

DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**THE SERVER-BASED AUTOMATIC FARE
COLLECTION AND MANAGEMENT SYSTEM IN
PUBLIC TRANSPORT AND ANALYSIS WITH
DATA MINING TECHNIQUES**

by
Ufuk DEMİR ALAN

June, 2018
İZMİR

THE SERVER-BASED AUTOMATIC FARE COLLECTION AND MANAGEMENT SYSTEM IN PUBLIC TRANSPORT AND ANALYSIS WITH DATA MINING TECHNIQUES

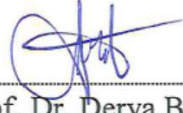
**A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of
Philosophy in Computer Engineering**

**by
Ufuk DEMİR ALAN**

**June, 2018
İZMİR**

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled “**THE SERVER-BASED AUTOMATIC FARE COLLECTION AND MANAGEMENT SYSTEM IN PUBLIC TRANSPORT AND ANALYSIS WITH DATA MINING TECHNIQUES**” completed by **UFUK DEMİR ALAN** under supervision of **ASSOC. PROF. DR. DERYA BİRANT** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.



Assoc. Prof. Dr. Derya BİRANT

Supervisor



Prof. Dr. Alp KUT

Thesis Committee Member



Assoc. Prof. Dr. Savaş ŞAHİN

Thesis Committee Member



Doç. Dr. Tüjba ÖZACAR ÖZTÜRK

Examining Committee Member



Doç. Dr. Galin Baştanlar

Examining Committee Member



Prof. Dr. Latif SALUM

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I would like to thank to my supervisor, Assoc. Prof. Dr. Derya BİRANT, for her support, patient guidance, supervision and useful suggestions throughout this study. She is more than the supervisor; she is my mentor also. I am also thankful to Prof. Dr. Alp KUT for his valuable advice and encouragement throughout this study. I would also like to thank my committee member, Assoc. Prof. Dr. Savaş ŞAHİN for his brilliant comments and suggestions.

Also, I would like to offer my special thanks to my family for their support, patience, and help. It would not have been able to complete this thesis without their support and help.

I would like to thank to Kent Kart Ege Elektronik San. Tic. A.Ş Company for believing in this project and supporting the implementation of the proposed system on their terminals and servers. Their support gave us the chance to install our infrastructure on the field and analyze the results.

This work was supported by the Scientific and Technological Research Council of Turkey under Grant TEYDEB 1501-3150065.

Ufuk DEMİR ALAN

THE SERVER-BASED AUTOMATIC FARE COLLECTION AND MANAGEMENT SYSTEM IN PUBLIC TRANSPORT AND ANALYSIS WITH DATA MINING TECHNIQUES

ABSTRACT

Mass transit systems have existed for over a century. In the traditional payment structure; tickets, cards, and tokens are used for public transportation. Balances, discounts and transfer fee benefits are generally kept on cards specially designed for each city, and the cost of the ride is calculated and deducted by the validators while getting on the vehicle. The present thesis proposes a novel server-based fare collection, calculation and user management system. The proposed system provides personalized exclusive privileges to riders without getting the special card of the province and allows passengers to benefit from all the advantages of the cities' transport pricing. In addition, it removes the obstacles that prevent the use of new technologies such as contactless bank cards and Near Field Communication (NFC) standard like a traditional province card in public transport.

The installation of the system with NFC usage was applied to several cities in Turkey. Real-world data was collected and then analyzed with two data mining techniques. First, the association rule mining technique was applied for the first time to find out NFC usage habits and user profile in public transportation. Second, classification algorithms were run on NFC-based transportation data and compared with each other in terms of accuracy rates.

In order to show the personalization capability of the proposed system, this thesis also proposes a novel passenger scoring model, namely RFLT (Recency, Frequency, Loyalty, and Time), for offering a free pass promotion in public transportation. RFLT is the modified version of the well-known marketing method RFM (Recency, Frequency, Monetary). Experimental studies demonstrate the applicability of RFM model on transportation for the first time and show the comparison results of RFLT and wRFLT (weighted version) on a real-world dataset.

The results of the study can be used to establish an efficient policy for increasing public transportation ridership.

Keywords: Public transportation, automatic fare collection, data mining, association rule mining, classification, RFM model, NFC-based mobile application



TOPLU TAŞIMADA SUNUCU TABANLI OTOMATİK ÜCRET TOPLAMA VE YÖNETİM SİTEMİ VE VERİ MADENCİLİĞİ TEKNİKLERİ İLE ANALİZİ

ÖZ

Toplu taşıma sistemleri yüzyılı aşkın süredir bulunmaktadır. Geleneksel ödeme yapısında; toplu taşıma araçlarında bilet, kart ve jeton kullanılmaktadır. Bakiye, indirimler ve aktarma ücret avantajları her il için özel olarak üretilen kartlar üzerinde tutulmakta, biniş ücreti hesaplaması ve kartın bakiyesinden düşülmesi validatörler aracılığıyla biniş esnasında yapılmaktadır. Bu tez, yeni bir sunucu tabanlı ücret toplama, hesaplama ve kullanıcı yönetim sistemi önermektedir. Önerilen sistem, seyahat edilen bölgeye ait toplu taşıma kartı almadan yolculara kişisel ayrıcalıklar sağlamak ve o ilin tüm ücretlendirme avantajlarından faydalanabilmesine imkan vermektedir. Buna ek olarak, temassız kredi kartı ve yakın alan iletişimi (NFC) gibi yeni nesil teknolojilerin geleneksel toplu taşıma kartı gibi kullanılmasının önündeki engelleri kaldırmaktadır.

NFC kullanımı ile sistemin kurulumu Türkiye'de çeşitli şehirlerde gerçekleştirilmiştir. Sahadan gerçek veriler toplamış ve iki farklı veri madenciliği tekniği ile analiz edilmiştir. İlk olarak, birliktelik kuralları madenciliği tekniği uygulanarak, toplu ulaşımda NFC kullanım alışkanlıkları ve kullanıcı profilleri ilk defa ortaya koyulmuştur. İkinci olarak, NFC tabanlı ulaşım verilerinde sınıflandırma algoritmaları denenmiş ve doğruluk oranları açısından karşılaştırılmıştır.

Bu tez aynı zamanda, önerilen sistemin kişiselleştirme yeteneğini ortaya koymak için, toplu ulaşımda serbest geçiş promosyonu sunan ve RFLT (Güncellik, Frekans, Sadakat ve Zaman) olarak adlandırılan yeni bir yolcu puanlama modeli önermektedir. RFTL, iyi bilinen RFM (Güncellik, Frekans, Parasal) pazarlama yönteminin değiştirilmiş bir versiyonudur. Deneysel çalışmalar, RFM modelinin ulaşımda uygulanabilirliğini ilk kez göstermekte ve gerçek dünya veri setinde RFLT ve wRFLT (ağırlıklı versiyon) karşılaştırma sonuçlarını sunmaktadır.

Çalışmanın sonuçları toplu taşıma kullanımının artırılmasında etkili bir politika oluşturmak için kullanılabilir.

Anahtar kelimeler: Toplu taşıma, otomatik ücret toplama, veri madenciliği, birliktelik kuralları madenciliği, sınıflandırma, RFM modeli, NFC tabanlı mobil uygulama



CONTENTS

	Page
Ph.D. THESIS EXAMINATION RESULT FORM	ii
ACKNOWLEDGMENTS	iii
ABSTRACT	iv
ÖZ	vi
LIST OF FIGURES	xi
LIST OF TABLES	xiii
 CHAPTER ONE - INTRODUCTION	 1
1.1 General	1
1.2 Purpose	2
1.3 Novel Contributions of this Thesis.....	4
1.4 Organization of the Thesis	5
 CHAPTER TWO - RELATED WORK.....	 6
2.1 Literature Review	6
2.2 Field Review	9
 CHAPTER THREE - SERVER-BASED SYSTEM ARCHITECTURE FOR PUBLIC TRANSPORTATION.....	 14
3.1 General Architecture	14
3.2 System Components	18
3.2.1 Fare Engine and Fare Calculation Algorithm	19
3.2.2 Automatic Fare Collection	23
3.2.3 Card Manager.....	24
3.2.4 Payment Engine.....	25
3.2.5 Mobil Engine.....	27
3.2.6 Terminal Engine	28

3.3 Security Infrastructure	29
3.3.1 Cloud-Based Secure Element.....	29
3.3.2 Token Security	31
3.3.3 Token-Based User Authentication	32
3.3.4 Network Security.....	33
3.3.5 Payment Security.....	34
3.4 Comparison of Proposed System and Traditional Systems	34
CHAPTER FOUR – DESIGN OF THE SYSTEM.....	37
4.1 Overall Flow.....	37
4.2 Use Cases	39
4.3 Sequential Diagrams	41
4.3.1 Terminal Operations.....	41
4.3.2 Mobile Engine Operations	46
4.3.3 Server Side Operations.....	49
4.4 Database Design.....	52
CHAPTER FIVE - MOBILE APPLICATION, INTEGRATION, AND TESTING	59
5.1 User Interfaces of NFC-Based Mobile Application.....	59
5.2 NFC-Based Mobile Application.....	61
5.3 Integration	63
5.4 Testing.....	64
5.4.1 Performance Test.....	65
5.4.2 Security Test.....	65
5.4.3 Bug Fix.....	66
CHAPTER SIX - APPLICATION OF DATA MINING TECHNIQUES.....	68
6.1 Association Rule Mining.....	68
6.1.1 Association Rule Mining Algorithms	69

6.1.2 Dataset Description	71
6.1.3 Data Preparation	73
6.1.4 User Profiling	74
6.1.5 Usage Mining	77
6.2 Classification	81
6.2.1 Classification Algorithms.....	81
6.2.2 Classification in Public Transportation	83
6.2.3 Dataset Description	85
6.2.4 Comparison of Classification Algorithms.....	85
 CHAPTER SEVEN - EXPERIMENTAL STUDY: PASSENGER SCORING .	89
7.1 Free Pass Promotion in Public Transport	89
7.2 Advantages of Free Pass Promotion in Public Transportation.....	90
7.3 RFM and Its Versions	91
7.4 RFLT (Recency, Frequency, Loyalty, Time)	93
7.5 Passenger Scoring Algorithm.....	100
7.6 Dataset Description	102
7.7 Comparison RFLT and wRFLT	103
7.8 Sensitivity Analysis	105
 CHAPTER EIGHT - CONCLUSION AND FUTURE WORK	110
8.1 Conclusion.....	110
8.2 Future Work	111
 REFERENCES.....	113
 APPENDICES	125
 APPENDIX: LIST OF ACRONYMS.....	125

LIST OF FIGURES

	Page
Figure 3.1 The system modules	15
Figure 3.2 General operation structure of SBPTS	18
Figure 3.3 The relationship between components.....	19
Figure 3.4 High level flowchart for fare calculation.....	22
Figure 3.5 Flow of the 3DS protocol	27
Figure 3.6 Cloud-based secure element	30
Figure 4.1 Overall operation of the system.....	37
Figure 4.2 General structure and connection of modules in SBPTS	39
Figure 4.3 Use case diagram of host operations	40
Figure 4.4 Use case diagram of passenger registration.....	40
Figure 4.5 Use case diagram of NFC usage on the public transport vehicle	41
Figure 4.6 APDU command structures	42
Figure 4.7 APDU exchange	43
Figure 4.8 Card usage on terminal	45
Figure 4.9 Sequence diagram of virtual card production.....	46
Figure 4.10 Sequence diagram of virtual card sale operations	47
Figure 4.11 Sequence diagram of mobile application synchronization	48
Figure 4.12 Lifecycle of the transaction	49
Figure 4.13 Flowchart of transmitting transactions from FE to NCAMS	50
Figure 4.14 Flowchart of the money withdrawal	51
Figure 4.15 ER diagram of fare module	55
Figure 4.16 Sample user interfaces of fare module.....	56
Figure 4.17 ER diagram of the NCAMS.....	57
Figure 4.18 Transaction history of NFC card produced by NCAMS	58
Figure 5.1 Screenshots of account creation steps on android phone.....	60
Figure 6.1 Single-item rules discovered from association rule mining	75
Figure 6.2 The most popular routes on which NFC virtual card used in the city of Turkey	79
Figure 6.3 The most popular bus-stop locations where NFC mobile application is used frequently in the city of Turkey.....	80

Figure 6.4 The average rank of classifiers	87
Figure 7.1 Comparison of RFLT and wRFLT with different weight values	105
Figure 7.2 Different wRFLT charts	107
Figure 7.3 Sensitivity analysis of RFLT model with different weight values	108
Figure 7.4 Comparison of RFLT and wRFLT by using IBM SPSS program	109



LIST OF TABLES

	Page
Table 2.1 Comparison of publications according to their contents.....	8
Table 2.2 The situation of the capital transport authorities in USA and UK according to research (Wallischek et al.) on February, 2015	12
Table 3.1 The scope of each module in the system	16
Table 3.2 Notation for fare calculation equation	21
Table 3.3 Comparison of card and server based systems.....	35
Table 4.1 Request-response of APDU structure	42
Table 4.2 Response APDU structure	42
Table 4.3 Virtual POS return codes and their explanation.....	52
Table 6.1 Horizontal data format	70
Table 6.2 Vertical data format	70
Table 6.3 Possible data values of dataset	72
Table 6.4 Sample data	74
Table 6.5 Frequent patterns discovered from NFC based public transportation data	76
Table 6.6 GPS coordinates that is used to divide regions	85
Table 6.7 Learning algorithm results on dataset	86
Table 7.1 Different versions of RFM model.....	93
Table 7.2 A sample dataset about passenger transactions.....	99
Table 7.3 The steps of RFLT analysis	99
Table 7.4 The passenger segments according to RFLT scores	100
Table 7.5 Notation table.....	101
Table 7.6 The attributes of the dataset with minimum and maximum values	104
Table 7.7 Comparison of RFLT and wRFLT techniques	104
Table 7.8 Weight RFLT and number of passengers that win 1, 2, 3, 4 and 5 free pass/passes.....	106

CHAPTER ONE

INTRODUCTION

1.1 General

Public transportation has played an important role in the development of countries and is necessary for the social and economic quality of citizens' lives. The benefits of public transportation affect citizens, even those who may never use public transport. Last ten years, authorities perceive its importance so much that they are making a great effort to broaden public transportation services.

Through the development of 21st-century technologies, rider's expectations and the capabilities of transportation systems are increasing day by day. New information technologies make rider's life easier by voice announcements on bus or train stops, so passengers know when to get off/on the vehicle. An alarm system alerts passengers when the next public transport vehicle is coming. Journey planner tools on websites help users to determine how to travel between two selected points by bus, train or a combination of both. All these improvements also affect the fare module of traditional transportation systems. Transport authorities are constantly trying to make passenger's life easier in terms of fare policy with flexible structure and cheap prices. Today, hundred millions of people use contactless Europay, MasterCard, Visa (EMV) based bank cards or NFC mobile phones which are able to communicate in contactless mode. In the near future, this number will be increased dramatically. It will be especially attractive to use a mobile phone as a payment instrument. In addition, the public transport is more personalized by touching passengers directly. For these reasons, usage of the NFC technology as a payment instrument appears to be a good solution for transport payments. NFC-enabled phones have a big advantage over paper tickets (Spanjaard, 2013). Tickets stored virtually in phones are inherently more durable and less likely to be lost. They are even more convenient than plastic cards as a greener solution and with no requiring any space in a wallet.

There are two primary contactless payment systems: closed-loop and open-loop (account-based ticketing). Today, the traditional transportation systems are generally

closed-loop systems, in which all data (i.e. personal information, balance, social rights etc.) is held on contactless smartcards issued by the operator. The validator in the vehicle reads information from the transit smart card and writes the result of the transaction back to the card. The fare for the desired trip is calculated by the validators at boarding time. However, this type of system is not flexible, and it is expensive to manage and integrate it with other corporations. It is possible to overcome such limitations with *account-based ticketing* (ABT) systems. In an ABT, validator sends transaction data to the server and fare is calculated at the server side. In addition, the users can pay for travel using their existing credit card or NFC mobile device, eliminating the need to carry a separate transit specific card. In this thesis, we designed and developed a new account-based ticketing system.

There are two major types of fare collection systems for public transportations: barrier and barrier-free (Sasaki, 2014). In the barrier system, there are gates to collect money from passengers before getting on the vehicle. The barrier-free system is based on mutual trust. There is no any physical gate to check users' payment but inspectors randomly check passengers for valid tickets. The present thesis focuses on barrier-based systems.

Personal norm, service quality and economic factors (Jomnonkwao, Sangphong, Khampirat, Siridhara & Ratanavaraha, 2016) are the significant elements that affecting usage of public transportation. All of them have a big effect on pushing up the usage of public transportation. In this thesis, we try to increase service quality by providing NFC-based mobile application and by eliminating the need to carry a separate transit specific card. In addition, we focus on the economic factor by proposing a new passenger scoring mechanism for free pass promotion on ABT system to encourage public transportation.

1.2 Purpose

The aim of this thesis is to develop a server-based fare collection, calculation and user management system to create an infrastructure to enable to use of new generation of contactless communication technology such as (i) near field communication(NFC)

(Coşkun, Ok & Ozdenizci, 2012), (ii) Bluetooth Low Energy-BLE (Bronzi, Frank, Castignani & Engel, 2016), (iii) Beacon, (iv) contactless credit card (Smart Card Alliance Report, 2011; Quibria, 2008), and (v) Biometric data in public transportation.

The existing NFC applications that have been used on public transport are generally about ticketing on closed-loop systems (Smart Card Alliance Report, 2011). However, the present thesis proposes an account-based system. Thus, our infrastructure provides personalized exclusive privileges to the riders that can be used anywhere in the world. Riders can be able to benefit from all the advantages of the cities' transport pricing without having special transport card.

The proposed system in this thesis (the establishment of server-based fare collection and user management) makes possible to take advantage of public transport in each province with a single payment. The passengers outside the province are able to benefit from the advantages of the city transport. The system is performed for the first time in Turkey. Transition to the single payment media for the provinces in Turkey is expected to be accomplished in view of this research. The application that combined with NFC technology enables mobile phone owners to use public transport more easily. In addition, the infrastructure proposed in this thesis enables municipalities to offer personalized campaigns and loyalty applications.

In contrast to the current systems, the proposed system in this thesis also provides economic advantages. Removal of the requirement for issuing a specific transportation card brings a reduction of the cost of card issue phase and provides an economic recovery. Especially, a simple and cheaper reader device plugged in transport vehicle is preferable when the subject is the African or Indian market. Therefore, there are many countries that wish to receive electronic toll collection services with a cheaper device. The installation cost of the proposed system is also reduced because of the cloud services.

Another aim of this thesis is to encourage people the use of public transportation over server-based public transportation system. For this purpose, it proposes a novel

passenger scoring mechanism (RFTL) that calculates the number of free pass rights to be given to passengers for the following month.

The proposed system in this thesis is named as SBPTS (Server-Based Public Transportation System) and this abbreviation will be used through the rest of the document to indicate the developed system.

1.3 Novel Contributions of this Thesis

The main contributions of this thesis are on four levels;

First, we developed a server-based public transportation system with all modules for the first time in Turkey. We presented the advantages of the proposed system over the existing systems. By this thesis, we show the realization of the national card project in Turkey.

Second, we focus on the usage of the NFC-enabled mobile phone to realize proposed system. The server-based architecture was sampled with a credit card at TFL (Transport for London) before, but this is the first real application with NFC in the world that applied to field with a server-centric concept.

Third, we gathered passenger-specific, multi-dimensional and province independent transportation data from the field. After that, the collected data was analyzed by using an association rule mining technique for the first time to find out about NFC usage habits and user profiles. In addition, classification algorithms were run on NFC-based transportation data and compared with each other in terms of accuracy rates.

Fourth, it proposes a novel passenger scoring mechanism, named RFLT (Recency, Frequency, Loyalty and Time), to be used in the public transportation field to encourage people to use public transportation. It introduces a modified version of RFM model, so it demonstrates the applicability of RFM model on public transportation for the first time.

As a result, in this thesis, (i) a novel system, SBPTS, was designed and applied to field, (ii) an account-based NFC-enabled mobile application was developed, (iii) data mining techniques were used to analyze collected data (i.e. passenger profiling) and (iv) a new passenger scoring technique, RFLT, is proposed.

1.4 Organization of the Thesis

This thesis includes eight chapters and the remaining of this thesis is organized as follows.

In Chapter 2, general information about related works, literature review and field research about server-based systems on public transportation are given.

In Chapter 3, SBPTS architecture and details about its modules are explained. The responsibilities of all modules and the relationships among them are described.

In Chapter 4, the design of SBPTS is given in terms of use cases, flowcharts and sequential diagrams. In addition, database design is explained.

In Chapter 5, mobile application details and usage of the system are given. In addition, the system testing and integration are explained.

In Chapter 6, first, real-world NFC usage data collected from SBPTS and used for analysis are described, data mining techniques (association rule mining and classification) applied on this dataset are explained in detail and experimental results are given.

In Chapter 7, a novel passenger scoring technique, RFLT, is described in detail. The applicability of RFM model in public transportation is discussed. In addition, the weighted version of RFLT is also implemented and the results are compared. Sensitivity analysis for finding optimal parameters is put forth.

Finally, in Chapter 8, the conclusion remarks and future works are presented.

CHAPTER TWO

RELATED WORK

In this chapter, technical research projects, literature and field reviews are explained and research results supporting desired goals are discussed.

2.1 Literature Review

Because the subject of the project is fairly new, the number of studies found in the literature in recent years is not high. The main articles are listed in Table 2.1. Such articles are generally studies on smart card data and the value of it in public transportation systems (Liu, Kauffman & Ma, 2015; Pelletier, Trépanier & Morency, 2011). Also, most of them are about case studies (Seaborn, Attanucci & Wilson, 2009; Song, Eom, Lee, Min & Yang, 2015; W. K. Tan & Y. J. Tan, 2013).

Smart cards are being used mostly by public transit agencies. The main purpose is to collect revenue so the money is held on these secure cards. Travel data is also held on cards in the traditional public transport approach (Pelletier et al., 2011). The new logic in our approach is holding nothing on the smart card.

Analyzing different models in public transport shows that the use of smart cards is still more popular. Journey planning data is obtained from smart cards (Seaborn et al., 2009). In our study, in comparison with the study of Seaborn et al. (2009), all data is directly transferred to big servers in a few minutes.

The new technologies and needs for change require systems to be more flexible, respond rapidly and perform real-time processing. However, there are many fare types linked to smart cards. This is a situation that prevents making big changes in the public transportation field (Munoz & Seekins, 2016). Nearly real-time systems are more feasible for today. Our infrastructure proposed in this study is designed as nearly real-time. Full real-time connection to the server is possible, but in some situations, for example, when the vehicle travels underground, the GPRS connection can be weaker

or lost. Our model works under all conditions. If full coverage exists, we can send transaction data to the server immediately. If it does not, the validator holds the transaction data until GPRS coverage becomes available. Monzón, Hernandez, Martínez, Kaparias & Viti (2016) are explained the same situation clearly.

Intelligent transport systems increase the expectations of passengers and municipalities in terms of journey management, vehicle management, and integration operations. During the last years (Monzón et al., 2016), the mobility of the passengers is increased not only in the city but also in terms of intercity. The same media usage is desired instead of regional transportation cards. The system proposed in the study satisfies this purpose by a server-based system with NFC mobile phones.

A more detailed report about future / next generation transit fare payment system is defined in Transit Cooperative Research Program (TCRP) report (Wallischeck, Weisenberger, Berthaume & Dinning, 2015). All fare types are also listed and compared in this report. It mentions that the payment media issued by both the transit agency as well as organizations outside the transit agency will be available in the future. This means that the transport media will be independent of agency-specific procedures, gaining global standards. In addition, the role of mobile payments in micropayment sector is analyzed, and the importance of it is not underestimated (Liu et al., 2015). In our study, we aim to place the NFC mobile phone as a global media that can be read by all operators in a standard manner. Our infrastructure also accepts EMV (Euro pay, MasterCard, and Visa) standard contactless cards or biometric value as a payment media. So it supports all types of payment, but in this study, we only sampled it with NFC. Mobile application for Android operating system was developed and used via NFC channel as the ticketing purpose in public transport.

Public transport is in a micropayment category. For directing people from micropayment scheme to multipurpose scheme, the possible starting point is to focus on encouraging more passengers to use the system (W. K. Tan & Y. J. Tan, 2013). To attract more users and motivate them to use the scheme for payment at the participating

retail outlets, discounts and campaigns can be offered. The sever-based infrastructure can allow retailers and public transport users to meet in different platforms.

Table 2.1 Comparison of publications according to their contents

Studies	Year	Description	Fare Collection and User Management		Transport Media			
			Terminal Based	Server Based	Smart Card	NFC Mobile	Contactless Credit Card	BLE Beacon
Seaborn et al.,	2009	Multimodal public transport journeys in London	√		√			
Kocur	2010	Open Payment for Regional Public Transportation Travel		√	√	√	√	
Keitel	2011	The electronification of transit fare payments					√	
Pelletier et al.,	2011	Smart card data use in public transit	√		√			
W. K. Tan & Y. J. Tan	2013	Single-purpose e-micropayment scheme to multi-purpose scheme	√		√			
Anders, Brakewood & Watkins	2014	Analysis of New Fare Payment Systems in Public Transit		√		√		
Song et al.	2015	Service evaluations utilizing transportation card data	√		√			
Tavilla	2015	Transit Mobile Payments				√	√	
Wallischeck et al.	2015	Comparison of future / next generation fare types		√	√	√	√	√

According to the evaluation result of the articles shown in Table 2.1, the server-based system is a requirement for public transport to pave the way for new technologies to be adopted. The type of intelligent transportation using one media in all places via small server-to-server integration is available. To the best of our knowledge, this is the first study that combines server-based structure with the NFC-enabled mobile phone application. We provide a system to the passengers where they

have the same rights (passenger type as a student, retired or model type as a transfer, check-in check-out (CICO), etc.) existing with a transport card but without having and carrying it. In addition, we show that interoperability of different cities is possible with our structure.

Within the last decade, data mining tasks have been applied on public transportation data for different purposes such as discovering traffic patterns (Kui-ying, Lin & Wen-sheng, 2012; Lipan & Groza, 2010), spatial rules (Yang & Liu, 2012), passenger travel time patterns (Cristóbal, Lorenzo & García, 2015; Sasaki, 2014), and service quality patterns for public transport passengers (J. Oña, R. Oña, & López, 2016; R. Oña, López, Rios & J. Oña, 2014). Several studies have been applied with data mining techniques on transport smart card data (Agard, Morency & Trépanier, 2007; Ma, Wu, Wang, Chen & Liu, 2013; Song et al., 2015). In our study, for the first time, NFC mobile transport data obtained from a server-based fare collection system is analyzed with data mining techniques.

2.2 Field Review

Server-based fare calculation process is the basis of this project. It allows giving transportation media to users in whatever form that is desired. Payment calculated not at the boarding time to the bus but on the host side has quite a different structure and advantages over the traditional fee collecting method. In recent years, emerging technologies, using a single card in all regions to provide access to the national card project, have begun to challenge this type of transmission systems. On the other hand, in geographies where the introduction of new technologies is slow, the host-based infrastructure installation is far from reality. Such areas still continue with the traditional method; even paying money by hand is inevitable. However, Turkey is a country where technology is consumed very quickly. This is our point of view; the system gives the opportunity to change the dynamics.

NFC is a standard for smartphones and similar devices to establish communication using radio frequency from a few centimeters (Coşkun, Ok & Ozdenizci, 2012). The number of NFC module preferred mobile phones is tending to rise in Turkey and

admissibility is high in recent years. Similarly, the use of contactless credit cards in this country found itself a market and began to develop. Typical applications include public transport contactless payment data exchange and more complex communication processes, including installation. Currently, in the states of America, the mag-stripe credit card is still being used. In recent years, the EMV chip has been adopted, and there has been an increase in card usage. Even though these countries are practicing with credit cards working in contactless mode, it is still quite a new issue for the banks and in terms of the use of prepaid cards. The NFC technology is being used in every country but in very limited areas. One of the main reasons for not using NFC as a payment channel is the lack of devices to accept contactless payments. But the countries have adopted EMV and are rapidly moving to close the gap. Furthermore, (actually a contradiction, however) the active share in the use of Apple Pay application belongs to the United States (Smart Card Alliance Report, 2011).

The issue of the use of contactless cards and derivatives in transportation that opens the way to server-based fare calculation is never named only as a host-based computing system. Currently, in many parts of the world, contactless credit card or credit card emulation via NFC devices are used in public transportation. However, the only difference between them is the media types. But the fixed fare amount is deducted from the credit card instantly on the bus upon boarding (First Data, 2010).

Public transport operations that take place in many cities on the world mentioned below are similar. The host-based system to create a base together with any of the server-based fare collection activity has to actually start. In 2012, transport authorities and operators in the states of the America launched a variety of pilot projects in order to have payments in public transport to be made with bank cards (First Data, 2010). Mobile phones are a must for contactless and NFC-enabled banking cards as they become more popular in the next evolution of the payment process.

According to research conducted by the "International Market Research (IMS Research)" company, the integration process will be a dual interface, or simply interface with contactless cards in the pocket of 40% of credit card users in the USA

in 2017 (Spanjaard, 2013). According to the "Federal Deposit Insurance Corporation Research" data, 8.2% of the population does not use any bank cards [<https://www.fdic.gov/householdsurvey/>]. These sources show that credit card usage is tending to rise in the USA also.

Utah Transport Agency (UTA) is the first place that tried to implement an open system in the USA in 2010 according to TCRP Report (Wallischeck et al., 2015). UTA has implemented an open payment, the account-based approach that requires only real-time authentication of the contactless medium on their validators. The price of each trip is calculated at the tap point.

Southeastern Pennsylvania Transportation Authority (SEPTA) started to apply open loop payment structure on their validators in 2013. If the ABT system accepts EMV-based bank cards, it is called open-loop system. It accepts prepaid bank cards. The SEPTA is working on new fare models in public transport. However, it is not yet EMV compliant and does not accept mobile payment.

The applications on Earth are presented and the situation of the field is reported in detail on the Smart Insights report (Spanjaard, 2013) and the Smart Card Alliance report (2011). According to the reports, in New York, New York City Transit (NYC Transit) made an agreement with MasterCard and Citigroup for 30 stations on the Lexington Avenue train line to bring contactless payment capable point of sale terminals. Pay Pass credit cards could be read by these terminals. Payment was held at the drop fee calculated terminals. The years from 2012 to 2015 passed with the introduction of contactless cards. The MTA (Metropolitan Transportation Authority) aims to fully complete the infrastructural changes until 2019. In October of 2011, New Jersey Transit (NJ Transit) tested mobile NFC with contactless terminals. For this purpose, a partnership was established with Google Wallet. The system was created so the boarding fee could be charged immediately. Chicago announced the Ventra testing system in 2013 (Anders et al., 2014). It laid the foundations of the systems for contactless prepaid card banks. Ventra has announced that it will build the infrastructure for NFC payment in the near future.

London is one of the areas where public transportation is used most heavily, so contactless credit cards started to be utilized a little more quickly than the others. As of 2014, contactless credit cards and prepaid cards began to appear in transportation. Apple Pay was announced to begin in a short time in the central railway. Table 2.2 shows a comparison of the existing situation in public transportation among important authorities. Although some of the studies described here are building the base for host-based systems, no one is configured to handle and run all the processes on the host side except London. The work we propose does all the operations on the server side along with the NFC-enabled device. The basic difference between our system and the TFL (Transport for London) system is that TFL enables the usage of Apple Pay. TFL is developing end-to-end server software for authentication services, authorization services, and a fare engine. In Turkey, Apple Pay is not active yet, and we support android-based NFC usage in public transportation with our specific mobile application.

