way that the rate can be enhanced as T decreases. Moreover, it is revealed in Fig. 11.8 that MARCO can provide significantly high communication rate, i.e., up to 45 (Kbps) for the setting in Fig. 11.8.

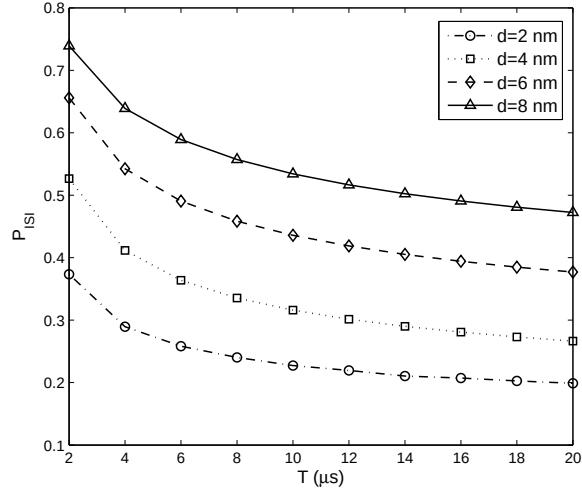


Figure 11.9: P_{ISI} is shown for varying T according to the different inter-node distance d values.

11.3.2 Effects of Inter-Symbol Time and Inter-Node Distance

In Fig. 11.9, P_{ISI} is shown for varying values of T and d . P_{ISI} increases with d due to the fact that longer movement of the molecules causes higher likelihood of ISI. As previously observed in Fig. 11.6, it is also clearly observed in Fig. 11.9 that P_{ISI} decreases with increasing T . Here, we can conclude that larger inter-node distance, d , requires larger inter-symbol time in order to provide smaller P_{ISI} .

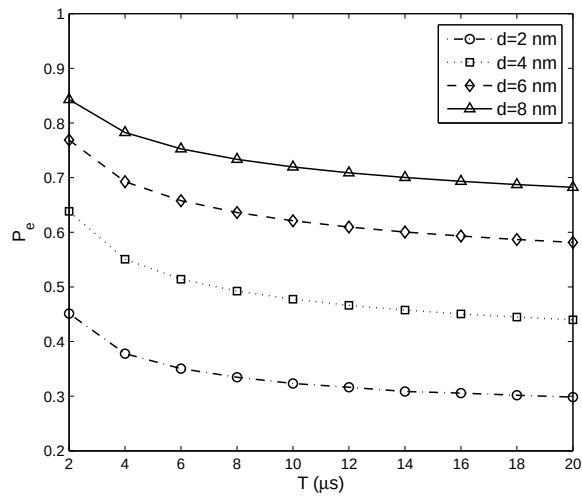


Figure 11.10: The error probability P_e is shown for varying T according to different d values.

In Fig. 11.10, the error probability P_e is shown for varying values of T and d . Similar to P_{ISI} in Fig. 11.9, P_e can be also reduced by increasing T . This is because the inter-symbol interference probability P_{ISI} decreases with increasing T . Furthermore, smaller P_e can be observed when d decreases. Hence, similar to P_{ISI} , larger d necessitates larger T to obtain smaller P_e .

Using the same setting in Fig. 11.9 and 11.10, the rate of MARCO, i.e., C/T in (bps), is shown in Fig. 11.11. Since the rate is directly a function of the error probability P_e , it clearly follows the similar characteristics with P_e in a way that it increases with T and decreases with d . In Fig. 11.11, it is also observed that MARCO can provide very high communication rate, i.e., up to approximately 275 Kbps.

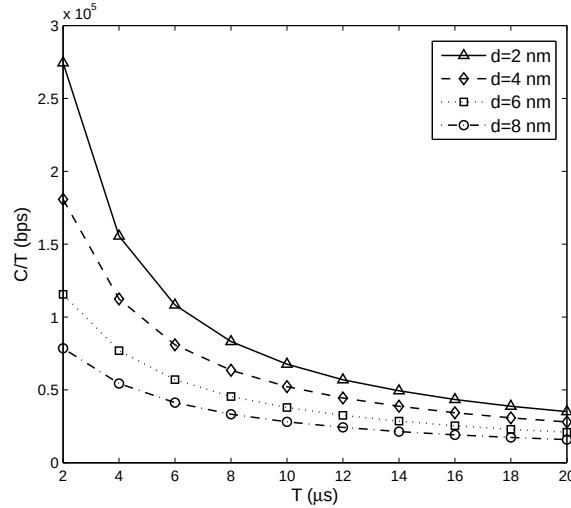


Figure 11.11: $\frac{C}{T}$ is shown for varying T according to the different d values.

11.3.3 Effects of Inter-Emission Time and Diffusion Coefficient

In Fig. 11.12, P_{ISI} is depicted for different t_e and diffusion coefficient D . The effect of t_e on P_{ISI} is clearly similar to the case in Fig. 11.6 such that as t_e increases, P_{ISI} decreases. However, P_{ISI} starts to increase as t_e gets closer to T . On the other hand, P_{ISI} can be also reduced when the diffusion coefficient D of the molecules is high because higher D decreases the expected value of random delay experienced by each molecule and random delay differences among the molecules. The error probability P_e also has the similar characteristics with

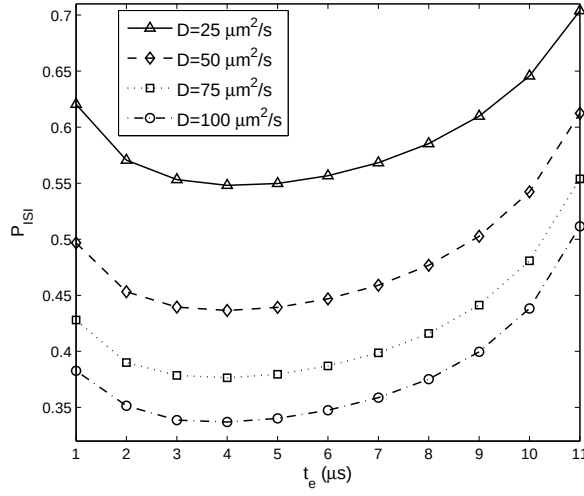


Figure 11.12: P_{ISI} is shown for varying t_e according to the different diffusion coefficient D values.

P_{ISI} as observed with Fig. 11.13. The rate of the MARCO channel, i.e., C/T , is shown with

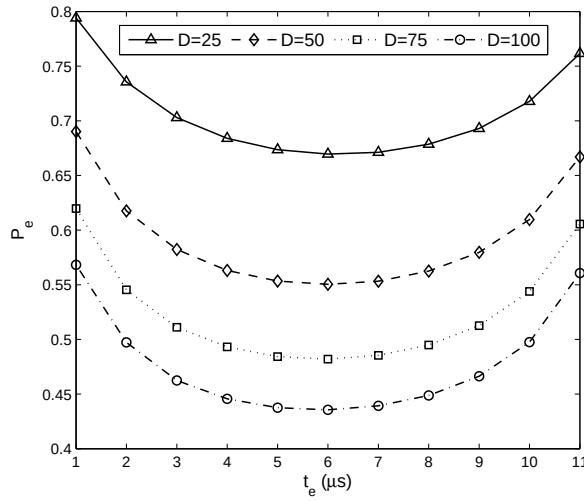


Figure 11.13: P_e is shown for changing t_e according to different D values.

varying t_e for different values of D in Fig. 11.13. Since $C = (1 - P_e)$, the rate easily follows an opposite characteristics with P_e shown in Fig. 11.12. While P_e has a convex shape, C/T forms a concave shape.

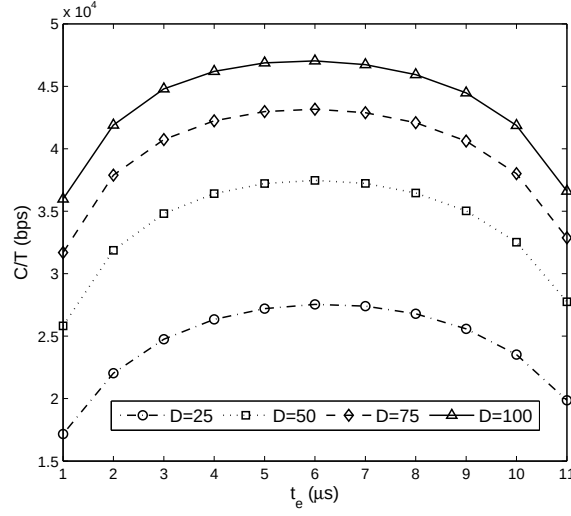


Figure 11.14: $\frac{C}{T}$ is shown for varying t_e according to the different D values.

11.4 Performance Comparison With Existing Molecular Communication Schemes

In this section, we present the analytical performance comparisons of MARCO with the previously proposed molecular communication schemes. Up to now, molecular communication channel has been considered as either concentration channel or timing channel. In the concentration channel, TN transmits a single or a number of molecules at the beginning of a time slot in order to deliver logic bit 1 and transmits no molecule to RN to deliver logic 0 during a slot duration. In the existing literature, such a concentration channel is widely investigated in [132], [181], [182], [183]. In the timing channel, random delay of the molecular arrival at the receiver side is considered as a channel input [184]. However, the conditional probability of channel output, given the channel input is analytically intractable in such a timing channel. Therefore, an approximate discrete time model in which the information carrying molecules are released at the end of the fixed-duration time slots, is proposed in [184]. This discrete time model renders the timing channel strictly synchronization-dependent that is similar to the concentration channel.

In the literature, there are three main modeling approach to the previously proposed molecular communication schemes, i.e., *Z-channel* [183], *binary channel* [132], [181], [182], [183], [184], and *4-input 2-output discrete memoryless channel approaches* [183]. Next, these approaches are briefly introduced and a rate expression is given for each of them for the perfor-

mance comparisons with MARCO.

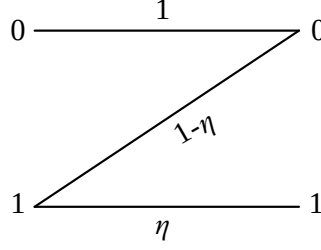


Figure 11.15: Z-channel approach to molecular communication.

11.4.1 Z-channel Approach

As shown in Fig. 11.15, in the z-channel approach, a molecule is transmitted by TN at the beginning of the time slot to deliver logic 1. If the molecule reaches RN within a slot duration, i.e., τ , RN can correctly receive the logic bit 1 with the probability η . Otherwise, RN erroneously receives bit 0 with probability $(1 - \eta)$. In order to deliver 0, TN transmits no molecule within a slot duration and all 0 transmissions are assumed to be successful in the Z-channel [183]. As previously introduced, the delay t experienced by any molecule to reach RN located at distance d obeys the probability density function given in (11.1). Hence, the probability η that TN successfully transmits bit 1 to RN can be obtained as

$$\begin{aligned}\eta &= \int_0^\tau f(t)dt \\ &= \int_0^\tau \frac{d}{\sqrt{4\pi Dt^3}} e^{-\frac{d^2}{4Dt}} dt\end{aligned}\quad (11.49)$$

Using η , the channel transition matrix of the Z-channel, i.e., $\mathbf{P}_z(y|x)$, can be given as

$$\mathbf{P}_z(y|x) = \begin{bmatrix} 1 & 0 \\ 1 - \eta & \eta \end{bmatrix} \quad (11.50)$$

where x and y denote the channel input and output, respectively. Based on the transition matrix $\mathbf{P}_z(y|x)$, the mutual information for the Z-channel, i.e., $I(X; Y)$, can be also derived as

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} Q(x) p_z(y|x) \log_2 \left(\frac{p_z(y|x)}{\sum_{x' \in \mathcal{X}} p_z(y|x')} \right) \quad (11.51)$$

where $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{0, 1\}$ are the input and output alphabets of the Z-channel, respectively. $Q(x)$ denotes the distribution of the input, i.e., $Q(0) = Pr(x = 0)$ and $Q(1) = Pr(x = 1)$.

Hence, the capacity of the Z-channel, i.e., C_z , can be given as

$$\begin{aligned} C_z &= \max_{Q(x)} I(X; Y) \\ &= \max_{Q(x)} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} Q(x) p_z(y|x) \log_2 \left(\frac{p_z(y|x)}{\sum_{x' \in \mathcal{X}} p_z(y|x')} \right) \end{aligned} \quad (11.52)$$

where C_z is in bits per transmission. Since a transmission lasts for a slot duration τ , a communication rate for the Z-channel in bits per second (*bps*) can be expressed as $\frac{C_z}{\tau}$ (*bps*).

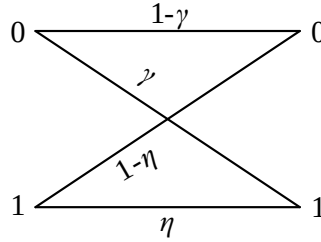


Figure 11.16: Binary channel approach to molecular communication.

11.4.2 Binary Channel Approach

In the Z-channel approach, all transmissions of bit 0 via emitting no molecule are assumed to be successful. However, this assumption is not realistic since the molecules emitted in the previous intervals may reach RN later when 0 bit is transmitted, which can clearly result in erroneous transmission. Therefore, such erroneous transmissions of logic bit 0 must be taken into account by employing a binary channel in Fig. 11.16 as previously introduced in [132], [181], [182]. Assuming that current transmission of logic bit 0 may be only affected by the logic bit 1 transmitted in the previous interval, the probability γ , that a logic bit 0 not successfully delivered to RN by TN, can be derived as

$$\begin{aligned} \gamma &= Q(1) \int_{\tau}^{2\tau} f(t) dt \\ &= Q(1) \int_{\tau}^{2\tau} \frac{d}{\sqrt{4\pi Dt^3}} e^{-\frac{d^2}{4Dt}} dt \end{aligned} \quad (11.53)$$

By using γ and η derived in (11.49), the channel transition matrix of the binary channel, i.e., $\mathbf{P}_b(y|x)$, can be given as

$$\mathbf{P}_b(y|x) = \begin{bmatrix} 1 - \gamma & \gamma \\ 1 - \eta & \eta \end{bmatrix} \quad (11.54)$$

Based on $\mathbf{P}_b(y|x)$, the capacity of the binary channel in bits per transmission, i.e., C_b , can be derived using the same steps followed in (11.51) and (11.52). Furthermore, a communication rate in *bps* for the binary channel can be also given as $\frac{C_b}{\tau}$.

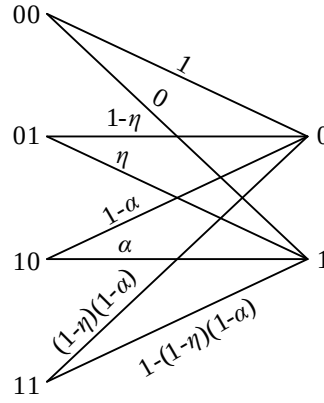


Figure 11.17: 4-input 2-output memoryless channel approach to molecular communication.

