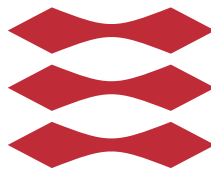


Non Identifying Data Management Systems

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 ...

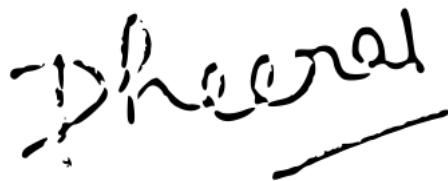
Preface

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	ix
1 Problem Statement and Background	1
1.1 Introduction	1
1.2 Background	1
1.3 Definitions	2
1.3.1 Privacy	2
1.3.2 Anonymity	2
1.3.3 Pseudonym	2
1.4 Case Study	2
1.4.1 Private Customers	2
1.4.2 Corporate Customers	3
1.5 Problem	3
2 State of the art Survey	5
2.1 Introduction	5
2.2 Technologies	5
2.3 Secure Multi-Party Computing	6
2.3.1 Homomorphic Encryption	6
2.3.2 Group Signature	7
2.4 Escrow Technologies	7
2.4.1 Secure Logging	7
2.4.2 Threshold Cryptography	7

2.5	Identity Management Systems	8
2.5.1	OpenID	8
3	Application Scenario	9
3.1	Introduction	9
3.2	Components	9
4	Zero Knowledge based solution	13
4.1	Introduction	13
5	IDM based solution	15
5.1	Introduction	15
6	Discussion/Analysis of solution	17
6.1	Introduction	17
7	Design of Prototype	19
7.1	Introduction	19
8	Implementation of Prototype	21
8.1	Introduction	21
9	Evaluation of Prototype	23
9.1	Introduction	23
10	Conclusion and Future work	25
10.1	Introduction	25
A	Appendix A	27
	Bibliography	29
	Bibliography	29

List of Figures

1.1	Identities in the system	3
2.1	Technology Overview	6

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

Following are some of the definitions used in the system

1.3.1 Privacy

1.3.2 Anonymity

1.3.3 Pseudonym

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

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.

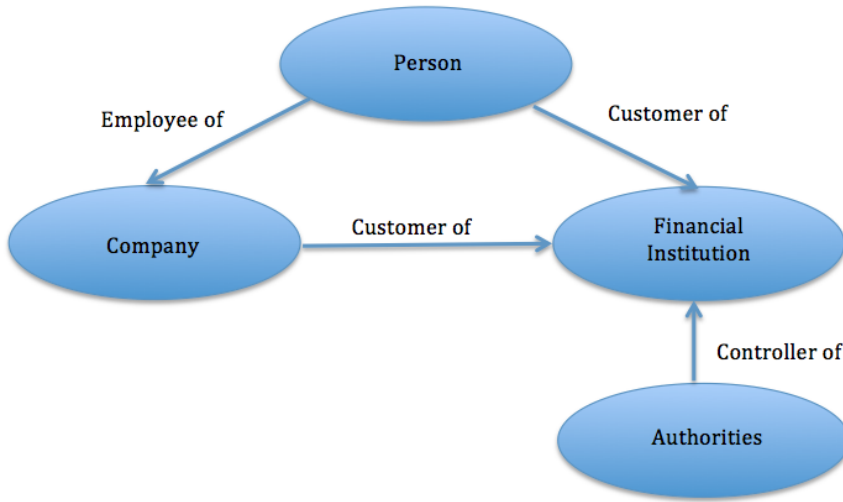


Figure 1.1: Identities in the system

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.