Table 2.2 The situation of the capital transport authorities in USA and UK according to research (Wallischeck et al.) on February, 2015

System	Description	Location	Supported Features			
			Card Based (CB)	Pilot (P)	Closed-Loop (CL)	Contactless Smart Card (CSC)
			Account Based (AB)	Field (F)	Open Loop (OL)	EMV Based Contactless Bank Card NFC Based Mobile Payment
SEPTA	Southeastern Pennsylvania Transportation Authority	Philadelphia	AB	P	OL	CSC - EMV
UTA	Utah Transport Agency	Salt Lake City, Utah	CB	F	OL	CSC - EMV
WTA	Whatcom Transportation Authority	Northwest Washington State	CB	F	CL	CSC
CTA	Chicago Transit Authority	Chicago	CB	F	CL	CSC - EMV
MTA	Metropolitan Transportation Authority	New York, New Jersey	AB	P	CL	CSC - EMV
WMTA	Washington Metropolitan Area Transit Authority	Washington, D.C.	CB	F	CL	CSC - EMV
TFL	London Transport Authority	London	AB	F	OL	CSC - EMV - NFC
TriMet	Tri-County Metropolitan Transportation District of Oregon	Portland	CB	F	OL	CSC - EMV

In addition, the systems described here are examples that only have to work in certain stations and in certain provinces. The infrastructure for independent provincial transportation has not yet started to be established. Today, the aim is to be able to calculate fare according to the distance covered (CICO) with the contactless credit card. It seems that NFC solutions will be widely used as well in the field of contactless credit cards. In the proposed new technology, the realization of the NFC secure element is actualized with the host card emulation (HCE) model on the server. HCE is a new technology that will be much used in this area. And with it, the process of fare calculation on the part of the host will form the infrastructure for open-loop payment systems, moving the system forward from there on.

As a result, server-based fare calculation and account management system generates excitement among the world's leading operators, and the world is rapidly heading in that direction. It is important for the project to be actualized earlier than other competitors in the sector and applied the first in the field.

CHAPTER THREE

SERVER-BASED SYSTEM ARCHITECTURE FOR PUBLIC TRANSPORTATION

This chapter explains the general architecture of our proposed system, Server-Based Public Transportation System (SBPTS), and its modules. SBPTS is a cloud-based fare collection system for public transportation. The validators that are installed in public transport vehicles no longer store any fare tariffs and calculate fares. The fare data for that specific trip is no longer a part of the transaction data anymore. Instead, all fare calculations take place at the host (server) side.

3.1 General Architecture

The main aims of the study are to develop SBPTS and account management infrastructure in public transport and then to analyze the real world data that collected from SBPTS. In this study, the server-based system is exemplified by the NFC technology. The modules that the system must have and their interactions are shown in Figure 3.1. The NFC technology integrated into the server-based fare collection system is applied through a cloud-based payment infrastructure. Payment with a physical transportation card is eliminated with this infrastructure, and instead, payment with NFC-enabled smartphones in public transport is allowed. Also with the Fare Management and Payment Management modules, the virtual card is charged periodically through credit card accounts. So physical recharging of transport cards (i.e. by the dealer or ticket vending machine) is also eliminated. For the solution exemplified by NFC phones, the transport operator creates a virtual Public Transport Card Number (PTCN) to be downloaded to the phone. NFC-enabled phones working in the field can use the system with this number.

The virtual transportation card on NFC is a form of token. The creation of token is processed at the server side, because of this reason, the host-based secure element is used. The asymmetric cryptographic algorithms can be preferred to provide security. The token is a reference between PTCN and account-based credentials.

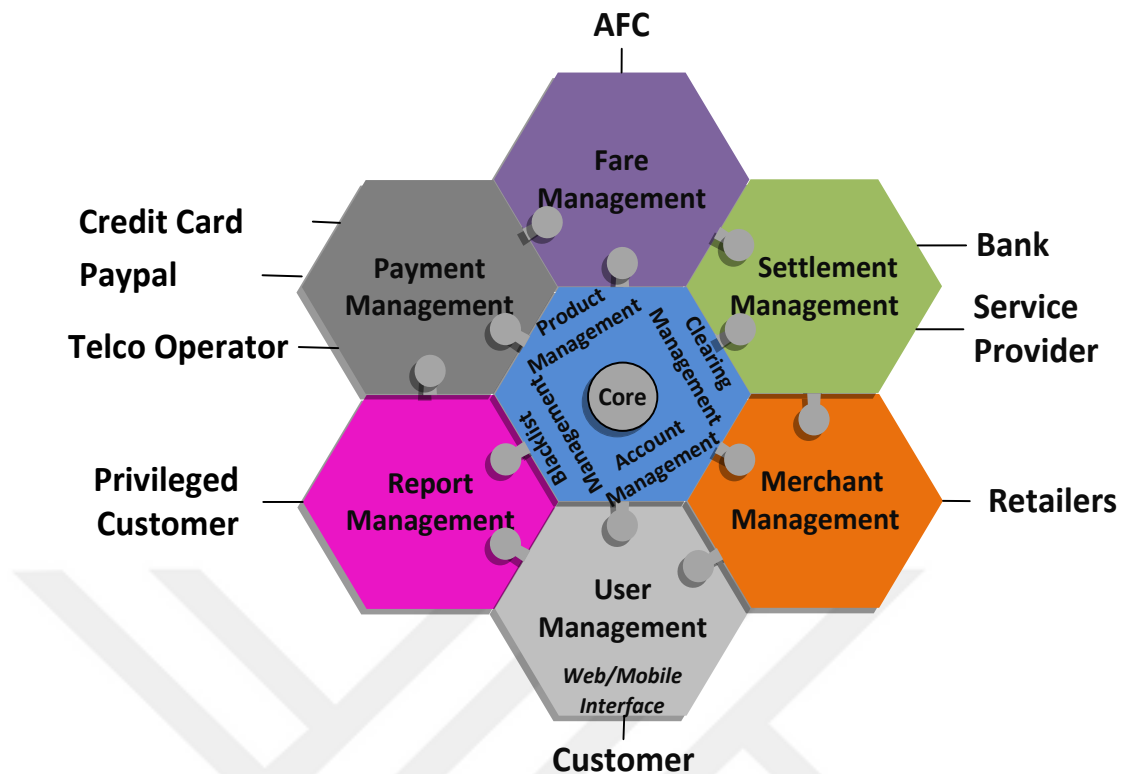


Figure 3.1 The system modules

The validators installed in public transport vehicles no longer store any fare tariffs and calculate fares. The fare data for that specific trip is not a part of the transaction data anymore. Instead, all fare calculations take place at the server-side. The riders have to define a pay channel for the basic host-based computing system. This requires the creation of a bank or PayPal account or fee collection through the integration of telephone operators.

The system is designed to operate as fast as a regular transportation card even if it is NFC-based mobile phone. This is enabled with the high performance terminal hardware. The NFC-enabled phone is read and approved in less than 500 msec.

SBPTS is a complete fare collection/management system and it has numerous modules. Modularity is a design approach to provide manageability of the complex systems. The scope of each module in the proposed system is given in Table 3.1. The modules can be implemented different physical or logical locations. The communication between them needs to be secure and robust.

Table 3.1 The scope of each module in the system

Module	Scope of Module
Fare Management Module	• Calculate fare with Product Management Module
Settlement Management Module	• Calculate remuneration of merchants • Distribute money to merchants
Merchant Management Module	• Create merchant account (name, address, bank account number etc.)
User Management Module	• Login procedure management • User account mapping • User data management
Report Management Module	• Prepare report for all modules
Payment Management Module	• User financial account management (Balance Management) • Trig AMM module ◦ Under limit (pre-paid) ◦ Periodic (post-paid) ◦ Immediately ◦ Retry operation • Account creation management • Membership monthly fee management
Clearing Management Module	• Manage end-of-day operation
Account Management Module (AMM)	• Draw money • Payment instrument management (Paypal, Credit Card, Telcom, etc.) • Payment instrument status management (expire date, lost card, etc.) • Create index on service provider side
Blacklist Management Module	• Manage blacklist on field • Put/Remove PTCN status
Product Management Module	• Create/Save PTCN for any media (NFC, credit card, biometric, etc.) • PTCN-User Management • PTCN-Region Management • PTCN-Right Management • Personalization

The SBPTS is a quite large. From the point of view of the passengers, the only mobile application has a user interface to interact with them. This is the entry point of

the system for the passengers. Passengers start to use the system by downloading mobile application from the store and registering himself/herself to the system. Defining the credit card information is required to start of using the phone as a transport card. For supporting different passenger types and providing personal opportunities, users have to define their identity numbers to the system via the mobile application.

Passenger gets a virtual card from host side by paying card price. When a request is reached to the server, a “token” is created for this request and it allows the passenger to ride in a public transport vehicle. Token has user-specific data and is encrypted with an asymmetric cryptographic algorithm, so only authorized devices can read the information in the token. Token consists of three main information

- Token validity period information
- Virtual card information
- User-specific information

Only virtual card number and simple virtual card information are holding on the mobile phone. There is no wallet on the mobile side. The type of the card and the tickets are holding on the server side. Validators only approve token validity and virtual card number authentication. If there is no problem, the system allows the passenger to ride a vehicle. Differently, from the current systems, fare calculation process is not carried on validator side. Instead of this, the transaction is sent to the server side, the fare structures and card properties are processed on the host immediately and the fare amount is deducted from users’ credit cards.

The mobile application installed on a mobile phone can be used on validators in offline mode which means that it is possible to use a virtual card without connecting to the host servers/internet. However, the server defines offline usage period during a card purchase. When purchasing a card, the host writes the validity period of this card into the token. In our study, the validity period was defined as three days. Thus, users can use their phones on buses or trains without connecting internet for three days. After three days, validator checks validity period and detects that the mobile offline period

is finished. In this situation, verification is not approved, so user has to connect internet on mobile application and synchronize his/her mobile application with the server side.

Host servers take all responsibility of the system, for example, they are responsible for the management of passenger types, fare structures, and extra loyalty issues. Validators always allow the ride and give information to the host immediately. Servers calculate fare price and get money from user's credit card. Figure 3.2 shows the general structure of the system.

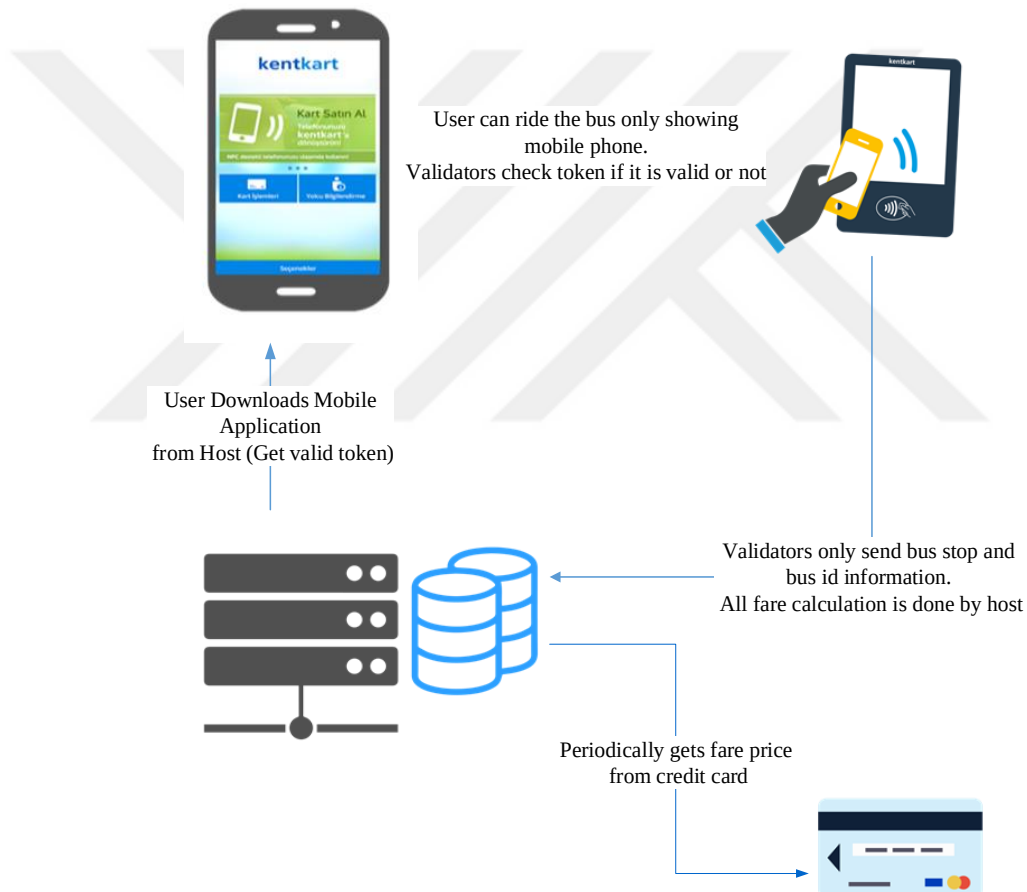


Figure 3.2 General operation structure of SBPTS

3.2 System Components

In this section, the components that should take place in the system are explained in detail. Figure 3.3 shows the components of the architecture and the relations among them roughly. The interactions among components are provided via services with a

Service-Oriented Architecture (Mankovski, 2009) in cloud computing (Erl & Puttini, 2013; Gong, Ying & Lin, 2012).

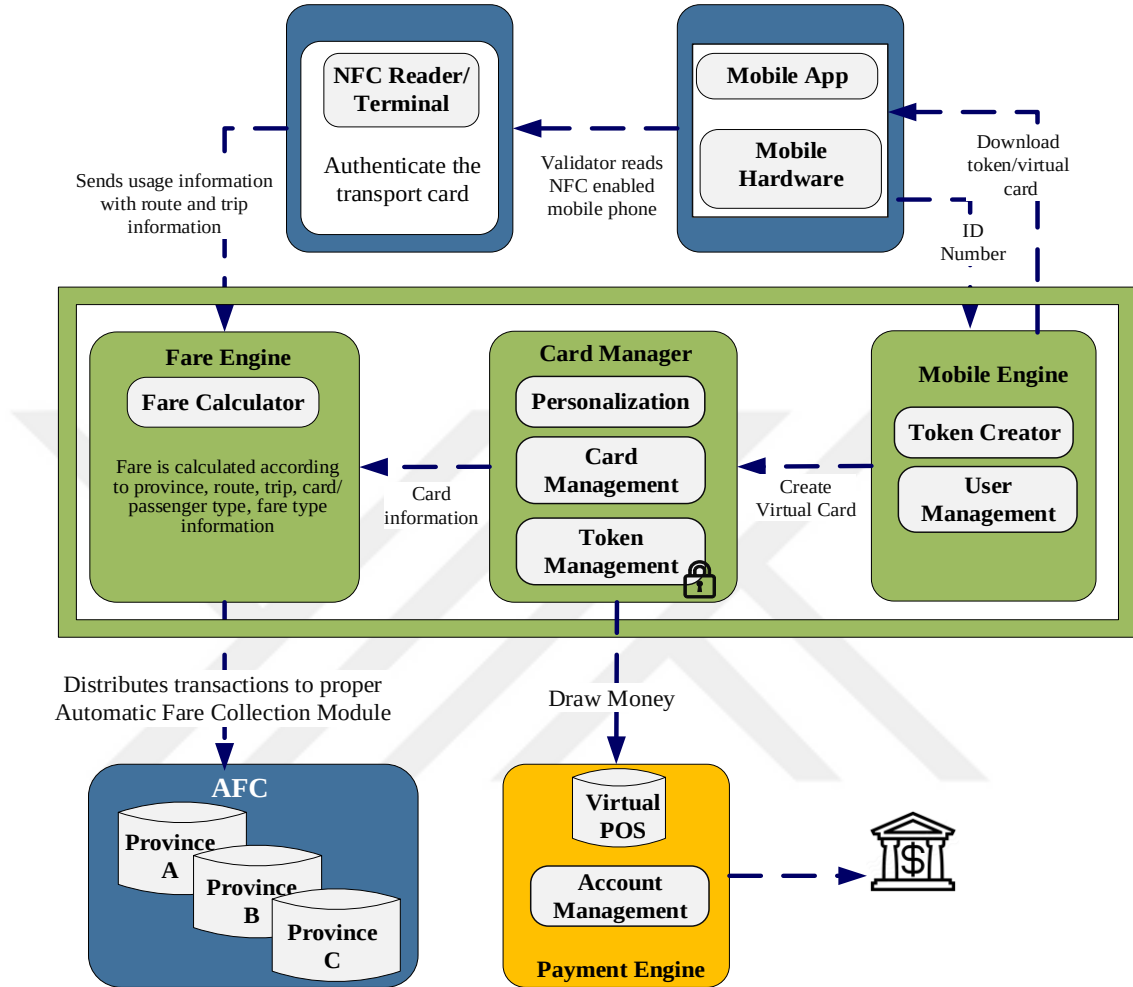


Figure 3.3 The relationship between components

3.2.1 Fare Engine and Fare Calculation Algorithm

The importance of this component precedes all others. Up until now, the transport operator companies always managed fare calculation in the terminal, and it continues to be done as such because all information required for calculating fare is held on the physical card. Fare Management Module is the core of our study that moves all fare calculation from the terminal to the server with all the parameters. The Fare Management Module enables the infrastructure to be used with an NFC phone media or a contactless credit card or maybe with a fingerprint in the future. The aim is to

determine a fare system that offers the same rights as with a physical transport card. The host-based fare calculating system will also determine the fare including all the rights of the user as before. In comparison to other existing infrastructures, the most important change lies in the following point. In the proposed context, the media used does not hold any data or the terminals cannot change any such data. All transactions have to be processed on the server side using information provided by the server.

Overall duties of this component to perform are:

- To calculate boarding fee by analyzing all the information that is possible in a short time
- Existing tariffs and fare modes (transfer, CICO, fixed fee, etc.)
- Passenger types and card types
- To allow easy modification of tariffs by the user with the help of an interface

The most important part of this module is to map fare structure of the province to the database tables. The fare tariff is stored in an XML (eXtensible Markup Language) file. In this file, card type and passenger type are combined to define customer type. For each route group and customer type, all necessary information (i.e. fee, transfer count, and transfer time) to calculate fare is defined in the fare tariff file. The routes, groups, transfer and all other parameters are stored in the database. The design of the database is explained in section 4.4 in detail.

Fare Calculation Algorithm

A complete fare calculation algorithm is essential for a successful public transport system. In the current systems, the reader at the transportation vehicle or at the barrier carries out the fare calculation. However, terminals are often limited, so they can only calculate one set of fare rules. Thus, in contrast to the current systems, this study proposes an intelligent fare calculation service that is working as stand-alone on the server side. Within the scope of this thesis, at first, the already existing fare lists of the provinces were transferred to the database tables by parsing XML. After this transfer operation, the necessary rules for fare calculation were defined at the server side. Upon getting the incoming data from the terminal, immediately after recording the data, an

answer is returned to the terminal in the minimum time period. With this method, the terminal continues to accept new transmissions while the system is processing the received data at the server side.

Fare calculation algorithm uses the fare table statements loaded into the database previously. The rules are executed one after another until a fee is calculated per trip of the individual rider depending on several variables such as payment type, date-time, and passenger and journey information.

As shown in Figure 3.4, fare calculation process begins by finding the necessary fare table version in the database when data first comes. After detection of the card type and passenger type on the table, the process continues on to the related database table to determine the customer type. After customer identification, charge mode identification is available. The table that holds the price of the customer types and criteria is used to determine whether the passenger has a transfer right or not. After finding the fare index, the price chart is obtained for travel at the final stage. The data to be returned such as availability and pricing information are stored in a table in the master engine part.

Mainly there are three steps for calculating fare correctly.

- Validators send virtual card usage to service. First of all, transactions coming from validator are saved to database. After that, Fare Engine (FE) is triggered to calculate fare.
- FE gets uncalculated data from database. For each transaction, it gathers other resources, creates input and sends it to FE. FE calculates fare and writes the result to the database.
- Transmit operation is triggered. All complete transactions are sent to automatic fare collection (AFC) module in XML format.

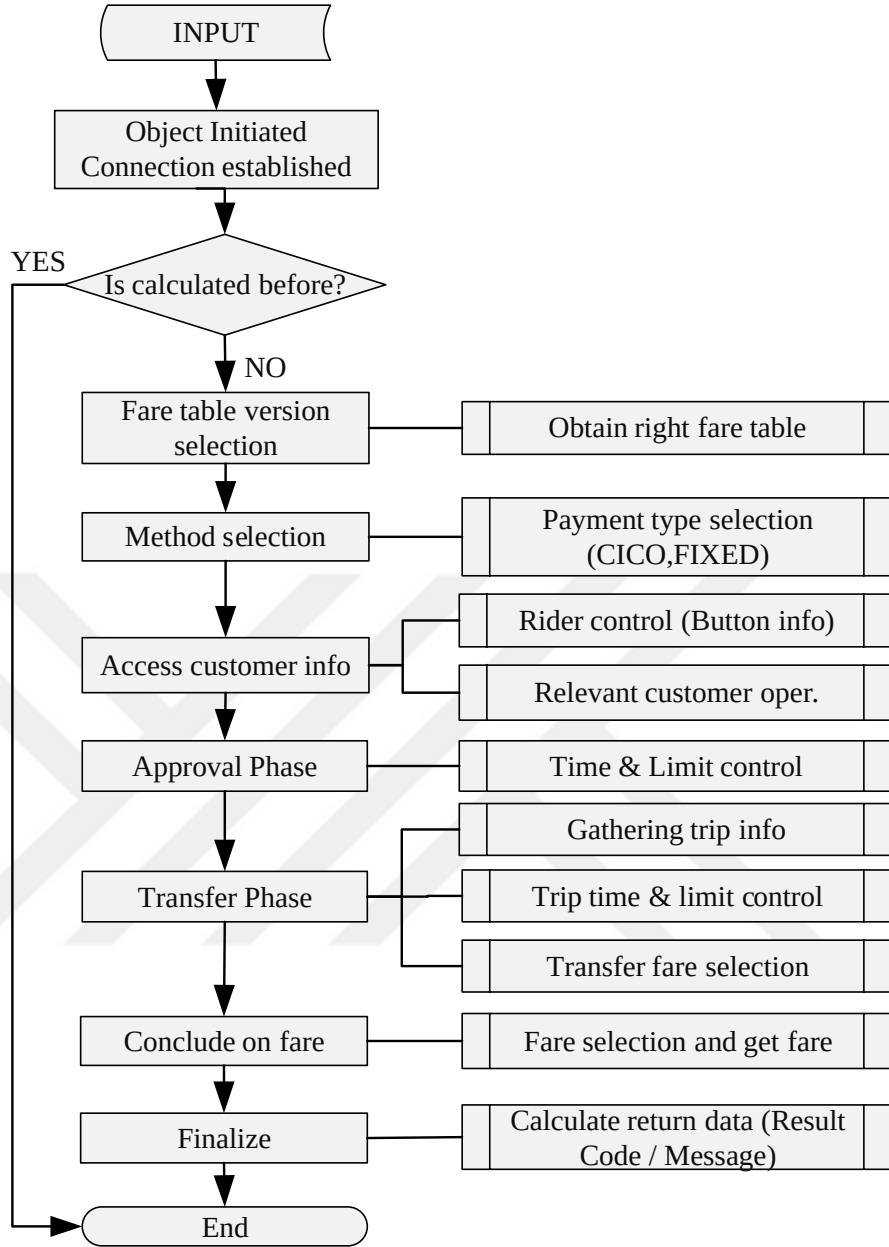


Figure 3.4 High level flowchart for fare calculation

FE has a timer task that runs threats in certain time circulation for getting uncalculated entries from the database, calculating and sending them to AFC.

The fare calculation equation is formalized as in (3.1). Table 3.2 shows the meaning of the symbols at the Equation (3.1).

$$Fare = (\alpha * P_{C_i}(m, e) + \Delta P_1 + \Delta P_2) * \left\lfloor \frac{\Delta T}{\max\{\Delta T, T_{limit}\}} \right\rfloor \quad (3.1)$$

Table 3.2 Notation for fare calculation equation

Symbol	Description
α	weight for passenger type (i.e. adult 1, discounted 0.8, student 0.6 etc.)
P	transportation price matrix for standard contactless fare media
C_i	i^{th} city or province
m	intermodal ticketing service (i.e. train, bus, ferry etc.)
e	fare tariff type (i.e. CICO, fixed fare etc.)
ΔP_1	fare media types (i.e. tickets, tokens, cash, NFC-enabled phone, contactless EMV cards etc.)
ΔP_2	fare increment that is differentiated for transportation policy (i.e. airport transfer, long paths etc.)
T_x	boarding time
T_c	current time
ΔT	$T_c - \max_{0 \leq x \leq N} T_x$ where N is the number of maximum transfer limit. $\Delta T > 0$
T_{limit}	transfer time limit for city C_i (i.e. one hour transfer limit)

3.2.2 Automatic Fare Collection

The main duty of this component is to manage province routes, line structure, bus stop id, names, etc. It is carried out with a database(s) that contains all transportation related data. Automatic fare collection (AFC) communicates with other modules via web services. The main tasks performed by this component are listed below:

- Keeping all card information belonging to the province
- Managing the card type of passengers
- Storing transaction data coming from field
- Holding all lines, routes, bus stops, and coordinates
- Performing vehicle management

This module has its own database schema on Oracle for each province. All transactions belonging to ride on the field are held on these schemas. Because of the huge data size, the management of this database is very important and critical.

AFC module posts the related NFC usages to the card management and clearing management modules. AFC module has already existed in the closed-loop system. For implementing SBPTS, some extensions were added to it. The interaction between new AFC module and other modules were provided by developing new web services. In

addition, a new PTCN number format was integrated into already existing database column format. However, this modification affected the available table structures of the database in many ways. For this reason, some parts of the database on the working system are redesigned.

3.2.3 Card Manager

The main duty of the component is to define the PTCN number and generate a token. In addition, management of the virtual card balance is also performed by this component. It is implemented on the cloud system. The main tasks performed by this component are listed below:

- Creating a virtual card
- For the pre-paid system
 - Monitoring and detection of specified limits according to the conditions of virtual card balance
 - Activating withdrawals from the user in case of being under limit
 - Deducting the fare amount from the balance of the virtual card
- Blacklist management.

For the pre-paid system, activating withdrawals from the user in case of being under limit as the following:

```
for each card in virtual_card_list
  if (card.balance < under_limit)
    if (WithdrawMoney(card.owner.account,
      predef_amount))
      card.balance += predef_amount
      Blacklist.remove(card)
    else
      Blacklist.add(card)
    end if
  end if
end if
end
```

The card manager is designed for all types of virtual cards. The functions are very important and it is located in the middle of SBPTS.

3.2.4 Payment Engine

This component is designed to perform two main operations: one is to manage money flow of the user accounts and the other is to distribute revenue to the operators.

Payment Engine is working with NFC Card Management System (NCAMS) and Financial Account Management (FAM). All transaction fees are calculated in NCAMS, and money is withdrawn from the credit card or deducted from the user account by FAM. The type of the system (pre-paid or post-paid) is important for the payment procedure. If the system is established as *pre-paid*, this means that the user has an account on the system, and the Payment Engine always checks this account whether enough money exists or not. If the account is under the limit, the predefined amount is withdrawn from the credit card or any other payment tool defined by the user. If the system is established as *post-paid*, the sum of the transport fee is calculated and withdrawn from the payment tool. In the post-paid model, NCAMS is designed to collect all usages. Then FAM tries withdrawal from the account of the user during certain periods of the day. The post-paid model is based on a risk-sharing model with agreed companies, as the risk is more obvious.

Overall duties of this component to perform are:

- Create user payment account in the system
- Keep payment tool information such as credit card or PayPal, accordance with Payment Card Industry Data Security Standard (PCI DSS, 2013) rules
- Be able to manage different payment channels: credit card, PayPal, telephone bills etc.
- Withdraw money on behalf of the user on instructions given by the user
- If necessary, process reversal operations
- Manage blacklist according to the result of the withdrawal
- Inform user about money operations by way of e-mail and short message service (SMS).

Public transport may be provided by more than one operator: private transport operators and/or local government authority. For example, in the London underground,

the metros are operated by many companies. There is a backend processing center in this engine for clearing transactions. At the end of each day, all transactions that are collected from the field are grouped, processed and reported according to the operators. The relevant operator is easily detected from the transaction because each transaction has vehicle id and associated information. The daily revenue is distributed to the operators by the next day and settled into their bank accounts. This operation is independent of the collection process of passenger boarding fares. In case a user journeys on a path where vehicles of more than one operator run, each progress payment is made to the relevant operating company without any problem. The main service provider is responsible for distributing revenue to the operators over transactions. At the end of each day, following algorithm is executed once.

```

for each province in province_list
  for each transaction where transaction.status=false
    for each operator in province.operator_list
      if (transaction.vehicle_id==operator.vehicle_id)
        transaction.operator = operator.id
        transaction.status = true
        operator.revenue+=transaction.fee
      end if
    end
  end
end
end

```

Withdrawing money from user account depends on 3D-Secure (3DS) model of virtual POS system. EMVCo (six member organization - American Express, Discover, JCB, MasterCard, UnionPay, and Visa) and the Payment Card Industry Security Standard Council (PCI SSC) are collaborating the 3DS 2.0 specification (Europay, MasterCard and Visa [EMV], 2016). 3DS is a messaging protocol that enables customers to authenticate with their card issues when making online transactions. Figure 3.5 shows the flow of 3DS used in our study.

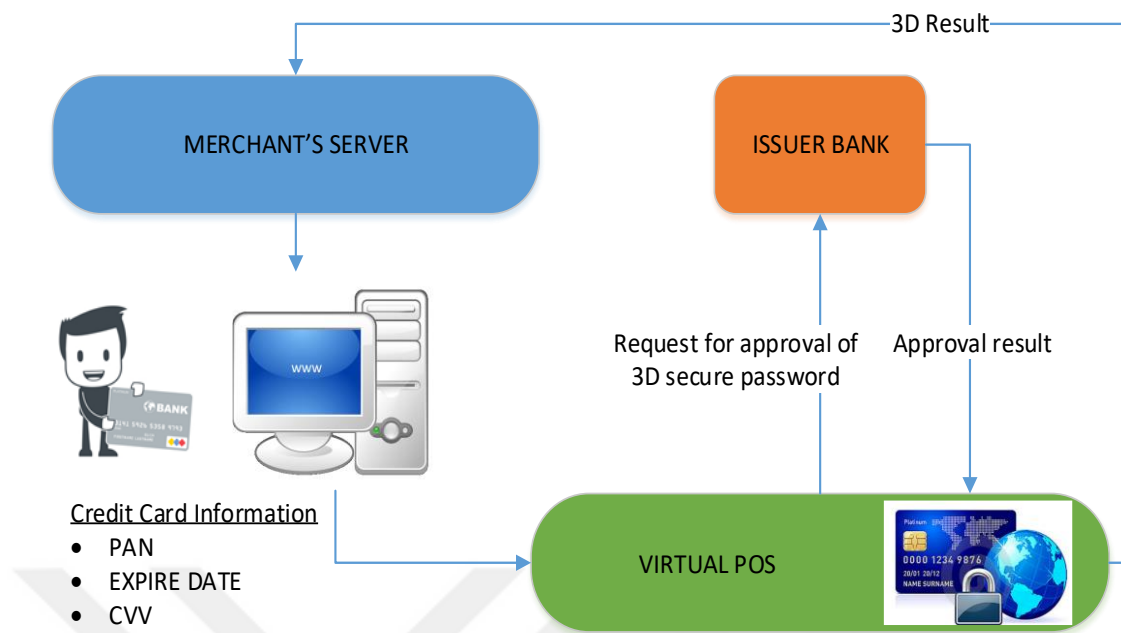


Figure 3.5 Flow of the 3DS protocol

3.2.5 Mobil Engine

Mobile Engine (ME) component serves to communicate and interact with mobile users through the mobile operator server systems. The main tasks of this module are:

- Allowing mobile users to login
- Keeping and managing SMS, email, and password information of the users
- In case of the NFC card purchase, provision of token information to the mobile phone
- Token management as in Host Card Emulation (HCE)
- Managing status of token in case of money not withdrawn (i.e. invalid token download)
- Conveying operations carried out by users via mobile application to relevant departments. For example, demand of detailed document for payment transactions.

