CBAC – A model for conflict-based access control

by

Marianne Loock

Thesis

submitted in fulfillment of the requirements for the degree

Doctor of Philosophy

in the subject of

Computer Science

in the

Faculty of Engineering, Built Environment and Information Technology

University of Pretoria

Pretoria

Supervisor

Prof JHP Eloff

September 2012

ABSTRACT

Organisations that seek a competitive advantage cannot afford to compromise their brand reputation or expose it to disrepute. When employees leek information, it is not only the breach of confidentiality that is a problem, but it also causes a major brand reputation problem for the organisation. Any possible breach of confidentiality should be minimised by implementing adequate security within the organisation and among its employees. An important issue to address is the development of suitable access control models that are able to restrict access not only to unauthorised data sets, but also to unauthorised combinations of data sets. Within organisations such as banks, clients may exist that are in conflict with one another. This conflict results from the fact that clients are functioning in the same business domain and that their information should be shielded from one another because they are in competition for various reasons. When information on any of these conflicting clients is extracted from their data sets via a data-mining process and used to their detriment or to the benefit of the guilty party, this is considered a breach of confidentiality.

In data-mining environments, access control usually strips the data of any identity so as to concentrate on tendencies and ensure that data cannot be traced back to a respondent. There is an active research field in data mining that focuses specifically on ‘preserving’ the privacy of the data during the data-mining process. However, this approach does not account for those situations when data mining needs to be performed to give answers to specific clients. In such cases, when the clients’ identity cannot be stripped, it is essential to minimise the chances of a possible breach of confidentiality. For this reason, this thesis investigated an environment where conflicting clients’ information can easily be gathered and used or sold, as to justify the inclusion of conflict management in the proposed access control model.

This thesis presents the Conflict-based Access Control (CBAC) model. The model makes it possible to manage conflict on different levels of severity among the clients of an organisation – not only as specified by the clients, but also as calculated by the organisation. Both types of conflict have their own cut-off points when the conflict is considered to be of no value any longer. Finally, a proof-of-concept prototype illustrates that the incorporation of conflict management is a viable solution to the problem of access control as it minimises the chances of a breach of confidentiality

PREFACE

Three peer-reviewed publications resulted from this study and the fourth is under review. The full papers are included in Appendix B of this document.

- Marianne Loock, Jan HP Eloff: Minimizing Security Risk Areas revealed by Data mining

- Marianne Loock, Jan HP Eloff: Investigating the usage of the Chinese wall security policy model for data mining,

- Marianne Loock, Jan HP Eloff: A New Access Control Model based on the Chinese Wall Security Policy Model

- Marianne Loock, Jan HP Eloff, Johannes Heidema: CBAC: Conflict-Based Access Control

SUMMARY

Title: CBAC – A model for conflict-based access control

Candidate: Marianne Loock

Supervisor: Prof J.H.P. Eloff

Department: Department of Computer Science, Faculty of Engineering, Built Environment and Information Technology, University of Pretoria

Degree: Doctor of Philosophy in Computer Science

Keywords: Conflict-based Access Control, Access Control, Breach of Confidentiality, Data-mining environment

ACKNOWLEDGEMENTS

To my Lord and Saviour, Jesus Christ – in whom all things are possible. Thank You for the opportunity to do this work.

A special word of thanks to my family, Johan, Corlia and Werner. You are very special to me!

My sincere gratitude to my mother and late father, for their loyalty and dedication towards me. Their belief in me never faded and was clear in everything that they did.

Thank you to Izak, Anne-Mari, Ester and their families; as well as to Professors Alta van der Merwe, Elmarie Kritzinger and Mariki Eloff, who were interested not only in my work, but also in me as a person during this time.

I would like to thank Isabel Claassen for the professional and efficient manner in which she did the language editing of this thesis.

Finally, I gratefully acknowledge the momentous role of my study leader, Professor Jan Eloff. He provided inspiration, constructive criticism and experienced guidance – a mentor in the true sense of the word.

