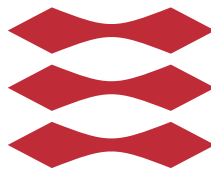


Non Identifying Data Management Systems

Dheeraj Kumar Bansal

DTU



Kongens Lyngby 2015

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

The goal of the thesis is to ...

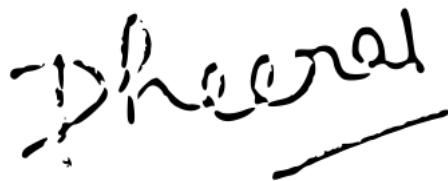
Preface

This thesis was prepared at DTU Compute in fulfilment of the requirements for acquiring an M.Sc. in Engineering.

The thesis deals with ...

The thesis consists of ...

Lyngby, 26-June-2015

A handwritten signature in black ink, appearing to read 'Dheeraj', with a long horizontal stroke underneath.

Dheeraj Kumar Bansal

Acknowledgements

I would like to thank my....

Contents

Summary (English)	i
Preface	iii
Acknowledgements	v
List of Figures	xi
1 Problem Statement and Background	1
1.1 Introduction	1
1.2 Background	1
1.3 Definitions	2
1.3.1 Privacy	3
1.3.2 Anonymity	3
1.3.3 Pseudonym	3
1.3.4 User	3
1.3.5 Bank	3
1.3.6 Service	3
1.3.7 Unlinkability	3
1.3.8 Revocation	3
1.3.9 Partial Information Disclosure	3
1.3.10 Conditional Anonymity Removal	3
1.3.11 Legal Requirement	3
1.4 Case Study	3
1.4.1 Private Customers	4
1.4.2 Corporate Customers	4
1.5 Problem	4
1.6 Summary	5

2	State of the art Survey	7
2.1	Introduction	7
2.2	Technologies	7
2.3	Secure Multi-Party Computing	8
2.3.1	Homomorphic Encryption	8
2.3.2	Group Signature	9
2.4	Escrow Technologies	9
2.4.1	Secure Logging	9
2.4.2	Threshold Cryptography	9
2.5	Identity Management Systems	10
2.5.1	OpenID	10
2.6	Zero Knowledge Technologies	11
2.6.1	IDEMIX	11
2.6.2	U-Prove	13
2.7	Summary	15
3	Application Scenario	17
3.1	Introduction	17
3.2	Information Flow	17
3.3	Separation of Identities	19
3.4	Current Banking System	22
3.5	summary	24
4	Proposed System	25
4.1	Introduction	25
5	Simple Prototype	27
5.1	Introduction	27
6	OpenID Based Solution	29
6.1	Introduction	29
7	IDEMIX Based Solution	31
7.1	Introduction	31
8	Analysis	33
8.1	Introduction	33
9	Conclusion and Future work	35
9.1	Introduction	35
10	Anything Extra	37
10.1	Introduction	37
A	Appendix A	39

CONTENTS	ix
Bibliography	41
Bibliography	41

List of Figures

1.1	Identities in the system	4
2.1	Technology Overview	8
3.1	Example of information maintained for each relationship	18
3.2	Current Banking System	23

CHAPTER 1

Problem Statement and Background

1.1 Introduction

This chapter introduce the topic of the thesis. It defines the problem statement and also give a brief background about it.

1.2 Background

Administrative data systems currently rely on the Central Personal Register Id (CPR Nr.) to link customer data with a real world identity. This means that almost all data managed by the institutions must be classified as personal identifiable information and therefore managed according to strict confidentiality requirements as well as integrity and availability requirements. The purpose of this thesis project is to examine ways to decouple transaction data from the underlying identities, so that the data used in the institution's day to day operations are decoupled from the underlying customer identity. This limits the confidentiality requirements for the data and the vulnerability to insider threats, such as the recent leak of celebrity data from NETS to the magazine "Se & Hør".

It must, however, be possible to link customer data to real world identities when reporting financial data to the Tax authorities or in connection with suspicions about criminal activity, e.g. fraud, insider trading, whitewashing, etc. Austria is already considering similar approaches in the public health care system, where the health records of the citizens are saved under pseudonyms, which are mapped one-to-one to the single citizen.