The process of converting NFC mobile phone to virtual card is as the following:

Step 1: Mobile user logs in to the application with GSM (global system for mobile) number, email, and password.

Step 2: Mobile application sends One Time Password (OTP) to the user

Step 3: User enters OTP from screen

Step 4: User creates account for withdrawals

Step 5: User selects to purchase virtual card and accepts user agreement.

Step 6: Money is withdrawn from user account

Step 7: Secure token is created and downloaded to the mobile application.

Step 8: Virtual card information is shown on the screen.

3.2.6 Terminal Engine

The terminals running on public transport vehicles have to support NFC communication. They approve the security of the token and create transaction information to send to the host servers. The terminal engine component is developed to run on terminal hardware. Overall duties of this component to perform are:

- Communicating on NFC Standard
- Blacklist control
- Reading token from phone and approving according to HCE standards
- Working in offline and online mode
- Fulfilling verification of token information via NFC phones
- Transmitting boarding information (date-time, bus code, bus stop code, coordinates, route code etc.) to the management server in near real-time information

After token file is read, the following algorithm is executed by the terminal.

```

if (IsValid(token))
  if ((phone.owner==token.card_owner) &
      (Blacklist.contains(token.card) &
       (CheckOrigin(token))))
    if (token.startdate<=sysdate<=token.enddate)
      transaction>CreateTransaction(token.card);
      Send(trasaction)
    end if
  else
    ShowMessage(error);
  end if
else
  ShowMessage(error);
end if

```

3.3 Security Infrastructure

In this thesis, the security concept is taken into consideration in five different subjects: (i) Secure element in the cloud, (ii) Token security, (iii) Token-based user authentication, (iv) Network security and (v) Payment security. The remaining part of this chapter explains each of them separately.

3.3.1 Cloud-Based Secure Element

Mobile applications have an important role to protect sensitive data such as credit card information, token, ticket, and personal information. To protect them against to internal and external attacks, the risk analysis should be performed and secure methods should be listed. In this study, the sensitive data consists of two main parts: payment media information and token.

There is a new solution, called the white-box library, for an android phone. This library protects sensitive data and holds the keys on encrypted database. It is a kind of small Hardware Security Module (HSM) that provides an encryption on specific keys. However, there is no any free white-box library. In addition, the support limitations of the library make the development difficult.

In order to store sensitive data on the phone securely, there is a structure called Secure Element (SE). SE provides secure access to the stored sensitive data. Confidential information can be safely stored on the SE. SEs can only access applications signed with pre-approved certificates. The SEs can be found on

- subscriber identity module (SIM)
- micro SD cards that are plugged into the mobile device
- mobile device
- server side

When considering holding SE on the SIM, the time and financial losses are very high, since it requires signing separate agreements with each SIM provider. In the

second and third choices, the phone can be hacked, information can be obtained and a virtual card or free pass right can be created as an illegal way. For these reasons, as in the case of the Google Wallet, we decided to keep the Secure Element at the server side, instead of the inside of the phone or SIM. Server-based SE provides independency from phone and SIM provider. At the server side, HSM device is used for all encryption processes. In this thesis, cloud-based SE was used, which has the similar structure with Host Card Emulation (HCE) solution.

Figure 3.6 shows the location and communication of cloud-based SE. Application provider installs all security infrastructures on the cloud. The main item of the SE is the HSM device. The mobile application only connects with SE over a secure network for all encryption and key management issues.

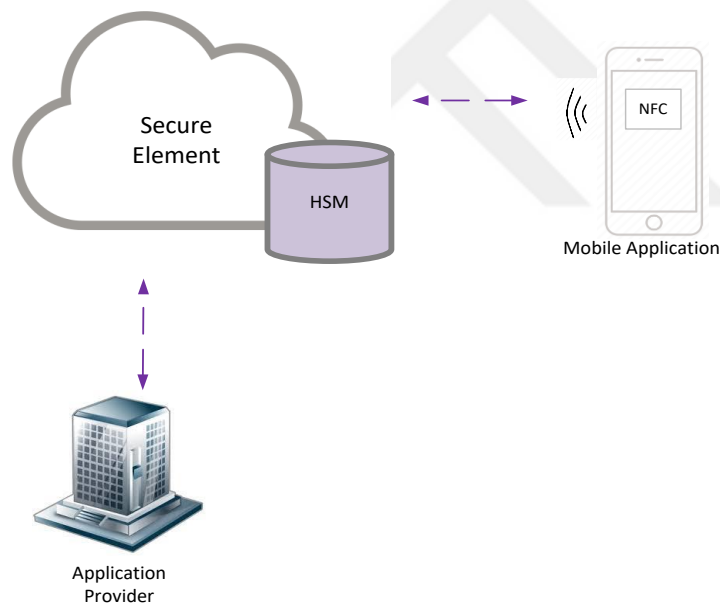


Figure 3.6 Cloud-based secure element

HSMs are special devices designed to increase the security of the encryption process. They are used to protect highly sensitive information such as a private key. HSM creates the cryptographic keys from the data getting from different sources. It holds the keys and never opens it. When there is a need to a cryptographic operation, the secure channel is established between client and server. Data exchanges are provided after the necessary authentication operations are performed on this secure

channel. The token parameters are transmitted to the HSM over this secure channel, and encrypted with RSA (Rivest-Shamir-Adleman) and then transmitted back to the requester. At this point, completely safe, unbreakable structure is obtained. By this way, the security risk is also reduced by defense systems against attack to the servers. As a result, high security infrastructure is provided by the host based SE and HSMs.

3.3.2 Token Security

In this thesis, the SBPTS was designed to install on cloud-based servers. Each system unit except mobile application was provided at the back-end (at the server side). The operations performed at the server side are black-box and they are not opened to outside world. The only mobile application is the point of interface with the users. Due to the hidden nature of sensitive data, it is necessary to protect them by physical and logical ways. The usage of cloud-based SE is the physical part of the protection. At the logical side, the token structure has to be manageable and controllable. SBPTS is intended to ensure that the token is created by the application provider and it is not edited after creation.

Token security is one of the responsibilities of Card Manager Module. This module creates the token and encrypts it. Each token has the following parameters:

- *Unique token identifier number*: Token ID is generated in a logical structure. When a token identifier is read, the creation date and the creator of the token are identified.
- *Application identifier number*: Each application on the store has an identifier. This identifier is also present in the token. By this information, SBPTS ensures that the token is created by approved mobile application.
- *Cryptography version*: The type of the encryption algorithm is defined by this parameter. Currently, the asymmetric RSA cryptography algorithm is used, however, in the future; the algorithm may be changed with another algorithm i.e. elliptic curve cryptography. For this reason, this parameter provides the flexible structure for possible future changes.
- *Key version*: This parameter holds the type of the key that is used on cryptographic operations. Currently, the 1024 bit RSA key is used, however,

in the future; the key length or value may be changed. Therefore, this parameter is necessary for changing cryptographic operations over time.

- *User-specific data:* The user specific data is very important for providing user specific token creation. If a user changes his/her phone or copies the token information to another phone, the token authorization process easily detects it.
- *Card information:* The card information is the largest one among all these parameters. It contains card creation date, card type, unique card number, and validity period. At the end of this container, the secure MAC (Message Authentication Code) that is generated by the symmetric encryption of card information is added.
- *Period information:* Period information defines validity period of the token because each token should be valid only in a limited period. It contains two-date information: start date of validity and end date of validity.

The parameters described above combined with TLV (type-length-value) format and encrypted with 1024 bit key RSA algorithm. The resulting data is called as the *token*. The token is read-only data and cannot be changed on the field. It can be only read when it is decrypted with the public key.

3.3.3 Token-Based User Authentication

SBPTS sends One Time Password (OTP) to the users when they are logging in. The user needs to specify his/her email address and sets a password for himself/herself. Any user who identifies the password and specifies the email address has an account on the system. The system uses mobile phone numbers, email addresses, and self-defined passwords to identify users uniquely. After this process, verification SMS is sent to the user. SMS approval is the basic authentication methods. For this reason, this method has been chosen for the security of user accounts.

The user information is stored in the database as a hash version. The SHA-2 (Secure Hash Algorithm 2) method is used for the hash operation. Session management is provided by the OAuth-2 protocol and tokenization. Users have to be authenticated to

SBPTS over the mobile application. Accounts are managed by the token-based structure. Username and password are entered one time to obtain an authentication token. This token is valid for a specific resource and specific time period. It is defined as 24 hours in our system. When the user opens the application, the already obtained token is passed the system at the background.

In this study, JSON Web Token (JWT, 2015) was used, which was encrypted with private/public key pair. The encrypted text can only be created by a private key. The receiver is only responsible for validating it by decrypting it with a public key. As a result, no one uses or recreates the token that belongs to another one.

3.3.4 Network Security

In this thesis, all services have been written in Web Service Definition Language (WSDL). Each service deployed after providing security conditions. There are two types of security preventions: server-based constraint and service-based constraint.

Server-based constraint means that the third person never reads the data that flow on a connection between server and application. This is provided by the encryption messaging protocol on HTTP by installing X.509 digital certificates on server. SSL (Secure Sockets Layer) is a security protocol developed by Netscape to ensure security and confidentiality during data transfer on the network. SSL ensures that the information sent is strictly decipherable only to the correct address. Information is automatically encrypted before it is sent and can only be decrypted by the correct recipient. Verification is done on both sides to protect the confidentiality and integrity of the process and the information. The strength of the encryption method used in the data stream depends on the key length used. Our system uses 2048 bit SSL certificates.

Service-based constraint means that only authenticated persons can call the WSDL. It is provided by coding in service. At each call, requester provides username and password. These are checked by HSM and then WSDL provides result after this control.

3.3.5 Payment Security

The security of the payment related data like credit card information is based on worldwide standards and certificates. PCI DSS (2013) (The Payment Card Industry Data Security Standard) is the standard for the protection of the payment process from start to end. It contains protection of the sensitive data on the mobile application side, server side, bank side and all communication channels. Fulfillment of PCI DSS requirements is very difficult. The certified persons audit the companies for PCI DSS compliance and this process can take months. Because of this reason, companies generally use certified third party processors.

In this study, we used the virtual POS infrastructure of one private bank provided by Asseco SEE (South-Eastern Europe). The sensitive payment data was stored on PCI DSS certified databases of the Asseco SEE. Virtual POS API (point of sale, application program interface) provided by them was embedded directly into the mobile application. Although virtual POS securely stores and processes the card data, at the client side we need to complete Self-Assessment Questionnaire (SAQ v1.1) in order to compliant with PCI. SAQ is a validation tool designed for merchants to demonstrate compliance with PCI.

3.4 Comparison of Proposed System and Traditional Systems

In traditional *closed-loop systems*, all sensitive data (i.e. personal data, balance, tickets, social rights etc.) is held on contactless smartcards issued by the operator. The fare is calculated by the validators at boarding time by getting the relevant information from the card. However, this type of system is not flexible, and it is expensive to manage and integrate it with other corporations. To overcome such limitations, this study proposes an *account-based ticketing system (ABT)* which means that fare is calculated at the server side and passengers may use NFC-enabled mobile phones or credit cards as payment media. The comparison is given in Table 3.3 to clarify the advantages of the proposed server-based system over the traditional card-based system from the perspectives of transportation agencies, passengers, and municipalities.

Table 3.3 Comparison of card and server based systems

	Traditional Card-Based System	SBPTS
Advantages to passengers	Obtaining a transportation media is necessary	Ease of use, e.g. fingerprint, NFC-enabled mobile phone, credit card etc.
	Several media options (card, ticket, coin)	More fare media options
	Transport card is designed to support all fare structures and passenger types on that province	System supports all fare structures and passenger types on any province
	Addresses only province local riders	Addresses also occasional riders like tourists
	Impossible to give discount to specific users, at specific conditions	Supports any type of discount (or campaigns) and free pass according to user-specific situations like user birthdate etc.
	Ticket are stored in plastic cards or papers	Tickets stored virtually in phones are inherently more durable and less likely to be lost
	Environmental problems with the usage of plastic cards, metal tokens or paper tickets	A greener solution without extra plastic, metal or paper usage
Advantages to transportation agency	Depended to fare media.	Independent from fare media.
	Special transport card issuing.	Card issued and managed by banks and other operators, not by the transit agency
	Cross transit agency interoperability is nearly impossible.	Achieving cross transit agency interoperability easy.
	Needs very complicated card and terminal based integration.	Only requires host based integration.
	Fare is calculated at terminal when entering or exiting.	Fare is calculated at the host side (so easier, cheaper and more flexible).
	Transaction time lower than 500 msec.	Transaction time lower than 500 msec.
	Fare equipment is important.	Improved reliability of fare equipment.
	Extra operations are necessary to upgrade the system.	Easy upgrading of the system.
Advantages to municipalities	Used in only one province	Used in all provinces that support host to host communication
	Impossible to apply country card project	Allows country card project
	Advanced reader device	Simple and cheaper reader device
	Data mining on traditional transportation data	Data mining on NFC mobile transportation data for the first time
	No loyalty applications	Loyalty applications
	System installation costly	Reduced installation cost because of the cloud services

Implementing a host-based system helps the transport agency to work with different transport agencies easily. Interoperability is not a difficult process anymore. Furthermore, accepting passengers as inhabitants of the province or municipality provides efficient public service to both the local travelers and the tourists. The new system makes it now possible to provide loyalty applications according to passenger specific transportation habits or special days, which flatter the pass passengers.

In contrast to the existing systems, applying tariffs of the city to each passenger whether he/she has a transport card or not is an invaluable service, and it constitutes the basis of the national card project. In addition, our system has economic advantages in terms of a cheaper reader device, reductions in the cost of system installation and the costs of card issuing phase.



CHAPTER FOUR

DESIGN OF THE SYSTEM

This chapter describes the design phase of SBPTS that contains the design of the database, use-cases, and flows between components. Each component has its own database design and each use-case (i.e. virtual card creation, usage, and synchronization) has different interactions between actors.

4.1 Overall Flow

Considering the user perspective, the utilization of the proposed system contains certain details. The overall flow of the operations is shown in Figure 4.1.

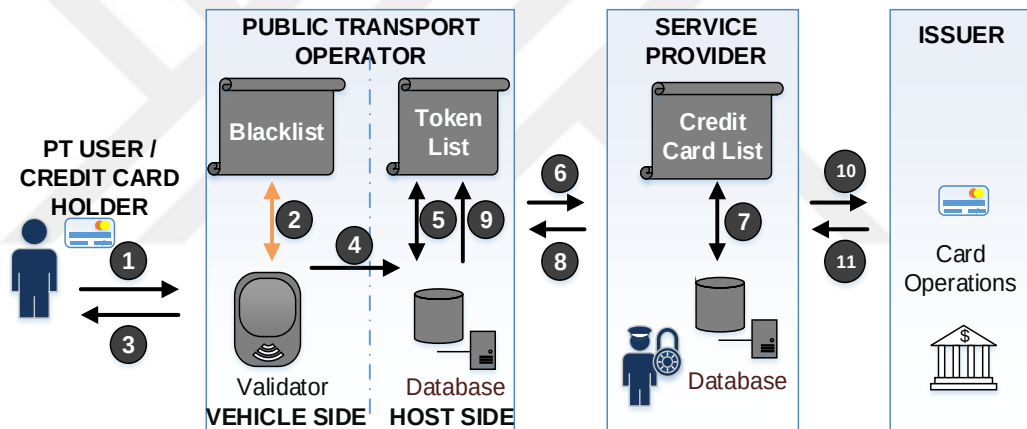


Figure 4.1 Overall operation of the system

The steps which are enumerated at the figure can be explained as the follow:

1. Public transport user uses NFC-enabled phone on the terminal in vehicle.
2. Validator receives the necessary information over the phone, identifies the PTCN number with a process and checks the blacklist with this information.
3. The system allows the passage of each user not in terminal blacklist.
4. Validator posts both the PTCN number and other required information to the server.
5. The server calculates the fee of the travel using passenger type and travel type, and determines whether transfer right or discount etc. exists or not.

6. The total fee and PTCN are passed to the service provider securely according to PCI DSS rules.
7. Service provider transmits PTCN to the database and gets real credit card information. This operation occurs in a more secure environment.
8. The result returns from the service provider to the public transport operator.
9. The result is written to the database and if something goes wrong, the user is blacklisted.
10. Fee calculation is performed and is charged by means of the withdrawal request of the service provider.
11. Results are returned via bank to the related service provider.

The ISO/IEC 14443 RFID communications standard defines contactless card communication with others. It defines the rule of communicating with a contactless card. NFC also supports this communication standard. By adoption of this standard, all terminals can interact with the contactless cards. This provides interoperability all over the world.

Figure 4.2 shows main components and interaction among them in general. Terminals are the usage point of NFC virtual card. The interaction of the virtual card is provided by ISO standards. SBPTS is triggered by terminal side when user getting on the public transportation vehicle. The route information, transaction date time information and virtual card information are sent to SBPTS.

Fare Engine (FE) is moved to the host side and before the transaction goes to the AFC, FE starts to work to calculate fare. This means that terminal and FE are combined and behave like a traditional validator. FE is sent ride transactions to the NCAMS. All usages and charges are managed by NCAMS. FAM is responsible to withdraw money from user account. It is integrated with third parties about credit card, bank card or Paypal account management. NUM (NFC User Management) manages both the interface of the application and user credentials.

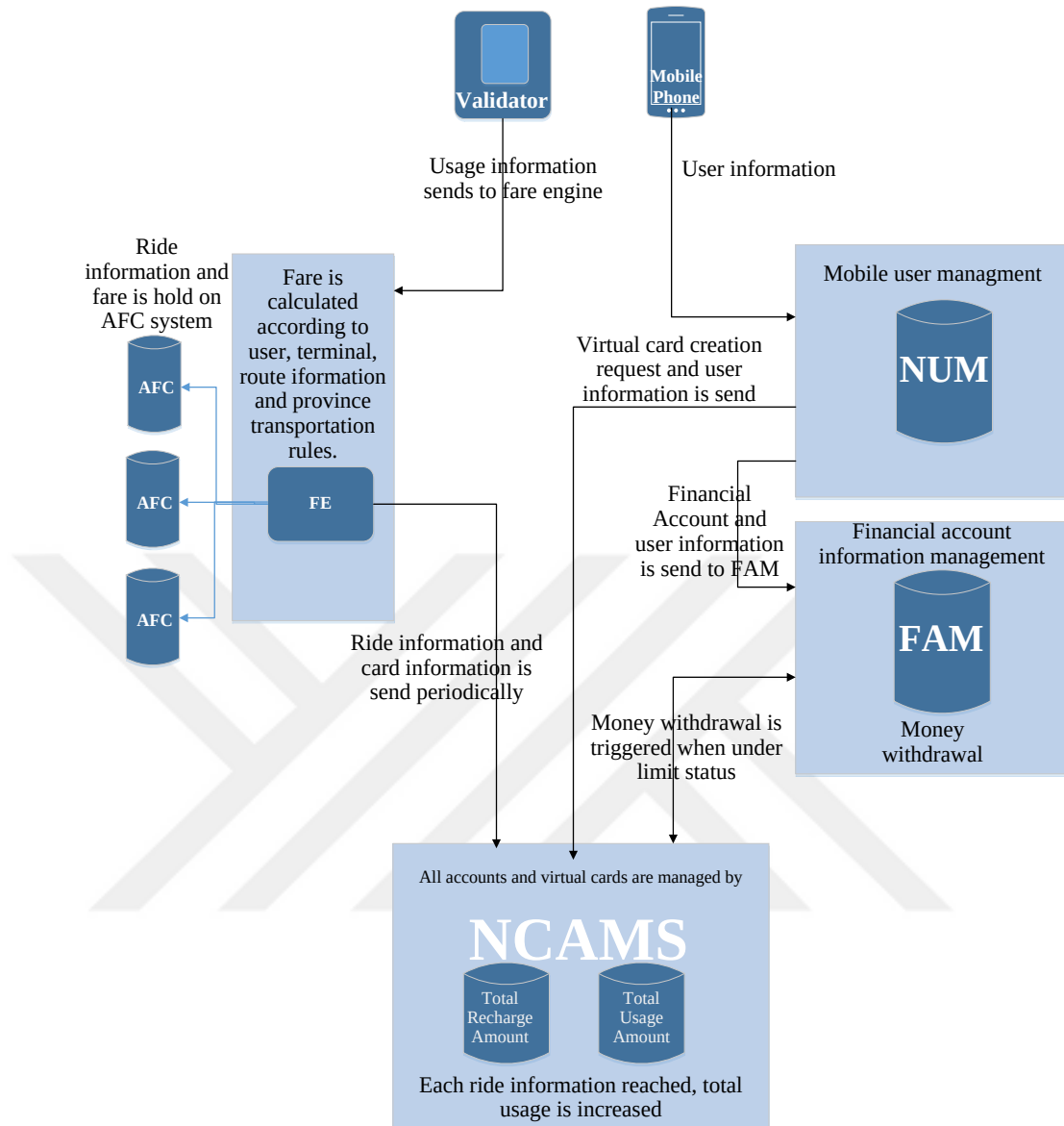


Figure 4.2 General structure and connection of modules in SBPTS

4.2 Use Cases

To clarify requirements, a use case diagram of the host side with actors and their main operations are given in Figure 4.3. This use-case diagram shows the relations and dynamics of our proposed system more clearly. The requirements and relations of four different host placed engines are represented: transport operator, fare engine, card manager and payment engine.

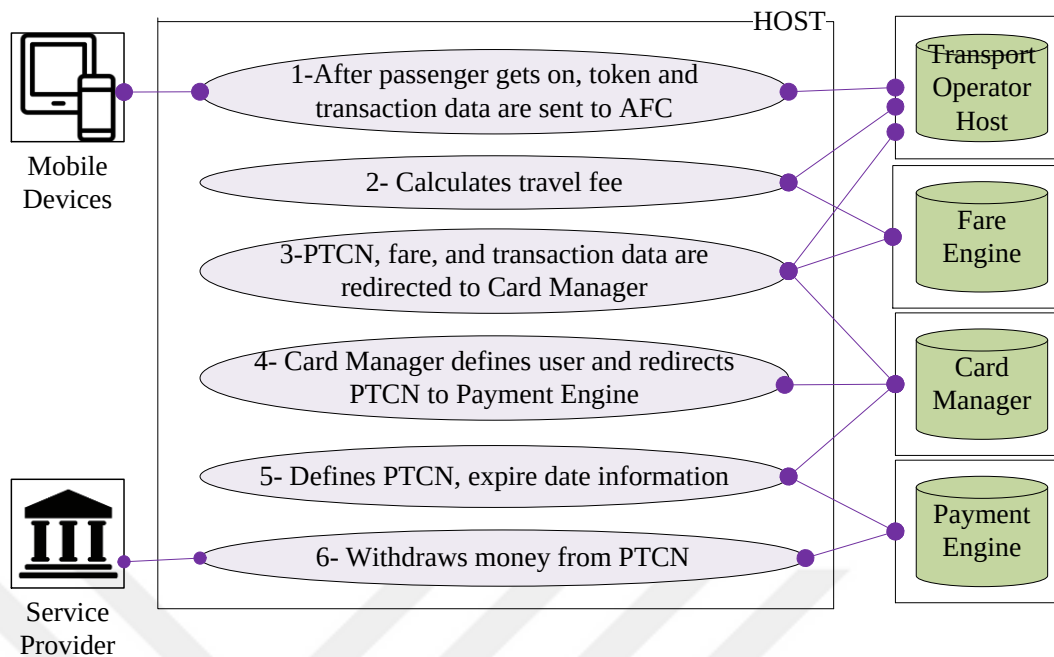


Figure 4.3 Use case diagram of host operations

In the proposed system, the most of the processes are carried on the backend. However, passengers are the users of the system and they also perform several actions. From the user perspective, the main actions are to become a member of the system and the usage of NFC-enabled phone on the vehicle. Figure 4.4 summarizes the registration step of the user.

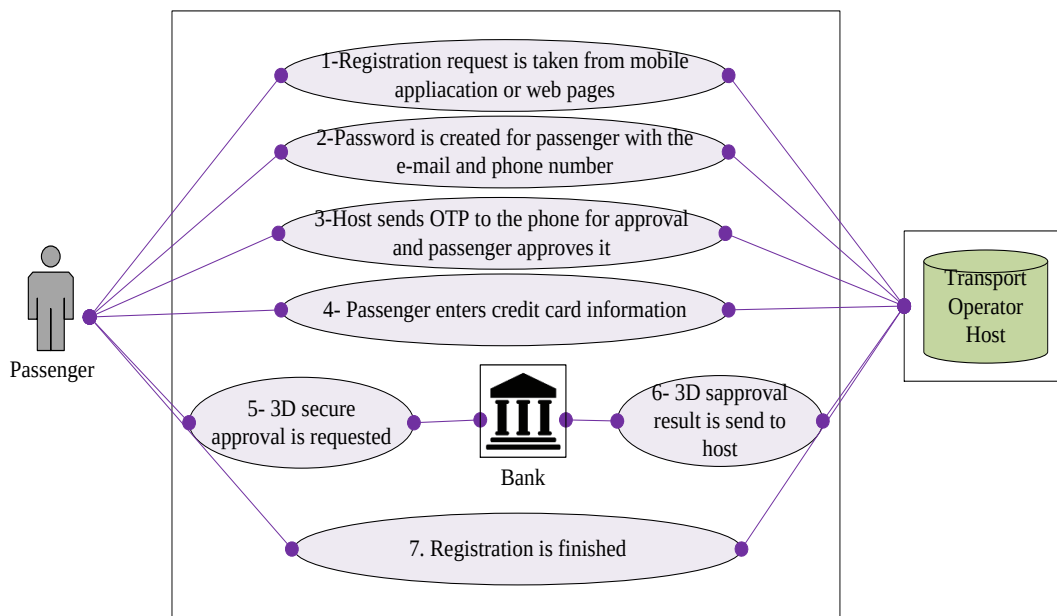


Figure 4.4 Use case diagram of passenger registration

Figure 4.5 shows the use case diagram drawn for the usage of the NFC-enabled phone at the vehicle. When NFC-enabled phone is touched to the validator, in contrast to the traditional transportation systems, the validator does not calculate the fare and deduct the money. Only token is approved by the validator.

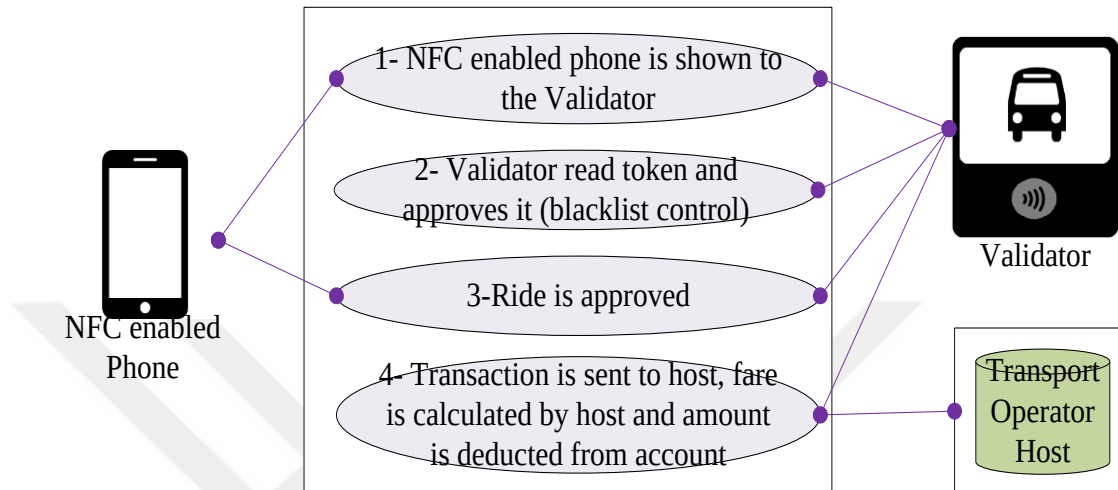


Figure 4.5 Use case diagram of NFC usage on the public transport vehicle

4.3 Sequential Diagrams

4.3.1 Terminal Operations

Terminals are the interaction points of the mobile devices. When a mobile phone is touched to the terminal, the terminal selects specific mobile application applet. The communication protocol that is used between mobile phone and validator is ISO/IEC 7816-4.

ISO/IEC 7816-4 (2013) is an international standard for identification cards. It contains organization information, security issue and commands for interchange. According to this standard, a physical interface between the card and the outside world shall be enabled to support the processing of command-response pairs. The syntax of command-response pair is called application protocol data unit (APDU). There are four APDU command structures as shown in Figure 4.6.

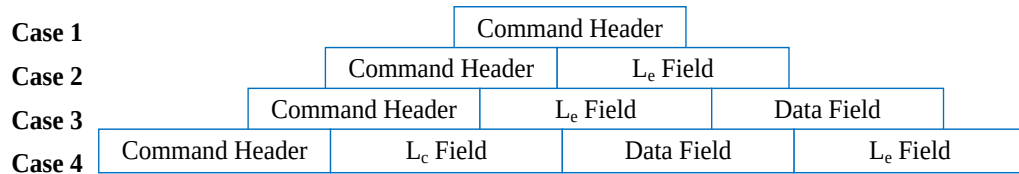


Figure 4.6 APDU command structures

In APDU structure, L_e is the length of the response data, while L_c is the length of the command data. Each APDU structure can be commented as shown in Table 4.1.

Table 4.1 Request-response of APDU structure

Case #	Request	Response
Case 1	No command data field	No response data field
Case 2	No command data field	Response data field
Case 3	Command data field	No response data field
Case 4	Command data field	Response data field

The terminal creates a response for APDU request which is called “response APDU”. The structure of it is shown in Table 4.2. *SW1* and *SW2* are the status words. The status word indicates the result of the operation as follows:

- 90(*SW1*) 00(*SW2*): successful execution
- Other values indicate error or warning condition

The length of the data field of APDU response is given implicitly. Short length can be maximum 256 bytes, while extended length can be maximum 32768 bytes.

Table 4.2 Response APDU structure

Response APDU		
Optional body	Mandatory trailer	
Data Field (variable length)	<i>SW1</i> (1 byte)	<i>SW2</i> (1 byte)

Figure 4.7 shows APDU exchange between reader and card on a contactless channel. The Android Kitkat version and higher versions support ISO 7816-4 standard on their NFC interfaces. In this study, the NFC-enabled phones that have minimum Kitkat version, have to be used on the field.