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

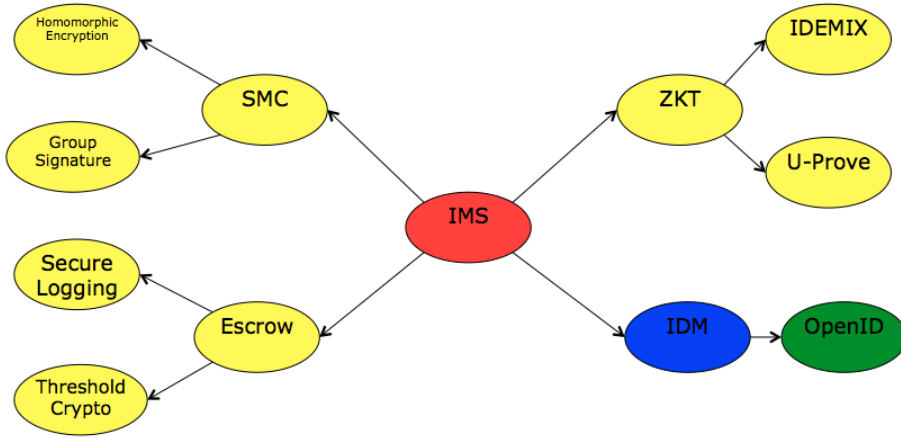


Figure 2.1: Technology Overview

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 authenticate by the method provided by OpenID provider
- OpenID provider ask 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

CHAPTER 3

Application Scenario

3.1 Introduction

Here we will discuss the application scenario of the technologies discussed in chapter two. Most of it will be based on the models we described in the requirement document and also banking document. In this chapter we will try to describe our understanding of the current banking system. We will give different types of data that exist in the current system and the different operations that it is necessary to support in the system.

3.2 Components

We have identified following four components that the bank needs to maintain in its relationship with its customers. Together, these four components define a customer engagement:

- ID This is the main identity of the user. The user is identified in the system using this ID. This ID can be anything from a pseudonym, as

in the numbered Swiss bank accounts, to the verified real world identity (CPR number) of the customer used in Danish banks.

- Basic Data Related to the ID is the basic data of the user. This data is the data that is required by the bank to identify the customer in real life and maintain its relationship with the customer. Basic data can consist of the following:
 - Name
 - Address
 - Email ID
 - Phone Nr.
 - CPR nr.
 - Marital Status
 - Gender
 - Date of Birth This is the most basic form of data which describes a single customer and which rarely changes. It can include some other data that might be crucial for the bank.
- Account Next component in the chain is the account of user with the bank. The account component holds all the static information regarding the account. Examples of such information are:
 - Account Nr.
 - Account Type
 - Owner ID
 - Interest rate
 - Balance
 - Account opening date
 - Overdraw limit As before it can include some other data that might be needed to operate the account or that might be crucial for the bank.
- Transaction History The transaction history includes all the dynamic data that the bank has on a particular account. Typical transactions are:
 - Deposits
 - Withdrawals

- Accruing Interests
- Authorization and Access Control In the following, we identify the most basic operations needed to maintain the information above and the customer/bank relationship. This includes the operations that are permitted on the accounts. We represent it in our system as an API, which takes an input, and perform the desired operation. Some examples can be:
 - Deposit (ID, Account Nr., Amount) This is the most basic operation. This will take user ID, Account Nr. and deposit the amount in the account.
 - Withdraw (ID, Account Nr., Amount) This will take user ID, Account Nr. and withdraw the amount from the account.
 - Transfer (ID, Account Nr. 1 Account, Nr. 2, Amount) This will take user ID of the person initiating the transfer, Account Nr. 1, Account Nr. 2 and transfer the amount from Account Nr. 1 to Account Nr. 2.
 - Close Account (ID, Account Nr.) This will take user ID, Account Nr. and close the account.
 - Open Account (ID) This will take user ID and open an account for the given user ID.
 - Actions (ID, Account Nr., ID1, Action1, ID2, Action2, ..., IDn, Actionn) This will take user ID, Account Nr. and other IDs and Actions that those IDs are allowed to do on the account and then will create a policy for those IDs in the database. For all above operations we assume that ID is authorized to perform such operations.

CHAPTER 4

Zero Knowledge based solution

4.1 Introduction

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

CHAPTER 5

IDM based solution

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

Discussion/Analysis of solution

6.1 Introduction

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

CHAPTER 7

Design of Prototype

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

Implementation of Prototype

8.1 Introduction

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

CHAPTER 9

Evaluation of Prototype

9.1 Introduction

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

CHAPTER 10

Conclusion and Future work

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
