

Towards practical privacy preserving frequent Itemset mining on Encrypted cloud data

Submitted in partial fulfillment of the requirements for the
award of Bachelor of Engineering Degree in Computer
Science and Engineering

By

Rishi. S

(39290085)

Lakshmi Srinivas. U

(39290055)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING**

**SATHYABAMA
INSTITUTE OF SCIENCE AND**

TECHNOLOGY

**JEPPIAAR NAGAR,
RAJIVGANDHI SALAI,
CHENNAI – 600119,
TAMILNADU.**

APRIL 2022



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC
(Established under Section 3 of UGC Act, 1956)

PIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI- 600119

www.sathyabama.ac.in



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **Rishi. S (Reg. No. 39290085)** and **Lakshmi Srinivas. U (Reg. No. 39290055)** who carried out the project entitled "**Towards practical privacy preserving frequent Itemset mining on Encrypted cloud data**" under my supervision from December 2021-May 2022.

Internal Guide

Dr. Cruz Antony, M.E., Ph.D.,

Head of the Department

Dr. S. VIGNESHWARI, M.E., Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I, **Rishi.S (Reg. No. 39290085)** hereby declare that the Project Report entitled "**Towards practical privacy preserving frequent Itemset mining on Encrypted cloud data**" done by me under the guidance of **Dr. Cruz Antony** is submitted in partial fulfillment of the requirements for the award of Bachelor of Science degree in Computer Science.

DATE:

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. SASIKALA, M.E., Ph.D.**, Dean, School of Computing and **Dr. S. VIGNESHWARI, M.E., Ph.D.**, and **Dr. L. LAKSHMANAN, M.E., Ph.D.**,

Head of the Department, Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. Cruz Antony, M.E., Ph.D.**, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Computer Science and Engineering who were helpful in many ways For the Completion of the project.

Table of Contents

Revision History

1. Introduction

- 1.1 Purpose
- 1.2 Project Scope
- 1.3 References

2. Overall Description

- 2.1 Product Perspective
- 2.2 Product Features
- 2.3 User Classes and Characteristics
- 2.4 Operating Environment
- 2.5 Design and Implementation Constraints

3. System Features

- 3.1 System Feature 1
- 3.2 System Feature 2

4. External Interface Requirements

- 4.1 User Interfaces
- 4.2 Hardware Interfaces
- 4.3 Software Interfaces
- 4.4 Communications Interfaces

5. Other Nonfunctional Requirements

- 5.1 Performance Requirements
- 5.2 Safety Requirements
- 5.3 Security Requirements
- 5.4 Software Quality Attributes

6. Other Requirements

Appendix A: Glossary

Appendix B: Analysis Models

1. Introduction

Purpose

The main aim of this project is to build a new framework for enforcing privacy in frequent itemset mining on encrypted cloud data.

Project Scope

Cloud computing has become an increasingly popular service for its flexibility and scalability, which motivates many organizations, institutions and companies to prefer to outsource data services to cloud platform. At the same time, much attention has been paid to cope with the special security and privacy problems in cloud. The easy accesses and low prices of public cloud services can significantly save mining costs on massive datasets. More importantly, since large-scale data from multiple and variable data sources are collected by cloud services are sensitive, such as locations, medical records and financial data, mining these confidential data inevitably brings important privacy issues. For example, a curious service provider can easily reveal confidential data in a transaction database while users and miners would like to keep the data in private during the entire mining process. To avoid this user data should be in encrypted format and we propose a privacy-preserving framework for secure frequent itemset mining on encrypted data.

References

1. S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic itemset counting and implication rules for market basket data," in ACM SIGMOD Record, vol. 26, no. 2. ACM, 1997, pp. 255–264.
2. 2.J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data," in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002, pp. 639–644.
3. 3.M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data," IEEE Transactions on Knowledge & Data Engineering, no. 9, pp. 1026–1037, 2004.
4. 4.J. Vaidya and C. Clifton, "Secure set intersection cardinality with application to association rule mining," Journal of Computer Security, vol. 13, no. 4, pp. 593–622, 2005.
5. 5.T. Tassa, "Secure mining of association rules in horizontally distributed databases," Knowledge and Data Engineering, IEEE Transactions on, vol. 26, no. 4, pp. 970–983, 2014.