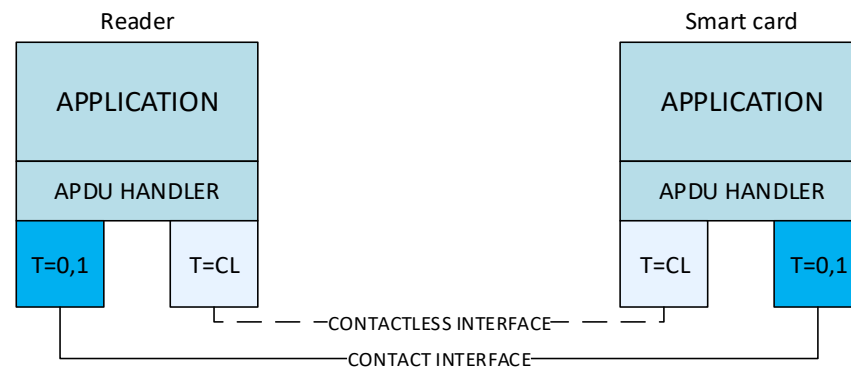
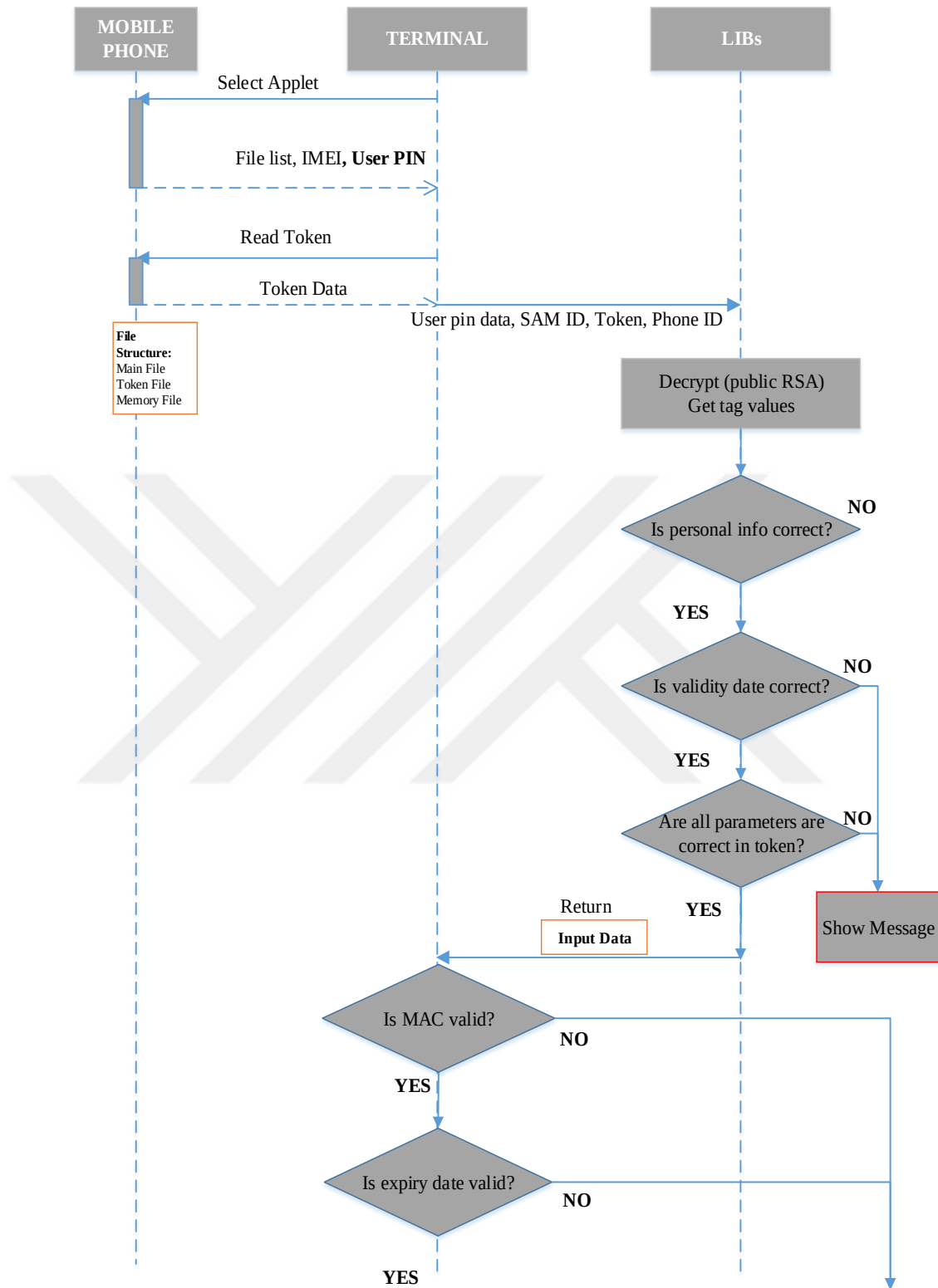


Figure 4.7 APDU exchange

The terminals interact with the NFC-enabled mobile phone on ISO 7816-4 standard. Figure 4.8 shows the sequence and flowchart diagram of card usage steps. Terminal selects proper applet that is written specially and loaded to the mobile phone by the mobile application. Some information about mobile device and user is returned to the terminal. After that, the secure and encrypted token is read. The token is decrypted by the public key and information in the token is compared and checked according to predefined rules. For example, the date and time values are used to check that the token is still valid or not. Similarly, the system checks that the mobile phone and token have the same phone ID or not. After the all conditions are satisfied, the validation procedure is successfully finished. If it is not, the proper message is shown on the terminal screen.



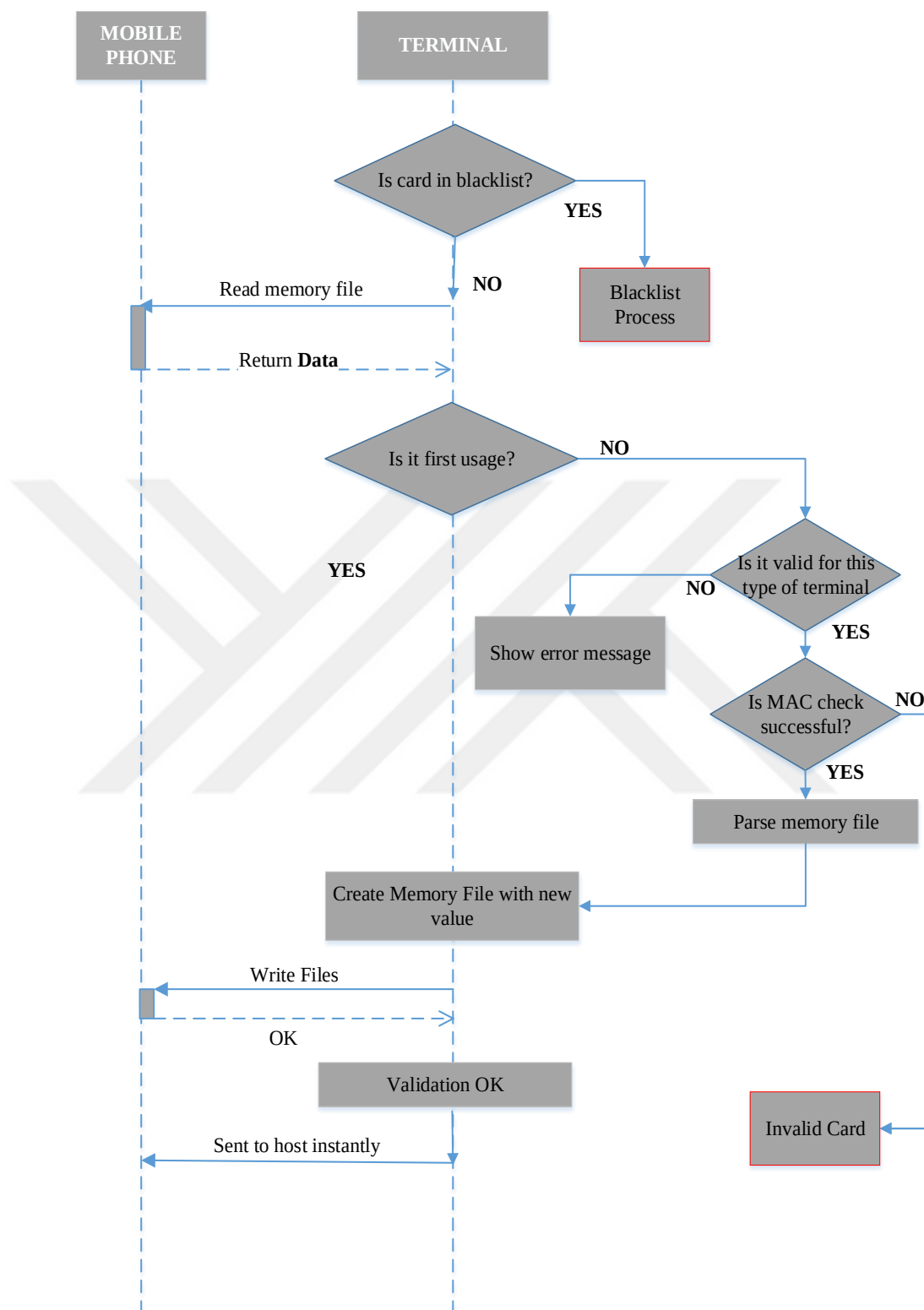


Figure 4.8 Card usage on terminal

4.3.2 Mobile Engine Operations

In the proposed system, when passengers enter a transport vehicle, they use their mobile devices. The mobile device is only connected to Mobile Engine (ME). All interactions between ME and other modules are defined for three basic operations: card production, card sale, and card synchronization.

Figure 4.9 shows the sequence diagram of the interaction between ME and NCAMS when a virtual card is produced. First, the user creates the request from the mobile device. The request is moved to ME. The virtual card is generated by the NCAMS. Each virtual card has a unique PTCN. Besides to PTCN; chip serial number, expire date and type are the basic parameters for defining one virtual card. All this information is packed into one secure data packet. The MAC is also inserted to the end. Finally, the used cryptography algorithm and key version are added to the packet as parameters, because the algorithm and key can be changed in the future.

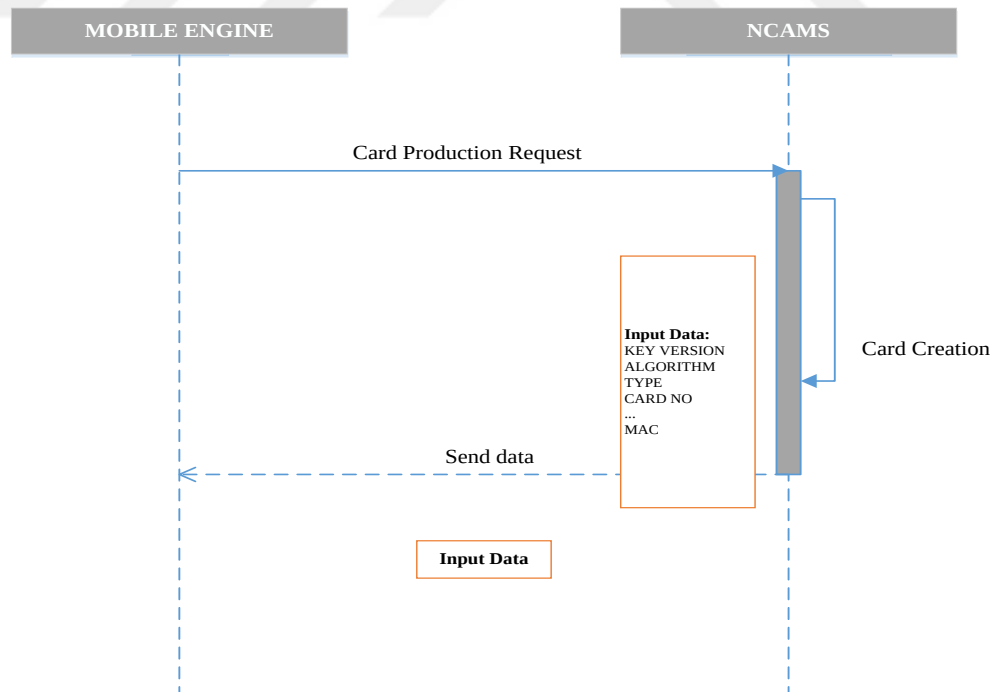


Figure 4.9 Sequence diagram of virtual card production

From the user point of view, the virtual card production is the part of virtual card sale operation. The user starts the operation for purchasing virtual card by giving credit card information. Figure 4.10 shows the steps necessary to follow for virtual card sale operations.

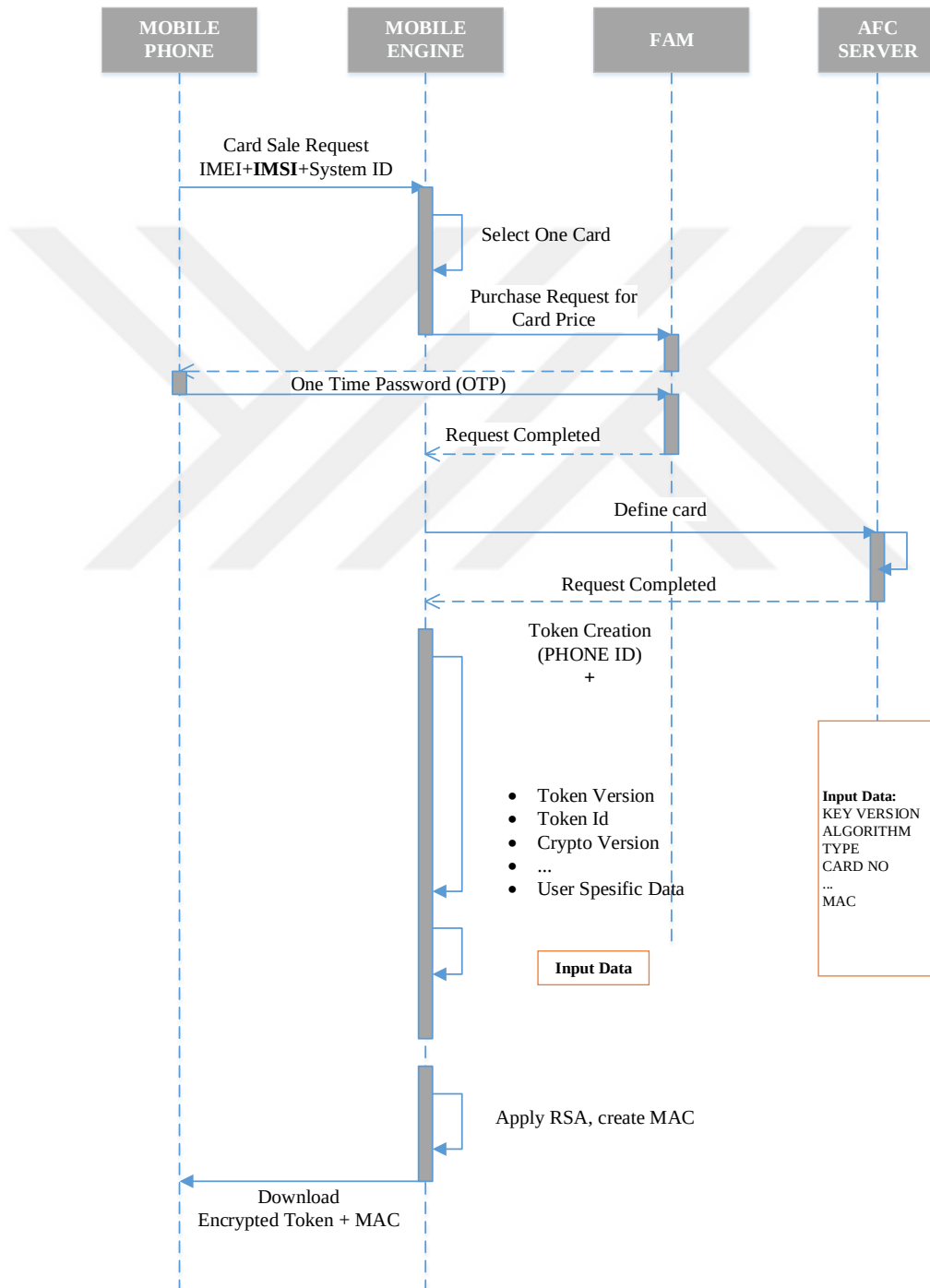


Figure 4.10 Sequence diagram of virtual card sale operations

ME communicates with FAM for all withdrawal operations. Virtual card sale operation contains getting money from user defined accounts. When money withdrawal is completed successfully, the ME creates a secure token that contains virtual card information and many different parameters inside. The general structure of this token is explained in section 3.3.2

ME needs to be synchronize with the server side periodically. The period was currently defined as maximum three days. Mobile application tries to connect server side when network connection is available on the mobile phone for one time in a day. Figure 4.11 shows the sequence diagram of synchronization operation.

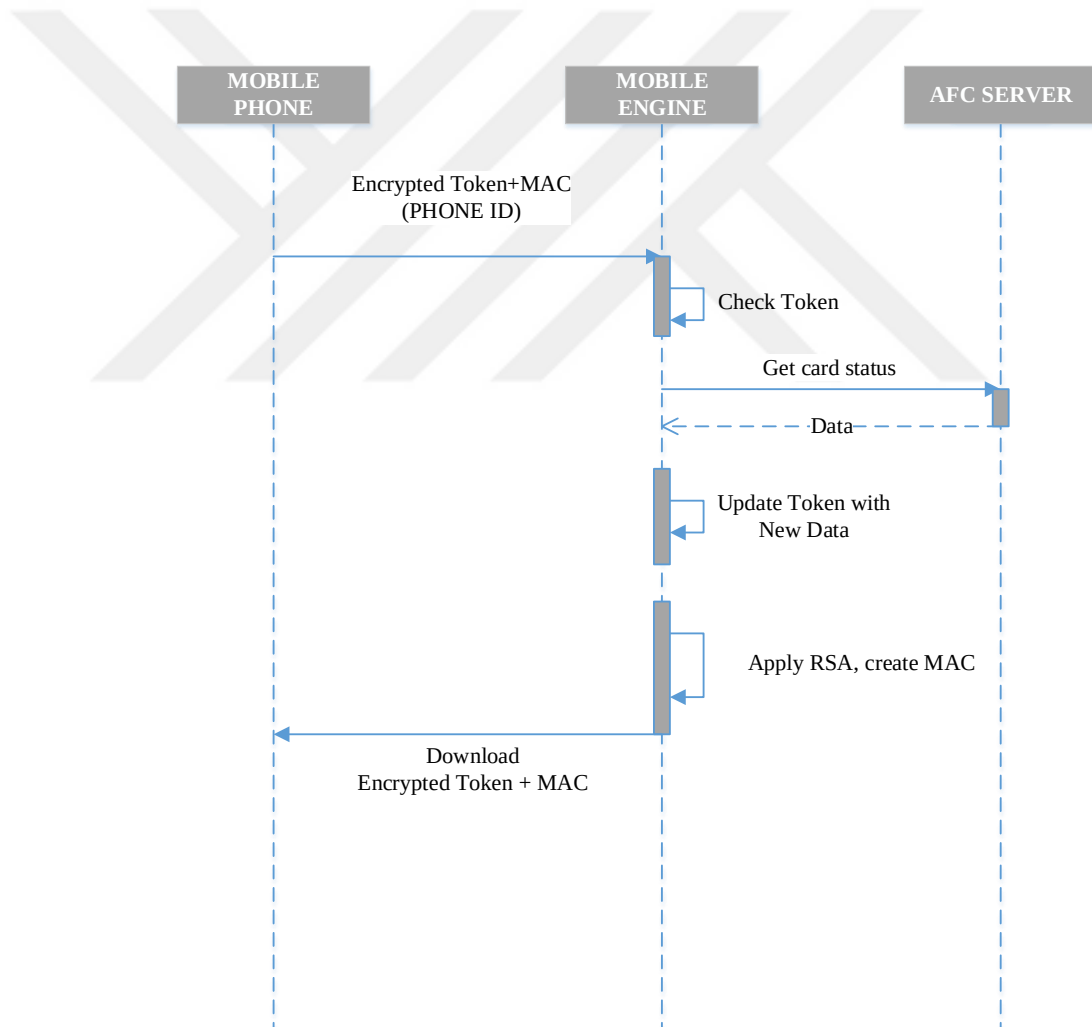


Figure 4.11 Sequence diagram of mobile application synchronization

In every sync operation, token in the mobile phone is checked by the server side. If there is no any new condition, no action is taken. If some conditions are changed (i.e.

The transaction is sent to the NCAMS when all necessary fields are completed. The most important field is fare amount. After the fare amount is calculated by FE based on many parameters such as passenger type, route, fare mode, tariffs and time; the transactions are moved to the NCAMS. There is a scheduled job to do moving transactions between modules. Figure 4.13 shows the flowchart of the decision process, which determines continuously that whether the transactions will be transmitted from FE to NCAMS or not.

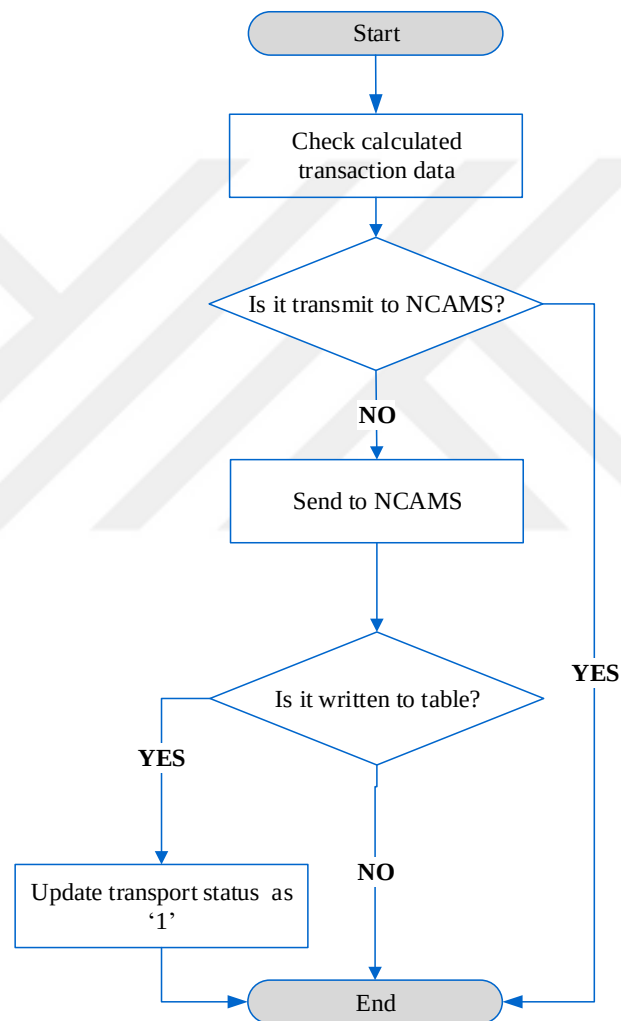


Figure 4.13 Flowchart of transmitting transactions from FE to NCAMS

The interaction between NCAMS and FAM is required during money withdrawal phase, which can be resulted in three statuses:

- *SUCCESS*: In this condition, the virtual card is going to live in a normal case.
- *FAIL*: In this condition, the virtual card is inserted into the blacklist.

- **UNKNOWN:** In this condition, the virtual card is not inserted into the blacklist directly. Instead of this, money withdrawal is tried again in thirty minutes. After three attempts, the card is inserted into the blacklist.

The flowchart of money withdrawal is shown in Figure 4.14. The main condition that should be controlled during money withdrawal is that whether account balance is under the limit or not. If money withdrawal is approved, there are three possible conditions as mentioned above: success, fail, and unknown.

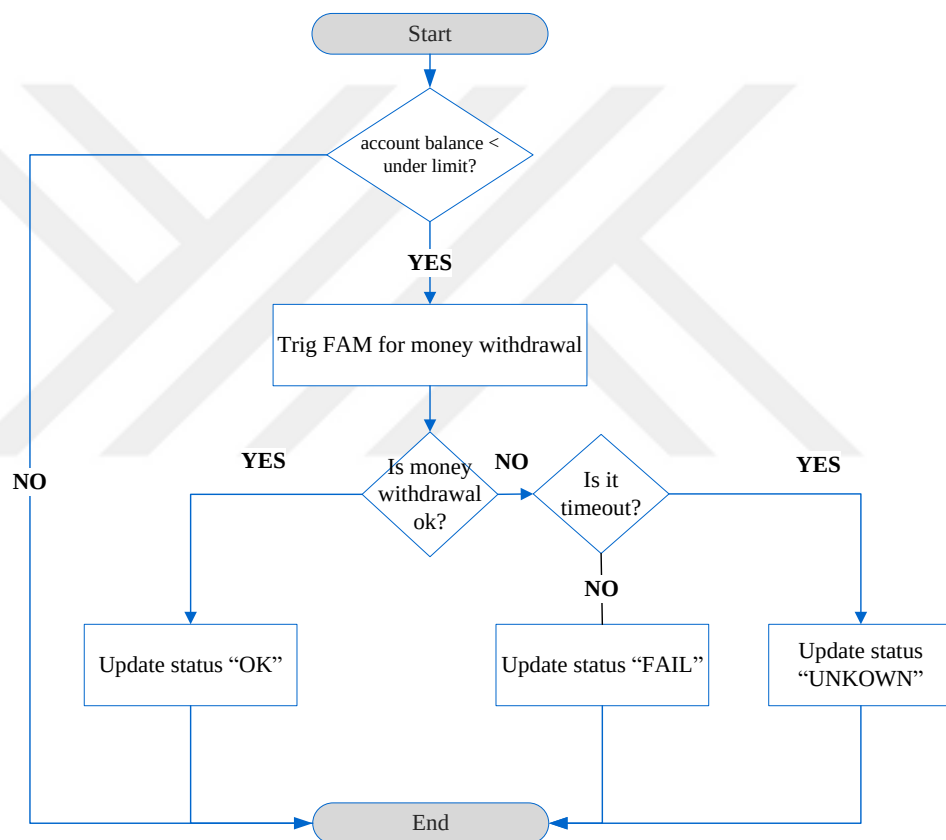


Figure 4.14 Flowchart of the money withdrawal

If the result is *FAIL* or *UNKOWN*, the retry condition is depended on the result code of the bank. Because there are many possible result codes like “expired card”, “stolen card” etc. and an action has to be defined for each condition.

Table 4.3 shows the error codes and retry conditions for each error code. For example, if the card status is *stolen*, the system blacklists the card and never tries again

to get money from the card. If card status is *under the limit*, the system tries two times in one month for checking the card limits.

Table 4.3 Virtual POS return codes and their explanation

Error Code	Description	Retry Condition
04	Stolen card	No retry
05	General rejection	Retry in 30 min
12	Invalid Transaction	No retry
19	Try Again	Retry in 30 min
28	Reject	No retry
41	Lost card	No retry
43	Conditional Stolen card	No retry
51	Under limit	Retry in 2 times in month
57	Un-allowed Transaction	No retry
80	Invalid expire date	Retry in 2 times in month
85	Reject	No retry
92	Time Out	Retry in 30 min
96	System Error	Retry in 30 min
99	General System Error	Retry in 30 min

Modular design helps us to divide the responsibilities between the different parties and makes management easy. For example, the management of problematic situations is the responsibility of FAM. Other modules are isolated from bankcard status problems.

4.4 Database Design

One of the most important points of the design works in this thesis is the design of the server side of the system and the planning of the important software modules. Especially, establishing the database and software infrastructures of the server-side modules is very important. Because, the fare calculation process is moved to the server side, so thousands of vehicles are passed their fare calculation issue to the server side. Therefore, the performance is of vital importance. In the case of server or network shut down, the servers will be kept in a few physically designated locations. It could be a good solution to implement the components of the system on the cloud servers. Cloud

technology is preferred due to its portability and access advantages. When selecting servers, the following criteria should be taken into consideration.

- Accessibility
- Safety
- Regular backup
- Maintenance
- Performance measurability

Within the scope of this thesis, three Amazon servers have been rented in the cloud:

- Oracle database server,
- Web service server,
- MySql database and web server

The part of the database was designated as a structural Oracle database to be able to do critical business fast. Some of the modules of the system were developed by using Oracle Real Application Clusters (RAC) technology, which is a premier technology that can be used with enterprise and standard editions of Oracle Database.

The MySql database was selected to use where the transaction speed is not critical and where there is no need to keep very large data. Database-level packages and functions were written in PL / SQL and SQL languages. There are several server database modules developed in this project to carry out the important issues. However, fare engine and virtual card management modules are the cores of the project and have a vital responsibility. The innovative aspect of the project is provided with these two modules. Figure 4.15 shows the ER (Entity Relationship) diagram of the fare engine module, which was implemented on Oracle database. There are mainly 22 tables in the fare engine module but only important ones are listed in the ER diagram. In the table *FARE_HEADER*, the price lists are held with their versions and region codes. The table basically contains the following attributes: region id, version, start date time, end date time and active days of the week. Each version of fare is valid for a predefined period. The structure of the table allows selecting the versions valid for specific days of the week.

Table *FARE_HEADER_GROUP* groups the lines in the region to enable fast fare calculation. Because, fares in a certain region in a city are the same, so these line groups are the main parameter for fare calculation. Region id, line group id and the count of each tariff are the other basic columns. Each line group has its own fare tariffs for each passenger type (i.e. student, adult, teacher, retired). The passenger type and line group information are not enough for the calculation of the fare. Each line and passenger type have its own fare tariff structure such as fixed, transfer or CICO. Especially, foreign countries use distance-based charging.

Table “*FARE_GROUP*” matches the passengers, line groups, and tariff types. It contains information about types of fare tariff structure. Passenger types can be calculated in different modes according to the line groups. Table *FARE_GROUP_DETAIL* allows passengers to be customized in the same tariff type. Sometimes the passengers of the same tariff mode may need to be calculated in a different way. In this case, different groups that hold in the table are used. Region id, version, line group, passenger type and tariff mode are the other basic parameters.

Table *PAX_GROUP_FARE* contains all necessary parameters to calculate the fare exactly. This table keeps the maximum possible times for the transfer of customers, daily limits required for the calculation of the fee. Tables *PAX_ORIGIN*, *CARD_ORIGIN*, and *TRAVEL_ORIGIN* are the master tables which are used to manage passenger types, card types, and travel types respectively. NFC card is read on the terminals at the public transportation vehicle. Terminals send the transactions of the NFC usages to the table *TRANSACTION* on the fare module schema. The main fields of this table are bus id, bus stop id, NFC card number, transaction date time, passenger type, terminal number, security code, and fare. Some of the columns of the table hold data coming from the terminal, some of the columns are filled with other internal tables. First, the values of all necessary parameters are collected, and then fare engine starts to process and calculates the fares according to the rules defined in the tables. After the fare is calculated, the transaction is completed successfully.