Opgedra aan:

Johan, die *liefde* in my lewe
Corlia en Werner, die *vreugde* in my lewe
Pa en Ma, die *bron van inspirasie* in my lewe.

Die meetsnoere het vir my in lieflike plekke geval, ja, my erfenis is vir my mooi.
Psalms 16:6, 1953-vertaling

TABLE OF CONTENTS

List of Figures

LIST OF TABLES

CHAPTER 1: INTRODUCTION - PROBLEM STATEMENT

1.1 Introduction

These days, large organisations are significantly affected by growing economic globalisation and the development in information technology, both of which result in the generation and storing of large amounts of data. With the correct tools, techniques and processes, data can be turned into information; information can be turned into knowledge and wisdom can be obtained from knowledge. What is needed to advance from one state to the next is firstly a reason and secondly, the ability to go to the next state.

The competitive business environment in which organisations find themselves provides a convincing reason for developing data into knowledge. Identifying those relations, correlations and patterns in business information and market prices that are not immediately obvious to managers due to the large volume of data will indisputably add to solving business problems in general.

In most cases the workforce has the ability to develop data into knowledge and eventually into wisdom – a skill that is often referred to as data mining in large organisations. This is a quality and ability that should be managed. According to the financial thriller writer, Linda Davies [1997], “[b]ankers who hire money hungry geniuses should not always express surprise and amazement when some of them turn around with brilliant, creative, and illegal means of making money”.

The banking sector is a typical case of the progression of data from information to knowledge. This sector also comprises organisations and customers varying between bodies with substantial debt to bodies with considerable wealth. The discussions to follow will refer to the banking section by means of an illustrative discussion.

1.2 PROBLEM STATEMENT

When taking a broad view on data mining, it is seen as the course of action involved in discovering appealing and exciting knowledge in large amounts of data stored in data bases, data warehouses or other information repositories [Han & Kamber, 2006]. Although data-mining technology can be seen as having many advantages, there are also some clear disadvantages that need to be addressed. A problem associated with data mining is the lack or loss of confidentiality associated with knowledge management, namely how to maintain security [Bertino, Khan, Sandhu, & Thuraisingham, 2006].

It is technologies such as data mining that make possible the nightmare of living in a world where all one's movements and purchases are tracked. Although data mining gives companies the information they need to market their products and services to customers, it also gives them access to a huge amount of personal information.

What stops these institutions from selling their data to others?

Even if legislation can stop institutions from selling data, what stops the individuals who work with this data from selling it to others once the data-mining process has changed the data to information?

Many consumers are completely unaware of data-mining technology and do not know that their lending habits, names, addresses and other information are being stored in a database. While data mining is a term that is well understood in some circles, it has not really entered the vocabulary of the general population.

Do the customers know that their information is put in a database?

Does the owner of the company tell customers that their information is put in a database specifically for data-mining purposes?

Customers should be given the right to choose whether or not they want to have their information placed in a database. Many clients may refuse because of the risk of a breach of confidentiality. Large companies that are fiercely competitive may avoid giving their customers an option because they don't want to lower their chances of having an edge on the competition. Because of this, they tend to refrain from informing customers that they have the option to allow or refuse their information to be placed in a database.

Customers again are cautious in giving formal permission to big and fiercely competitive companies to put their information in a database, and not without a reason! In the larger banking sector there are numerous examples of the illegal compilation of information on competing companies and the use of such information to the benefit of the compiler or to the detriment of the competing companies. Two familiar examples are discussed in the paragraphs that follow.

Jérôme Kerviel started working for the French bank Société Générale in 2000 [Gauthier-Villars, 2010; Treanor, 2011]. In 2008 he was accused of illegally accessing the bank’s computer system and of having conducted breach of trust and forgery between late 2006 and January 2008. It cost the bank $6,7 billion to unwind the unauthorised trades that had been made by Kerviel during this period. One of the reasons why this could happen is the fact that Kerviel had been allowed access rights to too much data, which gave him insight into ‘too much’ information. Because he could investigate, research and study trends in the data, he was in a position to manipulate the information at hand to his own benefit. As he moved money from one position to another for short periods of time, management was not able to pick it up. They did not expect the money to be lost within such a short timeframe. This specific knowledge helped him to disguise some of his actions and allowed him to work on specific actions that should not have been processed. In the aftermath to this scandal, Société Générale owned up to management failures and weaknesses in its risk control system and the French Banking Commission later fined Société Générale 4 million euros.