1.3 Definitions

Here we will give some of the definitions used in the system

1.3.1 Privacy**1.3.2 Anonymity****1.3.3 Pseudonym****1.3.4 User****1.3.5 Bank****1.3.6 Service****1.3.7 Unlinkability****1.3.8 Revocation****1.3.9 Partial Information Disclosure****1.3.10 Conditional Anonymity Removal****1.3.11 Legal Requirement****1.4 Case Study**

Nykredit is a major financial institution in Denmark providing different services, such as mortgages, retail banking, investment banking etc. They also are part of a big group of companies, which includes other financial institutions providing similar services. These financial institutions basically provide Nykredit services as their own services to the customers. Nykredit has mainly 2 types of customers:

- Private Customers
- Corporate Customers

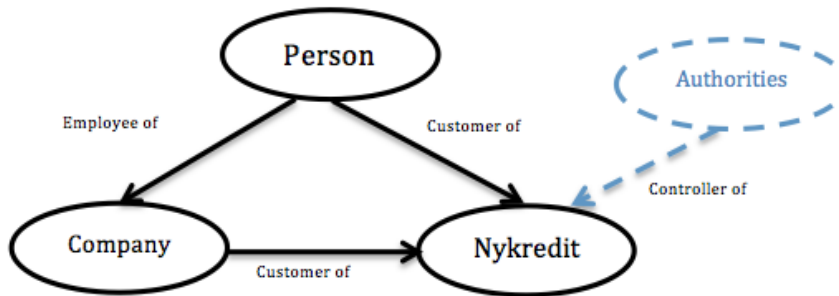


Figure 1.1: Identities in the system

1.4.1 Private Customers

Private customers are the individual customers who access Nykredit services on their own. Usually there is a single person accessing the services of the bank. These customers are usually people who get a personal bank account with Nykredit.

1.4.2 Corporate Customers

Corporate customers are either companies who are customers of Nykredit or other financial institutions which provide Nykredit services to their own private customers. Usually there are many people who access Nykredit services on behalf of the corporate customer.

1.5 Problem

We consider the case of a person, who may either be a private customer of Nykredit, or an employee of a company who is corporate customer of Nykredit. In this case, the person may also be responsible for managing the accounts of his employer with Nykredit.

Nykredit wants to setup an identity management system so that there is no need for the individual to disclose his personal identity to Nykredit to access the account on behalf of the company.

Nykredit, however, also have to comply with relevant legislation (KYC, AML,

“Hvidvaskningsloven”), e.g. in case the authorities (Tax, Police, etc.) find some suspicious transactions. Nykredit needs to provide the identity of the person responsible for these transactions. This means that it is required that Nykredit, in case of a legal request, is able to identify the individual employee from the institution, who is accessing the account on the corporate customer’s behalf. So the main goal of the system is:

- Nykredit should not learn identity of the individual person accessing the services on behalf of corporate customer.
- For complying with legislation Nykredit should be able to map real identity of the individual person with the transaction in case its required by the law.

The project will perform an initial analysis of a single business process from administrative data management, with respect to identifying the need to bind authenticated identities to actions at the different steps of the process; this analysis will be presented to stakeholders from the specific administrative domain. Based on the initial analysis of the selected business process, the project will develop a full identity model for the chosen business process with anonymisation and pseudonymisation of actors whenever possible. The feasibility of the proposed model will be evaluated through a prototype that implements the model using standard components from identity management infrastructures whenever possible.

1.6 Summary

Companies do not want to disclose the personal identity of their employees to Nykredit, but they still need the ability to access all services online. Managing all identities, while maintaining privacy, is not easy and provides different challenges. We have to design a system, which fulfill the entire privacy requirement and still enables Nykredit to provide its services to its customers and meet the regulatory requirements of the authorities without any problem.