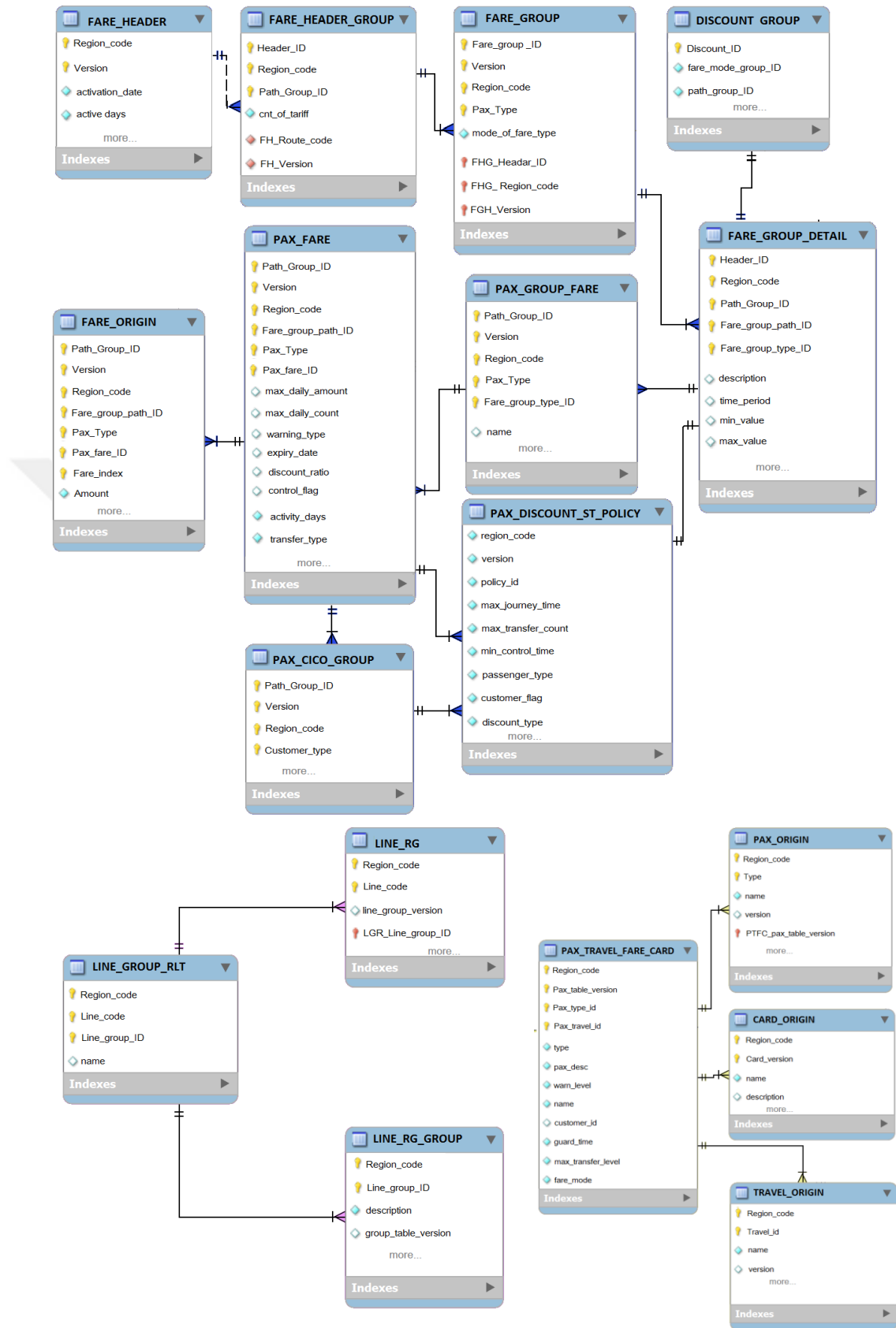


Figure 4.15 ER diagram of fare module

Fare management module has interfaces to manage related tables. Figure 4.16 shows sample administration interfaces for authorized persons.

Fare Lists
Customers
Passengers
Cards
Routes

Home / Customers / Edit

Customers

Assign Passengers to Customers
Save

Description: Tam - Normal
Customer: Tam
Passenger: Tam
Card: Normal
Restricted: True
Passback Time: 1800
Warning Guard Time: 2

Fare Lists
Customers
Passengers
Cards
Routes

Home / Fare Lists / Edit / Fare Group / Edit

Fare Groups

Manage Fare Group
Save

Name: Şehir İçi
Tariff Count: 1
Transfers: ☒ Enabled
Transfer Policy: Transfers Accepted From All Groups Customize...
Customers: Edit Customers...

☒ Tam
Activation Start: 00:00:00 Activation End: 00:00:00 Activity Days: MTWTFSS
Base Amount: 1.50 Warning Type: 1

☒ Transfer Enabled ☒ Allow Same Route
Transfer Amount: 1.20 Max Journey: 3600 Max Count: 2 Min Retrans: 300

Buttons: ☒ Button #1 ☒ Button #2 ☐ Button #3
Change Customer: Tam Öğrenci None
Tariff Amount #1: 2.00 1.50 None

☒ Öğrenci
Activation Start: 00:00:00 Activation End: 00:00:00 Activity Days: MTWTFSS
Base Amount: 1.20 Warning Type: 1

☒ Transfer Enabled ☒ Allow Same Route
Transfer Amount: 1.00 Max Journey: 3600 Max Count: 2 Min Retrans: 300

Buttons: ☒ Button #1 ☒ Button #2 ☐ Button #3
Change Customer: Tam Öğrenci None
Tariff Amount #1: 1.30 1.10 None

Fare Lists
Customers
Passengers
Cards
Routes

Home / Fare Lists / Edit

Edit Fare List

Manage Fare List
Add New Group Add New Customer Save as

Fare Groups	Customer	Mode	Base	Tariff 1	Transfer	...	Action
Şehir İçi	Tam	CICO	1.50	1.00		...	Edit Delete
Şehir Dışı	Tam	Fixed	1.50			...	View Clone
	Öğrenci	Fixed	1.20			...	View Clone

Figure 4.16 Sample user interfaces of fare module

Figure 4.17 shows the ER diagram of the virtual card schema created in the NFC Card Management System (NCAMS) module. This part is also implemented on Oracle database.

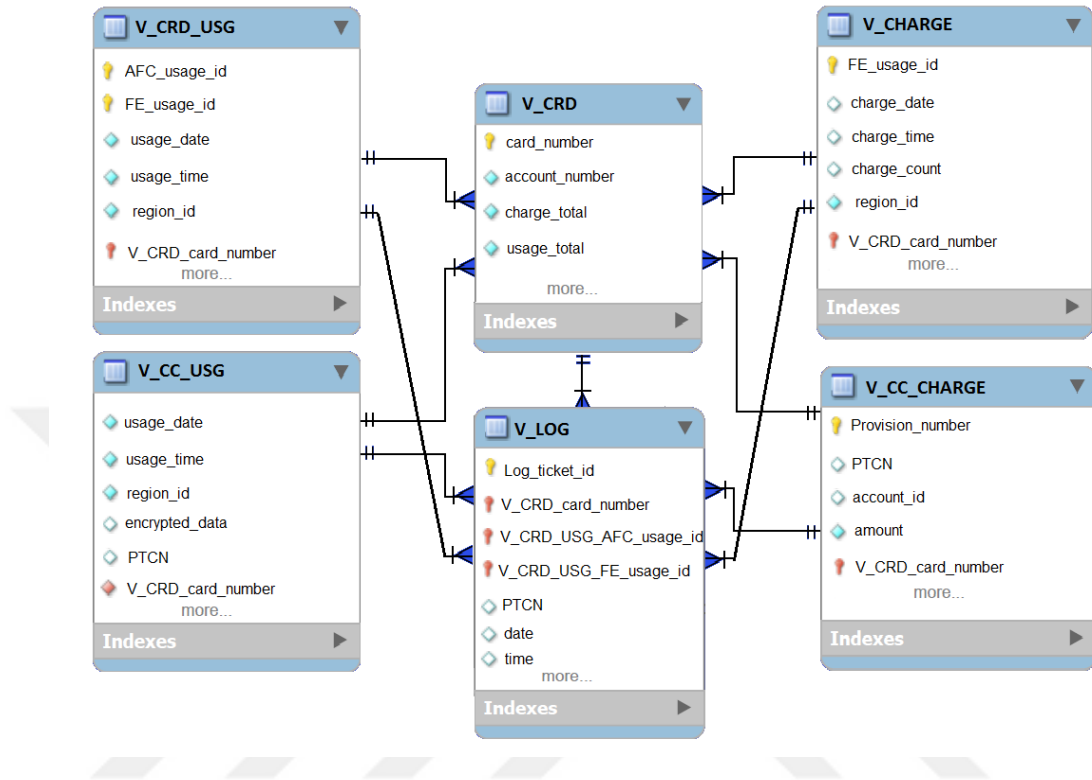


Figure 4.17 ER diagram of the NCAMS

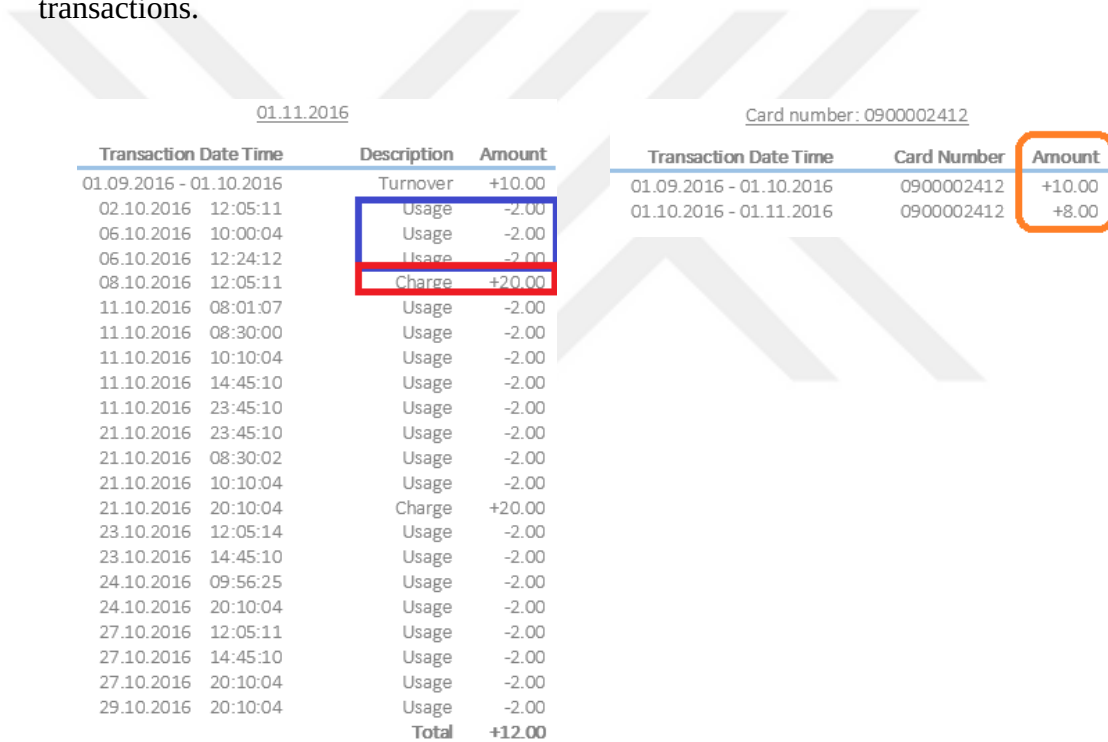
NCAMS module manages virtual card related issues. There are mainly six tables, which are required to integrate with already existing AFC system. Table *V_CRD* holds all fields of the virtual card, which is created for NFC system. The card number, account number, card type, passenger type, creation date, and supported region attributes are the basic columns of this table. When a virtual card is created, card related information is filled into the table properly. The up-to-date balance of the card is also kept on the table.

Tables *V_CRD_USG* and *V_CC_USG* are used for managing usage information of NFC cards and bankcards. Usage information contains bus id, bus stop id, usage date-time, region id, usage count and control codes. Tables *V_CHARGE* and *V_CC_CHARGE* are used for managing top-up information of NFC cards and bankcards. They contain charge date-time, account number, charge count, provision number and control codes.

Table *V_LOG* is used to trace all transactions on cards. Usage of the NFC card and charge of it are completely managed by NCAMS schema. NFC is a virtual card in fact. Because of this reason, users need to have information about card transactions in real-time. So this schema provides detailed information about NFC card holders.

BINLIST and *BLACKLIST* are other two important tables. Problematic cards are monitored by using these tables. According to the status of a card being blacklisted or not, NFC card usage is allowed or blocked.

Figure 4.18 shows the output of the NCAMS module for users to trace of the transactions.



01.11.2016			Card number: 0900002412		
Transaction Date Time	Description	Amount	Transaction Date Time	Card Number	Amount
01.09.2016 - 01.10.2016	Turnover	+10.00	01.09.2016 - 01.10.2016	0900002412	+10.00
02.10.2016 12:05:11	Usage	-2.00	01.10.2016 - 01.11.2016	0900002412	+8.00
06.10.2016 10:00:04	Usage	-2.00			
06.10.2016 12:24:12	Usage	-2.00			
08.10.2016 12:05:11	Charge	+20.00			
11.10.2016 08:01:07	Usage	-2.00			
11.10.2016 08:30:00	Usage	-2.00			
11.10.2016 10:10:04	Usage	-2.00			
11.10.2016 14:45:10	Usage	-2.00			
11.10.2016 23:45:10	Usage	-2.00			
21.10.2016 23:45:10	Usage	-2.00			
21.10.2016 08:30:02	Usage	-2.00			
21.10.2016 10:10:04	Usage	-2.00			
21.10.2016 20:10:04	Charge	+20.00			
23.10.2016 12:05:14	Usage	-2.00			
23.10.2016 14:45:10	Usage	-2.00			
24.10.2016 09:56:25	Usage	-2.00			
24.10.2016 20:10:04	Usage	-2.00			
27.10.2016 12:05:11	Usage	-2.00			
27.10.2016 14:45:10	Usage	-2.00			
27.10.2016 20:10:04	Usage	-2.00			
29.10.2016 20:10:04	Usage	-2.00			
	Total	+12.00			

Figure 4.18 Transaction history of NFC card produced by NCAMS

CHAPTER FIVE

MOBILE APPLICATION, INTEGRATION, AND TESTING

SBPTS has nearly ten modules and each of them has a connection with some others. The test phase of each module was performed when the module implementation had been finished. However, SBPTS is embroidered as a whole, so the integration of each module and integration tests are important to prevent the fatal errors on the field. This chapter explains testing, integration and installation phases of SBPTS on the field. In addition, it describes the mobile application developed within the scope of this thesis.

5.1 User Interfaces of NFC-Based Mobile Application

The mobile application interconnects the passengers with the SBPTS server side. From the passenger view, NFC based mobile application is the single point of communication. This is the user's face of the system. Account creation, login to the system, purchasing a virtual card, the usage of a virtual card, listing usage and charge transactions are the main actions in the mobile application. The detailed explanation of the each step is listed below.

New account creation is done by RESTFUL services on the server side. At the first stage, the user clicks on the "Login" button. This action allows the user to redirect to the login screen. After the user enters the phone number, this number is sent to the server side via the secure channel. If there is no account that belongs to the specified phone number, the One Time Password (OTP) code is sent to the user after the new account creation process is completed. If there is an account that has already been created, the OTP code is sent to the specified phone number. Once the user enters the OTP code on the screen and then clicks the *Confirm* button, and authentication service checks the validity of the entered code. If it is valid, the login is approved by generating an OAuth2 token. If not, the user is prompted to repeat previous operations. OAuth 2.0 is the industry-standard protocol for authorization (OAuth2, 2012).

The user creates a system registry by performing the steps shown in Figure 5.1. The logged-in user is redirected to the profile page. On the profile screen, it is expected to

update the missing information. The token validity period is limited to one day after the user enters the application. Every 24 hours, the refresh token service checks the validity of the token. If it is valid, a new token is created in the background to allow the user to remain logged on. To prevent agent style usage, OAuth2 authentication service is used on JSON Web Token (JWT) systems. An authorized access is restricted.

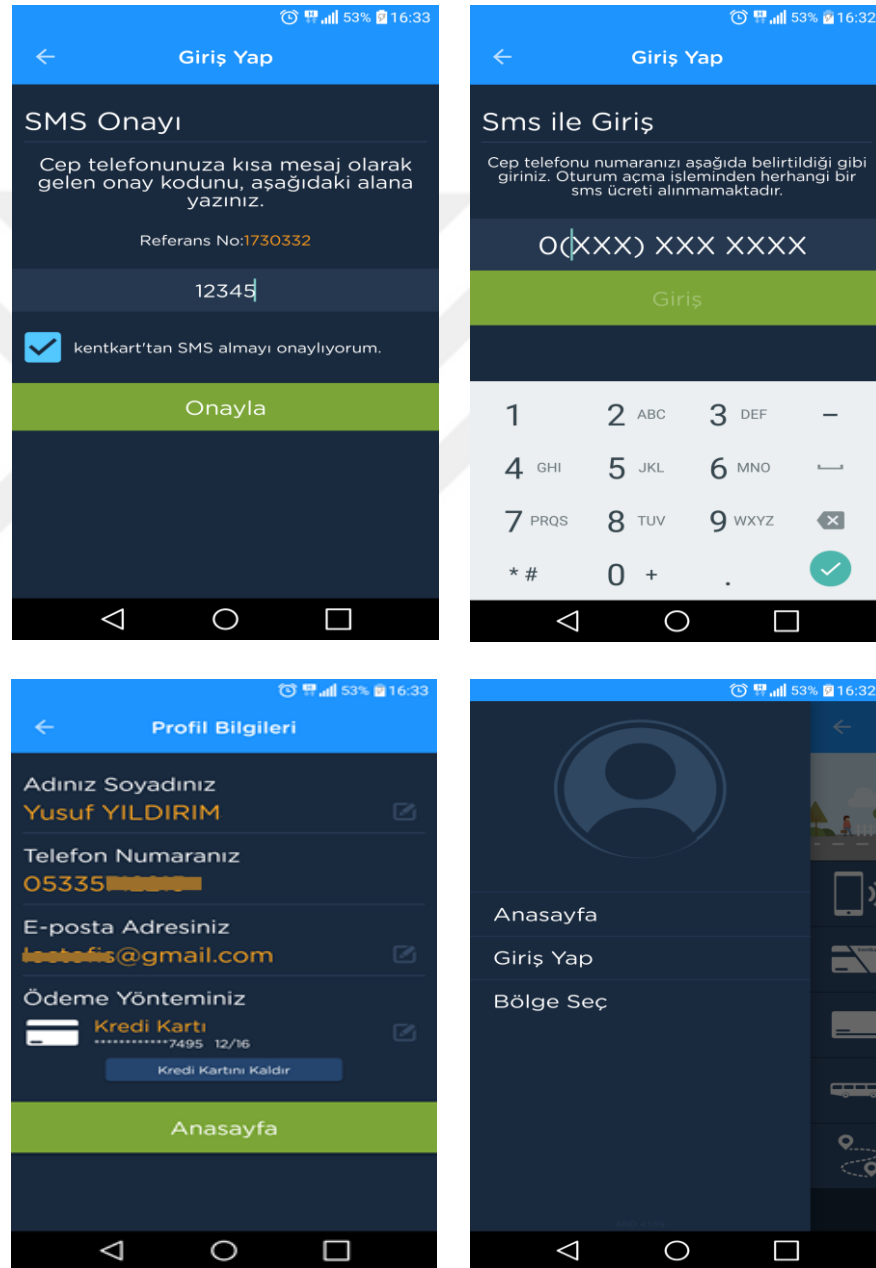


Figure 5.1 Screenshots of account creation steps on android phone

5.2 NFC-Based Mobile Application

The NFC mobile application developed for Android-based phones became the face for passengers. The security requirements of the application are highly vital when it comes to open source operating systems. To prevent fraud attacks, security element on the host-side approach is selected for the application. The servers at the host-side are managed by using high-level security standards (Madlmayr, Langer, Kantner & Scharinger, 2008). Host Card Emulation (HCE) and NFC technology is a strong combination in terms of security. Cloud-based HCE assumes that data stored on the mobile phone is vulnerable, and therefore it restricts the storage of sensitive data to host or cloud databases only. Taking this as a starting point, HCE-based on EMV standard is considered seriously, and the application is adapted to this standard (Balfe & Paterson, 2008).

The emulation of traveling card on mobile application requires passengers to define a tool for payment. An interface is developed for users to enter data such as a credit card or PayPal. Several screenshots of the mobile application are given in section 5.1.

NFC passengers are required to carry out the registration process over the web or with mobile applications. After this process, the token is downloaded to the mobile phone with certification information. The International Mobile Equipment Identity (IMEI), phone number and unique citizen number are taken to be saved on the host side.

The basic steps from the user perspective are listed as follows:

- User sends a request to purchase mobile card by opening the application.
- User is routed to the account creation screen.
- The password for the account is generated.
- One-time password (OTP) is sent to the user.
- User enters the OTP for the account approval.

Within the scope of the thesis, the NFC mobile application was developed to enable several operations like virtual card creation, purchase, and usage. There are multiple

ways to develop Android apps: Hybrid and Native. Since we are going to use NFC technology and keep performance on the front, we preferred to develop Native applications with Java. Since the most useful platform for developing Android applications with Java is Android Studio, it was used as the development environment. At the mobile application side, all services that communicate to host were implemented as RESTful web services. The data interchange format was defined in Javascript Object Notation (JSON).

The libraries used in this thesis can be grouped into three main categories.

- *UI Libraries:* These are libraries that affect the user-visible parts of the application. They are related with the appearance of the application directly.
- *Database Libraries:* They provide database operations required for the application.
- *Backend Libraries:* They are backplane libraries that are included in the functional parts of the application. So, they are used for operations in which the user does not see directly.

The list of the libraries used in the application is as the following:

- *Google PlayGoogle Play Services*
- *Calligraphy*
- *Materialish Progress*
- *View Pager Indicator*
- *DialogPlus*
- *Sliding Up Panel*
- *Easing Library*
- *Realm*
- *Paypal*
- *Crashlytics*
- *Parse*

The most important and difficult part of mobile application development is to construct the file structure of the virtual card. Android is an open operating system, for

this reason, it is more vulnerable to attacks. Therefore, the security part of the virtual card usage is essential part of this study.

5.3 Integration

The basic modules carried out within the scope of the project: payment module, fare collection module, fare calculation module, terminal management module, user module and virtual card module. All these modules are in communication with at least one other module.

A province was first selected for integration operations. We chose a city in Turkey for applying SBPTS. All schemas, database structures and services related to the province were brought into the production environment. The IPs used in the testing environment had been converted to production IPs. The Android application was activated for selected province. The system from start to end had been installed with all units.

Each module had been tested with unit tests during implementation. In order to ensure that the modules communicate with each other and work in a harmonious manner, the modules had been integrated into one another and integration tests had been carried out with scenarios that cover the whole system. Some of the integration tests performed are listed below.

- Fare deduction controls were made with NFC application. The selected lines are equipped with NFC-supported validators. Usage with NFC mobile application was checked whether the appropriate fare had been calculated on the server side. During the test, the personalized virtual cards had also been checked. For the virtual card that was signed as a student card, it was checked that whether the fare was calculated according to the student tariff, similarly, for the adult case, it was checked that whether the fare was calculated according to the adult tariff or not. The basic test cases were: the calculated fare was properly deducted from the user's prepaid account, whether the usage was reflected to the user account operations or not.

- It had been checked whether the user had have the right to transfer and whether the transfer transactions were calculated properly. All routes were analyzed in detail as transfer lines and non-transfer lines. The accuracy of the calculated fare and the reflections to the relevant account and user interface had also been checked.
- Controls of the buttons used on the terminal side were also checked. For example, users could change tariffs while using their phones by pressing -the buttons on the terminals. In these cases, the button information is checked on the server side to see if it is being processed properly.
- User account tests had been done. The first withdrawal from the account of the user was tested. It was checked whether the money could be withdrawn when the limit fell below the limit. If there was a problem with the following withdrawals, the card was inserted into the blacklist. The cases in which virtual cards needed to be removed from black list were also tested. The frequency of retry experiments was calculated during the test phase.
- Log operation was checked because all successful and unsuccessful transactions must be traceable. Log data generation by different units of all processes was validated.
- The warnings, which were given to the user, had been tested. Both the mobile side and the terminal side, messages had been tested by notation, by mail, and by screen message.
- Money clearing process was controlled. The case of using the same card at different provinces had been checked. In this case, the provinces were tested to see whether the correct payments were distributed to correct provinces by the reports.

Many other test scenarios had also been put forward and tested in the context of integration.

5.4 Testing

The test phase of the project has three aspects: performance test, security test, and bug fix operation. The system was thoroughly tested both in the office environment and in the field. After successfully passing through these stages, the project had been taken to the production environment.

5.4.1 Performance Test

There are two basic elements for measuring the performance of the system. The first one is the read time of NFC-enabled phones at the terminal. The second one is that the system has to provide uninterrupted service to all users when usage count is increased dramatically.

For effective usage, the phone reading and verification time on the terminal side should be under 500 milliseconds. In the developed system, it was observed that transactions were carried out under these periods.

Oracle Exadata was used in the system as a test database. With Oracle Exadata, which offers a high level of performance, many operations can be performed in milliseconds. In addition, the pool and parallel connection numbers were also set appropriately for the services in use. According to the results of the tests, 6 million user accounts can be managed and the transactions that are coming from 3000 vehicles in real time can be processed instantly.

In the tests on the fare calculation module of the server, it had been observed that the fare calculation took place in an average of one second when 28,000 real-world transaction data were used.

5.4.2 Security Test

Android is an open source operating system. Therefore, developed android applications are vulnerable to the attacks of hackers. The tests carried out in this section were mostly conducted on the mobile phone. The application source code is obfuscated so it cannot be opened by any user. Android application codes had been tested against reverse engineering. Test results showed that it is not possible to open it because the codes are mixed with the special software.

Another test is the possibility of copying of the token from one mobile phone to another. In this case, the terminal gives an invalid token warning when the phone is

touched to the validator because of the inconsistency of the phone ID. The ID of the mobile phone is held in the token.

It is possible to read the token by third parties but it is not possible to create it on the user side because it is created with a private RSA key in HSM device. The usage of HSM is only opened to the special IPs. All services are encrypted with secure HTTP, so messages are encrypted. Attack tests performed by the test unit are stuck to the IP limitation. In this regard, an application that does not raise doubt about safety has been developed.

It has been observed that for the same virtual card, which is used on more than one provinces, the revenue is smoothly distributed the correct provinces. To ensure this validation, the necessary tests and reports were compared. The money distribution has to be most reliable part of the system. Besides the usage of the system by the users, the distribution of the revenue for these usages to the institutions is very sensitive and important. The necessary tests had been completed in this regard and had been put into service with all reports.

5.4.3 Bug Fix

Some problems were encountered when the system was put into production stage and production test was performed in real life outside the office. For example, usage difficulty which was determined by user experience. Minor changes have been made on interfaces for easier use. On the other hand, communication between older generation terminals with NFC-enabled telephones has caused some problems. Reading did not take place even the phones were displayed on the screen of the validator for a long time. There are two reasons for this: (i) the location of the NFC antenna does not known on some NFC-supported phone. In this case, the user moves smoothly NFC-enabled phone on the validator screen to catch the communication frequency. (ii) Antenna level settings have been changed in older generation devices to allow NFC phones to be detected faster.

Another problem is the decryption of long data on the terminal side. Security keys on the terminal side are stored in secure access module (SAM). As long as the secure data is long, the decryption process in SAM does not work with length reason. The use of an alternative function in the SAM overcomes the length problem.

Another problem is that the fare calculation does not processed by the host in some cases. Due to the lack of data on the server side, fare is not calculated. To solve this problem, inadequate definitions had been removed and different fare tariffs from different provinces had been managed properly at the host side.

When these types of problems in the production tests were removed, the system was successfully launched to every user. At present, the system has been actively deployed in a few cities.

CHAPTER SIX

APPLICATION OF DATA MINING TECHNIQUES

In this chapter, the general information about data mining techniques is given and then the application of these techniques on a real-world transportation data and obtained results are explained in detail. The association rules were investigated for finding out NFC usage habits and user profiles by the FP-Growth algorithm. In addition, classification algorithms were used to construct classifiers and to determine significant features related to NFC usages.

Data mining is the process of discovering previously unknown, interesting and potentially useful knowledge (i.e. patterns, associations, anomalies, changes and significant parameters) from data. Data mining has a wide range of application areas such as marketing, education, healthcare, finance, telecommunication and social media. Techniques used in data mining can be grouped under three tasks: classification, clustering, and association rule mining.

6.1 Association Rule Mining

Association rule mining (ARM) is an essential part of the study. ARM is the extraction of interesting correlations, relationships and frequent patterns among sets of items in the transactions. Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of m distinct items, T be the transaction that contains a set of items such that $T \subseteq I$, D be a dataset $D = \{t_1, t_2, \dots, t_n\}$ that has n transaction records. A *pattern* $\{X, Y\}$ is a set of items that specifies co-existence of itemsets X and Y in the transactions, where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. A *frequent pattern* is a pattern whose frequency in D is larger than or equal to minimum support. *Support* of a pattern is defined as the percentage of records that contain both X and Y itemsets to the total number of transactions in the dataset D . Suppose the support of pattern $\{a, b, c\}$ is 40%, this means that forty percent of the transactions contain a , b and c items together. Support is calculated by the Equation as in (6.1).

$$Support(\{x, y\}) = \frac{|T(\{x, y\})|}{|D|} \quad (6.1)$$

6.1.1 Association Rule Mining Algorithms

The main strategy of the association rule mining is to divide the problem into two sections (Tan, Steinbach & Kumar, 2006):

- Frequent itemset generation: Finding itemsets that have support values higher than defined minimum *support* threshold.
- Rule generation: Finding association rules from the frequent itemsets that have confidence values higher than defined minimum *confidence* threshold.

In this study, association rule mining was performed on the collected real-world data. The aim was to find meaningful patterns from NFC usage data. The user profiles were found by finding frequent itemsets whose support values are higher than a specific value. Frequent itemset mining is computationally expensive. We used the FP-Growth algorithm because of its high performance.

There are three main frequent pattern mining algorithm (i) Apriori, (ii) Eclat, and (iii) FP-Growth.

Apriori is the oldest algorithm found by Agrawal and Sirikant (1994) for finding frequent patterns. This algorithm search database with Breath First Search approach to finding out all frequent items according to support value. It uses horizontal data format shown in Table 6.1.

The main idea of *Apriori* is that if one itemset is frequent then all subsets of it are also frequent. The weakness of *Apriori* algorithm is the producing of a large number of candidate itemsets and a large number of database scans (Garg and Kumar, 2013).

Eclat algorithm is an association rule-mining algorithm that improves the efficiency of Apriori. It uses vertical data format and Depth First Search algorithm. Table 6.2 shows an example of the vertical data format. “Frequent patterns are stored in a bit

matrix structure. Eclat is memory efficient because it uses prefix tree. The algorithm has good scalability due to the compact representation” (Garg and Kumar, 2013).

Table 6.1 Horizontal data format

Transaction Id	Itemset
1	Bus, Metro
2	Bus, Steamboat, Metro
3	Metro
4	Bus, Steamboat
5	Trolleybus
6	Steamboat
7	Bus, Steamboat
8	Bus, Trolleybus

Table 6.2 Vertical data format

Itemset	Transaction Ids
Bus	1, 2, 4, 7, 8
Steamboat	2, 4, 6, 7
Metro	1, 2, 3
Trolleybus	5, 8

FP-Growth is a tree based association rule mining algorithm proposed by Han, Pei and Yin (2000) for mining the complete set of frequent patterns. The main idea of the algorithm is to use a divide and conquer strategy: it firstly divides the set into small problems and then solves them recursively. FP-Growth algorithm performs better because of using compact data structure and eliminating repeated database operations.

According to the performance comparison in terms of execution time, the FP-Growth algorithm performs the best, while the Apriori algorithm performs the worst. “*FP Growth* performs well under all kinds of variations and therefore is the best among three algorithms. *Apriori* performs the worst among three algorithms and thus shows least scalability, whereas *Eclat* lies in the middle of *FP Growth* and *Apriori*.” (Garg and Kumar, 2013) (p.31).