11.4.3 4-input 2-output Memoryless Channel Approach

In order to mitigate the interference from the late molecules as in the Z- and binary channel approach, the 4-input 2-output discrete memoryless channel approach is introduced in [183] to comprehensively model the molecular communication channel as shown in Fig. 11.17. In this approach, the input and output alphabets are given as $\mathcal{X} = \{00, 01, 10, 11\}$ and $\mathcal{Y} = \{0, 1\}$, respectively. Furthermore, the channel transition probability α can be derived as follows:

$$\begin{aligned} \alpha &= \int_{\tau}^{2\tau} f(t) dt \\ &= \int_{\tau}^{2\tau} \frac{d}{\sqrt{4\pi D t^3}} e^{-\frac{d^2}{4Dt}} dt \end{aligned} \quad (11.55)$$

Hence, using α and η derived in (11.49), the channel transition matrix for the 4-input 2-output channel, i.e., $\mathbf{P}_f(y|x)$, can be given as

$$\mathbf{P}_f(y|x) = \begin{bmatrix} 1 & 0 \\ 1 - \eta & \eta \\ 1 - \alpha & \alpha \\ (1 - \alpha)(1 - \eta) & 1 - (1 - \alpha)(1 - \eta) \end{bmatrix} \quad (11.56)$$

Based on $\mathbf{P}_f(y|x)$, the capacity of 4-input 2-output channel in bits per transmission, i.e., C_f , can be derived using the same steps followed in (11.51) and (11.52). Note that since a transmission lasts for two slot duration, 2τ , a communication rate in *bps* for 4-input 2-output channel can be given as $\frac{C_f}{2\tau}$. Next, using the derived rate expression of Z-, binary, and 4-input 2-output channel approaches, the performance comparisons of MARCO with these channel approaches are presented.

Table 11.1: Rate Comparison With Respect To Varying Values of Diffusion Coefficient (D , $\mu m^2/s$)

Diffusion Coefficient	D=100	D=200	D=300	D=400	D=500	D=600	D=700	D=800	D=900	D=1000
Z-channel	1.1×10^{-41}	8.0×10^{-20}	2.8×10^{-12}	8.1×10^{-9}	1.3×10^{-6}	0.4×10^{-4}	0.4×10^{-3}	0.3×10^{-2}	0.0128	0.0411
Binary Channel	5.7×10^{-20}	5.7×10^{-9}	0.2×10^{-4}	0.2×10^{-2}	0.0290	0.1666	0.5828	1.4895	3.0765	5.4599
4-input 2-output Ch.	1.6×10^{-12}	4.0×10^{-9}	0.2×10^{-4}	0.0015	0.0205	0.1181	0.8279	1.0727	2.2498	4.0743
MARCO Channel	557.991	560.415	562.882	192.847	154.415	299.651	279.700	264.248	184.602	244.322

11.4.4 Performance Comparison

In this section, we numerically compare the communication rates of MARCO, Z-, binary, and 4-input 2-output channel approaches. Apart from the results given in Section 11.3, here, the inter-node distance d is selected on the scale of μ meters, that is, d is changed within the interval $d = 1 - 10 \mu m$. This also shows the relatively longer range performance of MARCO. The diffusion coefficient is changed from $D = 100$ to $D = 1000 \mu m^2/s$. The slot time τ and inter-symbol time T are changed from 1×10^{-4} to 2×10^{-6} . In Table 11.1, the communication rates are presented according to changing values of diffusion coefficient D . Here, the communication rates are computed by setting T , τ , and d as $T = \tau = 1 \times 10^{-4} s$ and $d = 2 \mu m$. Contrary to the Z-, binary, and 4-input 2-output channels, the communication rate

of MARCO decreases as D increases. However, MARCO significantly outperforms the other channel approaches.

Table 11.2: Rate Comparison With Respect To Varying Values of Slot and Inter-symbol Time (τ and T)

Slot Time	$\tau = 1 \times 10^{-4}$	$\tau = 2 \times 10^{-5}$	$\tau = 1 \times 10^{-5}$	$\tau = 2 \times 10^{-6}$
Z-channel	0.0411	4.0×10^{-19}	1.1×10^{-40}	4.7×10^{-214}
Binary Channel	5.4599	2.8×10^{-8}	5.7×10^{-19}	1.7×10^{-105}
4-input 2-output Ch.	4.0743	2.0×10^{-9}	3.2×10^{-18}	1.3×10^{-101}
Inter-Symbol Time	$T = 1 \times 10^{-4}$	$T = 2 \times 10^{-5}$	$T = 1 \times 10^{-5}$	$T = 2 \times 10^{-6}$
MARCO Channel	244.322	1168.56	2324.09	11568.50

In Table 11.2, the comparisons are given with respect to changing slot time τ and inter-symbol time T . Here, d and D are set to $d = 2$ and $D = 1000$. The rate of Z-, binary, and 4-input 2 output channels decreases as the slot time decreases. This stems from the erroneous transmissions coming from the late molecules such that as the slot time decreases, the number of late molecules increases. On the other hand, the rate of MARCO can be improved by increasing the inter-symbol time and provide significantly higher communication rate than Z-, binary, and 4-input 2 output channel approaches.

Table 11.3: Rate Comparison With Respect To Varying Values of Inter-Node Distance (d , μm)

Inter-Node Distance	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$	$d = 7$	$d = 8$	$d = 9$	$d = 10$
Z-channel	211.18	94.25	37.05	12.25	3.30	0.71	0.12	0.16	0.01	0.00
Binary Channel	155.06	45.10	10.44	1.88	3.59	3.06	1.73	0.74	0.25	0.07
4-input 2-output Channel	90.08	42.54	23.48	13.70	7.60	3.75	1.60	0.59	0.19	0.05
MARCO Channel	1913.80	1168.56	746.58	2783.86	2781.67	2780.48	2779.76	2779.30	2778.78	2778.78

The comparisons according to changing values of inter-node distance, i.e., d , are presented in Table 11.3. Here, T , τ , and D are set to $T = 2 \times 10^{-5}$, $\tau = 2 \times 10^{-3}$, and $D = 1000$, respectively. The rate of Z-, binary, and 4-input 2 output channel decreases as d increases. The rate of MARCO also decreases as d is changed from $d = 1$ to $d = 3$. However, as d is further increased, the rate of MARCO converges to nearly 2700 (*bps*). This result is due to the fact that the rate of MARCO becomes independent of the inter-node distance after some critical values of the inter-node distance. Furthermore, MARCO significantly outperforms Z-, binary, and 4-input 2 output channel approaches.

11.5 Conclusion

In this chapter, we introduce the molecular array-based communication (MARCO) scheme that uses the transmission order of different molecules to encode and exchange information symbols between nanomachines in nanonetworks. In order to quantify the MARCO performance, we first extensively derive the inter-symbol interference (ISI) and error probabilities based on the principles of Brownian motion. Then, using these probabilities we give a capacity expression for the MARCO channel in which two distinct molecules are used to encode the information. We also generalize the capacity expression for the case of arbitrary number of different molecules used for the encoding. Numerical results clearly show that MARCO incomparably outperforms the previously developed molecular communication schemes based on molecular concentration and timing channels. More specifically, it is revealed by the numerical results and performance comparisons that MARCO can provide significantly higher communication capacity, i.e., on the scale of 100 *Kbps.*, than the previously developed molecular communication schemes.

CHAPTER 12

Body Area NanoNetworks with Molecular Communications in Nanomedicine

The aim of this chapter is to introduce the concept of body area nanonetworks (BAN²) and highlight its unique research challenges for an early development stage of efficient and reliable communication and networking techniques in BAN². This work in this chapter was also presented in [185].

In Section 12.1, the motivation of this chapter is introduced. Nanomedicine and body area nanonetworks are discussed in Section 12.2. In Section 12.3, communication theoretical approaches to BAN² are discussed and the conclusions are given in Section 12.4.

12.1 Motivation

Within the last two decades, the outstanding progresses in nano and biotechnology have radically shifted how the human healthcare can be approached. Growing interest in the medical applications of these technologies has recently enabled the emergence of sophisticated nanomedicine applications [186]. Unlike the earlier applications with simple nano-structured materials and devices, these sophisticated applications include therapeutic NanoMachines (NM), e.g., engineered bacteria and nanorobots. For example, a group of bacteria can be genetically engineered to recognize and destroy tumors in human body [187]. Besides their great potential, very tiny size and unknown physics of the therapeutic NM, however, cause crucial unreliability and uncontrollability problems in these applications. For example, artificial cells used for a smart drug delivery application may easily expose to an intrinsic and extrinsic biological noise in gene expression. This makes the drug delivery process com-

pletely unreliable and uncontrollable [188].

Through a nanoscale communication paradigm, the network of therapeutic NM, i.e., Body Area NaNonetwork (BAN²), can be realized to enforce reliability and controllability of complex nanomedicine applications. More importantly, BAN² can also coordinate nanomedicine tasks among heterogeneous NM¹ populations to reach a highly sophisticated behavior and increase the number of design possibilities. For example, a group of non-communicating engineered cells behaves asynchronously and cannot coordinate to perform a predefined task. However, a BAN² of communicating synthetic cells can overcome the asynchronous behavior problem and coordinate the cell population for a tissue engineering application [188]. Furthermore, BAN² may be also required in the promising nanomedicine applications in order to

- coordinate cooperative NM tasks
- efficiently gather biosensor data
- correlate the biosensory inputs and make decision
- transmit information message to external entities

BAN² can also augment various other advancements in the emerging molecular computing systems where logic computations are performed via biological entities. Such computations could also enable fine grain detection of diseases and take action through smart drug release [189]. Therefore, BAN² can support communication of NM needed for such logic computations.

In the literature, there are four main nanoscale communication paradigms that may provide the interconnection of the therapeutic NM in BAN². These are nanomechanical, acoustic, electromagnetic, and molecular communication paradigms [12]. In nanomechanical communication, message is transmitted by a mechanical contact between transmitter and receiver. Acoustic energy such as pressure variation is used for conveying information in acoustic communication. Electromagnetic communication is enabled by the modulation of electromagnetic waves to transmit information. Furthermore, molecular communication allows NM to communicate with each other by using molecules as communication carrier. Among these, due to

¹ Throughout the chapter, we interchangeably use the term therapeutic nanomachine or nanomachine to call the therapeutic nanomachines that are used in the nanomedicine applications.

the lack of the communication equipments, the nanomechanical and acoustic communication may not be feasible for the realization of BAN². Although the electromagnetic communication equipments, e.g., carbon nanotube radio and antenna, are available, most of these devices are not biocompatible. This may render the electromagnetic communication inappropriate for the nanomedicine applications. On the other hand, molecular communication is the most promising approach for the realization of BAN² based on the following advantages:

- Molecular communication is biocompatible as it has been already employed in many natural phenomena. For example, in natural immune system, the white blood cells communicate with each other using molecules to cooperatively sense and eliminate the hazardous pathogen. The molecular communication among white blood cells forms the biological immune network that is an excellent defense mechanism of organism.
- The molecular communication equipments are currently available as the molecular communication systems have been already realized. For example, using nitric oxide signaling elements, an artificial cell-to-cell communication system in mammalian cells has been developed in order to govern a cell population. This system can also serve as a functional tool for complex nanomedicine applications such as gene therapy and artificial gene regulatory networks [17].

Apparently, BAN² with molecular communication has great potential in its operation size and capabilities to enable radically new and promising nanomedicine applications. However, this also brings many crucial communication challenges due to the salient characteristics of molecular communication outlined as follows:

- Communication carriers are messenger molecules whose propagation speed is extremely low.
- Communication devices and signals are severely prone to thermal noise, molecular noise, drift, and fading.
- Channel characteristics differs significantly from the characteristics of traditional wireless communication channel.
- Nanoscale materials have high manufacturing defect rates and operational uncertainties.

- Mobility of NM is governed by the rules of physics in this regime, e.g., the rules of Brownian motion.

In the existing literature, there also exist several research efforts toward addressing the unique challenges of molecular communications and nanonetworks. However, to the best of our knowledge, the concept of BAN² in nanomedicine, its associated challenges, and open research problems have not yet been investigated.

12.2 Nanomedicine and Body Area Nanonetworks

In this section, we first introduce a brief overview of nanomedicine. Then, we discuss the requirement of BAN² in nanomedicine applications and present the existing BAN² proposed or physically realized for a large set of nanomedicine applications.

12.2.1 A Brief Overview of Nanomedicine

Nanomedicine is the process that uses molecular tools and molecular knowledge of the human body to diagnose, treat, and prevent diseases. Nanomedicine is a continuously growing field as nanotechnology and biotechnology develop novel nanoscale tools that can be easily adopted for promising applications. The current and envisioned future nanomedicine applications can be categorized into three main groups according to their chronological evolution stages [186]. The first group includes the first generation nanomedicine applications with simple and currently available nanoscale tools such as nanopores [190] and quantum dots [191] as shown in Fig. 12.1. These tools are capable of directly interacting with biological phenomena, e.g., a single cell or a cell organelle. For example, to restore glucose feedback loop of diabetics, nanopores are used to encapsulate insulinoma cells [190]. The second group involves more remarkable synthetic biology-based tools, that are adopted or synthesized from nature, such as engineered bacteria [187] and synthetic cells [192] as seen in Fig. 12.1. For example, a group of bacteria can be genetically engineered to recognize and destroy tumors in human body [187]. The third group includes the nanomedicine applications with mostly futuristic tools, i.e., bio-hybrid or mechanical nanorobots, that may join nanomedicine equipments within a longer term. For example, wireless intracellular surgery are expected to be realized via mechanical nanorobots [186].

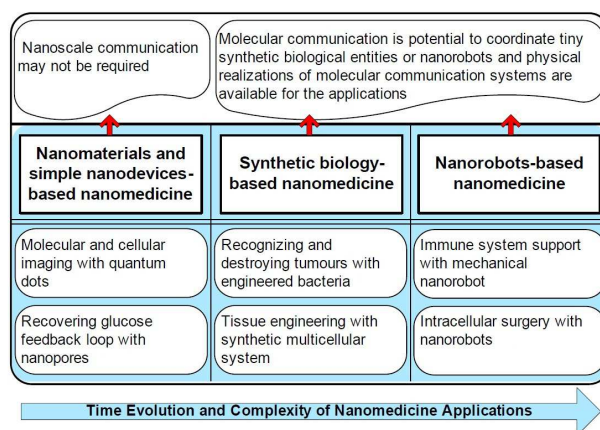


Figure 12.1: Three kinds of nanomedicine applications with appropriate nanoscale tools are categorized according to their time evolution and complexities.