CHAPTER 2

State of the art Survey

2.1 Introduction

This chapter introduce all the different technologies currently available which deal with anonymity and privacy. Some of the technologies are already commercially available and being used in the industry, while some are still in research phase. We will give a brief introduction for all of them and brief idea how they can be used.

2.2 Technologies

1. Secure Multi-Party Computing
2. Escrow Technologies
3. Identity Management Systems
4. Zero Knowledge Technologies

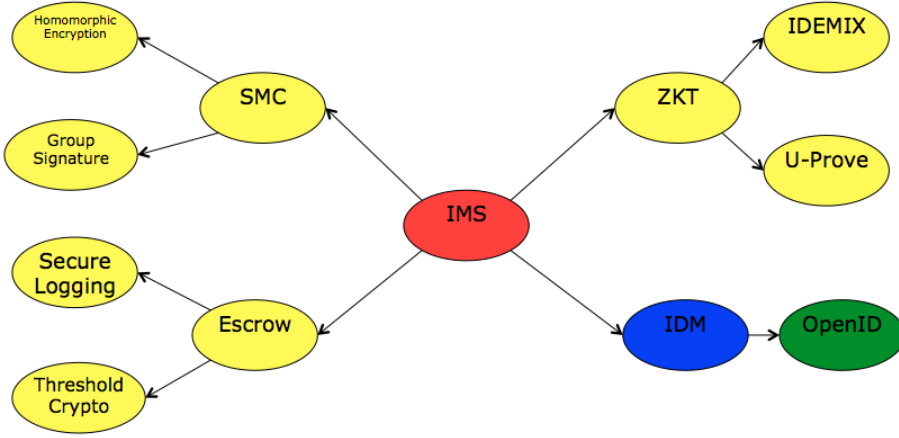


Figure 2.1: Technology Overview

2.3 Secure Multi-Party Computing

These technologies are the ones which involve multiple parties to do computations. It is a subfield of cryptography which involves multiple parties getting an input and compute a joint function on them while keeping these inputs private. We basically looked at 2 technologies of interest here:

2.3.1 Homomorphic Encryption

Homomorphic encryption is a type of encryption where certain arithmetic operations can be performed on the ciphertext such that when the resultant ciphertext is decrypted, the decrypted text is same as the operations were performed on the plaintext. This is a new field of cryptography and is very useful where we need some parties to perform such operations without revealing the underlying data to those parties. Homomorphic encryption is also useful for chaining of different services without exposing data to any of those services.

2.3.2 Group Signature

A group signature is a scheme which allows a member of the group to sign the message on behalf of the group but anonymously. To outsiders the message has been signed by someone from the group but the exact identity of the person is now known. Also if the same member signs 2 different messages its not possible to know if the message is signed by the same member. There is a notion of group manager in these scheme. Group manager is someone who manages the membership to the group. He can add/remove members from the group, find out who actually signed the message from the group. This scheme is useful where only thing that needs to be validated is that a certain person is part of the group, but his real identity is not required.

2.4 Escrow Technologies

Escrow technologies are the ones which are helpful in escrow purposes i.e. getting real data/identity later on in time from encrypted data if needed. We look at following 2 technologies.

2.4.1 Secure Logging

Secure logging is the process of saving the data in a secure manner. As saved data is really crucial and vulnerable to the attacks. We need to make sure that data is saved securely and its integrity is protected. This can be done in several ways. One way is to encrypt all the logs while storing them so that even if someone get hold of the logs, they can't use them without having access to the decryption key. Other way is to store logs at a third party after encrypting them. For escrow purposes, these logs can be decrypted later on with the decryption key.

2.4.2 Threshold Cryptography

Threshold cryptography is a field of public key cryptography where in order to decrypt an encrypted message, several parties must cooperate in the decryption. This message is encrypted using a public key and the corresponding private key is shared among different parties who will participate in the decryption process i.e. multiple parties hold the private key for a single public key. There is a term

called threshold, and if there are n parties who share the private key and at least t parties which are required to decrypt the message such system is called (t,n) threshold cryptosystem. Threshold cryptosystem is useful in escrow purposes where a minimum number of parties can be defined to decrypt the ciphertext in order to get the plaintext.