6.1.2 Dataset Description

The proposed system was installed and used in several cities at Turkey. Approximately for four months, the live transportation data was collected from the field. We combined data from two different sources. One is about user (passenger) information that was obtained from them via mobile application during registration and purchase operation of virtual traveling card on mobile. The other was obtained from real virtual traveling card usage on the field. The validators are sending NFC usage information to specific web service to collect them separately from other regular transportation data.

The dataset contains approximately 28,000 records and 9 attributes, including day, time, passenger type, age, gender, marital status, city, route code, and bus-stop code. Table 6.3 shows the possible attribute values obtained after applying *data integration*, *data selection*, and *data discretization* techniques on the database tables. For example; boarding date-time information is mapped to seven days of the week, and it is also categorized into four parts of the day including morning, afternoon, evening and night. Similarly, the age attribute was discretized into four intervals: young, middle-aged, adult and old.

Passenger type has four distinct values: adult-card, student-card, teacher-card and retired-card which are used when calculating the fare of each ride.

The basic parameters about boarding with NFC are as follows:

- PTCN: Unique public transport card number.
- Usage Date Time: Boarding date and time information.
- Bus Stop Information: The bus-stop id and bus-stop name that riding was realized.
- Route Information: The route id and route name where public transport vehicle went.
- Passenger Type: The type of the passenger such as student, teacher etc.
- City: The city name that the transaction was occurred.

- Gender: The gender of the passenger
- Marital Status: The marital status of the passenger.
- Age: The age of the passenger

Table 6.3 Possible data values of dataset

Parameter Name	Possible Values
Passenger Type	Adult-Card Student-Card Teacher-Card Retired-Card
Day of Boarding	Monday Tuesday Wednesday Thursday Friday Saturday Sunday
Marital Status	Single Married Divorced
Age	Young $\text{age} \leq 20$ Middle-Aged $20 < \text{age} \leq 35$ Adult $35 < \text{age} \leq 60$ Old $\text{age} > 60$
Time of Day	Morning 5am - 12pm Afternoon 12pm - 6pm Evening 6pm - 12am Night 12am - 5am
Gender	Female Male
City	Several cities in Turkey

In a traditional transportation system, the variety of data that can be obtained from standard transport media (i.e. tickets, cards, and tokens) is limited. However, a server-based transportation system enables us to obtain different kinds of data as the following:

- *Passenger-specific data*: The usage of a mobile phone or credit card in payment provides for the identification of passengers and for acquiring demographic data such as age, gender and marital status.

- *Multi-dimensional data*: Each transportation transaction has many dimensions such as date-time, location, passenger features, route, service type and company of operation.
- *Province independent data*: Server-based system enables the usage of single payment media in different provinces.

Such variety of data allows development of better data mining applications. Data mining techniques such as classification, clustering or frequent pattern mining can be used on data collected on the server side for different purposes i.e. personalization, outlier detection, scoring, and recommendation.

In this study, two analyses were carried out: NFC mobile user profiling and usage mining. The results of the analyses give detailed information about NFC usage on public transportation according to location, time, travel info and the demographics of passengers such as age, gender, and marital status.

6.1.3 Data Preparation

Data preparation is the process of collecting, cleaning, and consolidating data and is taken before data analysis is carried out.

Data Integration is the combination of data from separate sources. In this thesis, the data was obtained from three sources and merged to form a single dataset: (i) one source that contains automatic fare collection data, (ii) another data source that includes personal user information and (iii) the third data source of existing virtual card details.

Data Preprocessing is the preparation of data for further processing. Especially real-world data is often incomplete and have many errors. Data preprocessing is a proven method of resolving such issues. In this study, three steps were applied as listed below:

- *Data cleaning*: The noisy data was cleaned from original data.

- *Data transformation*: Different presentation of data was put together and the data is mapped and converted from one format to another to correct conflict formats.
- *Data discretization*: Data discretization is the process of converting continuous data attribute values into a finite set of intervals. Discretization process can be classified into different categories from different perspectives such as supervised or unsupervised, top-down or bottom-up, static or dynamic, local or global, nominal or ordinal, univariate or multivariate, direct or iterative. The discretization technique applied in this study is the type of unsupervised, top-down, static, global, nominal, univariate, and direct.

Within the scope of this thesis, boarding date attribute was formatted into seven days of the week. Similarly, boarding time attribute was converted into four different categories: morning, afternoon, evening and night. The age attribute was discretized into four segments: young, middle-aged, adult and old. Table 6.4 shows a sample data used in this research.

Table 6.4 Sample data

Day	Time	Passenger Type	Age	Gender	Marital Status	Route Code	Bus-Stop Code
Friday	Morning	Adult Card	Adult	F	Married	0011	45120
Sunday	Aftern.	Student Card	Middle-Aged	M	Single	0215	45490
Friday	Aftern.	Teacher Card	Adult	M	Married	0354	45370
Monday	Morning	Adult Card	Young	F	Single	0003	45790
Tuesday	Aftern.	Student Card	Young	F	Single	1509	45120
Friday	Aftern.	Student Card	Young	F	Single	0211	45340

6.1.4 User Profiling

In this study, the FP-Growth algorithm was selected among alternative ARM algorithms such as Apriori, FP-Growth, Eclat, Mafia; because of its performance improvement (Kosters, Pijls & Popova, 2003). Figure 6.1 shows the single-item

patterns as a bar chart when FP-Growth algorithm was applied on NFC-based public transportation data. According to the results, NFC owners use their mobile phones on vehicles mostly on Fridays and in the afternoons. In addition, approximately 62% of users are male and almost 80% of users are single. The probability of people using NFC under the age of 20 and between the ages 20-35 is nearly the same, and they share half of the total usage equally. So the fact is that people older than 60 don't prefer to use NFC virtual card. This may be because they don't have NFC-equipped smartphone.

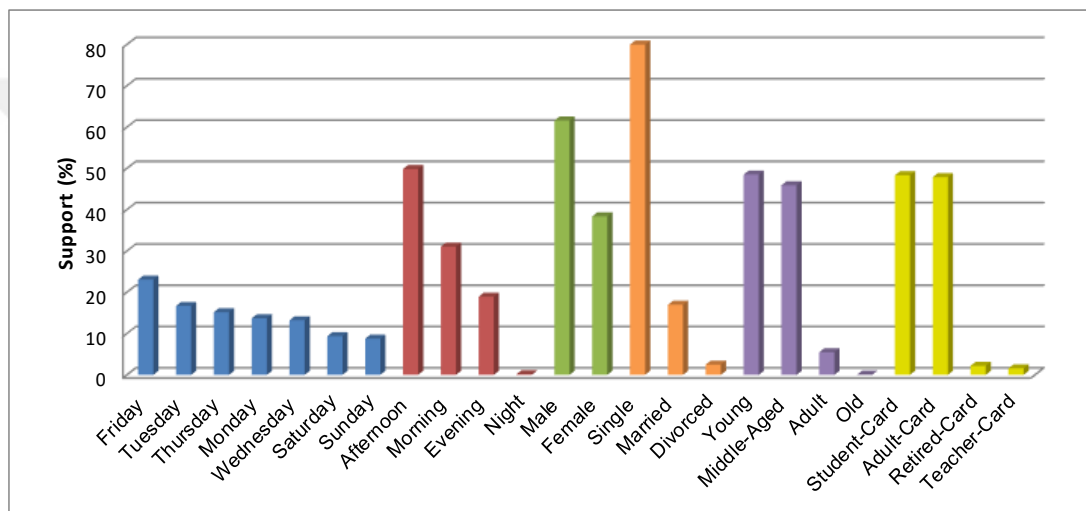


Figure 6.1 Single-item rules discovered from association rule mining

Table 6.5 shows a number of frequent patterns discovered by the algorithm when it was executed with 5% minimum support. According to the results, the “Karaçay-Horozköy Route” in the selected city is the most popular route. This is actually not surprising because the local university of the city is located there. It seems like the number of young females are higher than young males even if males use the system more than females. Female users count is relatively low, but most of them are young.

Age distribution of male users has shifted a little more to middle-age. Men rather than women usually use the system in the evening and those using in the evening are generally single and middle-aged. Additionally, we can say that women prefer to use NFC phones on public transport on Friday mornings.

Table 6.5 Frequent patterns discovered from NFC based public transportation data

Parameter Size	Frequent Patterns	Support (%)
1-Itemsets	{Friday}	23.07
	{Afternoon}	49.84
	{Male}	61.60
	{Single}	80.49
	{Young}	48.52
	{Student-Card}	48.34
	{Karaçay-Horozköy-Route}	12.42
2-Itemsets	{Female, Young}	26.42
	{Male, Young}	22.09
	{Female, Student-Card}	26.13
	{Male, Student-Card}	22.21
	{Young, Afternoon}	25.65
	{Middle-Aged, Afternoon}	21.02
	{Friday, Female}	12.85
	{Friday, Male}	10.21
	{Evening, Male}	15.63
	{Evening, Single}	14.54
	{Male, Married}	12.77
	{Friday, Afternoon}	11.63
3-Itemsets	{Yeni-Garaj-Bus-Stop, Female, Young}	8.16
	{Yeni-Garaj-Bus-Stop, Female, Student-Card}	8.18
	{Karaçay-Horozköy-Route, Young, Single}	7.77
	{Saturday, Male, Single}	5.47
	{Married, Adult-Card, Male}	8.38
	{Evening, Middle-Aged, Single}	8.64
	{Friday, Morning, Female}	8.82
4-Itemsets	{Female, Young, Student-Card, Single}	22.28
	{Female, Friday, Morning, Student-Card}	8.02
	{Female, Young, Student-Card, Afternoon}	11.36
	{Male, Middle-Aged, Afternoon, Single}	12.11
	{Young, Student-Card, Afternoon, Single}	15.69
	{Female, Student-Card, Afternoon, Single}	12.57
	{Male, Student-Card, Afternoon, Single}	8.82
5-Itemsets	{Yeni-Garaj-Bus-Stop, Female, Young, Student-Card, Single}	8.13
	{Friday, Morning, Female, Young, Single}	8.24
	{Friday, Female, Young, Student-Card, Single}	10.39
	{Tuesday, Female, Young, Student-Card, Single}	5.27
	{Adult-Card, Young, Afternoon, Male, Single}	7.87
	{Morning, Female, Young, Student-Card, Single}	10.72
	{Female, Young, Student-Card, Afternoon, Single}	11.36

When we look at the results from the location perspective, it is possible to say that young women mostly get on the bus at “Yeni Garaj” station. This is the location of the big intercity bus station of the city. It can be understood that young and single women use intercity bus station more than men with NFC. Adult cards generally belong to

married users, while student-cards are generally owned by females (Alan & Birant, 2018a).

6.1.5 Usage Mining

Usage mining was utilized in order to analyze transportation data to understand the impact of NFC-based application and to better serve the needs of the application. Figure 6.2 shows the most popular bus lines (routes) on the map of the city where NFC mobile phones are used for public transportation.

Figure 6.3 shows the most popular bus stops where NFC mobile application is used for public transportation on the map. The aim of this analysis is to understand the current situation and then to determine strategies for encouraging more passengers to use the system (Alan et al., 2018a).

Mobile user profiling and usage mining study results are useful for the improvement of decision making systems in public transportation. The results can be used for different purposes in different areas. Three main use cases are listed below.

- Internal (in the bus) and external (at the bus stop) advertisement tools can be positioned around. Defining an advertisement mechanism on usage profile and usage habits has vital for the third-party organization and the municipality itself. The passenger profile (age, marital status, etc.) gives more information about the users located on the routes and bus stops. For example; young and female passengers are the target for the cosmetic advertisements. This and many such inferences can be used for advertising purposes for different companies. On the other hand, the hospital or a shopping mall located on the routes can change their advertisement strategies based on the passenger profile.
- The study contains information about the owners of mobile phone which NFC-equipped. This is valuable information for the producer companies of NFC-equipped mobile phone or distributor companies of them. Target users analysis gives information about NFC on the field. This can be used for driven of the marketing and strategies of the production.

- The municipality may use the user profile information for defining cultural activities such as theater performance, festival and symposium. The content of these activities can be decided according to the user's profile on that route and around.





Figure 6.2 The most popular routes on which NFC virtual card used in the city of Turkey

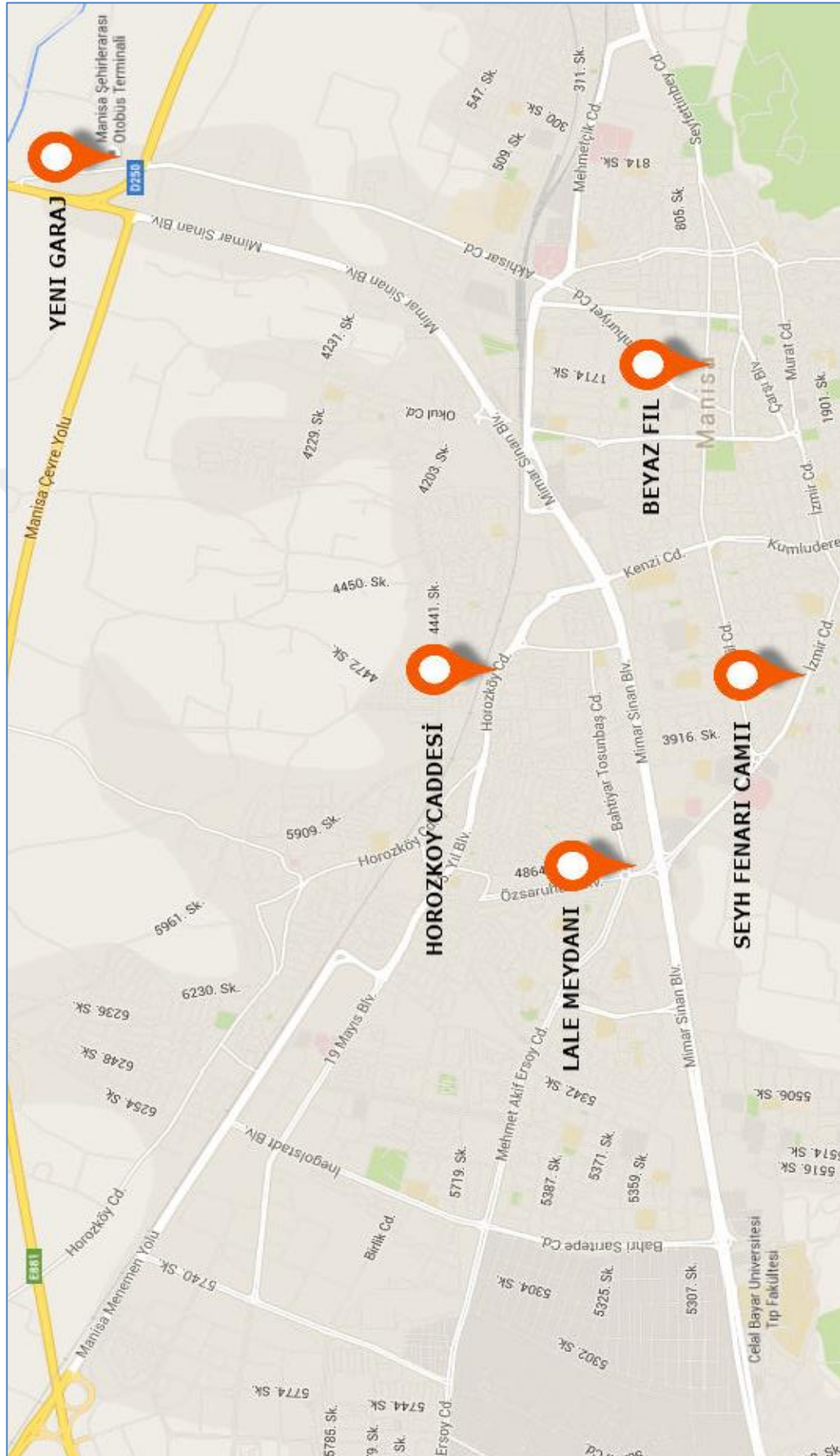


Figure 6.3 The most popular bus-stop locations where NFC mobile application is used frequently in the city of Turkey

6.2 Classification

Classification task in data mining is to develop an accurate model to categorize new input data into predefined labels. It can be used to make a prediction based on prior knowledge. In this study, classification algorithms were executed on NFC-based transportation data and compared with each other in terms of accuracy rates. In this study, the real-world data collected from smartphone users was analyzed to provide decision support. To the best of our knowledge, this is the first study that compares classification algorithms on public transport data collected by NFC enabled mobile phone application.

The novelty and main contributions of this study are as follows. First, it briefly explains the previous studies that apply classification algorithms on public transport data. Second, it compares five different classification algorithms on real-world data in terms of accuracy: Naive Bayes, C4.5, random forest, support vector machines, and k-nearest neighbors. The study can be useful for different purposes in public transportation i.e. personalization, outlier detection, scoring, and recommendation.

6.2.1 Classification Algorithms

There are many classification algorithms used for data mining problems, which are related to the prediction of the class of an object on the basis of its attributes (J. R. Quinlan, 1993). The most commonly used classification algorithms are C4.5 Decision Tree, Random Forest, Naive Bayes, Support Vector Machine and K-Nearest Neighbor.

Decision tree (DT) learning is a method widely used in data mining to create a classifier that predicts the value of a target parameter based on several input parameters. As one of the decision tree algorithms, C4.5 algorithm constructs a decision tree that has a flowchart-like structure in which each internal node splits the instance space into two or more sub-spaces based on a "test" on an attribute, each branch represents the outcome of the test, and each leaf node contains a class label. The C4.5 algorithm uses entropy and normalized information gain to construct a decision tree. Entropy is a formula to calculate the homogeneity of a sample data and

is the measure of uncertainty. Information gain is calculated by the taking difference between the target attribute entropy and the entropy of each branch. Both the discrete and continues attribute values, also missing values in the dataset can be handled by the C4.5 algorithm.

Random Forest (RF) builds more than one decision trees using different subsets of the data (bagging) and evaluating different subsets of features at each node. It is an ensemble learning method and uses bootstrap samples to train multiple classifiers. Each bootstrap sample creates a new tree that is different from the original tree because, at each tree generation, a random number of attributes are tested. The outputs of the classifiers are combined and the result is evaluated by a voting mechanism (Zhang & Ma, 2012). The random Forest algorithm generally produces highly accurate results because it uses many different characteristics to make a classification.

Naive Bayes (NB) is another classification algorithm that utilizes Bayes' theorem to attempt to predict class labels. It is a probabilistic and simple model that uses maximum-likelihood estimation. Naive Bayes classifier is based on assumption that each attribute in a class is independent of the presence of any other attribute.

Support Vector Machine (SVM) is a supervised machine learning algorithm that tries to find the optimal separating hyperplane as a decision surface in such a way that maximizes the margin of the training data (Ma & Guo, 2014). SVMs learn a linear binary classification boundary. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification by mapping their inputs into high-dimensional spaces by using the kernel trick.

K-Nearest Neighbor (kNN) algorithm is a technique in which an object is classified by a majority vote among the classes of its k neighbors in the training sample that are closest to this instance (Steele, Chandler & Reddy, 2016). It uses a distance measure (i.e. Euclidean distance) to compute the k nearest neighbors of a given point.

6.2.2 Classification in Public Transportation

Technical studies that use classification algorithms on public transport data are discussed in this section briefly.

Since the passengers are the key constituents of the public transportation, the public/private operators give the highest priority to meeting and satisfying passengers' demands. For example, passenger information services gain importance from day to day. Answering the following problems have begun to take an important place for the operators to provide high quality services to passengers.

- when the vehicle will arrive at the station (Kee, Wong, Khader & Hassan, 2017)
- which transport mode is preferred (Shafique, & Hato, 2017) and
- how to reach a destination (multi-model journey planning) and how long it takes (Kiss & Csiszar, 2014).

In recent years, the studies are centralized around the arrival time estimation (Kee et al., 2017), transport mode prediction (Guvensan, Dusun, Can & Turkmen, 2017), and quality of service classification (Tsami & Nathanail, 2017; Stottelaar, 2012) in public transportation. When addressing these problems, data mining algorithms have been widely used to make correct classifications and estimations.

To address the first problem (arrival time estimation of buses), Kee et al. (2017) compared several classification algorithms: decision tree, random forest, Naive Bayes, AdaBoost and neural network. If a bus will arrive in the first 15 minutes of the hour, it is classified as Q1; if a bus will arrive in the 16th to 30th minutes, it is classified as Q2; and the same applies for Q3 and Q4.

To deal with the second issue (transport mode detection), Guvensan et al. (2017) developed a mobile application to predict eight classes including stationary, walking, car, bus, tram, train, metro, and ferry. When the performances of the classifiers are compared, the results indicate that random forest has higher performance (80.62%) in terms of classification accuracy than C4.5, k-nearest neighbor, and Naive Bayes.

To handle the third issue (quality of service – QoS), several classification studies (Tsami & Nathanail, 2017; Stottelaar, 2012) have been done. Tsami and Nathanail (2017) applied C4.5 decision tree algorithm on questionnaire data related to QoS collected from public transport passengers in Greece. In their study, service quality in public transport was classified into five categories: very bad, bad, medium, good and very good. Stottelaar (2012) classified Twitter messages about public transport by sentiment (positive, neutral, negative), content and type of transport (bus, train) to detect problems, using C4.5 decision tree algorithm.

Besides to above researches, some studies (Karginova, Byttner & Svensson, 2012; Martinez, Heucke & Wang) have been performed on the classification of driving styles of the bus drivers. This problem is very important when the subject is public transportation in terms of fuel consumption and passengers' safety. Karginova et al. (2012) compared several classifiers (including neural networks, decision trees, random forests and k-nearest neighbor) to classify the driving styles into two categories: normal and hard.

Many attempts have been made to derive useful information about the behavior of the passengers traveling, using classification algorithms. For example, Ahlawat et al. (Ahlawat, Gupta & Vatsa, 2017) tried to estimate the trip purpose of the passengers which can be thought of a key attribute in the context of the behavior of the travelers. They applied Naive Bayes algorithm on the trip survey data to classify different trip purposes: commuting to work, commuting to school, leisure, business and returning home.

Data mining studies conducted in public transport shows that classification task plays an important role when addressing different types of public transport problems. Differently from the existing studies, our study focuses the classification of passengers based on their characteristics using real-world transport data collected through NFC-based mobile phone ticketing application.

6.2.3 Dataset Description

The dataset used in classification task in this study contains 24,485 NFC usage records on public transport vehicles collecting from five different cities in Turkey. There are eight attributes: gender, age, marital status, passenger types, region, day, time, and city. Section 6.1.2 and 6.1.3 explains the data and how prepared. Table 6.3 gives information about possible attributes and values used in classification study. The raw dataset contains global positioning system (GPS) coordinates. For better classification, the GPS coordinates were converted into region concept. The logic behind this conversion is to divide the each city into four regions according to the coordinates given in Table 6.6.

Table 6.6 GPS coordinates that is used to divide regions

City Code	Coordinates				
	Latitude 1	Longitude 1	Latitude 2	Longitude 2	Middle Point
M	38.646263	27.439563	38.602304	27.307137	38.620647, 27.409614
K	40.807224	30.017510	40.752596	29.851664	40.771613, 29.947201
A	36.978631	30.851122	36.813067	30.573608	36.911644, 30.720849
N	37.941617	34.629570	38.008193	34.729133	37.968282, 34.676433
E	36.720327	34.487919	36.882297	34.719783	36.821859, 34.619532

6.2.4 Comparison of Classification Algorithms

The application was developed by using Waikato Environment for Knowledge Analysis (WEKA) open source data mining library. In this study, five classification algorithms (DT, RF, NB, SVM, and kNN) were applied to transportation data collected through NFC-based mobile phone ticketing application. The algorithms were executed five times and a different attribute was selected as a target column in each try. The number of neighbors, k for kNN classifier was selected as 5. The input parameters of other classifiers were left as default Weka parameters.

The results were compared according to their correct estimation ratios (accuracy rates). Classification accuracies of the applied algorithms were obtained using 10-fold cross-validation. In addition, the most important parameters were determined by selecting the attributes at the high levels in the construct decision tree, because a decision tree algorithm recursively chooses "most significant" attributes as a root of tree/subtree.

The comparative results are given in Table 6.7 According to the results, Naive Bayes algorithm has lower performance when compared to other algorithms. The random forest algorithm has a little higher accuracy rate for some attributes than the C4.5 algorithm. Decision tree-based algorithms (C4.5 and random forest) have the highest performance in all cases. The experimental result also indicates that it is possible to predict the regions and marital statuses of the passengers with high accuracy rates.

According to the results, the most important features in NFC-based transportation data are generally age and passenger type.

Table 6.7 Learning algorithm results on dataset

Target Attribute	Naive Bayes	Random Forest	C 4.5	SVM	KNN	Significant attributes
	Correctly Classified Instance Ratio (%)					
Region	95.18	95.58	95.50	95.36	k=5 94.74 k=7 94.46 k=9 94.33	City, day, age
Marital Status	82.09	86.77	87.07	86.38	k=5 86.17 k=7 86.01 k=9 85.89	Age, gender, pax type
Gender	76.83	85.65	85.77	81.25	k=5 85.36 k=7 85.17 k=9 85	Age, pax type, marital status
Age	75.65	77.67	77.56	75.45	k=5 77.35 k=7 77.27 k=9 77.19	Marital status, pax type, gender
Passenger (Pax) Type	65.13	76.96	76.84	72.71	k=5 76.67 k=7 76.56 k=9 76.4	Age, gender, region

Feature selection is about identifying important attributes to reduce time consumption and to increase the classification accuracy. Relief feature selection algorithm, principal component analysis and information gain feature selection method were applied on the dataset. However, each time when removing the attributes from the dataset, the accuracy was decreased contrary to expectations. As the result of this experiment, we decided to use all attributes in the classification task.

Figure 6.4 shows the classifiers' average ranks. Classifiers are rated according to their accuracies. It is done by assigning rank 1 to the classifier with the best accuracy and increasing the rank for each evaluated classifier until giving rank 5 to the worst one. According to the results, when the random forest algorithm is applied, it obtains the lowest rank value which means that it outperforms the others in terms of accuracy.

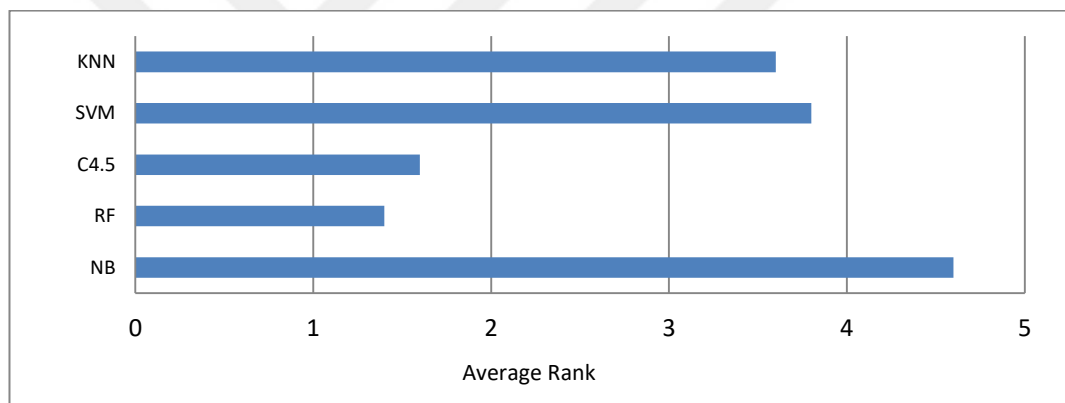


Figure 6.4 The average rank of classifiers

The goal of this study is to devise an accurate classification model and to determine the important features over real-world NFC-based transportation data. We collected data through NFC-based mobile ticketing application developed as a part of the server-based automatic fare collection and management system. Demographic features of the passengers and their mobile ticketing usage transactions were analyzed empirically by using five classification algorithms: Naive Bayes, C4.5 decision tree, random forest, support vector machines, and k-nearest neighbor. We observed that decision tree-based algorithms achieved high classification accuracies; especially random forest algorithm is the best one when compared with the others (Alan & Birant, 2018ba).

Classification algorithm results can be useful for the predicting the missing attributes for the future. For example, generally public transportation users' are not willing to give their personal information. The already existing values can be used as training data set and the missing part of the newly collected data can be predicted. For example, the age of the passengers can be predicted by using the passenger's usage habits and transport types like adult or student card.



CHAPTER SEVEN

EXPERIMENTAL STUDY: PASSENGER SCORING

The focus of promotions targeted to increase the use of public transportation concentrates on increasing the attractiveness of it, particularly by decreasing transportation fares. To serve the purpose, a novel passenger scoring model is proposed, namely RFLT (Recency, Frequency, Loyalty, and Time), for offering a free pass promotion in public transportation. It presents the comparison results of RFLT and wRFLT (weighted version) using real-world dataset obtained by NFC (near field communication) mobile payment application used in a city in Turkey. The study results can be used to establish an efficient policy for increasing public transportation ridership. This chapter gives information about RFLT in all aspects.

7.1 Free Pass Promotion in Public Transport

Public transportation plays an important role in everybody's life, especially in terms of their economic and social life quality. Observing developed provinces shows clearly how high the importance given to public transportation is. However, even though public transportation gains importance lately, car dependence in everyday travel has increased enormously in the last decades. This causes a series of problems in terms of human health and the environment (Brög, Erl & Mense, 2002). In order to overcome car dependency and motivate the use of public transportation, some promotions may be planned and offered for the use of everybody. This study gives an idea about the type of activities to be done in order to encourage people to use public transportation.

The most important factors affecting the intention to increase the use of public transportation are the personal norm, service quality and economic factors (Jomnonkwao et al., 2016). All three factors present a positive effect on encouraging the use of public transportation. This study focuses on the third factor (economic) through the offer of free pass promotion. Socioeconomic factors (i.e. gender, hometown, and education level) do not have a direct influence on the intention to increase the use of public transportation but have an indirect effect on vehicle ownership (Jomnonkwao et al., 2016).

The aim of this study is to develop a sustainable policy for promoting public transportation. The “free riding right” which has a great motivating effect on preferring public transportation (Cools, Fabbro & Bellemans, 2016) has been the subject of study for this purpose. The awarding of active public transportation riders was referenced as an incentive method. The active passengers to be given free passes were determined by a scoring method.

A novel passenger scoring method RFLT is proposed. The well-known marketing model RFM (Hughes, 2012) was extended by adding two other parameters, Loyalty (L) and Time (T), to form RFLT model. While Loyalty (L) defines the time period between the user’s first entry to the system and today, Time (T) measures roughly the times of daily public transport usage.

In the experimental studies, the proposed scoring technique RFLT was applied on the real-world dataset obtained by NFC (near field communication) mobile payment application used in a city in Turkey. RFLT and w-RFLT (weighted version) techniques were compared with each other to investigate the importance of the parameters. The experimental results show that the proposed scoring method is practical and can be used efficiently for promoting public transportation.

The main contributions of this study are threefold. First, it proposes a novel passenger scoring mechanism to be used in the public transportation field. It also introduces a new version of RFM model. Second, it demonstrates the applicability of RFM model on transportation for the first time. Third, the proposed approaches (RFLT and weighted-RFLT) are compared by using a real-world dataset for sensitivity analysis. To the best of our knowledge, it is the first time that sensitivity analysis has been applied to determine how variations in transportation parameters change passenger scores.

7.2 Advantages of Free Pass Promotion in Public Transportation

Making small discounts on riding prices has an impact on the preference of public transportation but the free pass is a strong tool to change the driving habits of citizens.