In Fig. 12.1, some outstanding nanomedicine applications that employ these three groups of nanomedicine tools are briefed with respect to their time evolution and complexities. As observed in the figure, controllability and reliability of the applications grow into a more challenging problem while the applications evolve and become more sophisticated. For example, in a tissue engineering application, an artificial cell population must be governed to reliably form a multicellular system having predefined patterns [192].

Such controllability and reliability issues can be addressed by forming BAN² through molecular communication of therapeutic NM. BAN² can also coordinate nanomedicine tasks among heterogeneous NM populations to reach a highly sophisticated behavior and increase the number of design possibilities. For example, in Fig. 12.2, artificial cells and nanorobots share molecular information to coordinate and accomplish a sophisticated drug delivery and immune system support task.

In fact, the nature-made nanonetworks are also indispensable for many vital functionalities. For example, the molecular communication of bacteria in quorum sensing forms a nature-made nanonetwork in which bacteria produce, emit, and receive hormone-like messenger molecules. This natural process permits bacteria to synchronize all colony activities and to change the colony state as a response to an external stimulus. Similar to nature, in nanomedicine, BAN² are required to make the applications more manageable and reliable while providing more sophisticated design possibilities.

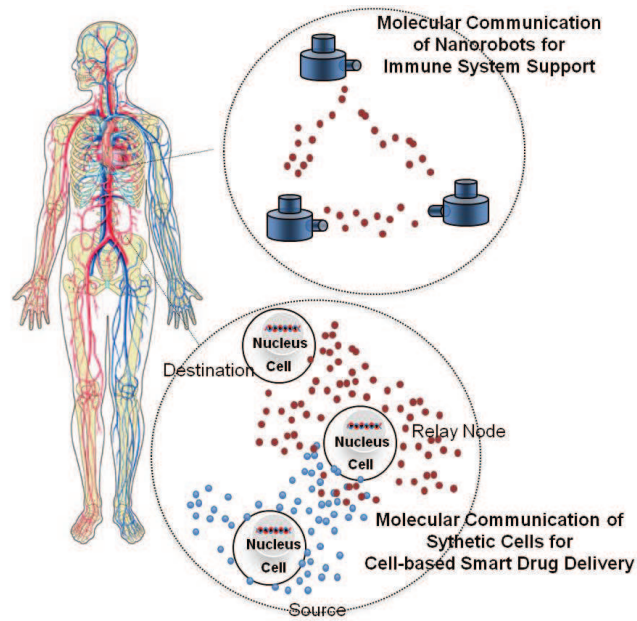


Figure 12.2: Two different body area nanonetworks of artificial cells and nanorobots with molecular communication to realize a sophisticated drug delivery and immune system support task.

12.2.2 Molecular Communications-enabled Body Area Nanonetworks in Nanomedicine Applications

As shown in Fig. 12.1, unlike the early nanomedicine applications, the synthetic biology- and nanorobots-based applications radically transforms treatment of complex diseases. In these applications, through molecular communication, BAN² of therapeutic NM can perform computation, logic operation, and make decisions for the treatment of complex diseases. Here, we review some outstanding examples of these BAN² proposals.

In synthetic biology, molecular communication systems are designed to control the behavior of synthetic cell populations in both time and space. With broadcasting and receiving messenger molecules, molecular communication can enable the robust and predictable population dynamics [188]. For example, in [192], a molecular communication system is proposed to allow genetically engineered cells to form some predefined spatial pattern. The chemical signal acyl-homoserine lactone (AHL) is synthesized and emitted by sender cells to the medium and the receiver cells with band-detect gene networks respond to a threshold AHL concentration by forming a ring-like topology. The given molecular communication system is expected to foster the promising nanomedicine applications such as tissue engineering, biomaterial fabri-

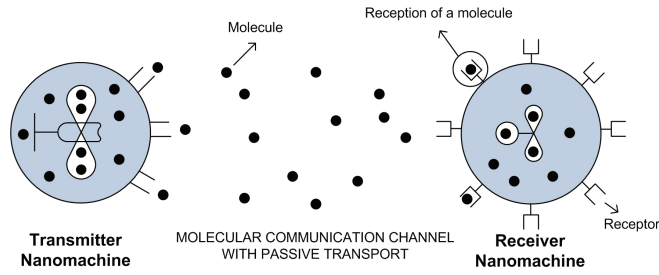


Figure 12.3: Molecular communication channel with passive transport of information carrying molecules.

cation, and biosensing. Furthermore, in [17], a cell-to-cell molecular communication system with nitric-oxide signaling elements is introduced. The given system aims to increase the functionality of synthetic genetic circuits used for coordinating and programming engineered cell populations. It may also serve as a versatile tool in gene therapy applications.

Similarly, nanorobots-based nanomedicine applications also requires BAN² of nanorobots to provide reliable and controllable operations. The collective efforts of the communicating nanorobots can improve the likelihood of successful operations. Molecular communication clearly helps the nanorobots to easily correlate its sensory observations with its close proximity and an external micro system designed for a health monitoring task [186].

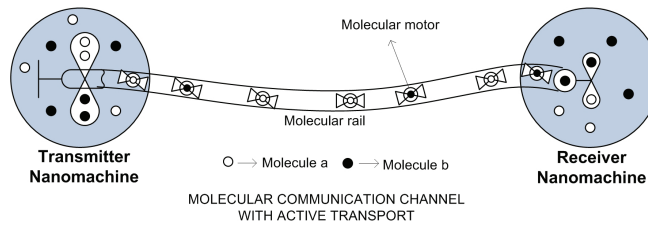


Figure 12.4: Molecular communication channel with active transport of information carrying molecules.

12.3 Communication Theoretical Approaches To BAN² For Nanomedicine

In this section, we present the current state-of-the-art channel models and communication theoretical analyses devised for molecular communication systems. We also introduce open research issues, new interests, and developments beyond these models and analyses.

12.3.1 Channel Model and Communication Rate

According to the propagation types of carrier molecules, molecular communication can be mainly categorized into two major groups, i.e., molecular communication with passive transport and with active transport. In the passive transport, information is conveyed from a transmitter nanomachine (TN) to a receiver nanomachine (RN) via free diffusion of carrier molecules as shown in Fig. 12.3. TN emits the molecules to the channel and the emitted molecules freely diffuse in the medium. Some of these molecules electrochemically interact with the receptors of RN. Such electrochemical contacts provide the reception of transmitted information. In the active transport, molecular information is communicated actively using molecular motors, e.g., kinesin, in a molecular rails or microtubules as shown in Fig. 12.4.

12.3.1.1 Molecular Communication with Passive Transport

In the traditional wireless communication, information messages are encoded via binary bit levels, i.e., bit 0 and 1. These bits also correspond to two different signal levels that are transmitted and received by transmitter and receiver, respectively. Similarly, molecular communication channel with passive transport is mostly modeled as *binary channel* with two bits, i.e., bit 0 and bit 1 [132]. In the binary channel approach, a single or a number of molecules are transmitted by TN at the beginning of the fixed-duration time slots to deliver bit 1. If the molecule or a threshold concentration reach RN within a slot duration, RN can correctly receive the bit 1. Otherwise, RN erroneously receives bit 0. In order to deliver 0, TN transmits no molecule within a slot duration. However, the molecules emitted in the previous slot intervals may reach RN later when 0 bit is transmitted. This can clearly result in error in the delivery of 0. In order to mitigate such errors, the binary channel approach is also modified by extending the transmission of single bit to two slot durations. More specifically, 00, 01, 10, and 11 are transmitted to deliver bit 0 and 1. Hence, the resultant channel is called as *4-input 2-output channel* [183].

In Fig. 12.5, the communication rates provided by the binary and 4-input 2-output channel approach are compared for the varying values of inter-node distance between TN and RN. 4-input 2-output channel approach clearly outperforms the binary channel approach since it mitigates the errors by wisely expanding the duration of single bit transmission. For the shorter

inter-node distance values, i.e., 0.1-0.5 μm , 4-input 2-output channel slightly enhances the rate while it considerably improves the rate for the relatively longer distances. This result reveals that the 4-input 2-output channel may be more appropriate for the nanomedicine applications in which NM are not relatively close to each other. Apart from the binary and 4-input 2-output channel approaches, passive transport is also analytically modeled by incorporating three different phases of molecular communication, i.e., molecular emission, diffusion, and reception, into a single channel model [40].

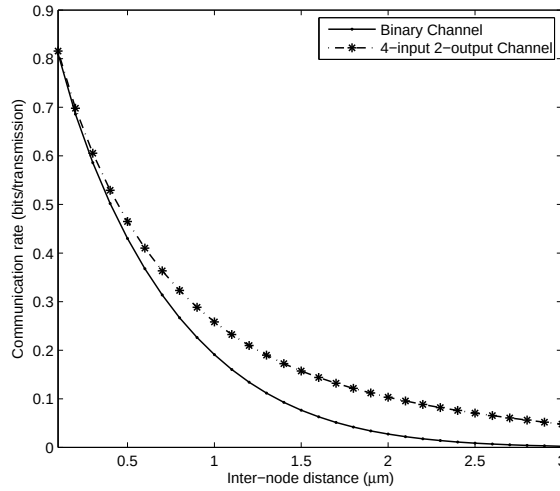


Figure 12.5: Communication rate provided by binary channel and 4-input 2-output channel with respect to varying inter-node distance.

12.3.1.2 Molecular Communication with Active Transport

The communication performance of the active transport has been mostly characterized by means of an error probability. The molecules, loaded by TN on molecular motors as shown in Fig. 12.4, are assumed to degrade and not to reach RN with the error probability p_e . If n distinct molecules are used, each of these n molecules is either successfully received with the probability $(1 - p_e)$ or cannot reach to RN with the probability p_e . Hence, such a molecular active transport channel can be modeled as a n -ary erasure channel and the communication rate of this channel can be given as $(1 - p_e) \log n$.

In Fig. 12.6, the communication rate of the n -ary erasure channel is shown with different p_e values according to varying number of distinct molecules, i.e., n , used in the channel. The

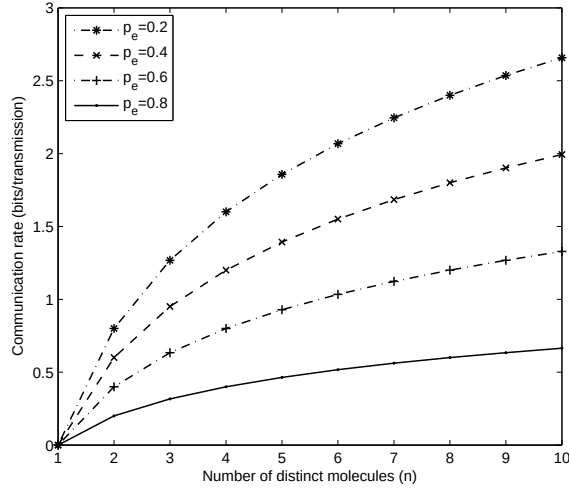


Figure 12.6: Communication rate provided by n – ary erasure channel with varying channel error probability p_e for varying number of distinct molecules used in the channel.

rate can be clearly increased with n . By comparing the numerical results in Fig. 12.5 and 12.6, it can be concluded that active molecular communication may be more proper than passive molecular communication in terms of providing higher communication rate. However, active molecular communication always necessitates an infrastructure, i.e., molecular rail and motors, while passive molecular communication are not affected by such restriction. In fact, therapeutic NM may not require a high communication rate to coordinate and to share a simple information content in nanomedicine applications. Therefore, active molecular communication may not be preferred if a preset infrastructure is not available.

12.3.2 Channel Noise

The characteristics of communication noise in the molecular communication channel also differ significantly from the traditional wireless communication channel. Unique theoretical model of molecular noise in diffusion-based molecular communication systems is comprehensively introduced in [176]. The effect of noise on the molecular communication performance is also investigated by considering various molecular transmission, propagation, encoding, and decoding mechanisms [177]. In addition to these efforts, biological noise [188], that is inherently observed in inter/intra cellular processes, must be jointly considered for a satisfactory communication performance in BAN². Furthermore, the exact probability distribution of

the noise process must be discovered. Using the probability distribution, information entropy of the noise must be also derived to comprehensively obtain the capacity of the molecular communication channel. In fact, the noise signal is statistically correlated with the channel input as introduced in [144]. Therefore, the channel input and noise must be jointly considered to investigate the information theoretical characteristics of the molecular communication channel. In addition to this, since the emitted molecules continuously diffuse in the medium until they are received, transient and asymptotic behavior of the noise must be also investigated.

12.3.3 Deployment and Topology

As observed in Fig. 12.5, the capacity of molecular communication is severely prone to communication distance such that the capacity decreases with the communication distance. Therefore, dense deployment of NM is imperative for the nanomedicine applications to reliably convey the message along highly delicate paths of BAN². On the other hand, some nanomedicine applications may require mobile NM. Unlike the traditional mobility models of ad hoc nodes, in BAN², mobility of NM is mainly governed by extremely different rules such as Brownian motion. For example, in [155], a communication rate is derived for mobile ad hoc molecular nanonetworks inspired by natural immune network consisting of mobile and communicating white blood cells. The molecular communication among NMs is assumed to be provided by means of random collisions of mobile NM. The analytical evaluations of the rate expression manifest that BAN² with mobile NM can be realized with a sufficiently high end-to-end communication rate. Furthermore, some nanomedicine applications may necessitate a predefined topology of BAN². For example, a BAN² may be deployed on some critical points of the human heart for the early detection of a possible heart attack. In such an application, all communication parameters must be regulated with respect to the predefined topology to minimize end-to-end latency so as to report a possible attack before it occurs.

12.3.4 Computation, Logic Operations, and Decision

In complex nanomedicine applications, BAN² of therapeutic NM performs computation, logic operations, and makes decision to treat complex diseases. For example, an autonomous biomolecular computer is designed, that analyzes the level of messenger RNA and produces

molecule so as to logically control the gene expression processes [189]. However, due to the lack of a central controller, in such applications, all operations must be managed by self-organization of NM through molecular communication. In these operations, the reliability of molecular communication is crucial in order to efficiently provide computations and logic operations with relatively low error rate. Therefore, a reliable and error-tolerant molecular encoding and decoding techniques must be developed for such nanomedicine applications. Furthermore, network information theory may be used to investigate ultimate computation capability of coordinating NM in BAN². Molecular information flow may be also formulated in order to deduce the computation speed. For example, deterministic molecular information flow in nanonetworks is formulated in [144] and this formulation can be used to investigate the computational aspects of BAN².