2.5 Identity Management Systems

These are traditional identity management systems. For our purposes we look at OpenID system.

2.5.1 OpenID

OpenID is open standard and decentralized protocol which can be used to authenticate to different cooperating sites with the use of a third party service. There is a notion of *Relying party* and *Identity Provider* in this system. The authentication steps in OpenID are as follows:

- User goes to the relying party service page
- Service page presents different OpenID providers to login to the service
- User chose the provider he has registered his openID with
- Relying party redirects the user to the OpenID provider url so that user can authenticate
- User can be authenticated by the method provided by OpenID provider
- OpenID provider asks user the permission to share the attributes with the relying party
- After user consent, user is redirected to the relying party website with user credentials
- Relying party can verify the credentials and then login the user to the service

2.6 Zero Knowledge Technologies

These are the technologies which use the concept of zero knowledge i.e. proving knowledge about something without divulging the information. For our purpose we focus on mainly 2 technologies – IDEMIX and U-Prove, which are based on the concept of zero knowledge and verifiable encryption.. They both have a lot in common and have been studied a lot in industry.

2.6.1 IDEMIX

U-Prove is a digital credential technology by IBM. it relies on anonymous credentials known as IDEMIX tokens. It is based on Camenisch-Lysyanskaya (CL) signature scheme which provides efficient zero-knowledge proofs. IDEMIX have different entities in the system

- **User** User is basically the entity, which is proving his identity in the system.
- **Verifier** Verifier is the entity, which verifies the identity of the prover.
- **Issuer** Issuer is the entity, which issues the credentials to the prover to prove his identity
- **Inspector** Inspector is the entity, which in case of discrepancy or legal requirement , can actually come and get the real identity of the prover.

For IDEMIX we need computing devices which work on behalf of each entity in the system.

2.6.1.1 IDEMIX Credential

An Idemix credential is CL Signature by issuer on user's private key and on attribute values. A user have independent public keys or pseudonyms for the same private key. These pseudonyms are IDEMIX tokens which are then used by the user to prove his identity to the different verifiers. IDEMIX has been studied a lot and many EU projects on anonymous credentials are based on it e.g. FutureID, ABC4Trust etc.

2.6.1.2 Issuance

The first step is the credential issuance. It involves following steps:

- User sends credential request to the issuer
- Issuer presents the issuance policy specifying
 - Which attributes to present
 - Which pseudonym/existing credentials to present
- Issuer also present a credential template specifying
 - Which attributes of the new credentials will be generated at random
 - Or carried over from existing credential or pseudonym
- User then present the issuance token satisfying the issuance policy
- Then multi-round cryptographic protocol ensues at end of which user get the IDEMIX credential

2.6.1.3 Presentation

Next step is to present the IDEMIX token for authentication to the verifier. It consists of following steps:

- User get the presentation policy from the verifier which specifies
 - Which credentials user must present
 - What information user should reveal from these credential
- User generate a presentation token in accordance with the presentation policy revealing only the attributes necessary
- User present this presentation token to the verifier
- Verifier can then verify the attributes

2.6.1.4 ID Escrow

IDEMIX provides the ID escrow ability in case it is required. Following steps need to be followed for ID escrow purposes in IDEMIX:

- The presentation policy can have following optional specifications for the purpose of ID Escrow
 - Public keys of inspectors
 - Attributes values to be encrypted using the keys
 - Inspection conditions under which these attributes can be revealed
- User can prove that he has put these values in the presentation token with verifiable encryption to the verifier
- Once the token is presented, the inspection conditions are fixed and cannot be changed
- In case of some discrepancy or legal requirement, an inspector can come and get the identity of the user from the token

2.6.2 U-Prove

U-Prove is a digital credential technology by Microsoft. It relies on anonymous credentials known as U-Prove tokens. It provides users a way to minimally disclose their personal information while interacting with different online services. U-prove have different entities in the system

- **Prover** Prover is basically the entity, which is proving his identity in the system.
- **Verifier** Verifier is the entity, which verifies the identity of the prover.
- **Issuer** Issuer is the entity, which issues the credentials to the prover to prove his identity
- **Auditor** Auditor is the entity, which in case of discrepancy or legal requirement, can actually come and get the real identity of the prover.