6. 6.C. Dong and L. Chen, "A fast secure dot product protocol with application to privacy preserving association rule mining," in *Advances in Knowledge Discovery and Data Mining*. Springer, 2014, pp. 606–617.
7. 7.R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *ACM SIGMOD Record*, vol. 22, no. 2. ACM, 1993, pp. 207–216.
8. 8.S. R. Oliveira and O. R. Zaiane, "Privacy preserving frequent itemset mining," in *Proceedings of the IEEE international conference on Privacy, security and data mining-Volume 14*. Australian Computer Society, Inc., 2002, pp. 43–54.
9. 9.Y. Saygin, V. S. Verykios, and A. K. Elmagarmid, "Privacy preserving association rule mining," in *Research Issues in Data Engineering: Engineering E-Commerce/E-Business Systems, 2002. RIDE-2EC 2002. Proceedings. Twelfth International Workshop on*. IEEE, 2002, pp. 151–158.
10. 10.R. Kharat, M. Kumbhar, and P. Bhamre, "Efficient privacy preserving distributed association rule mining protocol based on random number," in *Intelligent Computing, Networking, and Informatics*. Springer, 2014, pp. 827–836.

1. Overall Description

Product Perspective

In an existing system to preserve privacy of frequent itemset mining in public cloud services, some randomization-based approaches have been implemented; in randomization based approaches dramatically sacrifice the accuracy and utility of frequent itemset mining with limited privacy guaranteed for frequent itemset mining in public clouds, where all the encrypted transactions are centralized to the cloud and miners delegate all the mining tasks to the cloud. To enable frequent itemset mining, this work requires n aided semi-honest servers to perform distributed decryption during the evaluation on encrypted data. Moreover, the need of these additional n aided semi-honest servers slows down the running time of frequent itemset mining, and introduces huge interactions and communication overheads.

Product Features

In our proposed system we propose privacy-preserving framework for secure

frequent itemset mining on encrypted data, where only one aided server (referred to as Evaluator) is needed besides the Cloud Service Provider (CSP). The CSP collects encrypted transactions and maintains an encrypted transaction database, and the Evaluator to compute frequent itemset mining efficiently and correctly, and also preserve the privacy of transactions and the mining result. The Evaluator first generates a pair of public/private key. Then, each user encrypts its transactions with the public key of the Evaluator and contributes its encrypted transactions to the CSP, because in our model, users do not fully trust the CSP with the privacy of their transactions. Therefore, all their transactions are encrypted before contributing to the CSP.

User Classes and Characteristics

Cataloging of Users and Listing Channels

In this module user enter the personal information to registration, input field must validate and records are stored, User personal information are keep privacy After registration the user can be login into account then subscribe the channel . They are given list of channel user can choose interested channel and pay the amount using the bank account.

Bank Account creation

In this module user create a new account in our banking application to given an input for user detail in our new account registration .User detail must a valid information (ex:-phone-no, username, etc...). After account creation the initial amount is deposited into bank or the user may be deposit, user also possible to enquiry the account detail for knowing the account number and card number the cvv these detail are must to pay the amount in online.

User Behaviour Model

In this module without loss of generality, the user is assumed to switch between two devices (i.e., TV and phone), thus the model has two corresponding states. Once users subscribe their favourite channel then they will access the channel at any time. Now the user information will be encrypted using evaluator public key and stored in database, when ever user accessing the particular subscribed channel at that time channel hit count will be increased and updated in server.

Channel Recommendation

In this module data are both collected and mined in an encrypted form in a public cloud service. To guarantee data privacy and computation efficiency, we adopt two different homomorphic encryption schemes and design a secure and effective comparison scheme. Now we retrieve the user's data which is stored in server and make analyzing process,

during this process each and every users data will be analyzed one by one and provide recommended channels to the user based on the users behavior.