The study of Thogersen & Moller (2008) confirms that a free one-month transportation card effectively releases the habits of car drivers and increases their use of public transportation. Decreasing the number of cars on roads reduces congestion, pollution and parking challenges.

The study of Fearnley (2013) also shows that free pass in public transport is necessary because of its positive impacts on social, economic, health and environmental benefits. Galey (2014) also presents the benefits of free pass in relation to the city of Tallinn, the capital of Estonia, which recently became the world's largest municipality offering free public transportation.

Belter, Harten & Sorof (2012) explain the advantages of free public transportation services from four different points of view: aspects for public transport users, for the community, for the local authority, and for the public transport operator.

The study of Gase, Kuo, Teutsch & Fielding (2014) also presents the benefits of providing free pass rights to students in the Los Angeles County. According to the research, free transit passes for students enables them to increase school attendance, have more freedom and mobility, strengthen social networks, decrease contact with the justice system, and increase disposable income while reducing traffic volume, and allowing them to participate in sports and cultural events.

To encourage the use of public transportation and by this way, to provide the benefits aforementioned, a novel scoring mechanism is proposed here to determine the free pass rights of the inhabitants of a province in a fair way. The proposed scoring method can be implemented by the public transport operators and can be effectively used by the municipalities as a campaign tool in the scope of the free ride.

7.3 RFM and Its Versions

RFM is a scoring method for analyzing customer behavior using three variables stated as follows: recency (R) - time elapsed since the last purchase, frequency (F) -

the number of purchases, and monetary (M) - the value of the purchases. The main purpose of this method is to find customer segments and identify target customers for marketing campaigns and events. The RFM model is first proposed by Hughes (2012). This model has proven very effective (Blattberg, Kim & Neslin, 2008) when applied to marketing databases.

RFM analysis is a marketing technique generally used for customer selection, cross-selling and customer churn prediction. Several studies based on RFM concepts have also been proposed for different areas such as

- computer security (Kim, Im & Park, 2010; Vlasselaer et al., 2015),
- government (Ghodousi, Alesheikh & Saeidian, 2016),
- tourism (Dursun & Caber, 2016; Lumsden, Beldona & Morison, 2008),
- social media (Moreno, Lorente, Porcel & Viedma, 2015),
- automobile industry (Chan, 2008), education (Chang, 2010),
- banking (Bizhani & Tarokh, 2011; Wang, Di, Yu, Lei & Coenen, 2013),
- health care (Chen, Cheng, Lai, Hsu & Syu, 2012),
- dental care (Wei, Lin, Weng & Wu, 2012),
- hairdressing industry (Wei, Lee, Chen & Wu, 2013),
- telecommunication (Li, Shue & Lee, 2008) and the electronics industry (Chiu, Kuo & Chen, 2009).

The importance of our work comes from being the first study where RFM concept has been applied in the field of public transportation.

Table 7.1 presents different versions of RFM model which have been proposed by changing or adding some parameters. The reasons for these modifications are in order to adapt it to different fields or to achieve better results than the basic RFM model. In contrast to the previous studies, we used the RFM model in a different field (public transportation). Furthermore, the different is that we proposed the RFLT model by removing the "monetary" parameter while adding the "loyalty" and "time" parameters for accurate classification of public transportation users.

Table 7.1 Different versions of RFM model

Author	Year	Method	Description	Application Area
Daoud, Amine, Bouikhalene & Lbibb	2015	LRFM	Length, Recency, Frequency, and Monetary	Marketing (Customer segmentation)
Baradaran & Biglary	2015	CFM	Continuity, Frequency, and Monetary	Marketing (Customer segmentation)
Chen, Liu, Wang & Qin	2015	RFL	Recency, Frequency, and Length	Television User Segmentation
Cho & Moon	2015	FRAT	Frequency, Regency, Amount and Type of merchandise or service	Marketing (Recommender system)
Teng, He & Gu	2014	RFA	Recency, Frequency, Amount	Direct marketing
Golsefid, Turksen & Zarandi	2012	RFMCT	Recency, Frequency, and Monetary, Continuity, Trend	Global market analysis
Wei et al.	2012	LRFM	Length, Recency, Frequency, and Monetary	Dental care
Chang & Tsai	2011	GRFM	Group RFM	Marketing (Customer consumption behaviour)
Liu, Yong, Xue & Feng	2011	WRFM	Weighted RFM	Marketing
Bizhani et al.	2011	RF*M*	Recency, new definition of Frequency and Monetary	Banking
Yan & Chen	2011	RFD	Recency, Frequency, and Duration	Mobile application recommendation
Chang	2010	EL-RFM	RFM model for E-Learning	Education
Yeh, Yang & Ting	2009	RFMTC	Recency, Frequency, Monetary, the first purchase (T) and churn probability (C)	Direct marketing
Li, Lee & Liu	2006	TRFM	Timely RFM	Marketing (Product recommendation)

7.4 RFLT (Recency, Frequency, Loyalty, Time)

RFLT, which is a modified version of the RFM method, is a novel scoring technique proposed in this study for the public transportation field. RFLT model brings a framework to the table for objectively measuring the following four ideas on a numerical scale:

- Recency - how recently did the customers use public transportation?
- Frequency - How often do they use?

- Loyalty - How long do they use?
- Time - What time do they generally use?

Recency is the number of days since the passenger's last boarding. It measures the interval between the most recent transaction date and the analysing date. For example, if the last transaction date of the passenger is 12.16.2016 and the analysing date is 12.30.2016, the recency value will be 14 days. For instance, passenger P_2 given in Table 7.2 has a recency value of 1, which means that passenger P_2 took the public transportation yesterday, if the baseline is today. Recency is measured in days, that is, the lower the number of days, the higher the score of recency. According to RFLT, the passenger who has recently used public transportation will probably tend to use again, so he/she should be rewarded with more scores, more free-pass rights as well. In addition, it is a valuable measure to identify passengers at risk of churn (when recency is low). To calculate recency parameter, it is necessary to find the last riding date value of the related passenger. Recency is calculated by the last boarding date subtracted from the analysing date, as given in Equation (1).

Definition 1. (*Recency*). Recency is the number of elapsed days between the last date of the period being analysed and the last boarding date to the public transportation. The recency value (R) for the passenger P_i in days is calculated by Equation (7.1).

$$R_{(P_i)} = ED - \max_{SD \leq t \leq ED} TD_{(P_i)}^{(t)} \quad (7.1)$$

where SD and ED are the start and end dates of the analysing period t respectively, $TD_{(P_i)}^{(t)}$ includes the transaction dates recorded in the period t for the i^{th} passenger, $i = 1, 2, \dots, m$, and m is the number of passengers.

Frequency is a magic parameter to identify active and inactive passengers. It is the number of rides performed by the related passenger in a given time period. Frequency value is increased each time a passenger boards a public transport vehicle even though the boarding may be the result of a transfer from another route to complete the same one-way journey (i.e. a transfer between vehicles is counted as two unlinked trips). For instance, passenger P_2 given in Table 7.2 has a frequency value of 840, which means

that passenger P_2 used public transportation 840 times through the time period between January 2014 and September 2016. Frequency is measured in terms of ride counts. Some passengers might transit several times during the same day. However, the number of boarding is counted independently from the day, so each boarding to the public transport vehicle is considered as a separate trip even through occurring in the same day. Over time, frequency can be used to predict the future likelihood. According to RFLT, the passenger who has frequently used public transportation should be rewarded with more scores; so more free-pass rights as well.

Definition 2. (Frequency). Frequency is the number of boarding transactions made by a passenger during the analysis period. The frequency value (F) for the i^{th} passenger is calculated by Equation (7.2).

$$F_{(P_i)} = \sum_{t=SD}^{ED} |B_{(P_i)}^{(t)}| \quad (7.2)$$

where SD and ED are the start and end dates of the analysing period t respectively, and $|B_{(P_i)}^{(t)}|$ is the number of boarding transactions recorded on the t^{th} day related with the passenger P_i .

Loyalty is the time period of membership. Loyalty is measured in terms of day, which is calculated by the first boarding date subtracted from the analysing date. For instance, passenger P_2 given in Table 7.2 has a loyalty value of 620, which means that passenger P_2 has used public transportation for the first time 620 days before the analysing date. The higher the number of days means the higher the score of loyalty. In recent years, loyalty score has gained importance in public transportation. It is a key metric for many organizations and an inherently supported capability within most intelligent software systems. Generally, organizations tend to award former members. From this point of view, former users are more valuable and deserve to be awarded in RFLT method.

Definition 3. (Loyalty). Loyalty is the number of elapsed days up to now from the first boarding transaction date of the passenger. The loyalty value (L) for the i^{th} passenger P_i is calculated by Equation (7.3).

$$L_{(P_i)} = ED - \min_{SD \leq t \leq ED} TD_{(P_i)}^{(t)} \quad (7.3)$$

where SD and ED are the start and end dates of the analysing period t respectively, $TD_{(P_i)}^{(t)}$ includes the transaction dates recorded in the period t for the passenger.

Time parameter in RFLT model is the average non-peak hour travel times. In other words, it is the ratio of the number of trips passenger used the transportation service during the non-peak-hour period to total ride counts (frequency) of the passenger. For instance, passenger P_2 given in Table 7.2 has a time value of 0.86, because he used public transportation 720 times at non-peak hour period and 840 times at total (frequency value of P_2), so the ratio is $720 / 840 \sim 0.86$. Time parameter is important in order to provide a higher service quality, especially availability, in public transportation by encouraging riders to use public transportation outside of peak hours (Gwee & Currie, 2013). Peak hours can change from province to province but generally, there are two peak hour times in a day: from 7 am to 9 am, and from 16 pm to 18 pm. Travelling in out-of-peak hour is more valuable to reduce passenger volume (i.e. over-crowded boarding) at peak-hours. For this reason, in the RFLT model, the passengers who use public transportation at non-peak hours are awarded with more scores. Thus, free-pass opportunities will encourage passengers who can choose their departure time to make the trip at non-peak hours. In this way, the public transportation system can reduce the burden by peak spreading (Fearnley, 2013).

Definition 4. (Time). Time parameter in RFTL model is the ratio of non-peak hour ride counts to total ride counts of the passenger. The time value (T) for the passenger P_i is calculated by Equation (7.4).

$$T_{(P_i)} = |E_{(P_i)}^{(t)}| / F_{(P_i)} \quad (7.4)$$

where $|E_{(P_i)}^{(t)}|$ is the number of transactions recorded at out-of-peak hours in the analysing period t for the i^{th} passenger, $i = 1, 2, \dots, m$, and m is the number of passengers.

The monetary parameter in the original version of RFM is discarded because in public transportation, ticket prices are fixed; so cost information is not an accurate parameter for segmenting passengers. The monetary value of a passenger is directly related with frequency which already exists in the model.

In principle, RFM analysis can be performed by following one of two common scaling and scoring methods: (i) quintiles (equal-frequency binning) and (ii) simple fixed ranges (threshold-based binning) (Nair, 2017; Kohavi & Parekh, 2004). In the quintiles method, the records are sorted and divided into five equal groups, and then the top 20% values are given a score of 5, the next 20% values are given a score of 4, and so forth. In other words, grouping is done based on the number of passengers but passengers with equal values should be in the same group. In the second method (simple fixed ranges), the records are also sorted and then scored based on the range of numerical values. For instance, recency can be divided into five intervals according to the following ranges: 0-2 months, 3-6 months, 7-11 months, 12-24 months and 25+ months, which are scored as 5, 4, 3, 2, and 1, respectively. Different businesses may use different methods for ranking the RFM values on the scale of 1 to 5. Each approach has some advantages and some drawbacks compared to the other. The quantiles method yields equal number of customers in each segment, but it gives little insight about relationships. In the fixed ranges method, it is difficult to decide ideal boundaries and update them periodically as the data grows. Several segments can be very small, so it may not be worth targeting the segment if it contains a very small set of customers (Kohavi, 2004).

In this study, we chose quintiles method for RFTL model for public transport industry because of four reasons. First, dividing passengers into equal segments is an ideal way for giving free pass to passengers. In the range-based method, the bins vary significantly in the number of passengers. Giving too many free pass rights is not acceptable for the municipalities or transport authorities because this type of promotion leads to lost thousands of dollars for each day. Second, there are challenges with fixed range calculation for RFTL scores since the users should decide what range they consider ideal for recency, frequency, loyalty and time values. However, the

quintiles method is an easy way to manage since ranges are picked from data itself. Third, the quintiles method is more appropriate technique if segmentation schemes are generated periodically (i.e. free-pass rights of the passengers may calculated at the end of each month). However, in the range based method, the boundaries may need frequent adjustment when data grows. Fourth, the quintiles method is easy to understand for public transport passengers. Complex and dynamic range calculations are not easily explained to passengers, especially to old and low-educated passengers.

In order to demonstrate RFTL scoring, an example dataset about passenger transactions is given in Table 7.2. The first column, public transportation card number (PTCN), is a unique number to identify passengers. Other columns are recency, frequency, loyalty and time values respectively. Table 7.3 shows the steps of RFLT analysis which involves scaling of passengers based on each RFLT factor separately. The segmentation starts with sorting passengers in ascending order based on recency, i.e. period since last boarding. The passengers are then split into five equal groups and given the top 20% a recency score of 5, the next 20% a score of 4 and so on. Passengers are then sorted by frequency in decreasing order and then scored as similar. This process is then undertaken for loyalty and time columns as well. Finally, all passengers are ranked by concatenating R, F, L and M values and scored by the average of them.

There is an inverse relationship between recency and passenger score. This means that the recent usage has positive effect on the method; so if recency is small, the score of the passenger is increased. For this reason, recency column is sorted in ascending order while others are sorted in descending order. Frequent, former and non-peak hour travellers are more valuable than the infrequent, newer and peak hour travellers.

RFLT analysis assigns a quadruple value to each passenger on the basis of his past behaviour. To deal with 625 ($=5 \times 5 \times 5 \times 5$) different score combinations, it is possible to homogeneously group RFLT scores into segments that have similar characteristics, such as best, good, valuable, ideal, medium-level, uncertain, new, non-valuable and churned passengers.

Table 7.2 A sample dataset about passenger transactions

PTCN	Recency (day)	Frequency (number)	Loyalty (day)	Time (number)
P1	6	720	1000	288/720 = 0.40
P2	1	840	620	720/840 = 0.86
P3	24	512	90	412/512 = 0.80
P4	2	53	27	10/53 = 0.19
P5	5	464	58	455/464 = 0.98
P6	7	57	222	30/57 = 0.53
P7	35	75	150	30/75 = 0.40
P8	3	892	45	695/892 = 0.78
P9	150	834	450	576/834 = 0.69
P10	10	576	556	46/576 = 0.08
P11	14	345	360	87/345 = 0.25
P12	34	30	185	20/30 = 0.67
P13	87	854	859	700/854 = 0.82
P14	8	90	156	54/90 = 0.60
P15	38	5	40	1/5 = 0.20

Table 7.3 The steps of RFLT analysis

PTCN	Recency	R	PTCN	Frequency	F	PTCN	Loyalty	L	PTCN	Time	T	PTCN	RFLT	Score
P2	1	5	P8	892	5	P1	1000	5	P5	0.98	5	P2	5555	5
P4	2	5	P13	854	5	P13	859	5	P2	0.86	5	P13	1555	4
P8	3	5	P2	840	5	P2	620	5	P13	0.82	5	P8	5514	4
P5	5	4	P9	834	4	P10	556	4	P3	0.80	4	P1	4452	4
P1	6	4	P1	720	4	P9	450	4	P8	0.78	4	P5	4325	4
P6	7	4	P10	576	4	P11	360	4	P9	0.69	4	P10	3441	4
P14	8	3	P3	512	3	P6	222	3	P12	0.67	3	P9	1444	3
P10	10	3	P5	464	3	P12	185	3	P14	0.60	3	P11	3342	3
P11	14	3	P11	345	3	P14	156	3	P6	0.53	3	P6	4233	3
P3	24	2	P14	90	2	P7	150	2	P1	0.40	2	P14	3233	3
P12	34	2	P7	75	2	P3	90	2	P7	0.40	2	P3	2324	3
P7	35	2	P6	57	2	P5	58	2	P11	0.25	2	P12	2133	2
P15	38	1	P4	53	1	P8	45	1	P15	0.20	1	P4	5111	2
P13	87	1	P12	30	1	P15	40	1	P4	0.19	1	P7	2222	2
P9	150	1	P15	5	1	P4	27	1	P10	0.08	1	P15	1111	1

As shown in Table 7.4, the *best passengers* are in quintile 5 for each factor (5555) who has used public transportation most recently, most frequently in non-peak time period. While passengers with score 5111 are *new passengers*, 1111's are *non-valuable*. The passengers that have the RFLT pattern x51x are *valuable* ones because they have used public transportation frequently in a short time period. The passengers with score xxx5 can be evaluated as *ideal passengers*, because they generally use public transportation at out-of-peak hours. Transportation operators or governments can use RFLT values in developing different strategies for different passenger

segments. For example, they can identify *churned passengers* who have 1555 value and may contact with them to return.

Table 7.4 The passenger segments according to RFLT scores

Segment	Example RFLT Patterns	Example Passengers
Best Passengers	5555, 5554, 5545, 5455, 4555	P2
Good Passengers	4444, 4445, 4454, 4544, 5444	P1
Valuable Passengers	5514, 4514, 4515, 5415, 5414	P8
Ideal Passengers	4325, 4425, 3425, 4235, 3345	P5
Medium-level Passengers	3333, 3334, 3324, 3343, 3433	P11
Uncertain Passengers	2222, 2223, 2232, 2322, 3222	P7
New passengers	5111, 5112, 5113, 5114, 5115	P4
Non-valuable Passengers	1111, 1112, 1211, 2111, 1122	P15
Churned Passengers	1555, 1554, 1553, 1455, 1454	P13

7.5 Passenger Scoring Algorithm

Scoring in public transportation must be performed in order to determine how many free pass rights will be given as promotion to passengers in a particular month. This study proposes the new scoring methods: RFLT and wRFLT (weighted version of RFLT). In RFLT method, each parameter has the same effect on the result, $W_R = W_F = W_L = W_T = 1$; while in the wRFLT model, each parameter has its own weight which is equal or higher than 1. In these methods, the score of the passenger P_i is computed by Equation (7.5).

$$wRFLT_{(P_i)} = \frac{W_R * R_{(P_i)} + W_F * F_{(P_i)} + W_L * L_{(P_i)} + W_T * T_{(P_i)}}{W_R + W_F + W_L + W_T} + \alpha \quad (7.5)$$

where $R_{(P_i)}$, $F_{(P_i)}$, $L_{(P_i)}$ and $T_{(P_i)}$ represent the recency, frequency, loyalty, and time values of passenger P_i and W_R, W_F, W_L , and W_T represent the weights of each variable according to their importance respectively. The average RFLT score can vary between 1 and 5. The last parameter α is used to express “special day effect”. In public transportation, municipalities want to give extra rides to citizens on the special days such as birthdays, the first day of schools, international women's day, and teachers' day. This parameter can be tailored according to the decisions of the municipality where scoring is done. Table 7.5 lists the meanings of symbols used to define.

Table 7.5 Notation table

Symbol	Description
t	Time period
D_t	Dataset (transactions recorded in the period t)
$ D_t $	The number of transactions in D_t
k	The number of partitions
m	The number of passengers
a	The number of attributes, i.e. $a=4$ for RFLT (recency, frequency, loyalty, time)
P_i	i^{th} passenger, where $i=1, 2, \dots, m$
$R_{(P_i)}$	The recency value for the passenger P_i in days, where $R_{(P_i)} = ED - \max_{SD \leq t \leq ED} TD_{(P_i)}^{(t)}$
$F_{(P_i)}$	The total count of transactions of passenger P_i , where $F_{(P_i)} = \sum_{t=SD}^{ED} B_{(P_i)}^{(t)} $
$L_{(P_i)}$	The number of days up to now from the first ride of P_i , where $L_{(P_i)} = ED - \min_{SD \leq t \leq ED} TD_{(P_i)}^{(t)}$
$T_{(P_i)}$	The ratio of non-peak hour usage to total usage, where $T_{(P_i)} = E_{(P_i)}^{(t)} / F_{(P_i)}$
B	The boarding transactions recorded in a day, where $B \subseteq D_t$
E	The transactions recorded at out-of-peak hours, where $E \subseteq D_t$
W_R	Weight for recency R
W_F	Weight for frequency F
W_L	Weight for loyalty L
W_T	Weight for time T
α	Special day effect such as birthday, international women's day, etc.

The coding of recency, frequency, loyalty and time value is arbitrary. Quintiles (5x5x5x5, creating 5-recency, 5-frequency, 5-loyalty and 5-time value divisions) are frequently used and hence, 625 different score combinations are assumed. However, as given in Table 7.4, RFLT codes can be grouped into several segments, similar to the study (Olson, 2012). In addition, depending on the application, RFLT coding can be changed with the parameter k , i.e. quartile ($4 \times 4 \times 4 \times 4 = 256$ cells) or thirds ($3 \times 3 \times 3 \times 3 = 81$ cells), in the case of insufficient number of passengers. Which records will be included in each segment is depicted by Equation 7.6.

$$D_t = \cup_{i=0}^{k-1} (|D_t|/k * (i) - |D_t|/k * (i + 1)) \quad (7.6)$$

where D_t is the dataset in the period t and k is the number of partitions. The top and bottom limits of each segment, recording number and order are determined with this equation, e.g. for the dataset including 40 records, where $k=4$, the resulting groups will be as follows: [0-10), [10-20), [20-30) and [30-40]. The value of k parameter will be defined by the organization, i.e. local government, public transportation operator.

In RFLT model, free-pass right of a passenger is calculated as the weighted average of R, F, L and T values of the passenger as given in Equation (5), so 625 cells are reduced into five scores only, between 1 and 5. For instance, while passenger P2 with 5555 RFLT value in Table 7.3 is scored as 5 on average, passengers P13, P8, P1, P5 and P10 are scored as 4, and so forth.

Algorithm Description

RFLT algorithm is applied by defining the scales of R, F, L and T attributes. This process is divided into four steps as below:

Step 1: Sort the data according to R attribute.

Step 2: Partition the dataset into k equal parts and the records in each part are assigned as $k, k-1, \dots, 1$ score referring to passenger contribution to the use of public transportation.

Step 3: Repeat the previous two steps for other attributes F, L and T individually.

Step 4: Calculate free pass rights by dividing the sum of multiplied R, F, L and T values with their weights to total weight value.

The time complexity of the RFLT algorithm is composed of sorting, scaling and merge operations as follows: $O(n \log n * a + n * a + n)$, where n is the number of transactions in the dataset and a is the number of attributes.

7.6 Dataset Description

In this study, the proposed model RFLT and its weighted version wRFLT were applied on the real-world dataset to demonstrate their applicability in public transportation. The number of free pass rights to be given to passengers for the

following month in order to encourage the use of public transportation was calculated by RFLT scoring technique.

The real-world data obtained through the use of NFC-enabled mobile phones in public transportation in a city in Turkey is used. This is the first study in the related literature where NFC usage data in public transportation has been analyzed.

This particular data was collected between January 2014 and September 2016. The dataset is composed of 35,230 transactions belonging to 520 different users. So, the values of input parameters are $t=32$ months, $k=5$, $a=4$, $m=520$ and $|D_t|=35,230$.

The dataset covers five attributes drawn from different tables in the database by queries. They are PTCN, Last Boarding, Membership, Total Usage Count and Total non-peak Hour Usage Count. Table 7.6 shows the attribute name, their meanings, and min./max. values of the dataset. Peak hours can change from province to province but generally, there are two peak hour times in a day: from 7 am to 9 am, and from 16 pm to 18 pm.

7.7 Comparison RFLT and wRFLT

RFLT and w-RFLT techniques were compared with each other to investigate the importance of parameters. First, using the attributes shown in Table 7.6; R, F, L and T values were calculated with the proposed algorithm. Afterward, using the formula stated in Equation (7.1), RFLT score values and consequently, the “free pass right” numbers were figured. The number of free pass rights is minimum 1 and maximum 5. The extra free pass right (α) to be given on special days was considered as zero (0) in this study. However, it can be added to the scoring results upon demand.

Weight values are determined as 1 ($W_R = W_F = W_L = W_T = 1$) initially, and afterwards, passenger scores are calculated with different weight values. The weight values in this study were figured according to priority values of the variables following

discussions with the municipality. For example, R4F3L2T1 means that RFLT index weights are $W_R=4$, $W_F=3$, $W_L=2$ and $W_T=1$.

Table 7.6 The attributes of the dataset with minimum and maximum values

Attribute Name	Description	Min Value	Max Value
PTCN	Public Transport Card Number is a unique virtual card number to identify passenger.	16 digit numeric value.	
Last Boarding	Date and time value of the last transaction before analyzing time (dd.mm.yyyy hh:mm).	20.10.2014 06:35 pm	09.09.2016 11:34 am
Membership	Date and time value of the first transaction (dd.mm.yyyy hh:mm).	20.01.2014 06:32 pm	06.09.2016 09:25 am
Total Usage Count	Total transaction count of the user in period of t .	1	1473
Total Non-Peak Hour Usage Count	Total transaction count of the user in non-peak hours in period of t .	0	1095

Table 7.7 and Figure 7.1 compare RFLT and wRFLT techniques and show how many passengers have won how many free pass rights. According to RFLT results, 108 people have won 2 free rides, 254 people 3 free rides, 145 people 4 free rides and only 13 people 5 free rides. Score distribution in the RFLT technique is observed to be imbalanced. There is nobody winning just one free pass right while almost 50% of the users win 3 riding rights.

Table 7.7 Comparison of RFLT and wRFLT techniques

		Number of Passengers				
Score		1	2	3	4	5
RFLT		0	108	254	145	13
wRFLT	R2F3L1T3	18	113	228	152	9
	R4F3L2T1	2	155	205	137	21
	R1F1L4T1	10	171	186	135	18
	R2F4L1T1	9	159	172	150	30

It is possible to make score distributions more balanced by changing the weight values in the wRFLT technique. It is observed that the most balanced situation arises when the frequency parameter is emphasized more, i.e. with R2F4L1T1 weight values.

As a result, using the weighted RFLT technique is more suitable and creates more balanced distribution for this dataset.

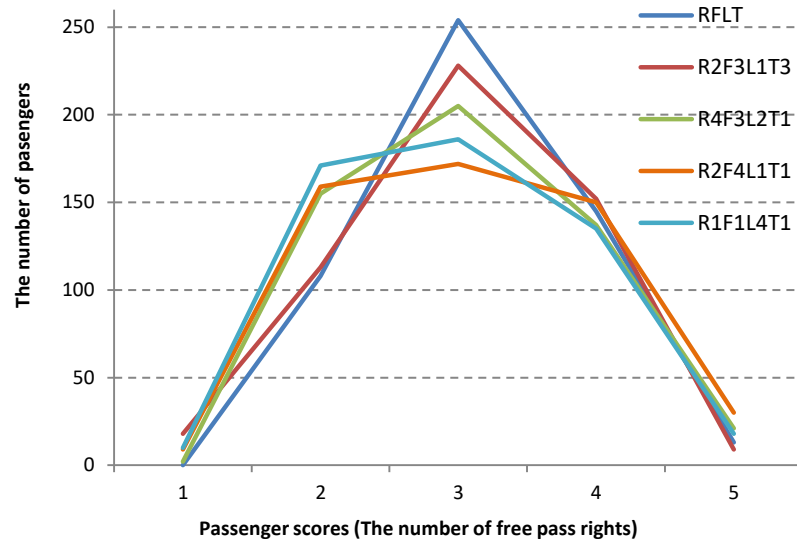


Figure 7.1 Comparison of RFLT and wRFLT with different weight values

7.8 Sensitivity Analysis

Sensitivity analysis of RFLT model was performed to find out how sensitive passenger scores are to any change in R, F, L and T weights. This study is applied to determine how different weight values impact free pass promotion in public transportation. Four types of analysis were performed for this purpose. In each analysis, we changed only one weight while keeping other weights constant.

Table 7.8 shows the R, F, L and T weight values ranged from 1 to 5 and the number of passengers who win free pass rights. For each analysis, only one parameter is changed with weight value while all others are keeping constant. For example

- “2111” means only weight value of R is equal to two
- “1311” means only weight value of F is equal to three
- “1141” means only weight value of L is equal to four
- “1115” means only weight value of T is equal to five

The weight values of each parameter are changed sequentially at each try.

Table 7.8 Weight RFLT and number of passengers that win 1, 2, 3, 4 and 5 free pass/passes

	1	2	3	4	5
RFLT	0	108	254	145	13
2111	2	141	239	127	11
1211	6	150	210	144	10
1121	2	131	257	120	10
1112	6	134	237	137	6
3111	2	144	201	154	19
1311	6	161	172	157	24
1131	2	132	229	139	18
1113	6	134	183	185	12
4111	7	165	178	151	19
1411	27	153	161	156	23
1141	10	171	186	135	18
1114	26	144	156	183	11
5111	7	163	167	151	32
1511	27	149	152	143	49
1151	10	165	169	145	31
1115	26	143	146	189	16

Figure 7.2 shows four different analyses on graphical representation. Y-axis shows the number of passengers, X-axis shows the weight values.

For each analysis, the balanced distribution points are important. At the Figure 7.2-a, all lines are close to each other when weight value is equal to “1211”. This means that the number of passengers who win 1, 2, 3, 4, and 5 free pass rights is close to each other. Balanced pass right distribution is achieved at “1211” point. According to the Figure 7.2-b, the most balanced distribution is provided at the point “1311”. At the Figure 7.2-c, all lines are close to each other when weight value is equal to “1411”, while it occurs at the point “1511” in Figure 7.2-d. Due to the analyses results, the second parameter affects the distribution well-balanced when weight value is changed.

By this study, we can easily say that the most sensitive parameter is frequency and the second one is time attribute.