For U-Prove we need computing devices which work on behalf of each entity in the system.

2.6.2.1 U-Prove Token

U-Prove token is basically cryptographically protected information of any kind e.g. attributes. These are issued by issuer to the prover by issuance protocol. These tokens are then presented by prover to the verifier. Issuance and presentation of U-Prove tokens is unlinkable.

2.6.2.2 Issuance

The first step is the credential issuance. It involves following steps:

- Prover invoke U-Prove issuance protocol
- Prover provides the attributes to be encoded

Using *Collaborative Issuance* property user can make sure that issuer doesn't actually knows the attributes
- Then multi-round cryptographic protocol ensues at end of which user get the U-Prove token from the issuer

2.6.2.3 Presentation

Next step is to present the U-Prove token for authentication to the verifier. It consists of following steps:

- Prover invoke the U-Prove presentation protocol
- User generate a presentation token in accordance with the presentation policy revealing only the attributes necessary
- User present this presentation token to the verifier
- Verifier can then verify the attributes

It must be noted that a revocation check can be added if needed before verifying the token.

2.6.2.4 ID Escrow

ID Escrow in U-Prove is actually an extension to existing U-Prove technology. It uses a type of ElGamal encryption which is verifiable.

- During the presentation protocol, prover proves that his ID is encrypted in the token by the use of verifiable encryption technology
- De-anonymization cannot be done by verifier or issuer

- A special entity called Auditor is responsible for de-anonymization in case of some discrepancy or legal requirement
- Threshold cryptography can be used in case of auditors and key can be split among multiple auditors.

2.7 Summary

All these different technologies provide different level of anonymization in the system. Some of them are easy to integrate in existing technology, while some are still not mature enough. For our purposes from now on we focus on mainly 2 technologies:

- OpenID
- IDEMIX

CHAPTER 3

Application Scenario

3.1 Introduction

This chapter will discuss the information flow of the current system. We will present our understating of the current banking system. we will also give different types of data that exist in the current system and the different operations that are necessary to support the system.

3.2 Information Flow

A person has different information associated with him.
To Nykredit it can be:

- Personal
This Personal ID can be one of the following identifiers:
 - External ID (e.g. NEMid)
 - Internal ID (e.g. login credentials of the bank)

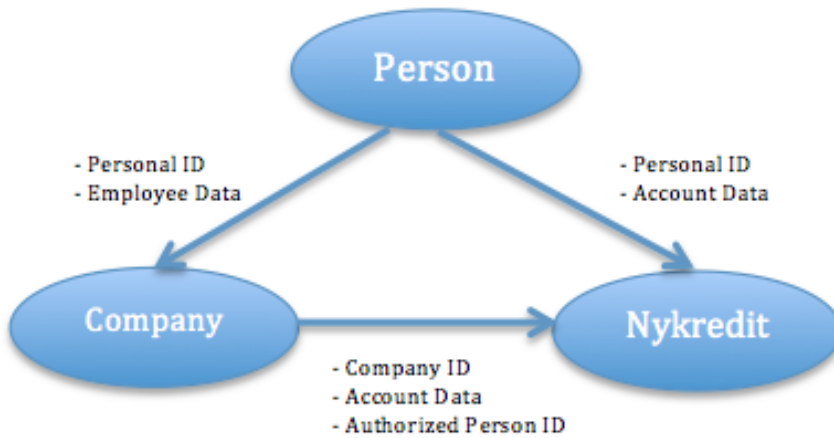


Figure 3.1: Example of information maintained for each relationship

- Account Data

To Company it can be:

- Personal ID
- Employee Data

A company has following information that it might share with Nykredit:

- Company ID
- Account Data
- Authorized Person ID

This Authorized Person ID is the identifier that is used by the authorized person on behalf of the company. This can be stored in some database where it is matched to Personal ID of the person.