Operating Environment

2.4.1 Software Requirements:

- Windows Operating System XP and Above
- JDK 1.7
- MySql 5.0
- Tomcat 6.0
- XAMPP
- Eyeos2.5

2.4.2 Hardware Requirements

- Hard Disk : 200 GB and above
- RAM : 2GB and above
- Processor : i3 and above
- Phone : Android v2.3 and Above

2.4.5. Design and Implementation Constraints

2.4.5.1 Constraints in Analysis

- Constraints as Informal Text
- Constraints as Operational Restrictions
- Constraints Integrated in Existing Model Concepts
- Constraints as a Separate Concept
- Constraints Implied by the Model Structure

2.4.5.2 Constraints in Design

- Determination of the Involved Classes
- Determination of the Involved Objects
- Determination of the Involved Actions
- Determination of the Require Clauses
- Global actions and Constraint Realization

2.4.5.3 Constraints in Implementation

A hierarchical structuring of relations may result in more classes and a more complicated structure to implement. Therefore it is advisable to transform the hierarchical relation structure to a simpler structure such as a classical flat one. It is rather straightforward to transform the developed hierarchical model into a bipartite, flat model, consisting of classes on the one hand and flat relations on the other. Flat relations are preferred at the design level for reasons of simplicity and implementation ease. There is no identity or functionality associated with a flat relation. A flat relation corresponds with the relation concept of entity-relationship modeling and many object oriented methods.

3. System Features

In our proposed system we propose privacy-preserving framework for secure frequent itemset mining on encrypted data, where only one aided server (referred to as Evaluator) is needed besides the Cloud Service Provider (CSP). The CSP collects encrypted transactions and maintains an encrypted transaction database, and the Evaluator to compute frequent itemset mining efficiently and correctly, and also preserve the privacy of transactions and the mining result. The Evaluator first generates a pair of public/private key. Then, each user encrypts its transactions with the public key of the Evaluator and contributes its encrypted transactions to the CSP, because in our model, users do not fully trust the CSP with the privacy of their transactions. Therefore, all their transactions are encrypted before contributing to the CSP.

4. External Interface Requirements

4.1 User Interfaces

1. All the contents in the project are implemented using Graphical User Interface (GUI) in Java.
2. Every conceptual part of the projects is reflected using the Java.
3. System gets the input and delivers through the GUI based.

4.2 Hardware Interfaces:

Ethernet

Ethernet on the AS/400 supports TCP/IP, Advanced Peer-to-Peer Networking (APPN) and advanced program-to-program communications (APPC).

ISDN

You can connect your AS/400 to an Integrated Services Digital Network (ISDN) for faster, more accurate data transmission. An ISDN is a public or private digital communications network that can support data, fax, image, and other services over the same physical interface. Also, you can use other protocols on ISDN, such as IDLC and X.25.

4.3 Software Interfaces

This software is interacted with the TCP/IP protocol, Socket and listening on unused ports. Server Socket and listening on unused ports and JDK 1.6

4.4 Communications Interfaces

1. TCP/IP protocol.
2. LAN Settings.

5. Other Nonfunctional Requirements

5.1 Performance Requirements

The performance of the wireless sensor network, to execute this project on LAN or wifi communication channel. So we need to one or more than machine to execute the demo. Machine needs the enough hard disk space to install the software and run our project.

5.2 Safety Requirement

1. The software may be safety-critical. If so, there are issues associated with its integrity level.
2. The software may not be safety-critical although it forms part of a safety-critical system. For example, software may simply log transactions.
3. If a system must be of a high integrity level and if the software is shown to be of that integrity level, then the hardware must be at least of the same integrity level.
4. There is little point in producing 'perfect' code in some language if hardware and system software (in widest sense) are not reliable.
5. If a computer system is to run software of a high integrity level then that system should not at the same time accommodate software of a lower integrity level.
6. Systems with different requirements for safety levels must be separated.
7. Otherwise, the highest level of integrity required must be applied to all systems in the same environment.

5.3 Security Requirements

Do not block the some available ports through the windows firewall.

5.4 Software Quality Attributes

Functionality: are the required functions available, including interoperability and security.

Reliability: maturity, fault tolerance and recoverability

Usability: how easy it is to understand, learn, and operate the software system

Efficiency: performance and resource behavior.

Maintainability: Maintaining the software.