12.3.5 Medium-Access and Routing

In order to achieve a communication performance closer to the information theoretical bounds, efficient networking techniques are also required for the realization of BAN². Due to the very low-end memory and processing capabilities of self-organizing NM, the developed algorithms entail the minimum level of computational complexity. In [144], the communication performance of a molecular multiple-access channel is investigated. The analytical evaluations reveal that different NM must access the medium using different types of molecules that can be distinguished at the receiver side. This is because communication bandwidth allocated to each NM reduces almost zero if the same molecule type is used to access the same channel by a set of NM [144]. Furthermore, the directivity of emitted molecules must be investigated to develop an efficient routing algorithm for BAN². In fact, a single molecule may diffuse toward all directions according to the rules of Brownian motion. However, a drift velocity may be used to direct a molecule to a specific direction. Therefore, drift velocity and routes may be jointly considered to develop an efficient routing algorithm. Since the propagation speed of the molecules is extremely low, routing algorithms must be also delay-tolerant. For example, mobility of NM may be exploited to tolerate high delay in a way that transmission strategies of NM are regulated according to the mobility patterns of NM. Such a transmission strategy can reduce the delay.

12.4 Conclusion

The field of nanomedicine continues to evolve and advance, resulting from extensive research and development efforts in the field of nanotechnology. However, the pace of this development as well as increased opportunities in nanomedicine can be further realized through molecular communication. In this chapter, we have presented the concept of BAN² which exploits molecular communication to enable NM to communicate. The chapter discussed current solutions used in nanomedicine, and these applications are further discussed from the perspective of BAN². Finally, the communication theoretical efforts such as channel capacity, deployment and topology, as well as networking issues, are reviewed. These issues are further discussed through open research challenges that introduce new interests and development opportunities. Our investigations reveal that BAN² with molecular communication is a necessary and promising tool for future nanomedicine in order to provide controllable, reliable, and efficient applications.

CHAPTER 13

Carbon Nanotube-based Nanoscale Ad Hoc Networks

The aim of this chapter is to introduce the concept of carbon nanotube-based nanoscale ad hoc networks (CANET) and highlight its unique research challenges for an early development stage of efficient and reliable communication and networking techniques for CANET. This work was also presented in [156].

In Section 13.1, the motivation behind this chapter is discussed. Hardware components of a CANET nanonode are described in Section 13.2. Then, the fundamental communication and networking issues in CANET are also introduced in Section 13.3 and 13.4, respectively. Finally, the concluding remarks are discussed in Section 13.5.

13.1 Motivation

Nanotechnology enables the practical realization of very low-end nanomachines that have tiny components to accomplish a simple specific task such as communication, computation, and sensing. Nanomachines can also be interconnected via nanoscale communication links to form a nanonetwork to fulfill complex tasks such as collaborative drug delivery, health monitoring, and biological or chemical attack detection in nanoscale environments [12]. On the other hand, in nanotechnology, one of the most important unsolved problems is how to make electrical contact from nanoelectronic devices to the macroscopic world without any performance degradation. One potential solution is nanoscale communication links, i.e., wireless network-on-chip networks [193]. Hence, the nanoscale communication necessitates the integration of nano and communication technologies to develop frontier nanonetwork applications.

In the literature, there are four main nanoscale communication techniques called nanomechanical, acoustic, molecular, and electromagnetic communication [13]. In nanomechanical communication, message is transmitted by a mechanical contact between transmitter and receiver. In acoustic communication, acoustic energy such as pressure variation is used for information transmission. Molecular communication provides nanoscale communication using molecules as communication carrier [132]. Among these, due to the availability of many micro or nanoscale communication equipments and high propagation velocity of electromagnetic waves, electromagnetic communication is the most viable and appropriate technique for the realization of nanoscale communication.

Promising potential of nanomaterials for transmission and processing electromagnetic signals provides a growing interest in their electromagnetic response. A new branch of nanoelectromagnetics is currently emerging to reveal favorable electromagnetic properties of nanoscale materials [194]. Especially, carbon nanotubes (CNT) are of special interest due to their unusual electronic and electromagnetic properties. Theoretical and experimental studies recently show that a single CNT can be designed as the four fundamental components of a radio circuitry, i.e., antenna, tuner, modulator, and demodulator, to listen radio broadcast [195]. Electromechanical vibrations of nanotubes can also be harnessed to design a nanoscale transmitter circuitry [196]. Moreover, using CNT it is also possible to fabricate conducting wires to enable nanoscale antennas that have the closest physical realization to dipole antennas [193].

CNT-based nanoscale communication has great potential in its size and capabilities to extremely shift current wireless technologies to radically new applications such as radio controlled tiny devices to operate in bloodstream, radically simpler, smaller, and cheaper wireless devices [195]. Apparently, all of these frontier communication propositions necessitate the realization of Carbon nanotube-based nanoscale Ad hoc NETWORKS (CANET). However, this brings many new research challenges, which render the current solutions for traditional wireless ad hoc networks inapplicable, due to the salient characteristics of CANET outlined as follows:

- Scale of the communication devices is on the order of micrometer.
- Wireless communication is based on electromechanical vibrations in CNT receiver and transmitter [195].

- Communication signals are severely prone to thermal noise and fading.
- Molecular composition of the communication medium is crucial to model the path loss and noise terms [197].
- Signal power generated by transmitter circuitry is considerably insufficient [196].
- Dense deployment of devices is imperative for network connectivity.
- Nanoscale battery lifetime is significantly lower than the existing solid-state batteries.
- Nanoscale memory and processor are considerably inefficient in data storage and computation.
- Nanoscale materials have high manufacturing defect rates and operational uncertainties.
- Mobility of nanoscale ad hoc nodes is governed by the rules of physics in this regime.

In the literature, there exist several research efforts on the definition, analysis, modeling, and realization of various nanonetworks [198], [199], [200], [201]. However, to the best of our knowledge, an infrastructureless ad hoc nanoscale networks composed of nanomachines, its associated challenges, and open research problems have not yet been investigated.

13.2 Hardware Components of a CANET Nanonode

In CANET, performance of nanoscale communication among network nodes, i.e., nanonode, is affected by the capabilities of four fundamental hardware components, i.e., nanotransceiver, nanopower, nanoprocessor, and nanomemory units as shown in Fig. 13.2. Thus, in this section, we give the state-of-the-art of these nanoscale hardware components and discuss their communication, computation, and data storing capabilities for nanoscale communication in CANET.

13.2.1 Nanotransceiver

Here, we review the carbon nanotube-based receiver, transmitter, and antenna circuitries, and inspect their communication capabilities.

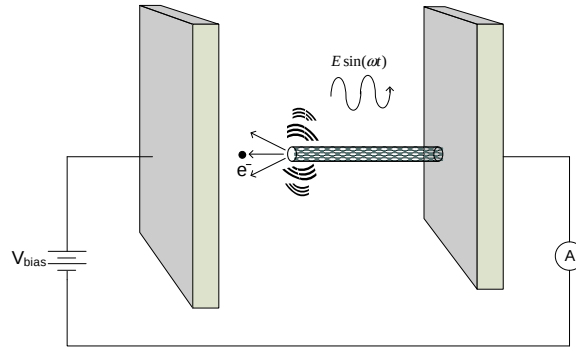


Figure 13.1: Illustration of nanotube radio circuit [195].

13.2.1.1 Carbon Nanotube Radio

The mechanical resonance frequencies of CNT are in the range of 50 MHz - 5 GHz and this range clearly overlaps with the microwave communication spectrum used in traditional wireless communication systems. Hence, this overlapping gives inspiration for developing many innovative nanoelectronic devices [202]. Recently, a single CNT is designed as the four fundamental components of a radio circuitry, i.e., antenna, tuner, modulator, and demodulator, to receive radio broadcast [195]. The operation principles of nanotube radio, i.e., CNT radio or CNT receiver, are extremely different from traditional radios since RF signal reception, tuning, amplification, and demodulation are electromechanical processes rather than completely electrical. If an incoming radio wave induces on the nanotube, it causes physical vibrations on the charged tip of nanotube as shown in Fig. 13.1. When the frequency of incoming wave matches with the resonance frequency of nanotube, the vibrations tune to the incoming wave. Hence, this electromechanical process allows the nanotube to receive the incoming signal.

Resonance frequency of nanotube, i.e., f , is affected by the length of the nanotube, i.e., L , such that $f \propto 1/L^2$ can be given [195]. To receive a specific radio broadcast, resonance frequency of nanotube must be initially set by regulating the length of the nanotube. However, the length of the nanotube may be exposed to a possible degradation due to the field emission current that traverses in the nanotube. The degradation of the nanotube may also randomly change its resonance frequency and the frequency tuning to carrier frequency of incoming wave can be no longer possible.

For the amplification and modulation of the incoming signal, DC bias voltage between the

electrodes is used as shown in Fig. 13.1. The field-emission current generated by the bias voltage is demodulated via the mechanical vibrations. However, degradation of the nanotube would be unavoidable for relatively excessive bias voltage levels, and the frequency tuning to the incoming signal would be no longer possible. Hence, the amplification of the incoming signal and the degradation of the nanotube must be jointly considered to efficiently receive the incoming signal. This is one of the most challenging barriers toward the realization of CANET.

In addition to the electromechanical vibrations, thermal vibrations may also be imposed on the nanotube tip. When the nanotube tip experiences some thermal vibrations, amplitude of the minimum detectable electric field, i.e., E , can be written as [195]

$$E = \frac{1}{q} \sqrt{\frac{4k_B T m f_0 B}{Q}} \quad (13.1)$$

where q is the charge on the nanotube tip, m is the effective mass of the nanotube, Q is the quality factor, B is the achieved bandwidth, k_B is the Boltzmann constant, f_0 is the fundamental vibration frequency of carbon nanotube, and T is the temperature [195]. Clearly, (13.1) may be used to derive a signal detection mechanism for nanoscale communication in CANET. Similarly, it may be possible to extract some analytical and theoretical issues from the field of physics and chemistry to derive theoretical models for nanoscale communication.

13.2.1.2 Tunneling Nanotube Radio

Beside its efficacy as a nanoscale receiver circuitry, the present form of CNT radio also has severe restrictions outlined as follows [202].

- It needs 200V of power supply to operate.
- It can only tune to 4 MHz which is 5 times lower than the FM bandwidth of 88-108 MHz.

Tunneling nanotube radio is proposed to address above challenges [202]. To reduce the required power level, a tunneling detection mechanism is proposed instead of field emission obtained from the vibrating charged tip of CNT as in [195]. Based on quantum tunneling,

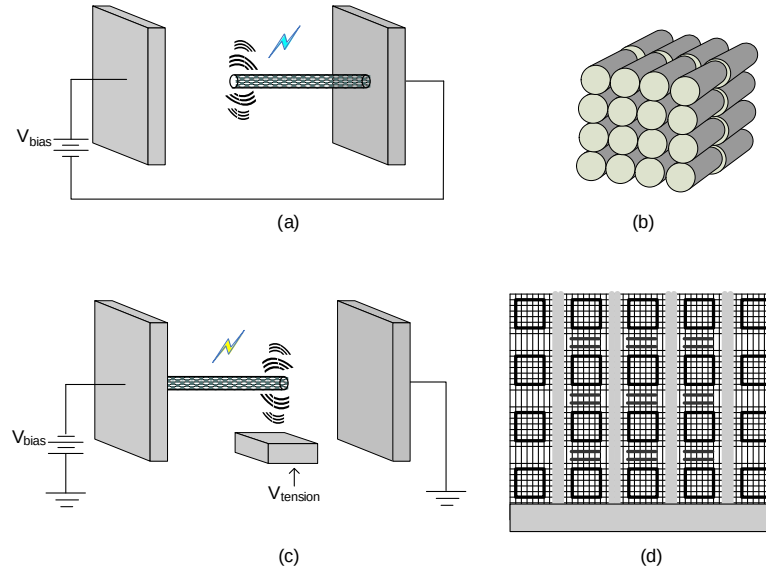


Figure 13.2: Illustration of integrated hardware required for nanoscale communication in CANET, including nanotransceiver, nanopower, nanoprocessor and nanomemory units.

bias voltage of the nanotube radio is decreased on the order of 10V. Moreover, based on the electrostatic actuation of the nanotube with another gate electrode, nanotube radio can be tuned using the entire FM and AM bands. Tunneling nanotube radio is promising to provide nanoscale communication with relatively low power level and more extensive spectrum range.

13.2.1.3 Nanotransmitter

Electromechanical vibrations of nanotubes can also be harnessed to design a nanoscale transmitter circuitry. In [196], using a CNT, an electromechanical transmitter circuitry is devised. The four fundamental components of a transmitter circuitry, i.e., modulator, oscillator, antenna, and amplifier, are implemented using a CNT. Frequency modulation (FM) in the nanotube transmitter is realized by modulating mechanical resonance frequency of the CNT. This mechanical modulation can be performed by an external electrode fed by another power source ($V_{Tension}$) as shown in Fig. 13.2(c). Information signal can be applied to this electrode for the modulation. The nanotube also acts as an antenna to allow the radiation of the modulated information signal. Finally, the power of the radiated information signal can be regulated by changing the oscillation amplitude of the nanotube, or increasing the charge in the nanotube, or using an array of nanotubes [196]. Similar to the nanoreceiver, nanotransmitter is also subject to the degradation of nanotube and considerably low power source of

Table 13.1: Estimated circuit, antenna, and system size (m^3) for various radios complied from the literatures [203]

Radio	Circuit size	Antenna size	System size
<i>Hitachi</i>	1E-14	1E-08	1E-08
<i>France-Telecom</i>	1E-09	1E-09	1E-09
<i>Smart Dust</i>	3.125E-09	1E-06	1E-06
<i>Single-chip radio</i>	-	-	1E-14
<i>Nanotube Radio</i>	-	-	1E-21
<i>Volume of living cell</i>	-	-	1E-18

nanopower unit. Hence, the degradation of nanotube and the amplification of the radiated power must be jointly considered to efficiently amplify and transmit the information signal.

13.2.2 Carbon Nanotube Antenna

Due to their promising electromagnetic characteristics, CNT-based antennas are favorable in different frequency regimes ranging from microwave to visible. One of the important parameters of a CNT is its quantum resistance that is much smaller than the normal metal wire with the same geometry in nanometer scale. The skin effect in CNT can also be ignored when the operating frequency reaches terahertz. Therefore, the power dissipation of a CNT antenna is low and this also leads to high antenna efficiency with respect to a metal wire with the same size [204]. However, theoretical estimates in [205] and [193] show that due to the estimated -90 dB of losses imposed by ohmic currents, the radiation efficiency of a single-walled CNT antenna, i.e., $\eta = P_r / (P_r + P_t)$, is very low, i.e., on the order of 10^{-4} and 10^{-5} , where P_r is the radiated power and P_t is the power of the thermal losses. Thus, CNT bundles are used to provide higher efficiency [204].