The Los Angeles Times [Popper, 2011] reported on a criminal trial in New York against Raj Rajaratnam, founder of the Galleon Group. Rajaratnam, who was arrested in October 2009, was found guilty of making more than $50 million in illicit profits by trading on inside information given to him by associates at some of America’s corporate standard bearers, including Google, Hilton and Goldman Sachs. The important question here is: How was this inside information generated? Is it possible that people who were given too many access rights to too much data were able to generate this information and make it available to Rajaratnam? Even more significantly: Was Rajaratnam the only receiver of this information? Although it is important to investigate and punish the actions of people like Raj Rajaratnam, one should also consider the accountability of those person/s who made the inside information available. These people are just as guilty and should not be overlooked when punishment is considered [Kandias, Mylonas, Virvilis, Theoharidou, & Gritzalis, 2010]. It is equally important for modern companies to make sure that their structures and corporate governance are such that this kind of inside information becomes hard to generate and owned by unidentified or non-traceable people.

This complex situation, where the possibility of a breach of confidentiality exists because the management of data and information (for example the results of a data-mining process) is not secure enough, can be addressed in the following two ways:

- Acknowledging the uncertainty of authorisation services (commonly known as access control) as was the problem experienced by the different companies in the examples referred to above
- Allowing companies to specify their conflicting companies and adding more detailed parameters to this authorisation services environment as to be safeguarded against the possibility of taking data or information from two or more conflicting companies and using it to the detriment of one or more company/ies

The secure management of data and information so as to minimise a breach of confidentiality forces one to concentrate on conflicting relationships. These conflicting relationships will therefore play an important role in the proposed model for conflict-based access control.

1.3 RESEARCH QUESTIONS

The current research recognises that authorisation services attempt to address conflict between companies/customers and that such services are a fundamental requirement for applications created for data-mining environments.

The aim of this research is thus to minimise a possible breach of confidentiality made probable by the hidden knowledge revealed through for example the data-mining process. Any person who leaks sensitive (positive or negative) hidden knowledge of one company to a conflicting company or to any other third party and, by doing this, adversely affects the first company, also puts the reputation of the company for whom he/she works at risk.

The aim of the study in hand will be accomplished by introducing a Conflict-Based Access Control (CBAC) model for a data-mining environment. The CBAC model introduces the 'spheres-of-conflict' concept that enables users to work on the data of different companies while avoiding confidentiality problems between the companies – thereby limiting potential breach-of-confidentiality problems. Furthermore, CBAC enables different companies whose data is used within the same data-mining environment to exercise some control over their level of exposure to other (or rival) companies.

The accomplishment of CBAC starts with proposing a model for access control that uniquely addresses specific requirements of minimising confidentiality breaching within data-mining environments. Existing mechanisms for implementing authorisation services do not address conflict between entities but rather provide a mapping between entities with regard to access rights. In other words, it is generally assumed that existing access control models map entities (subjects) onto other entities (objects) with regard to access rights, for example (subject_i, object_i, read). The handling of conflict between the entities cannot be solved by this triplet but it needs to be extended to include aspects such as the commercial relationship between these entities, for example if the entities are in a commercial conflict by trading in the same market segment. The triplet might therefore be extended to look like this: (subject_i, object_i, read, severity of conflict) to be able to accommodate the 'severity-of-conflict' commercial relationship.

The problem domain is addressed by considering the following research questions:

What are the security requirements for a data-mining environment?

The data-mining process needs to be examined and emphasis needs to be placed on each step of the process. Confidentiality issues during each step also need to be explained. The security requirements that should be present to ensure confidentiality during every step of the data-mining process also need to be studied.