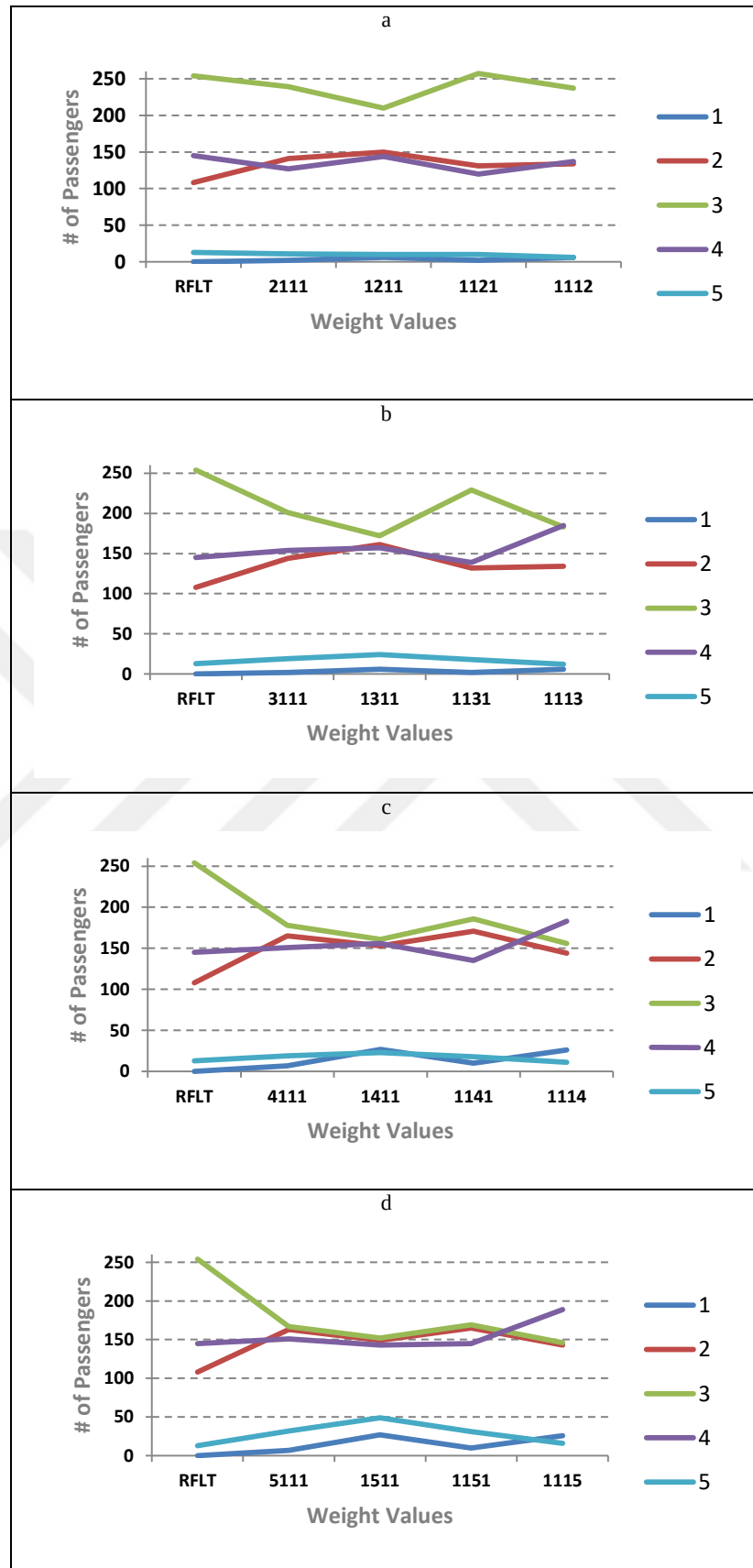


Figure 7.2 Different wRFLT charts

The sensitive attribute of the RFLT is defined according to testing several different weight values on each attribute. When the volume of the data became huge, the cost function can be defined by the user to calculate the expense of the production. This approach decreases the time for the predicting the sensitivity of the attributes by removing trial-and-error method.

Figure 7.3 gives the changing curve for each riding right (from 1 to 5) obtained by different weight values in the wRFLT method. When weight values differ, the number of passengers winning 1 and 5 pass rights does not quite change while those winning 2, 3 and 4 pass rights are considerably influenced. The minimum decrease in the number of those winning 3 pass rights becomes when the loyalty weight value is increased. The most balanced score distribution is realized when frequency weight value is increased. That loyalty, recency, time and frequency attributes respectively affect passenger scoring results is displayed with this study.

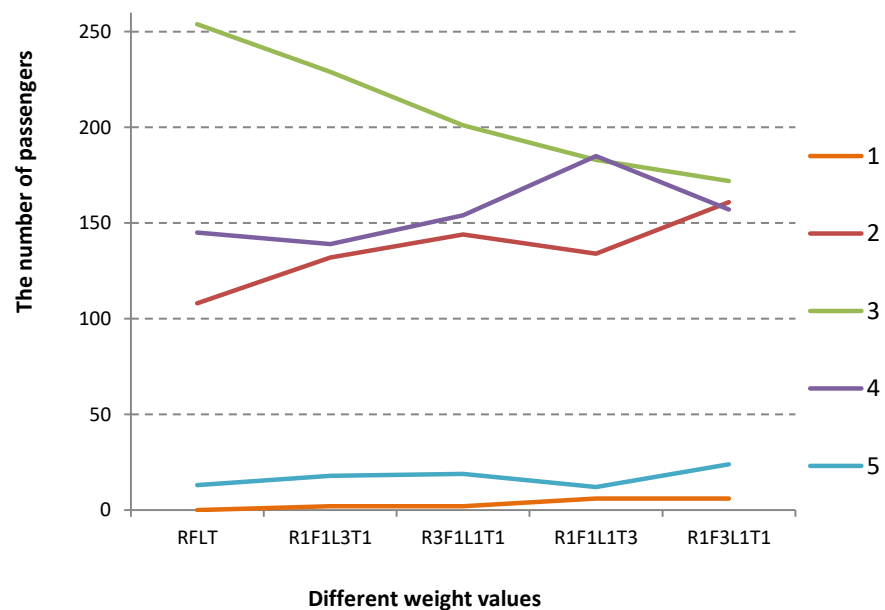


Figure 7.3 Sensitivity analysis of RFLT model with different weight values

Figure 7.4 represents the comparative result of each weight values. The graphics were drawn by using IBM SPSS statistic program. It was observed that the weighted version provides a more balanced score distribution.

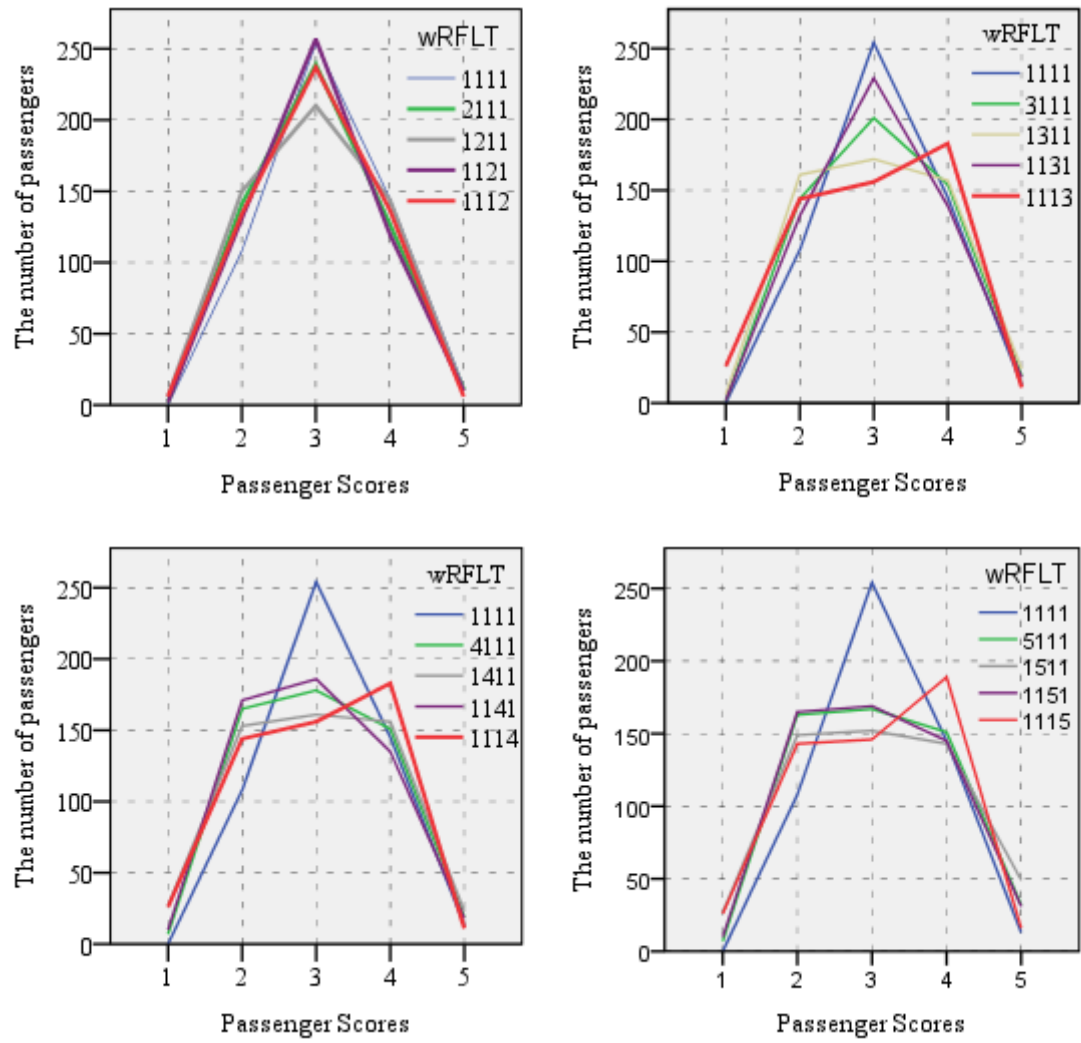


Figure 7.4 Comparison of RFLT and wRFLT by using IBM SPSS program

CHAPTER EIGHT

CONCLUSION AND FUTURE WORK

8.1 Conclusion

This thesis presents a complete infrastructure for a novel server-based fare calculation and user management system for public transportation. In the proposed system, all types of fare calculations are transferred to the server by applying the open-loop payment model. In contrast to the existing systems, the proposed system enables the use of new generation technologies (NFC, contactless credit card, beacon, etc.), instead of traditional tickets, cards and tokens.

When compared with the current systems, the proposed system has numerous advantages such as ease of use, more fare media options, greener solution, user specific loyalty support, ease of integration, interoperability, easy upgrade and others (section 3.4). The proposed system enables a passenger to get a public transportation media that can be used in any city and allows passengers to benefit from all the advantages of the cities' transport pricing.

Ten different modules were implemented while the infrastructure of the system was established. These are Fare Management Module, Settlement Management Module, Merchant Management Module, User Management Module, Report Management Module, Payment Management Module, Clearing Management Module, Account Management Module, Blacklist Management Module, and Product Management Module.

The installation of the system with NFC usage was applied in several cities in Turkey. Real-world transportation data was collected and analyzed using different techniques of data mining. Two different techniques were demonstrated on public transportation dataset: association rule mining to find out NFC usage/user profile and classification to support decision making. The proposed system in this thesis removes the barrier in front of the user-specific campaign in public transport, because, in the

server-based structure, the user and usage information are collected just before fare calculation.

One of the important policies to increase the use of public transportation is to make a free-fare promotion. In order to determine how many free pass rights the passengers will win the next month; this study proposes a novel scoring method, namely RFTL. RFTL is the modified version of a well-known marketing method RFM and is based on four key metrics: Recency, Frequency, Loyalty and Time. *Recency* is the number of days between the most recent travel time date and today. *Frequency* measures the number of travels made in a given time period. *Loyalty* is the number of days from the first transport ride until today. *Time* is the ratio of the number of non-peak-hour travels to total usage counts.

Experimental studies done over real-world data collected through NFC-enabled mobile payment application used in a city in Turkey to show that the proposed RFLT method is a suitable technique for public transportation and can be applied effectively. However, it is observed that the weighted version provides a more balanced score distribution. The sensitivity analysis performed has shown that the frequency parameter (F) among 4 parameters (R, F, L, and T) influences the score results the most. The study results can be used to establish an efficient policy to increase public transportation ridership.

As a result, in this thesis, (i) a novel system, SBPTS, was designed and applied to field, (ii) account-based NFC-enabled mobile application was developed for the first time, (iii) data mining techniques were used to analyze collected data (i.e. NFC usage habits and user profiles) and (iv) a new passenger scoring technique, RFLT, is proposed.

8.2 Future Work

By applying the project on the field, intercity ticket integration has been accomplished among cities in Turkey. The proposed infrastructure may be established for international interoperability in the future.

In addition, different data mining tasks such as clustering or outlier detection can be applied in future studies. For example, discovering both homogenous and non-homogenous groups of public transportation users/usages, extracting relevant clusters, and finding high skewness distribution are possible. In the future, classification and association rule mining may be applied on credit card usage and users. It provides municipalities to understand the passenger profile and their expectations better. Accordingly, transport operators will have a chance to put target better incentives and services. Municipalities may improve transportation policies.

Furthermore, different parameters can be added to the RFLT model such as passenger types (adult, student, discounted etc.), distance covered, or type of vehicle (metro, sea transportation, bus etc.). The parameters in question can be tailored according to the purposes of the municipalities. Another future study can cover differing prizes such as theater/movie tickets instead of free pass rights in order to encourage the use of public transportation.

REFERENCES

- Agard, B., Morency, C., & Trépanier, M. (2007). Mining public transport user behaviour from smart card data. *IFAC Proceedings*, 39 (3), 399-404.
- Ahlawat, G., Gupta, A., & Vatsa, A. K. (2017). Predicting behavior of passengers using data collected through smart cards. *Innovative Applications of Big Data in the Railway Industry*, 6, 127-156.
- Alan, U. D., & Birant, D. (2018a). Server-based intelligent public transportation system with NFC. *IEEE Intelligent Transportation Systems Magazine*, 10 (1), 30-46.
- Alan, U. D., & Birant, D. (2018b). Comparison of classification algorithms on NFC-based public transport data. *International Conference on Advanced Technologies, Computer Engineering and Science ICATCES '18*, 1-5.
- Anders, J. D., Brakewood, C., & Watkins, K. (2014). A case study analysis of new fare payment systems in public transit. *Transportation Research Board 93rd Annual Meeting*. Retrieved August 9, 2016, from TRB and TRID databases.
- Agrawal, R., & Sirikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of the 20th VLDB Conference Santiago, Chile*.
- Balfe, S., & Paterson, K. G. (2008). e-EMV: emulating EMV for internet payments with trusted computing technologies. *STC '08 Proceedings of the 3rd ACM workshop on Scalable Trusted Computing*, 81-92.
- Baradaran, V., & Biglary, M. (2015). Customer segmentation in fast moving consumer goods (FMCG) industries by using developed RFM model. *Journal of Business Management*, 7 (1), 23-42.

- Belter, T., Harten, M., & Sorof, S. (2012). Advantages and disadvantages of free public transport services. *SustraMM - Sustainable Transports for Managing Mobility EnergitEE*, 1-14.
- Bizhani, M., & Tarokh, M. J. (2011). Behavioural rules of bank's point-of-sale for segments description and scoring prediction. *International Journal of Industrial Engineering Computations*, 2 (2), 337-350.
- Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). Market basket analysis. In *Database marketing: analyzing and managing customers* (1st ed.) (323-337). New York: Springer Science & Business Media.
- Bronzi, W., Frank, R., Castignani, G., & Engel, T. (2016). Bluetooth Low Energy performance and robustness analysis for inter-vehicular communications. *Ad Hoc Networks*, 37 (1), 76-86.
- Brög, W., Erl, E., & Mense, N. (2002). Individualised marketing changing travel behaviour for a better environment. *Communicating Environmentally Sustainable Transport: The Role of Soft Measures*, 83-97.
- Chan, C. C. H. (2008). Intelligent value-based customer segmentation method for campaign management: a case study of automobile retailer. *Expert Systems with Applications*, 34 (4), 2754-2762.
- Chang, H. C. (2010). Developing EL-RFM model for quantification learner's learning behavior in distance learning. *2nd International Conference on Education Technology and Computer (ICETC)*, 452-454.
- Chang, H. C., & Tsai, H. P. (2011). Group RFM analysis as a novel framework to discover better customer consumption behavior. *Expert Systems with Applications*, 38 (12), 14499-14513.

- Chen, Y. S., Cheng, C. H., Lai, C. J., Hsu, C. Y., & Syu, H. J. (2012). Identifying patients in target customer segments using a two-stage clustering-classification approach: a hospital-based assessment. *Computers in Biology and Medicine*, 42 (2), 213-221.
- Chen, Y., Liu, Z., Wang, Y., & Qin, Y. (2015). Research on user segmentation based on RFL model and k-means clustering algorithm. *International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2015)*, 419-423.
- Chiu, C. Y., Kuo, I. T., & Chen, P. C. (2009). A market segmentation system for consumer electronics industry using particle swarm optimization and honey bee mating optimization. *Global Perspective for Competitive Enterprise, Economy and Ecology, Proceedings of the 16th ISPE International Conference on Concurrent Engineering*, 681-689.
- Cho, Y. S., & Moon, S. C. (2015). Recommender system using periodicity analysis via mining sequential patterns with time-series and FRAT analysis. *Journal of Convergence*, 6 (2), 9-17.
- Cools, M., Fabbro, Y., & Bellemans, T. (2016). Free public transport: a socio-cognitive analysis. *Transportation Research Part A*, 86, 96-107.
- Coşkun, V., Ok, K., & Ozdenizci, B. (2012). NFC business ecosystem. In *Near Field Communication (NFC): from theory to practice*. (567–613) Chichester: John Wiley & Sons, Ltd.
- Cristóbal, T., Lorenzo J. J., & García, C. R. (2015). Using data mining to improve the public transport in Gran Canaria Island. *Computer Aided Systems Theory, EUROCAST 2015*, 9520, 781-788.

- Daoud, R. A., Amine, A., Bouikhalene, B., & Lbibb, R. (2015). Customer segmentation model in e-commerce using clustering techniques and LRFM model: the case of online stores in Morocco. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 9 (8), 1949-1959.
- Dursun, A., & Caber M. (2016). Using datamining techniques for profiling profitable hotel customers: an application of RFM analysis. *Tourism Management Perspectives*, 18, 153-160.
- EMV 3-D Secure Protocol and Core Functions Specification v2.0.0* (2016). Retrieved Jan 10, 2016, from <https://www.emvco.com/specifications.aspx?id=299>.
- Erl, T., & Puttini, R. (2013). Cloud computing mechanisms. In *Cloud Computing: Concepts, Technology & Architecture* (4th ed.) (139-169). New Jersey: Prentice Hall.
- Fearnley, N. (2013). Free fares policies: impact on public transport mode share and other transport policy goals. *International Journal of Transportation*, 1 (1), 75-90.
- First Data (2010). *Transit payment systems: a case for open payments*. A First Data White Paper. Retrieved October 24, 2016, from https://www.firstdata.com/downloads/thought-leadership/transit-payment-systems_wp.pdf
- Galey, D. (2014). License to ride: free public transportation for residents of Tallinn. *Critical Planning Journal*, 21, 19-33.
- Garg, K., & Kumar, D. (2013). Comparing the performance of frequent pattern mining algorithms. *International Journal of Computer Applications*, 69 (25), 29-32.
- Gase, L. N., Kuo, T., Teutsch, S., & Fielding, J. E. (2014). Estimating the costs and benefits of providing free public transit passes to students in Los Angeles County:

- lessons learned in applying a health lens to decision-making. *International Journal of Environmental Research and Public Health*, 11 (11), 11384-11397.
- Ghodousi, M., Alesheikh, A. A., & Saeidian, B. (2016). Analyzing public participant data to evaluate citizen satisfaction and to prioritize their needs via K-means. *Cities*, 55, 70-8.
- Golsefid, S. M. M., Turksen, I. B., & Zarandi, M. H. F. (2012). A systematic type-2 fuzzy optimization model for global market analysis and its application. *Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS)*, 1-6.
- Gong, Y., Ying, Z., & Lin, M. (2012). A survey of cloud computing. *Proceedings for the International Conference on Green Communications and Networks (GCN 2012)*, 225, 79-84.
- Guvensan, M. A., Dusun, B., Can, B., & Turkmen, H. I. (2017). A novel segment-based approach for improving classification performance of transport mode detection. *Sensors for Transportation (MDPI)*, 18, 1-19.
- Gwee, E., & Currie, G. (2013). Review of time-based public transport fare pricing. *Journeys*, 59-68.
- Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. *2000 ACM SIGMOD International Conference Management of Data*, 29 (2), 1-12.
- Hughes, A. (2012). *Strategic database marketing* (4th ed.). Chicago: Probus Publishing Company.
- ISO/IEC 7816-4:2013 (2013). Retrieved April 12, 2015, from http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=54550

- Jomnonkwao, S., Sangphong, O., Khampirat, B., Siridhara, S., & Ratanavaraha, V. (2016). Public transport promotion policy on campus: evidence from Suranaree University in Thailand. *Public Transport*, 18 (2), 185-203.
- JSON Web Token Internet Engineering Task Force RFC 7519* (2015). Retrieved Feb 15, 2016, from <https://tools.ietf.org/html/rfc7519>.
- Karginova, N., Byttner, S., & Svensson, M. (2012). Data-driven methods for classification of driving styles in buses. *SAE 2012 International World Congress & Exhibition*, 1-6.
- Kee, C. Y., Wong, L., Khader, A. T., & Hassan F. H. (2017). Multi-label classification of estimated time of arrival with ensemble neural networks in bus transportation network. *IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, 150-154.
- Keitel, P. (2011). *The electronification of transit fare payments: examining the case for partnerships between payments firms and transit agencies*. Federal Reserve Bank of Philadelphia, Retrieved July 27, 2016, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1908339
- Kim, H. K., Im, K. H., & Park, S. C. (2010). DSS for computer security incident response applying CBR and collaborative response. *Expert Systems with Applications*, 37 (1), 852-870.
- Kiss, D. E., & Csiszar, C. (2014). Evaluation of multimodal journey planners and definition of service levels. *International Journal of Intelligent Transportation Systems Research*, 13 (3), 154-165.
- Kocur, G. (2010). *Open payment for regional public transportation travel*. Leveraging Technology for Transportation, Retrieved June 16, 2016, from <https://wagner.nyu.edu/files/rudincenter/Kocur.pdf>

- Kohavi, R., & Parekh, R. (2004). Visualizing RFM segmentation. *2004 SIAM International Conference on Data Mining*, 391-399.
- Kosters, W. A., Pijls, W., & Popova, V. (2003). Complexity analysis of depth first and FP-Growth implementations of apriori. *Machine Learning and Data Mining in Pattern Recognition*, 2734, 284-292.
- Kui-ying, Y., Lin, J., & Wen-sheng C. (2012). Research of the data mining based public traffic transfer algorithm. *AASRI Procedia*, 2 (1), 293-299.
- Li, L. H., Lee, F. M., & Liu, W. J. (2006). The timely product recommendation based on RFM method. *Proceedings of the International Conference on Business and Information*, 1-17.
- Li, S. T., Shue, L. Y., & Lee, S. F. (2008). Business intelligence approach to supporting strategy-making of ISP service management. *Expert Systems with Applications*, 35 (3), 739-754.
- Lipan, F., & Groza, A. (2010). Mining traffic patterns from public transportation GPS data. *Intelligent Computer Communication and Processing (ICCP) IEEE International Conference*, 1, 123-126.
- Liu, J., Kauffman, R. J., & Ma, D. (2015). Competition, cooperation, and regulation: understanding the evolution of the mobile payments technology ecosystem. *Electronic Commerce Research and Applications*. 14 (5), 372-391.
- Liu, W. J., Yong, D. S., Xue, Y., & Feng, W. X. (2011). Determination of customer value measurement model RFM index weights. *African Journal of Business Management*, 5 (14), 5567-5572.
- Lumsden, S. A., Beldona, S., & Morison, A. M. (2008). Customer value in an all-inclusive travel vacation club: an application of the RFM framework. *Journal of Hospitality Marketing & Management*, 16 (3), 270-285.

- Ma, Y. & Guo, G. (2014). *Support Vector Machines Applications*. Switzerland: Springer.
- Ma, X., Wu, Y. J., Wang, Y., Chen, F., & Liu, J. (2013). Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, 36, 1–12.
- Madlmayr, G., Langer, J., Kantner, C., & Scharinger, J. (2008). NFC devices: security and privacy. *Availability, Reliability and Security, ARES 08 Third International Conference*, 642-647.
- Mankovski, S. (2009). Service oriented architecture. *Encyclopedia of Database Systems*, 2633-2636.
- Martinez, C. M., Heucke, M., & Wang, F. Y. (2018). Driving style recognition for intelligent vehicle control and advanced driver assistance: a survey. *IEEE Transactions on Intelligent Transportation Systems*, 19(3), 666-676.
- Monzón, A., Hernandez, S., Martínez, A. G., Kaparias, I., & Viti, F. (2016). ITS technologies for public transport. In *Modelling public transport passenger flows in the era of intelligent transport systems*. (85-92). Gewerbestrasse: Springer International Publishing.
- Moreno, J. B., Lorente, A. T., Porcel, C., & Viedma, E. H. (2015). A new model to quantify the impact of a topic in a location over time with Social Media. *Expert Systems with Applications*, 42 (7), 3381-3395.
- Munoz, J. C., & Seekins, L. P. (2016). Operations and design. In *Reconstructing public transport through bus rapid transit*. (281-299) Bristol: Policy Press at the University of Bristol.

- Nair, A. (2017). *RFM analysis for successful customer segmentation* (2017). Retrieved December 12, 2017, from <https://www.putler.com/rfm-analysis/>
- OAuth Open Standard for Authorization 2.0 RFC 6749* (2012). Retrieved March 25, 2016, from <https://tools.ietf.org/html/rfc6749>
- Olson, C. (2012). *Adding email engagement to RFM scoring* (2012). Retrieved May 03, 2017, from <https://marketingland.com/adding-email-engagement-to-rfm-scoring-10481>
- Oña, R., López, G., Rios, F. J. D., & Oña, J. (2014). Cluster analysis for diminishing heterogeneous opinions of service quality public transport passengers. *Procedia Social and Behavioral Sciences*, 162, 459–466.
- Oña, J., Oña, R., & López, G. (2016). Transit service quality analysis using cluster analysis and decision trees: a step forward to personalized marketing in public transportation. *Transportation Springer*, 43 (5), 725-747.
- Payment Card Industry Data Security Standard (PCI DSS) 2.0* (2013). Retrieved January 23, 2017, from https://www.pcisecuritystandards.org/pdfs/PCI_DSS_v2_Cloud_Guidelines.pdf
- Pelletier, M., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: a literature review. *Transportation Research Part C: Emerging Technologies*, 19 (4), 557-568.
- Quibria, N. (2008). *The contactless wave: a case study in transit payments*. Federal Reserve Bank of Boston. Retrieved August 15, 2016, from <https://www.bostonfed.org/economic/cprc/publications/briefings/transit.pdf>
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning* (1st ed.). San Francisco: Morgan Kauffmann, 1993.

- Sasaki, Y. (2014). Optimal choices of fare collection systems for public transportations: barrier versus barrier-free. *Transportation Research Part B: Methodological*, 60, 107-114.
- Seaborn, C., Attanucci, J., & Wilson, N. (2009). Analyzing multimodal public transport journeys in London with smart card fare payment data. *Transportation Research Record: Journal of the Transportation Research Board*, 2121 (06), 55-62.
- Shafique M. A., & Hato E. (2017). Classification of travel data with multiple sensor information using random forest. *Transportation Research Procedia*, 22 (1), 144-153.
- Smart Card Alliance (2011). *Transit and contactless open payments: an emerging approach for fare collection*. A Smart Card Alliance Transportation Council White Paper. Retrieved August 10, 2016, from http://www.smartcardalliance.org/resources/pdf/Open_Payments_WP_110811.pdf
- Song, J. Y., Eom, J. K. , Lee, K. S., Min, J. H., & Yang, K. Y. (2015). Public transportation service evaluations utilizing Seoul transportation card data. *Procedia Computer Science*, 52, 178- 185.
- Spanjaard, T. (2013). Open loop in US transportation systems. *Smart Insights White Paper by Intelling*. Retrieved May 20, 2016, from <http://www.cartes-america.com/files/siwp.pdf>
- Steele, B., Chandler, J., & Reddy, S. (2016). *Algorithms for Data Science*. Switzerland: Springer.
- Stottelaar, B. (2012). Detection of public transport problems using location extended community based sensing. *17th Twente Student Conference on IT*, 45-54.

- Tan, P.N, Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Retrieved Sep 13, 2016, from <http://www-users.cs.umn.edu/~kumar/dmbook/ch6.pdf>
- Tan, W. K., & Tan, Y. J. (2013). Transformation of smart-card-based single-purpose e-micropayment scheme to multi-purpose scheme: a case study. *Expert Systems with Applications*, 39 (3), 2306-2313.
- Tavilla, E. (2015). *Transit mobile payments: driving consumer experience and adoption*. Federal Reserve Bank of Boston, Retrieved July 24, 2016, from <https://www.bostonfed.org/-/media/Documents/PaymentStrategies/publications/2015/transit-mobile-payments.pdf>
- Teng, G., He, C., & Gu, X. (2014). Response model based on weighted bagging GMDH. *Soft Computing*, 18 (12), 2471-2484.
- Thogersen, J., & Moller, B. (2008). Breaking car use habits: the effectiveness of a free one-month travelcard. *Transportation*, 35 (3), 329-345.
- Tsami, M., & Nathanail, E. (2017). Guidance provision for increasing quality of service of public transport. *Procedia Engineering*, 178, 551-557.
- Vlasselaer, V. V., Bravo, C., Caelen, O., Rad, T. E., Akoğlu, L., Snoeck, M., et al. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, 75, 38-48.
- Wallischeck, E. Y., Weisenberger, T., Berthaume, A., & Dinning, M. G. (2015). Transit Cooperative Research Program (TCRP) report 177: *preliminary strategic analysis of next generation fare payment systems for public transportation*. Retrieved June 21, 2016, from http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_177.pdf

- Wang, Y. J., Di, G., Yu, J., Lei, J., & Coenen, F. (2013). Feature representation for customer attrition risk prediction in retail banking. *Advances in Data Mining: Applications and Theoretical Aspects, ICDM 2013*, 7987, 229-238.
- Wei, J.T., Lee, M. C., Chen, H. K., & Wu, H. H. (2013). Customer relationship management in the hairdressing industry: an application of data mining techniques. *Expert Systems with Applications*, 40 (18), 7513-7518.
- Wei, J. T., Lin, S. Y., Weng, C. C., & Wu, H. H. (2012). A case study of applying LRFM model in market segmentation of a children's dental clinic. *Expert Systems with Applications*, 39 (5), 5529-5533.
- Yan, B., & Chen, G. (2011). AppJoy: personalized mobile application discovery. *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, 113-126.
- Yang, XF., & Liu, Z. (2012). The application of association rules mining in building intelligent transportation systems. *Journal of Convergence Information Technology (JCIT)*, 7 (20), 575-582.
- Yeh, I. C., Yang, K. J., & Ting, T. M. (2009). Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, 36 (3), 5866-5871.
- Zhang, C. & Ma, Y. (2012). *Ensemble Machine Learning: Methods and Applications*. New York: Springer.

APPENDICES

APPENDIX: LIST OF ACRONYMS

Acronym	Definition
AFC	Automatic Fare Collection
AMM	Account Management Module
APDU	Application Protocol Data Unit
ARM	Association Rule Mining
BLE	Bluetooth Low Energy
CICO	Check-In Check-Out
EMV	Euro pay, MasterCard, and Visa
ER	Entity Relationship
FE	Fare Engine
FAM	Financial Account Management
GPS	Global Positioning System
GSM	Global System for Mobile
HCE	Host Card Emulation
HSM	Hardware Security Module
JSON	Javascript Object Notation
JWT	JSON Web Token
KNN	K-Nearest Neighbor
NCAMS	NFC Card Management System
NFC	Near Field Communication
NUM	NFC User Management
MAC	Message Authentication Code
ME	Mobile Engine
ORAC	Oracle Real Application Clusters
OTP	One Time Password
PTCN	Public Transport Card Number
RFLT	Recency, Frequency, Loyalty, and Time
RFM	Recency, Frequency, and Monetary

PCI DSS	Payment Card Industry Data Security Standard
PCI SSC	Payment Card Industry Security Standard Council
SBPTS	Server-Based Public Transportation System
RSA	Rivest-Shamir-Adleman
SAM	Secure Access Module
SAQ	Self-Assessment Questionnaire
SE	Secure Element
SHA-2	Secure Hash Algorithm 2
SIM	Subscriber Identity Module
SMS	Short Message Service
SSL	Secure Sockets Layer
SVM	Support Vector Machine
TCRP	Transit Cooperative Research Program
TLV	Type-Length-Value
wRFLT	Weighted Version of RFLT
WSDL	Web Service Definition Language
XML	eXtensible Markup Language
3DS	3D-Secure