Portability: can the software easily be transferred to another environment, including install ability

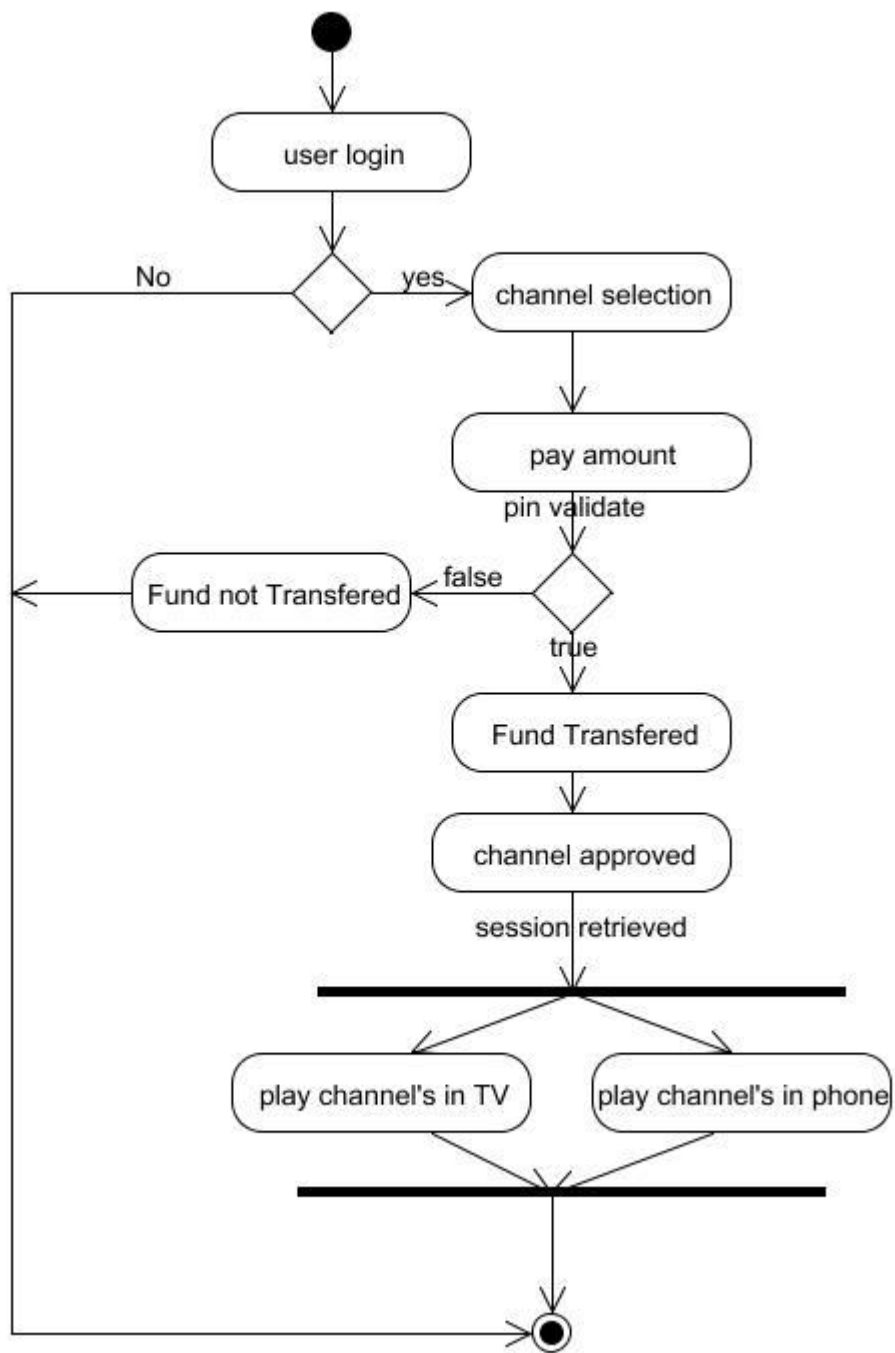
Appendix A: Glossary

IEEE - Institute of Electrical and Electronic Engineers.