The Authorized Person ID can either be same identifier as the Personal ID, the Company ID or a different identifier that can be authenticated.

In case the Personal ID is used as the Authorized Person ID, it gives Nykredit some additional capabilities:

- Nykredit can use this information to recruit new customers. If the authorized person is not a customer before, Nykredit can use this info to contact them.
- If the person is already a customer, then Nykredit can use this information to provide additional services to him on his personal account, so as to influence the authorized person for his decisions regarding the company account (similar to the way airlines reward frequent flyers).
- As Nykredit already knows about company accounts, the performance of the company might influence their decision regarding the private account of the employee (e.g. it may be difficult for a person to take out a mortgage if Nykredit knows that the company they work for is in financial difficulties).
- In the case where the customer is accessing Nykredit services on behalf of a smaller financial institution, the capability of matching the authorized ID to Personal ID gives a chance to Nykredit to recruit this customer away from the smaller financial institutions.

While good corporate governance at Nykredit will prevent these issues, it is desired to completely and demonstratively remove the link between Personal ID and Authorized Person ID. This, however, creates difficulties with respect to regulatory requirements for accountability at Nykredit and KYC, AML, "Hvidvaskningsloven, etc., so there must be some way to map the identifier used in a financial transaction to a real person, i.e. map a given Authorized Person ID to a Personal ID.

3.3 Separation of Identities

A way to remove the link between Personal ID and Authorized Person ID is to use separate IDs and to maintain a database which links the Authorized Person ID back to the Personal ID. This database can be protected in different ways, so that the information can be used only in case legal authorities need to link the two IDs.

This database can be maintained at 3 places :

- Company
- Nykredit
- Trusted 3rd party

3.3.0.5 Database on Company Side

Advantages

- Nykredit doesn't have to invest extra in IT infrastructure
It is expensive to maintain the entire IT infrastructure by Nykredit, so it is easier and cheaper for Nykredit to let the company maintain the database.
- Nykredit can easily prove that it cannot link different Identities
Nykredit does not have access to the database, so it can easily be proved that Nykredit cannot link the different identities.
- Company maintain their own private data
Companies can be sure that Nykredit does not have access to the personal data of their employees

Disadvantages

- There is no way to retrieve data if the company stops existing
In this case the entire mapping database may be lost.
- Authorities have to go to the Company to get the data
Nykredit does not have access to the database, so the authorities have to go to Nykredit first to obtain the Authorized Person ID and then to the individual companies next to get the Personal ID.
- Company might tamper with the database
In the case of a rogue employee at the company, which is exactly the case that the legislation is intended to identify, this employee will have easy access to this database and hence the ability to tamper with the database and remove the authorities ability to identify him.
- It might prove too difficult for new customers to fulfill all the technical requirements
Larger companies can have their own IT infrastructure, but for smaller companies it might prove to be a difficult task to become a new Nykredit customer if they have to invest extra in IT infrastructure just for this purpose.

3.3.0.6 Database on Nykredit Side

Advantages

- In case the company stops existing, the data can still be retrieved
The database is always with Nykredit, so if some company stops existing, it can still be accessed.
- Authorities have a single place to obtain all the data
Authorities do not have to go to individual companies to get the relevant data as everything is at one place.
- Company cannot tamper with database
As companies have no direct access to the database, they cannot tamper with it.

Disadvantages

- Nykredit has to invest extra in IT infrastructure
Nykredit has to invest extra to keep this system in place.
- Company does not have control over their own private data
The database is on Nykredit side, so companies have to store the data there and hence they do not have control over their own private data.
- It is difficult for Nykredit to prove that they cannot link different identities when they are managing the database
Nykredit will be managing everything in-house, so it is difficult to prove that they cannot access the database and link the identities.
- It may be difficult for customers to adhere to the Nykredit technological standards
Nykredit may not be able to support all available technologies for their customers, so some customers, who are using a different setup than Nykredit, may find it difficult to comply with the Nykredit standard.