When a CNT antenna carries several microamperes of current under an applied voltage of a few volts, it emits the maximum of a few microwatts of power to its surrounding environment. As the required communication range increases, it may be possible to amplify the transmission power using a number of nanotubes in parallel. In addition to amplifying the power, an antenna array also enables directionality properties of transmission, that leads to less power consumption when point-to-point communication is needed [206].

CNT antennas may be incorporated into available micro or nanoradio circuitries to provide great efficiency in the size of the overall system volume. In Table 13.1, some radio systems in the literature are compared with respect to their circuit, antenna, and entire radio system size. The aim of this comparison is to illustrate the relative importance of antenna volume in total system size. The comparison clearly reveals that the small circuit size is possible, however having small antenna size is more challenging [203]. Hence, due to the availability of CNT antenna that is on the order of 100 times smaller than a classical dipole antenna, CNT antennas are clearly promising alternative for nanoscale communication requirements in frontier nanotechnology applications of CANET. In Table 13.1, the system sizes of the given radios are also compared with the volume of a single living cell. This also reveals that carbon nanotube radio and tunneling nanotube radio [195], [202] with $1\text{E-}21\text{m}^3$ of average system volume are feasible alternatives for nanoscale communication in envisioned biomedical nanotechnology applications in a living cell [203].

13.2.3 Nanobattery

Power supply at nanoscale is one of the most critical challenges to realize future CANET nodes. In [207], nickel-zinc batteries having 0.02cm^2 footprint and 0.555 mWh/cm^2 of energy supply are manufactured and their open circuit voltage ranges from 1.7V to 1.8V. In [208], micro-battery arrays with nanoscale anodes and cathodes are introduced using commercially available nanomaterials with extremely small diameter size on the order of nanometers. The capacity of these battery arrays ranges from 3 mAh/g to 18 mAh/g with $1.5 - 2\text{V}$ nominal voltage. Consequently, current technologies may support nanoscale power source for the realization of nanoscale communication in CANET. However, overall power budget analysis including the energy consumption in transceiver, memory, and processor units must be performed. This analysis can be critical to assess whether the existing nanobattery technologies allow both point-to-point and multi-hop nanoscale communication or not.

13.2.4 Nanomemory and Nanoprocessor

Memory unit is one of the most critical hardware components to provide an efficient store-and-forward mechanism in CANET. In [209], nanowire crossbar circuit is used to design a nanoscale memory that operates with $0.5 - 3.5\text{V}$. By switching and setting the resistance of

the crossbars, each cross point is used as an active memory cell. Crossbar circuits are also configured as multiplexer and demultiplexer circuits. An 8×8 crossbar circuit is inserted into an area of $1\mu\text{m}^2$ with a density of $64\text{ Gbits}/\text{cm}^2$. In [210], another addressable nanomemory is designed using aligned carbon nanotubes with cross geometry. Some nanoprocessor designs can also be found in the current literature. In [211], using semiconductor nanowire blocks, functional processor components are constructed. Nanowire junction arrays are configured to build OR, AND, and NOR logic-gates and to enable simple computations. In [212], sequential nanomemory and processor with clocked operations are devised.

Consequently, nanomemory, nanoprocessor, and nanopower units are feasible to enable required nanoscale communication functionalities in CANET. However, their extremely limited power, computation, and storing capabilities also severely restrict nanoscale communication in terms of transmission power, communication range, data storage, and computational complexity. This also radically shifts the conventional network functionalities to extremely resource-constrained and uncontrolled domain. In Fig. 13.2, nanopower, nanomemory, and nanoprocessor units are illustrated.

13.3 Fundamental Communication Issues in CANET

Here, we discuss the fundamental research challenges and open issues of nanoscale communication in CANET.

13.3.1 Operating Frequency Band

In nanoscale communication, it may be possible to use a frequency band selected from a large range of electromagnetic spectrum, i.e., from megahertz (MHz) to terahertz (THz). The MHz frequency range may be used with CNT receiver and transmitter, whereas THz ranges require CNT antennas. Clearly, if the operating frequency is selected from the megahertz range, it may be feasible to provide higher communication range than the case with CNT antenna. However, as introduced in Section 13.2, CNT receiver and transmitter need on the order of 100V of power source. Therefore, CNT antennas with an applied voltage of a few volts may be more feasible alternative for nanoscale communication. In fact, tunneling nanotube radio reduces the power need of CNT receiver and transmitter circuitry to on the order of 10V and

rapidly developing nanotechnology might be expected to further decrease this power need, which may make CNT receiver and transmitter more feasible.

13.3.2 Path Loss and Noise

Due to extremely low transmission power level in nanoscale communication, signal propagation is severely prone to spreading and molecular absorption loss that constitute the path loss term. The spreading loss stems from the expansion of the electromagnetic wave as it propagates through the medium. The molecular absorption loss is an environmental phenomenon by which a part of the wave energy is absorbed by some of the molecules in the medium [213]. Each molecule type has different absorption characteristics in frequency and amplitude. This imposes an ambiguity on the received signal and may severely reduce the achievable communication rate in nanoscale communication [197].

The molecules also disperse electromagnetic energy that is captured from the medium [213], which incurs an ambient noise. Each molecule type in the medium has different characteristics of electromagnetic energy dispersion. This also imposes a critical irregularity on the characteristics of noise power whose statistical properties may completely differ from traditional wireless communication. Hence, in order to comprehensively analyze nanoscale communication channel, the electromagnetic energy absorption and dispersion characteristics of molecules found in the medium should be carefully taken into account.

Nanoscale communication is also prone to some environmental factors such as thermal noise and interference from other wireless systems. If CNT receiver and transmitter that operate in MHz range are used, it may be more likely to be exposed higher interference level from other wireless systems using the MHz range. However, CNT antennas that operate in THz range that is still an unlicensed band might be a better choice to avoid high interference level.

13.3.3 Communication Rate and Range

In nanoscale communication, communication rate and range are function of molecular composition in the medium, upper bound of transmission power, and operating frequency. Among them, the molecular composition of the medium may vary according to molecular mixture of the medium. Thus, achievable communication rate and range also differ from a medium to an-

other. For a medium with high level of molecular absorption loss, achievable communication rate and range may be improved by increasing upper bound of the transmission power.

Due to the lack of the precise analytical models to determine the upper bound of transmission power provided by CNT transmitter and antennas, it may not be possible to estimate achievable communication rate and range even if some environmental assumptions on molecular composition of the medium are made. Actually, the analysis of them is an important open research issue that must be addressed. However, recently, in terahertz nanoscale communication channel, communication range from $10\mu m$ to $10m$ and $10^{10} - 10^{14}$ (*bits/s*) of communication rate are shown to be possible under a set of certain assumptions such as transmission power [197].

13.3.4 Analog vs. Digital Communication

In CNT receiver and transmitter circuitry, transmission and reception of information signal are purely analog. The carrier frequency of information signal is absolutely specific to the length of the nanotube. By using CNT receiver and transmitter circuitry, nanonodes that are in close proximity of each other can communicate if they tune in the same frequency. Therefore, half-duplex communication is only feasible and available communication bandwidth is inefficient and static for CNT receiver and transmitter. This static frequency tuning causes also considerably high interference among nanonodes. However, CNT antennas that are integrated with a sufficiently small radio system can allow nanonodes to use more than one communication channels, and thus, provide full-duplex digital nanoscale communication [206]. This can also provide higher communication bandwidth.

13.3.5 Encoding of Information

None of the available encoding mechanisms in digital communication may be feasible if CNT receiver and transmitter are used in CANET nodes. Coded transmission of information may be possible if CNT antenna is integrated with a sufficiently small radio system. However, due to considerably low computational power in nanoprocessor, available encoding mechanisms may not be practical for nanoscale communication in CANET. Similarly, none of the digital error detection and correction mechanisms may be applicable. Instead, novel encoding and

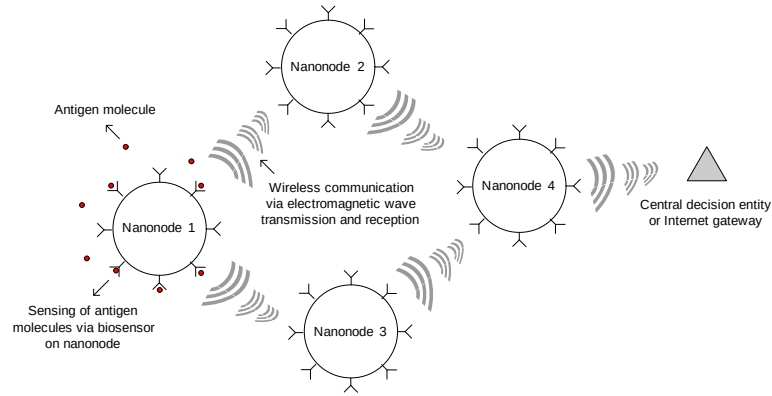


Figure 13.3: Illustration of a CANET application example, i.e., nanoscale bio-sensor network.

error detection and correction mechanisms with considerably low computational complexity must be developed.

13.4 Fundamental Networking Issues in CANET

Here, we consider an application example of CANET, i.e., nanoscale bio-sensor network (NBSN) to detect hazardous antigen concentration as an immune system support, and then highlight the corresponding networking challenges and open issues to be addressed. In addition to the hardware components given in Section 13.2, we assume that nanonodes in NBSN have also a sensing unit [214] to detect the presence of antigen molecules as shown in Fig. 13.3.

In NBSN, we assume that nanonodes are deployed to detect a concentration of antigens and to forward the concentration information toward a central decision entity to make decision such as the dosage of the taken drug or type of the detected antigen to fire an action potential. During the network operation of CANET, there might be three operational phases for nanonodes as follows:

- If a nanonode detects a concentration of antigens, it measures, encodes, and transmits the concentration information to its surrounding environment. This phase is called as *transmitter phase* of nanonodes. In Fig. 13.3, nanonode 1 is in the transmitter phase.
- If a nanonode receives a communication signal from another nanonode, it is in *forwarder phase* and it amplifies and forwards the received transmission. Nanonode 2, 3,

and 4 are in the forwarder phases as seen in Fig. 13.3.

- If a nanonode neither has information to transmit nor receives a communication signal from its surrounding environment, it is in *idle phase*.

Clearly, all of the communication challenges and open issues previously outlined must be addressed as the prerequisite for the deployment of CANET toward the realization of this sample NBSN application. Furthermore, the realization of NBSN necessitates intelligent control of these three phases in terms of energy consumption, reliability, and robustness. This mandates the development of novel communication networking techniques for CANET, which addresses the challenges outlined next.

13.4.1 Deployment and Topology

Because of severely energy constrained nanobatteries, transmission power generated by nanotransmitter is significantly low and therefore, nanonodes have very limited radio range. This necessitates the dense deployment of nanonodes for a network-wide connectivity with smaller energy consumption. Due to high manufacturing defect rates and operational uncertainties in CNT, CANET must also have a resilient topology to maintain normal network operations despite some node failures or mobility.

13.4.2 Multi-hop Reliability

Upon the transmission of information signal, some additional functionalities and challenges should be also overcome to route the signal among CANET nodes. A possible routing scheme may be similar to flooding-based routing mechanisms if CNT receiver and transmitter are used in nanotransceiver circuitry since they naively generate analog signal that radiates along the network. However, CNT antennas provide the capability of directivity for nanonode transmission [206]. This may enable a spatial routing scheme in which each nanonode can forward its transmission toward the direction of its destination.

A simple store-and-forward mechanism may be attainable for CANET nodes. However, power consumption required by data storage and computation may seriously reduce the network lifetime. Therefore, an efficient store-and-forward scheme should also take into account

the power consumption in data storage and computation.

13.4.3 Medium Access Control

Due to uncoded analog data transmission, code-division multiple access (CDMA) is not feasible for CANET with CNT receiver and transmitter. Similarly, due to their high computation complexity requirement, CDMA mechanism may not also be appropriate for the configuration with CNT antenna. Instead, a considerably simple form of frequency-division multiple access (FDMA)-based or time-division multiple access (TDMA)-based approaches might be suitable to efficiently access the communication medium. Using more than one CNT receiver and transmitter in each nanonode and each of which has different CNT resonance frequency, nanonodes might use different frequency channels that do not interfere with each other. This might provide a simple form of FDMA mechanism in which each nanonode selects a frequency to transmit and some of its neighbors receive its transmission if there is no other simultaneous transmission on the same frequency channel. The realization of such an access mechanism clearly entails the efficient selection of transmission frequencies. However, it might be feasible to initially select and set transmission frequency of all nanonodes to provide the minimum level of interference among nanonodes.

Moreover, a TDMA-based mechanism might be also possible as long as nanoprocessor provides the synchronized operations [212]. However, these clocked and synchronized operations are subject to some synchronization errors due to low computational power of nanoprocessor. For such kind of asynchronous channel access case, it might be also feasible to use pulse-width of the nanonode transmissions to provide an efficient collision detection mechanism. Let assume that every nanonode located in the same collision domain uses a different pulse-width that is integer multiple of a predefined minimum pulse-width. If the transmissions of more than one nanonodes collide and the pulse-width of these transmissions overlap then, it is highly likely that pulse-width of the received signal is not an integer multiple of the predefined minimum pulse-width and this pulse-width cannot be identified at the receiver side. This mechanism may clearly lead to a collision detection mechanism.

13.5 Conclusion

In this chapter, we introduce the concept of Carbon nanotube-based nanoscale Ad hoc NETWORKS (CANET). By investigating some existing nanotransceiver, power, memory, and processor circuitries, we first explore some hardware components required for nanoscale communication. After discussing the suitability of these hardware components, we present fundamental communication and networking issues and research challenges in CANET. Our investigations reveal that CANET significantly shift the traditional wireless communication and networking paradigms to an unobservable, uncontrollable, and extremely challenged domain.