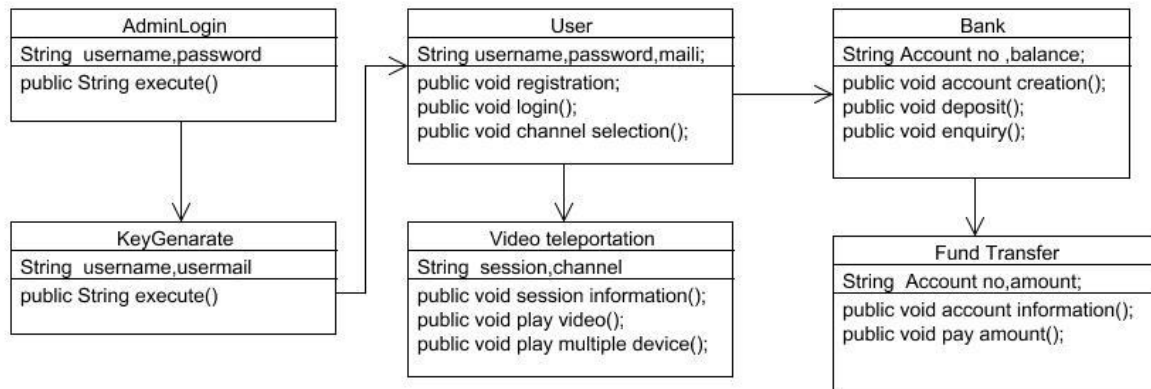
LAN - Local Area Network.

TCP - Transmission Control Protocol.

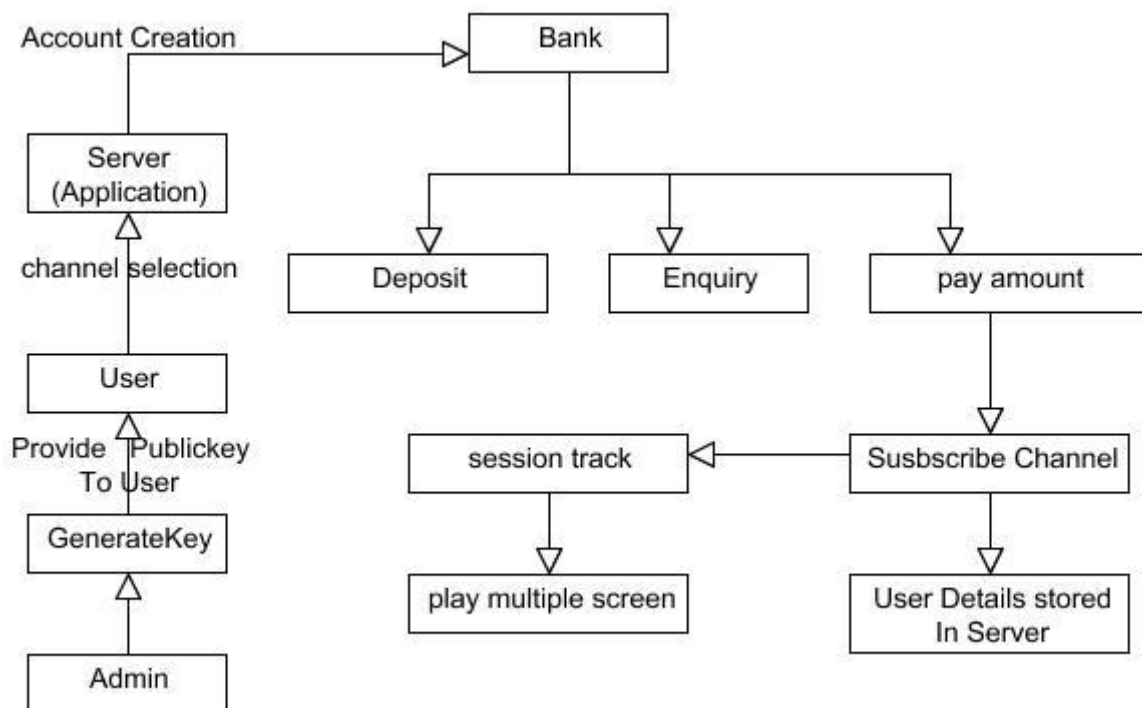
ACTIVITY DIAGRAM:



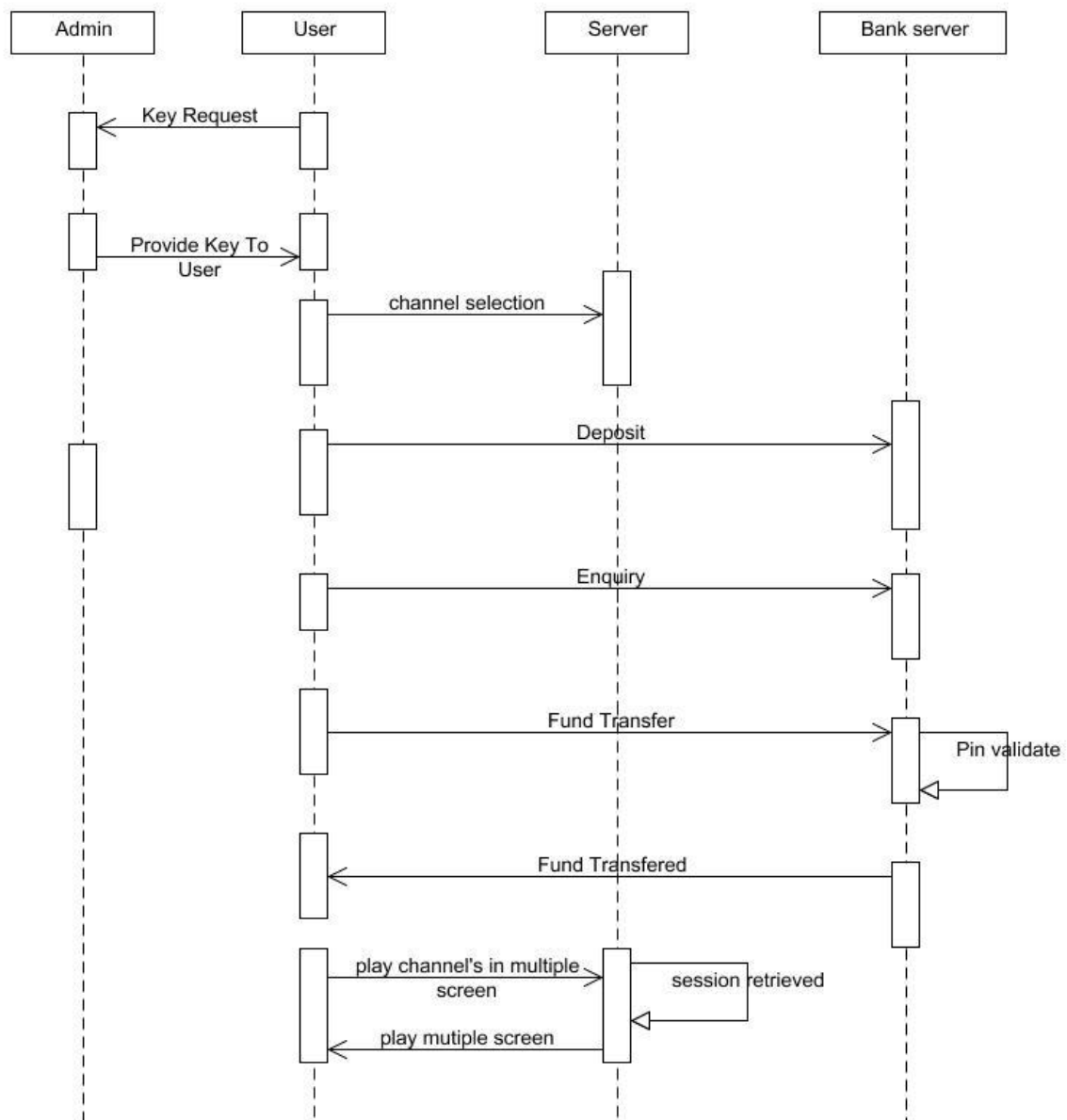
CLASS DIAGRAM:



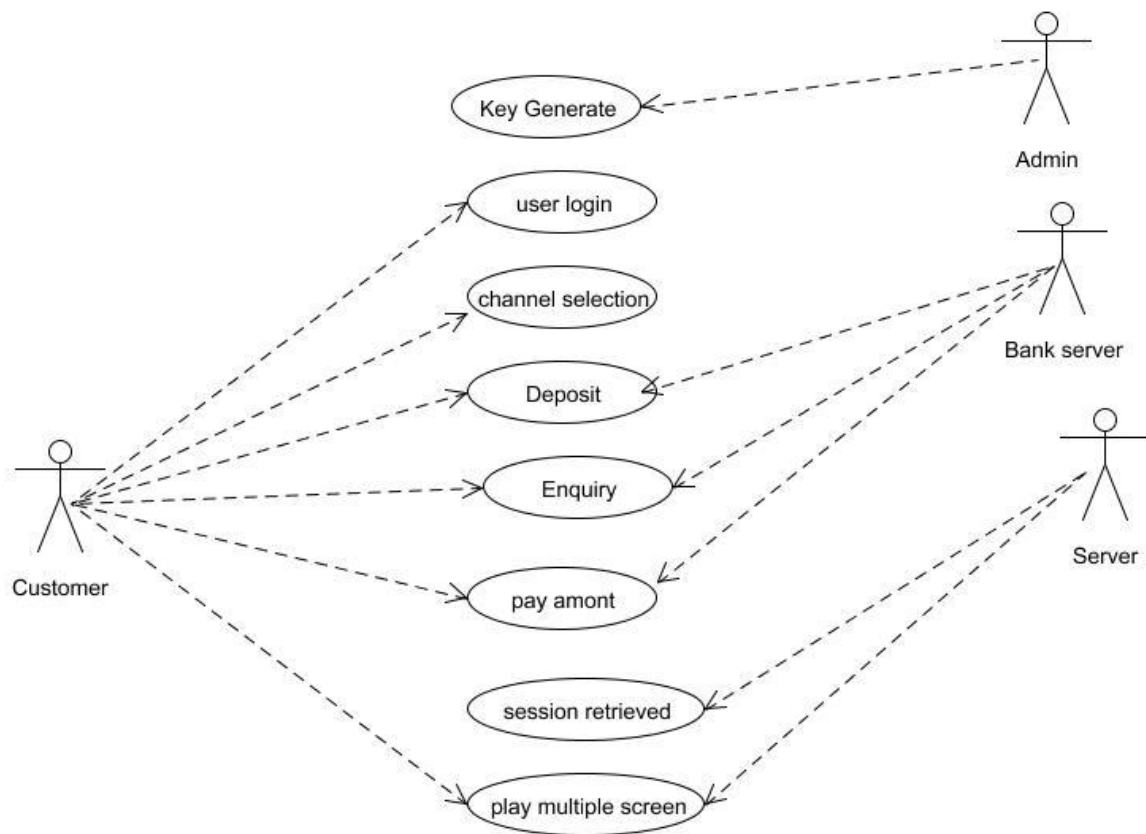
COLLABORATION DIAGRAM:



SEQUENCE DIAGRAM:

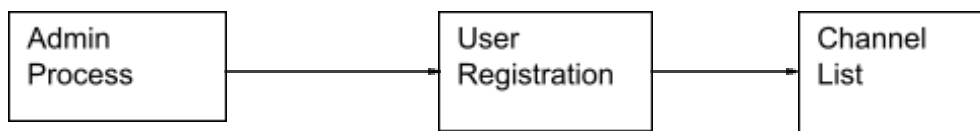


USECASE DIAGRAM:

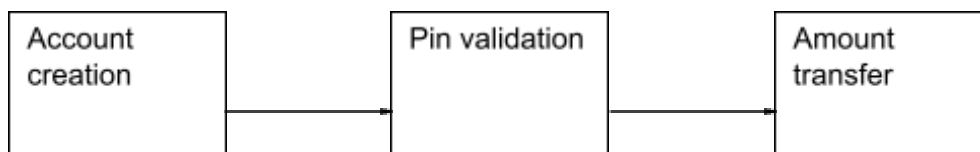


DFD

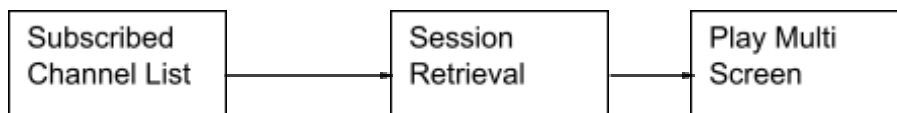
Level 0:



Level 1:



Level 2:



Level 3:

