Non Identifying Data Management Systems

Dheeraj Kumar Bansal

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

Dheeraj Kumar Bansal

Acknowledgements

I would like to thank my....

Contents

Chapter 1

Problem and Background

1.1 Introduction

In this chapter we will talk about the problem that is the topic of the thesis.

1.2 Problem

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

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 need to provide the identity of the person responsible for these transactions. This means that it is required that Nykredit, in caase 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.

1.3 Information Flow

A person has different information associated with him.

To Nykredit it can be:

- Personal ID
  - External ID (e.g. NEMid)
  - Internal ID (e.g. login credentials of the bank)

  This Personal ID can be one of the following identifiers:
- 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.

One way to do this 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 at 3 places :

- Company Side
- Nykredit Side
- 3rd Party

The arrangements have their own advantages and disadvantages:

1.3.1 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.

1.3.2 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.

1.3.3 Database on 3rd Party Side

There is also a possibility if either Nykredit or the Company transfers the database handling to a trusted 3rd party.

- 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.

1.4 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 will provide an insight about the anonymous identity management systems currently available. We will talk about the technologies currently in market. This will mostly be our technology overview document but in an elaborate manner.

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

This appendix is full of stuff ...

Bibliography

[1] C. A. R. Hoare, "Communicating sequential processes," *Communications of the ACM*, vol. 26, no. 1, pp. 100–106, 1983.

[2] E. Emerson and E. Clarke, "Characterizing correctness properties of parallel programs using fixpoints," in *Automata, Languages and Programming* (J. de Bakker and J. van Leeuwen, eds.), vol. 85 of *Lecture Notes in Computer Science*, pp. 169–181, Springer Berlin / Heidelberg, 1980.

[3] O. M. Group, *Business Process Model and Notation (BPMN) 2.0*. Needham MA, USA: Object Management Group, 2011.

[4] R. M. Dijkman, M. Dumas, and C. Ouyang, "Formal semantics and analysis of bpmn process models." 2007.

[5] C. Ouyang, M. Dumas, and A. H. M. T. Hofstede, "Pattern-based translation of bpmn process models to bpel web services," *International Journal of Web Services Research (JWSR)*, vol. 5, no. 1, pp. 42–62, 2007.

[6] P. Y. Wong and J. Gibbons, "A process semantics for bpmn," in *Proceedings of the 10th International Conference on Formal Methods and Software Engineering*, ICFEM '08, (Berlin, Heidelberg), pp. 355–374, Springer-Verlag, 2008.

[7] M. Kwiatkowska, G. Norman, and D. Parker, "Prism: probabilistic model checking for performance and reliability analysis," *SIGMETRICS Perform. Eval. Rev.*, vol. 36, pp. 40–45, March 2009.

[8] M. Kwiatkowska, G. Norman, and D. Parker, "PRISM 4.0: Verification of probabilistic real-time systems," in *Proc. 23rd International Conference on*

Computer Aided Verification (CAV'11) (G. Gopalakrishnan and S. Qadeer, eds.), vol. 6806 of *LNCS*, pp. 585–591, Springer, 2011.

[9] E. Clarke, O. Grumberg, S. Jha, Y. Lu, and H. Veith, "Counterexample-guided abstraction refinement," in *Computer Aided Verification* (E. Emerson and A. Sistla, eds.), vol. 1855 of *Lecture Notes in Computer Science*, pp. 154–169, Springer Berlin / Heidelberg, 2000.

[10] R. Mardare, C. Priami, P. Quaglia, and O. Vagin, "Model checking biological systems described using ambient calculus," in *Computational Methods in Systems Biology* (V. Danos and V. Schachter, eds.), vol. 3082 of *Lecture Notes in Computer Science*, pp. 85–103, Springer Berlin / Heidelberg, 2005.

[11] J. A. Fisteus, L. S. Fernández, and C. D. Kloos, "Applying model checking to bpel4ws business collaborations," in *SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*, (New York, NY, USA), pp. 826–830, ACM, 2005.

[12] M. Kwiatkowska, G. Norman, and D. Parker, "Prism: probabilistic model checking for performance and reliability analysis," *SIGMETRICS Perform. Eval. Rev.*, vol. 36, no. 4, pp. 40–45, 2009.

[13] K. L. McMillan, *Symbolic Model Checking.* Norwell, MA, USA: Kluwer Academic Publishers, 1993.