BIBLIOGRAPHY

- [1] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks Journal (Elsevier)*, vol. 50, no. 13, pp. 2127-2159, September 2006.
- [2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201-220, 2005.
- [3] B. Atakan, O. B. Akan, "Biologically-inspired Dynamic Spectrum Access in Cognitive Radio Networks," *Bio-inspired Computing and Communication Networks*, Edited by Y. Xiao and F. Hu, Auerbach Publications, CRC Press, 2008.
- [4] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor Networks: A Survey," *Computer Networks (Elsevier)* vol. 38, pp. 393-422, 2002.
- [5] B. Atakan, T. Tugcu, O. B. Akan, "Bio-inspired Communications in Wireless Sensor Networks," in *Handbook of Wireless Sensor and Ad Hoc Networks*, Edited by S. Misra, I. Woungang, and S. C. Misra, Springer (London), 2008.
- [6] E. Drexler, "Engines of creation: The coming era of nanotechnology", *Doubleday*, New York, 1986.
- [7] V. Balzani, A. Credi and M. Venturi, "Molecular devices and machines", *Nanotoday*, vol. 2, no. 2, pp. 18-25, April 2007.
- [8] G. A. Ozin, I. Manners, S. Fournier-Bidoz, A. Arsenault, "Dream Nanomachines" *Advanced Materials*, vol. 17, no. 24, pp. 3011-3018, December 2005.
- [9] G. M. Whitesides, "The Once and Future Nanomachine", *Scientific American*, September, 2001.
- [10] B. Bhushan, "Biomimetics: lessons from nature-an overview", *Philosophical Trans. of The Royal Soc.*, vol. 367, no. 1893, pp. 1445-1486, April 2009.
- [11] R. K. Soong, D. Bachand, H. P. Neves, A. G. Olkhovets, H.G. Craighead, C. D. Montemagno, "Powering an inorganic nanodevice with a biomolecular motor", *Science*, vol. 290, no. 5496, pp. 1555-1558, 2000.
- [12] I. F. Akyildiz, F. Brunetti, C. Blazquez, "Nanonetworks: A New Communication Paradigm," *Computer Networks*, vol. 52, no. 12, August 2008, pp. 2260-2279.
- [13] R. A. Freitas, "Nanomedicine, Volume I: Basic Capabilities", *Landes Bioscience*, Texas, USA, 1999.
- [14] C. M. Waters, B. L. Bassler, "Quorum Sensing: Cell-to-Cell Communication in Bacteria", *Annual Review of Cell and Developmental Biology*, vol. 21, pp. 319-346, Nov. 2005.

- [15] T. A. Springer, "Adhesion receptors of the immune system", *Nature*, vol. 346, pp. 425-434, August 1990.
- [16] M-T. Chen, R. Weiss, "Artificial cell-cell communication in yeast *Saccharomyces cerevisiae* using signaling elements from *Arabidopsis thaliana*", *Nature Biotechnology*, vol. 23, pp. 1551-1555, 2005.
- [17] W-D. Wang, Z-T. Chen, B-G. Kang, R. Li, "Construction of an artificial inter-cellular communication network using the nitric oxide signaling elements in mammalian cells," *Experimental Cell Research*, vol. 314, no. 4, pp. 699-706, Feb. 2008.
- [18] V. Brik, E. Rozner, S. Banarjee, P. Bahl, "DSAP: a protocol for coordinated spectrum access," in Proc. *IEEE DySPAN 2005*, pp. 611-614, November 2005.
- [19] M.M. Buddhikot, P. Kolody, S. Miller, K. Ryan, J. Evans, "DIMSUMNet: new directions in wireless networking using coordinated dynamic spectrum access," in Proc. *IEEE WoWMoM 2005*, pp. 78-85, June 2005.
- [20] J. Zhao, H. Zheng, G.-H. Yang, "Distributed coordination in dynamic spectrum allocation networks," in Proc. *IEEE DySPAN 2005*, pp.259-268, November 2005.
- [21] J. Huang, R.A. Berry, M.L. Honig, "Spectrum sharing with distributed interference compensation," in Proc. *IEEE DySPAN 2005*, pp.88-93, November 2005.
- [22] I. F. Akyildiz, W. Y. Lee, K. R. Chowdhury, "CRAHNS: Cognitive radio ad hoc networks," *Ad Hoc Journal (Elsevier)*, vol. 7, no. 5, pp. 810-836, July 2009.
- [23] Qing Zhao, B. M. Sadler, "A Survey of Dynamic Spectrum Access," *IEEE Sig. Proc. Mag.*, vol. 24, no. 3, pp. 79-89, May 2007.
- [24] L. Ma, X. Han, C. C. Shen, "Dynamic open spectrum sharing MAC protocol for wireless ad hoc networks," in Proc. *IEEE DySPAN'05*, Baltimore, USA, Nov. 2005.
- [25] S. Sankaranarayanan, P. Papadimitratos, "A bandwidth sharing approach to improve licensed spectrum utilization," *IEEE Comm. Mag.*, vol. 43, no. 12, pp. suppl.10-suppl.14, 2005.
- [26] C. Cordeiro, K. Challapali, "C-MAC: A Cognitive MAC Protocol for Multi-Channel Wireless Networks," in Proc.*IEEE DySPAN*, Dublin, April, 2007.
- [27] J. Jia, Q. Zhang, X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Area Comm.*, vol. 26, no. 1, pp. 106-117, January 2008.
- [28] O. B. Akan and I. F. Akyildiz, "Event-to-Sink Reliable Transport in Wireless Sensor Networks," *IEEE/ACM Transactions on Networking*, vol. 13, pp. 1003-1016, 2005.
- [29] C. Wan, S. B. Eisenman, and A. T. Campbell, "CODA: Congestion Detection and Avoidance in Sensor Networks," in Proc. *Proc. ACM Sensys 2003*, Los Angeles, USA, November 2003.
- [30] F. Stann and J. Heidemann, "RMST: Reliable Data Transport in Sensor Networks," in Proc. *ACM MobiHoc 2004*, Roppongi, Japan, May 2004.

- [31] S. Pudlewski, T. Melodia, "A Distortion-minimizing Rate Controller for Wireless Multimedia Sensor Networks," *Computer Communications (Elsevier)*, vol. 33, no. 12, pp. 1380-1390, 2010.
- [32] C. Y. Wan, A. T. Campbell, L. Krishnamurthy, "PSFQ: a reliable transport protocol for wireless sensor networks," in Proc. *ACM WSNA 2002*, pp. 1-11, Atlanta, Georgia, USA, 2002.
- [33] T. Melodia, D. Pompili, C. V. Gungor, I. F. Akyildiz, "A Distributed Coordination Framework for Wireless Sensor and Actor Networks," in Proc. *ACM, MOBIHOC 2005*, pp. 99-110, Urbana-Champaign, IL, USA, 2005.
- [34] V. C. Gungor, O. B. Akan, and I. F. Akyildiz, "A Real-Time and Reliable Transport Protocol for Wireless Sensor and Actor Networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 2, pp. 359-370, April 2008.
- [35] K. R. Chowdhury, M. Di Felice, "Search: a routing protocol for mobile cognitive radio ad hoc networks," *Computer Communication Journal*, vol. 32, no. 18, pp. 1983-1997, Dec. 2009.
- [36] K. R. Chowdhury, I. F. Akyildiz, "CRP: A Routing Protocol for Cognitive Radio Ad Hoc Networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 794-804, April 2011.
- [37] Q. Wang, H. Zheng, "Route and spectrum selection in dynamic spectrum networks," in Proc. *IEEE CCNC 2006*, pp. 625-629, Feb. 2006.
- [38] Q. Wang, H. Zheng, "Efficient Routing Algorithms for Multi-Channel Dynamic Spectrum Access Networks," in Proc. *IEEE DySPAN 2007*, pp. 288-291, April 2007.
- [39] S. Hiyama, Y. Moritani, T. Suda, R. Egashira, A. Enomoto, M. Moore and T. Nakano, "Molecular Communication", in Proc. *NSTI Nanotech 2005*, Anaheim, California, USA.
- [40] M. Pierobon, I. F. Akyildiz, "A Physical Channel Model for Molecular Communication in Nanonetworks", *IEEE J. Selected Areas Commun. (JSAC)*, vol. 28, pp. 602-611, May 2010.
- [41] K. V. Srinivas, R. S. Adve, A. W. Eckford, "Molecular communication in fluid media: The additive inverse Gaussian noise channel", submitted to *IEEE Transactions on Information Theory*, Dec. 2010.
- [42] B. Atakan, O. B. Akan, "Immune System Based Distributed Node and Rate Selection in Wireless Sensor Networks," in Proc. *IEEE/ACM BIONETICS 2006*, Cavalese, Italy, December, 2006.
- [43] J. Kusuma, L. Doherty, K. Ramchandran, "Distributed compression for sensor networks," in Proc. *IEEE Image Processing* vol. 1, pp.82-85, 2001.
- [44] H. Gupta, V. Navda, S. R. Das, V. Chowdhary, "Efficient Gathering of Correlated Data in Sensor Networks," in Proc. *ACM MOBIHOC*, pp. 402-413, Urbana-Champaign, Illinois, May 2005.
- [45] P. V. Rickenbach, R. Wattenhofer, "Gathering Correlated Data in Sensor Networks," in Proc. *ACM DIALMPOMC'04*, Philadelphia, Pennsylvania, USA, October 2004.

- [46] R. Cristescu, B. Beferull-Lozano, M. Vetterli, "On network correlated data gathering," in Proc. *IEEE INFOCOM 2004*, pp. 2571-2582, March 2004.
- [47] R. Cristescu and M. Vetterli, "On the optimal density for real-time data gathering of spatio-temporal processes in sensor networks," in Proc. *IPSN 2005*, April 2005.
- [48] D. Ganesan, R. Cristescu, B. Beferull-Lozano, "Power-efficient sensor placement and transmission structure for data gathering under distortion constraints, " in Proc. *IPSN 2004*, pp. 142-150, April 2004.
- [49] M. C. Vuran, O. B. Akan, "Spatio-temporal Characteristics of Point and Field Sources in Wireless Sensor Networks," in. Proc. *IEEE ICC 2006*, Istanbul, June 2006.
- [50] M. C. Vuran, I. F. Akyildiz, "Spatial Correlation-based Collaborative Medium Access Control in Wireless Sensor Networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 2, pp. 316-329, June 2006.
- [51] S. S. Pradhan, K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Transactions on Information Theory*, vol. 49, pp. 626-643, March 2003.
- [52] C. Guestrin, P. Bodi, R. Thibau, M. Paski, S. Madde, "Distributed regression: an efficient framework for modeling sensor network data," in. Proc. *IPSN'04*, pp. 26-27, April 2004.
- [53] J. Timmis, M. Neal, J. Hunt, "An artificial immune system for data analysis," *BIOSYSTEMS, Elsevier*, vol. 55, no. 1-3, pp. 143-150, 2000.
- [54] J. H. Jun, D. W. Lee, K. B. Sim, "Realization of Cooperative Strategies and Swarm Behavior in Distributed Autonomous Robotic Systems using Artificial Immune System," in. Proc. *IEEE SMC'99*, pp. 614-619, 1999.
- [55] J. Timmis, M. Neal, J. Hunt, "Data analysis using artificial immune systems, cluster analysis and Kohonen networks: some comparisons" in. Proc. *IEEE Systems Man and Cybernetics*, 1999.
- [56] S. Forrest, A. S. Perelson, L. Allen, R. Cherukuri, "Self-Nonself Discrimination in a Computer," in Proc. *IEEE Symposium on Security and Privacy*, pp. 202-212, 1994.
- [57] J. D. Farmer, N. H. Packard, A. S. Perelson, "The immune system, Adaptation, and Machine Learning," *Physica D: Nonlinear Phenomena*, vol. 22, no. 1-3, pp. 187-204, 1986.
- [58] N.K. Jerne, "Idiotypic Network and Other Preconceived Ideas," *Immunological Rev*, vol. 79, pp. 5-24, 1984.
- [59] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-Temporal Correlation: Theory and Applications for Wireless Sensor Networks," *Computer Networks Journal (Elsevier)*, vol. 45, no. 3, pp. 245-261, June 2004.
- [60] M. Gastpar, M. Vetterli, "Source-Channel Communication in Sensor Networks," in Proc. *IPSN'03*, Palo Alto, USA, April 2003.
- [61] J. Hightower, G. Borriello, "Location systems for ubiquitous computing," *IEEE Computer*, vol. 34, no. 8, pp. 57-66, Aug. 2001.

- [62] J.O. Berger, V. de Oliveira, B. Sanso, "Objective bayesian analysis of spatially correlated data," *J. Am. Statist. Assoc.*, vol. 96, pp.1361-1374, 2001.
- [63] S. Santini, K. Römer, "An Adaptive Strategy for Quality-Based Data Reduction in Wireless Sensor Networks," in Proc. *3rd International Conference on Networked Sensing Systems (INSS'06)*, 2006.
- [64] B. Atakan, O. B. Akan, "BIologically-inspired Spectrum Sharing in Cognitive Radio Networks," in Proc. *IEEE WCNC 2007*, pp. 43-48, Hong Kong, March 2007.
- [65] Q. Zhao, L. Tong, A. Swami, "Decentralized cognitive MAC for dynamic spectrum access," in Proc. *IEEE DySPAN 2005*, pp.224-232, November 2005.
- [66] H. Zheng, L. Cao, "Device-centric spectrum management," in Proc. *IEEE DySPAN 2005*, pp.56-65, November 2005.
- [67] E. Bonabeau, M. Dorigo, G. Theraulaz, "Swarm Intelligence, From Natural to Artificial System," *Oxford University Press*, 1999.
- [68] B. Wild, K. Ramchandran, "Detecting primary receivers for cognitive radio applications," in Proc. *IEEE DySPAN 2005*, pp. 124-130, November 2005.
- [69] B. Atakan, B. Gulbahar, O. B. Akan, "Immune System-inspired Evolutionary Opportunistic Spectrum Access in Cognitive Radio Ad Hoc Networks," in Proc. *IFIP/IEEE MED-HOC-NET 2010*, Juan-les-pins, France, June 2010.
- [70] J. Kim, P. Bentley, "Towards an artificial immune system for network intrusion detection: An investigation of clonal selection with negative selection operator," in Proc. *IEEE Evolutionary Computation*, pp.1244-1252, Seoul, Korea, May 2001.
- [71] L. N. Castro, F. J. Von Zuben, "Learning and optimization using the clonal selection principle," *IEEE T. Evol. Comput.*, vol. 6, no. 3, pp. 239-251, 2002.
- [72] N. C. Cortes, C. A. C. Coello, "Multiobjective Optimization Using Ideas from the Clonal Selection Principle," *LNCS*, vol. 2723, pp.200, 2003.
- [73] D. Bradley, A. M. Tyrrell, "Hardware fault tolerance: An immunological solution," in Proc. *IEEE Syst. Man Cybern.*, pp.107-112, Oct. 2000.
- [74] Q. Zhao, L. Tong, A. Swami, Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Area Comm.*, vol. 25, no. 3, April 2008.
- [75] L. Ma, C. C. Shen, B. Ryu, "Single-radio adaptive channel algorithm for spectrum agile wireless ad hoc networks," in Proc. *IEEE DySPAN'07*, Dublin, April, 2007.
- [76] T. Renk, C. Kloeck, D. Burgkhardt, F. K. Jondral, D. Grandblaise, S. Gault, J. C. Dunat, "Bio-Inspired Algorithms for Dynamic Resource Allocation in Cognitive Wireless Networks," *Mobile Networks and Applications*, vol. 13, no. 5, pp. 431-441, October 2008.
- [77] Y. W. Lee, I. F. Akyildiz, "Optimal Spectrum Sensing Framework for Cognitive Radio Networks," *IEEE Trans. Wirel. Commun.*, vol. 7, no. 10, pp. 3845-3875, October 2008.
- [78] V. Cutello, M. S. Alouini, M. K. Simon, P. S. Oliveto, "On the Convergence of Immune Algorithms," in Proc. *IEEE FOCI 2007*, pp. 409-415, April 2007.