3.3.0.7 Database on 3rd Party Side

Advantages

- Neither companies nor Nykredit have to invest extra in IT infrastructure
The database is managed by the trusted 3rd party, who will invest in the infrastructure, so neither Nykredit nor the companies will have to invest extra in IT infrastructure

- It is easier for new and old customers to be a customer at Nykredit
The trusted 3rd party can support a wide range of technologies, so it is easier for customers to use their existing technology when becoming a new customer at Nykredit
- Nykredit can easily prove that it cannot link different Identities
Nykredit is not hosting the database, so it is easier for them to prove that they cannot link the identities.
- Data can still be retrieved in case the company stops existing
The database is always with the trusted 3rd party, so it does not matter if some company stops existing, the data can still be accessed. Special arrangements have to be made in case the trusted 3rd party ceases to exist, but this will be rare and in that case, Nykredit may decide to take over that part of the trusted 3rd party.
- Company cannot tamper with database
The companies do not have access to the database, so they cannot tamper with the data.
- Authorities only have to go to the trusted 3rd party to get the data in case its needed.

Disadvantages

- The 3rd party must be trusted by both Nykredit and its customers
The database is neither with the company nor Nykredit, so the external service provider should be trusted by both parties to hold their sensitive data.
- In case the trusted 3rd party goes out of business it might be difficult to retrieve the data.
- Companies do not have control over their own data
The database is maintained by an external service provider, so the companies have to store the data there and hence they do not have control over their own private data.

3.4 Current Banking System

The current banking system can be seen as following. There are 2 parts of the system:

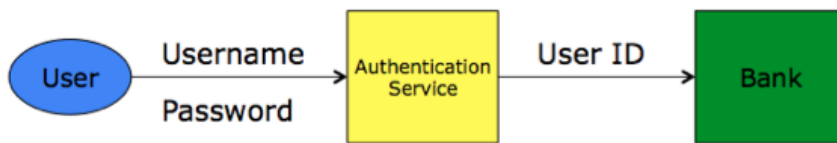


Figure 2: Current Banking System

Figure 3.2: Current Banking System

- Authentication Service
- Bank

The end user interact with the service as follows:

- User goes to the authentication service and enters his credentials
- Authentication service authenticate the user and gives the bank User ID of the user
- All other details of the user are stored at bank side in relation with his user ID
- Bank provide services to the user

The authentication service can either be controlled by bank or a 3rd party (e.g. NemID)

In this system Bank is the most powerful entity. It has all the mappings

- User ID -> Account ID
- User ID -> Policy
- User ID -> User Personal Data
- User ID -> Transactions

This makes the bank most powerful identity in the system. Its very easy for bank to get any private data of customers using the data stored on bank side.

3.5 summary

In nutshell current banking system gives bank a lot of power over user data. Also we have seen that traditional methods are not sufficient to provide proper anonymity to the users or they don't fit all the requirements properly.

CHAPTER 4

Proposed System

4.1 Introduction

This chapter will present our proposed system as the solution. We will define our system and show how it solves the problem of privacy of the users.

CHAPTER 5

Simple Prototype

5.1 Introduction

This chapter will deal with our Identity management based solution. Mainly we will talk about OpenID solution here. And how this solution will apply to the application scenario discussed in chapter 3.

CHAPTER 6

OpenID Based Solution

6.1 Introduction

In this chapter we will discuss and analyse the solutions presented in chapter 4 and 5.

CHAPTER 7

IDEMIX Based Solution

7.1 Introduction

From chapter 6 we will pick up one solution and will design a prototype based on the advantages. In this chapter we will explain the design.

CHAPTER 8

Analysis

8.1 Introduction

This chapter will explain our implementation of the prototype from the design as discussed in chapter 7.

CHAPTER 9

Conclusion and Future work

9.1 Introduction

This chapter will deal with the evaluation of the prototype implemented in chapter 8.

CHAPTER 10

Anything Extra

10.1 Introduction

Final chapter will conclude the thesis and will give some directions to the future work that can be done in the field.

APPENDIX A

Appendix A

This appendix is full of stuff ...

Bibliography