- [79] A. C. C. Hsu, D. S. L. Wei, C.-C. J. Kuo, "A cognitive MAC protocol using statistical channel allocation for wireless ad-hoc networks," in Proc. *IEEE WCNC*, pp. 105-110, March 2007.
- [80] X. Liu, Z. Ding, "ESCAPE: A channel evacuation protocol for spectrum-agile networks," in Proc. *IEEE DySPAN*, Dublin, Ireland, April 2007.
- [81] J. Zhao, H. Zheng, G. H. Yang, "Distributed coordination in dynamic spectrum allocation networks," in Proc. *IEEE DySPAN*, USA, Nov. 2005.
- [82] G. Auer, H. Haas, P. Omiyi, "Interference aware medium access for dynamic spectrum sharing," in Proc. *IEEE DySPAN*, Dublin, April 2007.
- [83] B. Hamdaoui, K. G. Shin, "OS-MAC: An efficient MAC protocol for spectrum agile wireless networks," *IEEE Trans. Mob. Comput.*, vol. 7, no. 8, pp.915-930, August 2008.
- [84] P. Pawelczak, S. Pollin, H. W. So, A. Bahai, R. V. Prasad, R. Hetmat, "Performance Analysis of Multichannel Medium Access Control Algorithms for Opportunistic Spectrum Access," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, 2009.
- [85] B. Atakan, O. B. Akan, "Distributed Audio Sensing with Homeostasis-inspired Autonomous Communication," *Ad Hoc Networks Journal (Elsevier)*, vol. 9, no. 4, pp 552-564, June 2011.
- [86] I. F. Akyildiz, T. Melodia, K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol.51, no.4, pp. 921-960, 2007.
- [87] E. Gurses, O. B. Akan, "Multimedia Communication in Wireless Sensor Networks," *Annals of Telecommunications*, vol. 60, no. 7-8, pp. 799-827, July-August 2005.
- [88] S. Misra, M. Reisslein, G. Xue, "A Survey of Multimedia Streaming in Wireless Sensor Networks," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, pp. 18-39, 2008.
- [89] M. Neal, J. Timmis, "Once more unto the breach towards artificial homeostasis," in *Recent Advances in Biologically Inspired Computing*, Edited by L. N. De Castro, F. J. Von Zuben, pp. 340-365, 2005.
- [90] A. V. Oppenheim, R. W. Schaffer, J. R. Buck, "Discrete-Time Signal Processing" *Prentice Hall*, 1999.
- [91] B. Atakan, O. B. Akan, "On Event Signal Reconstruction in Wireless Sensor Networks," in Proc. *IFIP/TC6 NETWORKING 2007*, Atlanta, GA, May 2007.
- [92] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short modified periodogram", *IEEE Transaction on Audio and Electroacoustics*, vol. AU-15, no. 2, 1967.
- [93] M. Zuniga, B. Krishnamachari, "Analyzing the transitional region in low power wireless links", in Proc. *IEEE SECON 2004*, pp. 517-526, October 2004.
- [94] B. Atakan, O. B. Akan, "Bio-inspired Cross-layer Communication and Coordination in Sensor and Vehicular Actor Networks," *IEEE Transactions on Vehicular Technology*, under revision, 2011.

- [95] I. F. Akyildiz, I. H. Kasimoglu, "Wireless sensor and actor networks: research challenges," *Ad Hoc Networks (Elsevier)*, vol. 2, no. 4, pp. 351-367, 2004.
- [96] R. Madan, S. Cui, S. Lall, A. Goldsmith, "Cross-Layer Design for Lifetime Maximization in Interference-Limited Wireless Sensor Networks," *IEEE Transactions on Wireless Comm.*, vol. 5, no. 11, pp. 1964-1975, 2006.
- [97] E. C.-H. Ngai, J. Liu, M. R. Lyu, "An Adaptive Delay-Minimized Route Design for Wireless Sensor-Actor Networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 5083-5094, Nov. 2009.
- [98] T. Melodia, D. Pompili, C. V. Gungor, I. F. Akyildiz, "Handling Mobility in Wireless Sensor and Actor Networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 2, pp. 160-173, Feb. 2009.
- [99] T. H. Labella, F. Dressler, "A Bio-Inspired Architecture for Division of Labour in SANETs," *Advances in Biologically Inspired Information Systems*, vol. 69, pp. 211-230, Springer, 2007.
- [100] A. A. Abbasi, M. Younis, K. Akkaya, "Movement-Assisted Connectivity Restoration in Wireless Sensor and Actor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 9, pp. 1366-1379, September 2009.
- [101] X. Li, N. Santoro, I. Stojmenovic, "Localized Distance-Sensitive Service Discovery in Wireless Sensor and Actor Networks," *IEEE Transactions on Computers*, vol. 58, no. 9, pp. 1275-1288, September 2009.
- [102] F. Barsi, A. A. Bertossi, F. B. Sorbelli, R. Ciotti, S. Olariu, M. C. Pinotti, "Asynchronous Corona Training Protocols in Wireless Sensor and Actor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 8, pp. 1216-1230, August 2009.
- [103] A. E. Gil, K. M. Passino, S. Ganapathy, A. Sparks, "Cooperative scheduling of tasks for networked uninhabited autonomous vehicles," in *Proc. IEEE Conference on Decision and Control*, vol. 1, pp. 522-527, Dec. 2003.
- [104] V. Gazi, K. M. Passino, "Stability analysis of swarms," *IEEE Transactions on Automatic Control*, vol. 48, no. 4, pp. 692-697, April 2003.
- [105] K. Langendoen, N. Reijers "Distributed localization in wireless sensor networks: a quantitative comparison," *Computer Networks*, vol. 43, pp. 499-518, 2003.
- [106] I. Dietrich, F. Dressler, "On the lifetime of wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 5, no. 1, 2009.
- [107] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660-670, Oct. 2002.
- [108] S. Boyd, L. Vandenberghe, "Convex Optimization," *Cambridge University Press*, 2004.
- [109] D. Corne, M. Dorigo, F. Glover, D. Dasgupta, P. Moscato, R. Poli, K. V. Price, "New Ideas in Optimization," *Mcgraw Hill-Advanced Topics in Computer Science Series*, 1999.
- [110] D. Stephens, J. Krebs, "Foraging Theory," *Princeton University Press*, 1986.

- [111] A. Houston, J. McNamara, "Models of Adaptive Behavior," *Cambridge University Press*, 1999.
- [112] K. M. Passino, "Biomimicry of bacterial foraging for distributed optimization," *IEEE Control Systems Magazine*, vol. 22, no. 3, pp. 52-67, 2002.
- [113] D. W. Stephens, E. L. Charnov, "Optimal Foraging: Some Simple Stochastic Models," *Behavioral Ecology and Sociobiology (Springer)*, vol. 10, no. 4, pp. 251-263, 1982.
- [114] B. P. Kotler, J. S. Brown, S. R. X. Dall, S. Gresser, D. Ganey, A. Bouskila, "Foraging games between gerbils and their predators: temporal dynamics of resource depletion and apprehension in gerbils," *Evolutionary Ecology Research*, vol. 4, pp. 495-518, 2002.
- [115] B. W. Andrews, K. M. Passino, T. A. Waite, "Foraging theory for decision-making system design: Task-type choice," in *Proc. IEEE Conference on Decision and Control*, vol. 5, pp. 4740-4745, Dec. 2004.
- [116] B. W. Andrews, K. M. Passino, T. A. Waite, "Social foraging theory for robust multi-agent system design," *IEEE Transactions on Automation Science and Engineering*, vol. 4, no. 1, pp. 79-86, Jan. 2007.
- [117] N. Quijano, K.M. Passino, B.W. Andrews, "Foraging theory for multizone temperature control," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 18-27, Nov. 2006.
- [118] D. R. Cox, "Renewal Theory," *Methuen, London*, 1962.
- [119] C. W. Therrien, "Discrete Random Signals and Statistical Signal Processing," *Prentice Hall, USA*, 1992.
- [120] B. Atakan, O. B. Akan, "Biological Foraging-inspired Communication in Intermittently-Connected Mobile Cognitive Radio Ad Hoc Networks," *IEEE Transactions on Vehicular Technology*, under review, 2011.
- [121] Z. Zhang, "Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and challenges," *IEEE Communications Surveys and Tutorials*, vol. 8, no. 1, pp. 24-37, March 2007.
- [122] A. Vahdat, D. Becker, "Epidemic routing for partially-connected ad hoc networks," *Technical Report (CS-2000-06) Duke University*, July 2000.
- [123] M. Grossglauser, D. Tse, "Mobility Increases the Capacity of Ad Hoc Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp.477-486, 2002.
- [124] T. Spyropoulos, K. Psounis, C. S. Raghavendra, "Spray and wait: an efficient routing scheme for intermittently connected mobile networks," in *Proc. SIGCOMM Workshop on Delay-Tolerant Networking (WDTN'05)*, pp. 252-259, Philadelphia, USA, 2005.
- [125] W. Zhao, M. Ammar, E. Zegura, "A message ferrying approach for data delivery in sparse mobile ad hoc networks," in *Proc. MobiHoc 2004*, pp. 187-198, 2004, Tokyo, Japan.
- [126] H. Khalife, N. Malouch, S. Fdida, "Multihop cognitive radio networks: to route or not to route," *IEEE Network*, vol. 23, no. 4, pp. 20-25, 2009.

- [127] G. Marsaglia, "Ratios of normal variables and ratios of sums of uniform variables," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 193-204, March 1965.
- [128] T. Spyropoulos, K. Psounis, C. S. Raghavendra, "Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks," in Proc. *ACM SIGCOMM'05*, pp.252-259, 2005.
- [129] T. Small, Z. Haas, "Resource and performance tradeoffs in delay tolerant wireless networks," in Proc. *ACM SIGCOM Workshop on Delay Tolerant Networking*, 2005.
- [130] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Performance analysis of mobility-assisted routing," in Proc. *ACM/IEEE MOBIHOC*, 2006.
- [131] B. Atakan, O. B. Akan, "An Information Theoretical Approach for Molecular Communication", in Proc. *ACM BIONETICS 2007*, Budapest, Hungary, December 2007.
- [132] B. Atakan, O. B. Akan, "On Channel Capacity and Error Compensation in Molecular Communication", *Springer Trans. on Computational System Biology*, vol. 10, pp. 59-80, February 2008.
- [133] T. Suda, M. Moore, T. Nakano, R. Egashira, A. Enomoto, "Exploratory Research on Molecular Communication between Nanomachines", in Proc. *GECCO 2005*, June 25-29, 2005, Washington, DC, USA.
- [134] M. Moore, A. Enomoto, T. Nakano, R. Egashira, T. Suda, A. Kayasuga, H. Kojima, H. Sakakibara, K. Oiwa, "A Design of a Molecular Communication System for Nanomachines Using Molecular Motors", in Proc. *IEEE PERCOMW 2006*, Italy, 2006.
- [135] Y. Moritani, S. Hiyama, T. Suda, "Molecular Communication for Health Care Applications", in Proc. *IEEE PERCOMW 2006*, Italy, 2006.
- [136] T. Nakano, T. Suda, M. Moore, R. Egashira, A. Enomoto, K. Arima, "Molecular Communication for Nanomachines Using Intercellular Calcium Signaling", in Proc. *IEEE Conference on Nanotechnology 2005*, Nagoya, Japan, July 2005.
- [137] S. Hiyama, Y. Isogawa, T. Suda, Y. Moritani, K. Sutoh, "A Design of an Autonomous Molecule Loading/Transporting/Unloading System Using DNA Hybridization and Biomolecular Linear Motors", in Proc. *European Nano Systems 2005*, Paris, France, December 2005.
- [138] J. P. Rospars, V. Krivan, P. Lansky "Perireceptor and receptor events in olfaction. Comparison of concentration and flux detectors: a modeling study.", *Chem. Sens.*, vol. 25, pp. 293-311, 2000.
- [139] V. Krivan, P. Lansky, J. P. Rospars, "Coding of periodic pulse stimulation in chemoreceptors", *Biosystem (Elsevier)*, vol. 67, pp. 121-128, 2002.
- [140] M. J. Saxton "Anomalous Diffusion Due to Binding: A Monte Carlo Study", *Biophysical Journal*, vol. 70, pp.1250-1262, March 1996.
- [141] M. Long, S. Lu, G. Sun, "Kinetics of Receptor-Ligand Interactions in Immune Responses", *Cell. & Mol. Immuno.*, vol. 3, no. 2, pp. 79-86, 2006.

- [142] C. J. Camacho, S. R. Kimura, C. DeLisi, S. Vajda “Kinetics of Desolvation-Mediated Protein Binding”, *Biophysical Journal*, vol. 78, pp. 1094-1105, March 2000.
- [143] G. I. Bell, “Models for the specific adhesion of cells to cells”, *Sciences*, vol. 200, pp. 618-627, 1978.
- [144] B. Atakan, O. B. Akan, “Deterministic capacity of information flow in molecular nanonetworks,” *Nano Communication Networks (Elsevier)*, vol.1, no. 1, pp. 31-42, March 2010.
- [145] A. W. Eckford, “Molecular communication: Physically realistic models and achievable information rates”, submitted to *IEEE Transactions on Information Theory*, 2008.
- [146] B. Atakan, O. B. Akan, “On Molecular Multiple-Access, Broadcast and Relay Channels in Nanonetworks”, in Proc. *ACM BIONETICS 2008*, Hyogo, Japan, December, 2008.
- [147] B. Atakan, O. B. Akan, “Single and Multiple-Access Channel Capacity in Molecular Nanonetworks”, in Proc. *ICST/ACM Nano-Net 2009*, Luzern, Switzerland, October 2009.
- [148] D. T. Gillespie, “Exact stochastic simulation of coupled chemical reactions”, *Journal of Physical Chemistry*, vol. 81, no. 25. pp. 2340-2361, 1977.
- [149] H. C. Berg and E. M. Purcell , “Physics of chemoreception”, *Biophysical Journal*, vol. 20, no. 2, pp. 193-219, 1977.
- [150] T. E. Turner, S. Schnell, K. Burrage, “Stochastic approaches for modelling in vivo reactions”, *Computational Biology and Chemistry (Elsevier)*, vol. 28, no. 3. pp. 165-178, 2004.
- [151] D. T. Gillespie, “The chemical Langevin equation”, *Journal of Chemical Physics*, vol. 113, no. 1. pp. 297-306, 2000.
- [152] D. T. Gillespie, “Markov Processes: An Introduction for Physical Scientists”, *Academic, San Diego*, 1992.
- [153] S. Avestimehr, S. Diggavi, D. Tse, “A deterministic approach to wireless relay networks,” in Proc. *Allerton Conference*, 2007.
- [154] L. R. Ford, D. R. Fulkerson, “Flows in Networks,” *Princeton University Press*, New Jersey, Princeton, 1962.
- [155] A. Guney, B. Atakan, O. B. Akan, “Mobile Ad Hoc Nanonetworks with Collision-based Molecular Communication”, *to appear in IEEE Transactions on Mobile Computing*, 2011.
- [156] B. Atakan, O. B. Akan, “Carbon nanotube-based nanoscale ad hoc networks”, *IEEE Comm. Magazine*, vol. 48, pp. 129-135, 2010.
- [157] T. Nakano, J.-Q. Liu,, “Information Transfer through Calcium Signaling”, in Proc. *Nanonet 2009*, Switzerland 2009, pp. 29-33.
- [158] Y. Moritani, S. Hiyama, T. Suda, “Molecular Communication among Nanomachines Using Vesicles”, in Proc. *NSTI Nanotechnology Conference and Trade Show*, 2006.

- [159] M. Pierobon, I. F. Akyildiz, "Information Capacity of Diffusion-based Molecular Communication in Nanonetworks", *to appear in Proc. of IEEE International Conference on Computer Communication, INFOCOM, Miniconference*, Shanghai, China, April 2011.
- [160] M. S. Kuran, H. B. Yilmaz, T. Tugcu, I. F. Akyildiz, "Modulation Techniques for Communication via Diffusion in Nanonetworks", in *Proc. IEEE ICC 2011*, Kyoto, Japan, June 2011.
- [161] M. Gregori, I. F. Akyildiz, "A New NanoNetwork Architecture using Flagellated Bacteria and Catalytic Nanomotors," *IEEE J. Selected Areas Commun. (JSAC)*, vol. 28, pp. 612-619, May 2010.
- [162] J. Wiedermann, L. Petru, "Communicating Mobile Nano-Machines and Their Computational Power," in *Proc. ICST Conference on Nano-networks*, Boston, USA, 2008.
- [163] R. L. Juliano, "Signal transduction by cell adhesion receptors and the cytoskeleton: functions of integrins, cadherins, selectins, and immunoglobulin-superfamily members", *Annu. Rev. Pharmacol. Toxicol.* vol. 42, pp. 283-323, 2002.
- [164] S. E. Chesla, P. Selveraj, C. Zu, "Measuring Two-Dimensional Receptor-Ligand Binding Kinetics by Micropipette", *Biophysical Journal*, vol. 75, no. 3, September, pp. 1553-1572, 1998.
- [165] D. A. McQuarrie, "Kinetics of small systems I", *Chemical Physics*, vol. 38, pp. 433-436, 1963.
- [166] M. L. Dustin, D. R. Colman, "Neural and Immunological Synaptic Relations", *Science*, vol. 298, pp. 785-789, 2002.
- [167] A. Manwani, "Information-theoretic analysis of neuronal communication", *Thesis(Phd), California Institute of Technology*, 2000.
- [168] L. J. DeFelice, "Introduction to membrane noise", *Plenum Press*, 1981.
- [169] S. P. Strong, R. Koberle, R. R. de R. van Steveninck, William Bialek, "Entropy and Information in Neural Spike Trains", *Physical Review Letters*, vol. 80, no. 1, 1998.
- [170] T. Spyropoulos, K. Psounis, C. S. Raghavendra,, "Efficient routing in intermittently connected mobile networks: the multiple-copy case", *IEEE/ACM Trans. Networking* , vol. 16, pp. 77-90, 2008.
- [171] Z. J. Haas, T. Small, "A new networking model for biological applications of ad hoc sensor networks", *IEEE/ACM Trans. Networking*, vol. 14, no 1, pp. 27-40, 2006.
- [172] X. Zhang, G. Neglia, J. Kurose and D. Towsley, "Performance modeling of epidemic routing", *Computer Networks*, vol. 51, no. 10, pp. 2867-2891, 2007.
- [173] A. Khelil, et. al., "An epidemic model for information diffusion in MANETs", in *Proc. ACM MSWIM*, Sept. 2002.
- [174] W. O. Kermack, A. G. McKendrick, "A contribution to the mathematical theory of epidemics", *Proc. R. Soc. Lond.*, pp. 700-721, 1927.
- [175] B. Atakan, O. B. Akan, "Nanoscale Communication with Molecular Arrays in Nanonetworks," *IEEE Transactions on Nanobiosciences*, to be appear in 2011.

- [176] M. Pierobon, I. F. Akyildiz, "Diffusion-based Noise Analysis for Molecular Communication in Nanonetworks," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2532-2547, June 2011.
- [177] M. Moore, T. Suda, K. Oiwa, "Molecular Communication: Modeling Noise Effects on Information Rate," *IEEE Transactions on Nanobioscience*, vol. 8, pp.169-180, June 2009.
- [178] T. Nakano, L. Jian-Qin, "Design and Analysis of Molecular Relay Channels: An Information Theoretic Approach", *IEEE Transactions on Nanobioscience*, vol. 9, no. 3, pp. 213-221, Sept. 2010.
- [179] T. Nakano, M. Moore, "In-sequence molecule delivery over an aqueous medium ", *Nano Communication Networks (Elsevier)*,vol.1, no. 3, pp. 181-188, September 2010.
- [180] I. Karatzas, S. E. Shreve, "Brownian Motion and Stochastic Calculus", *Springer (New York)*, 1991.
- [181] M. Ş. Kuran, H. B. Yılmaz, T. Tuğcu, B. Özerman, "Energy model for communication via diffusion in nanonetworks ", *Nano Communication Networks (Elsevier)*,vol.1, no. 2, pp. 86-95, June 2010.
- [182] M. U. Mahfuz, D. Makrakis, H. T. Mouftah, "On the characterization of binary concentration-encoded molecular communication in nanonetworks ", *Nano Communication Networks (Elsevier)*,vol.1, no. 4, pp. 289-300, Dec. 2010.
- [183] D. Arifler, "Capacity analysis of a diffusion-based short-range molecular nano-communication channel," *Computer Networks (Elsevier)*,vol.55, no. 6, pp. 1426-1434, April 2011.
- [184] A. W. Eckford, "Nanoscale Communication with Brownian Motion", in *Proc. Information Sciences and Systems 2007, (CISS' 07)*, pp. 160-165, Baltimore, MD, September 2007.
- [185] B. Atakan, S. Balasubramaniam, O. B. Akan, "Body Area NanoNetworks with Molecular Communications in Nanomedicine," to appear in *IEEE Communications Magazine*, 2011.
- [186] R. A. Freitas, "Current Status of Nanomedicine and Medical Nanorobotics," *Journal of Computational and Theoretical Nanoscience*, vol. 2, pp. 1-25, 2005.
- [187] J. C. Anderson, Clarke E. J., A. P. Arkin, C. A. Voigt, "Environmentally controlled invasion of cancer cells by engineered bacteria", *Journal of Molecular Biology*, vol. 355, pp. 619-627, 2006.
- [188] P. E. M. Purnick, R. Weiss, "The second wave of synthetic biology: from modules to systems," *Nature Reviews Molecular Cell Biology*, vol. 10, pp. 410-422, 2009.
- [189] Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, E. Shapiro, "An autonomous molecular computer for logical control of gene expression," *Nature*, vol. 429, pp. 423-429, May 2004.
- [190] L. Leoni, T. A. Desai, "Nanoporous biocapsules for the encapsulation of insulinoma cells: biotransport and biocompatibility considerations," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1335-1341, Nov. 2001.

- [191] X. Gao, L. Yang, J. A. Petros, F. F. Marshall, J. W. Simons, S. Nie, "In vivo molecular and cellular imaging with quantum dots," *Current Opinion in Biotechnology*, vol. 16, no. 1, pp. 63-72, Feb. 2005.
- [192] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, pp. 1130-1134, April 2005.
- [193] P. J. Burke, S. Li, Z. Yu "Quantitative Theory of Nanowire and Nanotube Antenna Performance," *IEEE Transactions on Nanotechnology*, vol. 5, no. 4, pp. 314-334, 2006.
- [194] C. Rutherglen, P. Burke, "Nanoelectromagnetics: Circuit and Electromagnetic Properties of Carbon Nanotubes," *Small*, vol. 5, no. 8, pp. 884-906, April 2009.
- [195] K. Jensen, J. Weldon, H. Garcia, A. Zettl "Nanotube Radio," *Nano Letters*, vol. 7, pp. 3508-3511, Oct. 2007.
- [196] K. Jensen, J. Weldon, H. Garcia, A. Zettl "Nanomechanical Radio Transmitter," *Physica Status Solidi (b)*, vol. 245, no. 10, pp. 2323-2325, Sep. 2008.
- [197] J. M. Jornet, I. F. Akyildiz, "Channel Capacity of Electromagnetic Nanonetworks in the Terahertz Band," in Proc. *ICC 2010*, Cape Town, South Africa, May 2010.
- [198] J. Pei M. She, J. T. W. Yeow, "Nanotechnology-Enabled Wireless Sensor Networks: From a Device Perspective," *IEEE Sensor Journal*, vol. 6, no. 5, pp. 1331-1339, Oct. 2006.
- [199] S. F. Bush, S. Goel, "Graph Spectra of Carbon Nanotube Networks," in Proc. *ICST NANO-NET 2006*, pp. 1-10, Sep. 2006.
- [200] B. Atakan, O. B. Akan, "Carbon Nanotube Sensor Networks," in Proc. *IEEE Nanocom 2009*, San Francisco, USA, August 2009.
- [201] J. M. Jornet, I. F. Akyildiz, "Electromagnetic Wireless Nanosensor Networks," *Nano Communication Networks Journal (Elsevier)*, vol. 1, no. 1, pp. 3-19, 2010.
- [202] D. Dragoman, M. Dragoman, "Tunneling Nanotube Radio," *Journal of Applied Physics*, vol. 104, no. 7, Oct. 2008.
- [203] P. Burke, C. Rutherglen "Towards a Single-chip, Implantable RFID System: is a single-cell radio possible?," *Biomedical Microdevices (Springer)*, Jan. 2009.
- [204] Y. Huang, W.-Y. Yin, Q. H. Liu, "Performance Prediction of Carbon Nanotube Bundle Dipole Antennas," *IEEE Transactions on Nanotechnology*, vol. 7, no. 3, pp. 331-337, Sep. 2008.
- [205] G. W. Hanson, "Fundamental Transmitting Properties of Carbon Nanotube Antennas," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 11, pp. 3426-3435, Nov. 2005.
- [206] A. Nojeh, P. Pande, A. Ganguly, S. Sheikhaei, B. Belzer, A. Ivanov, "Reliability of Wireless On-chip Interconnects Based on Carbon Nanotube Antennas," in Proc. *IEEE IMS3TW'08*, pp. 1-6, Vancouver, Canada, June 2008.
- [207] P. H. Humble, J. N. Harb, R. LaFolletteb, "Microscopic Nickel-Zinc Batteries for Use in Autonomous Microsystems," *Journal of the Electrochemical Society*, vol. 148, pp. 1357-1361, 2001.

- [208] C. Dewan, D. Teeters, "Vanadia Xerogel Nanocathodes Used in Lithium Microbatteries," *Journal of Power Sources (Elsevier)*, vol. 119-121, pp. 310-315, June 2003.
- [209] Chen et al, "Nanoscale Molecular-switch Crossbar Circuits," *Nanotechnology*, vol. 14, pp. 462-468, 2003.
- [210] T. Rueckes, K. Kim, E. Joselevich, G. Y. Tseng, C. L. Cheung, C. M. Lieber, "Carbon Nanotube-Based Nonvolatile Random Access Memory for Molecular Computing," *Science*, vol. 289, no. 5476, pp. 94-97, July 2000.
- [211] Y. Huang, X. Duan, Y. Cui, L. J. Lauhon, K. H. Kim, C. M. Lieber, "Logic Gates and Computation from Assembled Nanowire Building Blocks," *Science*, vol. 294, no. 5545, pp. 1313-1317, Nov. 2001.
- [212] G. S. Rose, M. R. Stan, "Memory Arrays Based on Molecular RTD Devices," in *Proc. IEEE NANO*, vol. 2, pp. 453-456, 2003.
- [213] R. M Goody, Y. L. Yung, "Atmospheric Radiation: Theoretical basis", *Oxford University Press*, 1989.
- [214] X. He, R. Yuan, Y. Chai, Y. Shi, "A Sensitive Amperometric Immunosensor for Carcinoembryonic Antigen detection with porous nanogold film and nano-Au/chitosan composite as immobilization matrix", *Journal of Biochemical and Biophysical Methods*, vol. 70, no. 6, April 2008, pp. 823-829.
- [215] A. Grakoui, et. al., "The Immunological Synapse: A Molecular Machine Controlling T Cell Activation", *Science*, vol. 285, no. 5425, pp. 221-227, 1999.
- [216] D. M. Blough, G. Resta, P. Santi, "A statistical analysis of the long-run node spatial distribution in mobile ad hoc networks", *Wireless Networks*, Vol 10, pp. 543-554, 2004.

VITA

Barış Atakan was born on February 20, 1977, in Ankara, Turkey. He received his B.Sc. degree in Electronics Engineering from Ankara University, Ankara, Turkey, in 2000. He received his M.Sc. degree in Electrical and Electronics Engineering from Middle East Technical University (METU), Ankara, Turkey, in 2005. He attended the doctoral program at the Department of Electrical and Electronics Engineering in METU, and Koç University, İstanbul, Turkey. He was a research assistant in the Next-generation and Wireless Communications Laboratory (NWCL) during his doctoral study in METU and Koç University. In December 2011, he received his Ph.D. degree in Electrical and Electronics Engineering from Koç University.