What are the access control requirements for a data-mining environment?

After investigating security requirements in a data-mining environment, a study should be made of the access control requirements for data mining that should be present to ensure confidentiality during the steps of the data-mining process.

How can we introduce conflict-based concepts into an access control model for a data-mining environment?

Research has to be conducted into the possibility of minimising the risk of a breach of confidentiality in a data-mining environment by managing the access control in such an environment based on the conflicts that exist between the different subjects that form part of this environment.

The thesis statement for this research can thus be formulated as follows: It is possible to describe and explain a conflict-based access control model (CBAC) that will minimise the breach of confidentiality in a data-mining environment.

1.4 RESEARCH HIERARCHY AND METHODOLOGY

The current research is based on an extensive literature review about access control models in a data-mining environment and therefore the *research design* is non-empirical. The latter focuses on the intended end product, namely a conflict-based access control (CBAC) model. The *research methodology* focuses on the research process and the kind of tools and procedures that will be used [Mouton, 2001].

Firstly, the researcher conducted an extensive study of the literature on existing access control models and their ability to manage the data and information of conflicting entities. Secondly (and based on the knowledge gathered during the literature study), the new conflict-based access control model was created – a model that will help to prevent a breach of confidentiality between conflicting entities during data mining. Lastly, a proof-of-concept prototype was designed to demonstrate a subset of the access control model's ability to make access request decisions based on conflict.

The research in hand was aimed at developing a new model – an access control model – to explain a particular phenomenon, namely the managing of information of conflicting entities in such a way that a breach of confidentiality does not occur. This is called model-building studies [Mouton, 2001].

1.5 SCOPE AND CONTEXT OF THE STUDY

The proposed access control model is not responsible for the implementation of any read/write concepts for the purposes of this research. If considered to be applicable, it will form part of further research.

The proposed access control model has been formulated for a data-mining environment and has not been tested in any other environment. It was discussed, developed and tested (with the help of a prototype) in a commercial environment. The examples that are used all come from a banking perspective.

1.6 TERMINOLOGY USED IN THE THESIS

In order to avoid any misunderstanding, it is important to correctly interpret the terminology used in this thesis. The researcher now provides a brief definition of what is meant by the terms access control, data mining, conflict and model.

Access control is a fundamental element in computer security where every requested access must be governed by an access policy stating who is allowed access to what; i.e. the request must be mediated by an access policy enforcement agent [Pfleeger & Pfleeger, 2003].

Data mining (sometimes called data or knowledge discovery) is the process of analysing data from different perspectives and extracting hidden predictive and useful information from large databases. It is a process that consists of six steps [Daimler-Benz, ISL, & NCR]: the Business-Understanding Step; the Data-Understanding Step; the Data Preparation Step; the Data-Mining Step (also called the Modelling Step); the Evaluation Step; and the Deployment Step.

Conflict between different companies refers to the competition that exists between companies for any possible business reason for example companies that have to compete for the same market segment, such as petrol companies that compete for clients from the same area.

A *model* is the implementation of a policy. A security model implements the directions that are set by the security policy.

1.7 LAYOUT OF THESIS

This thesis consists of eight chapters. Figure 1.2 provides a graphical depiction of its layout. The current chapter, **Chapter 1**, provides an introduction to the research problem.

In **Chapter 2**, a case study is provided to form a background for further explanations regarding the proposed access model.

Chapter 3 provides information on access control in general. Special reference is made in this chapter to the different access control models for confidentiality, conflict and in a data-mining environment.

In the next chapter, **Chapter 4**, the data-mining process is discussed and security requirements for data mining are investigated.

The final list of requirements for the proposed access control model is established in **Chapter 5** and a conceptual framework for the proposed model is explained.

Chapter 6 is an in-depth explanation of the proposed conflict-based access control model – CBAC.

In the penultimate chapter, **Chapter 7**, the design of a proof-of-concept prototype is explained. This proof-of-concept prototype demonstrates a subset of the model.

Chapter 8 concludes the thesis.

THESIS LAYOUT

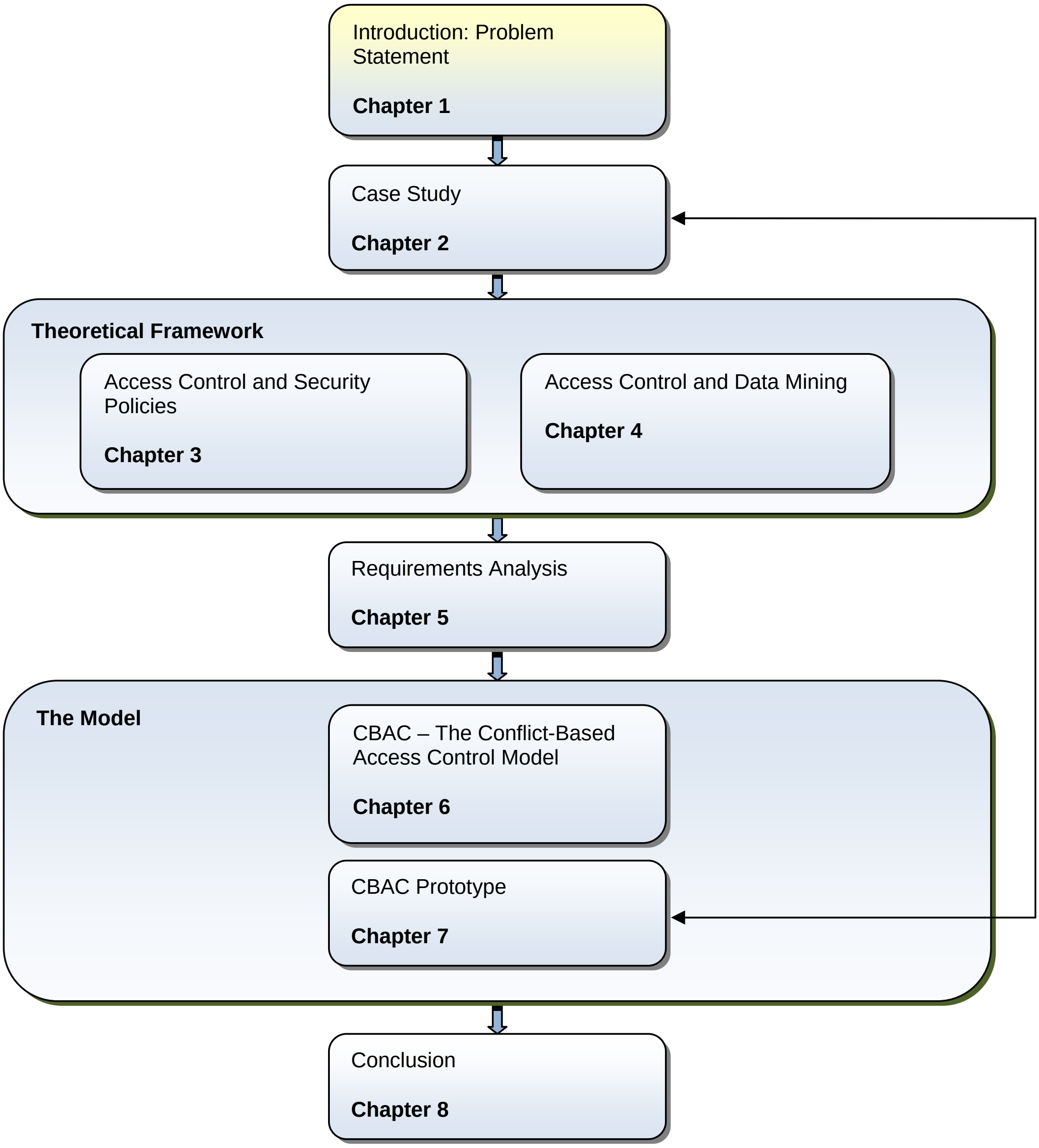
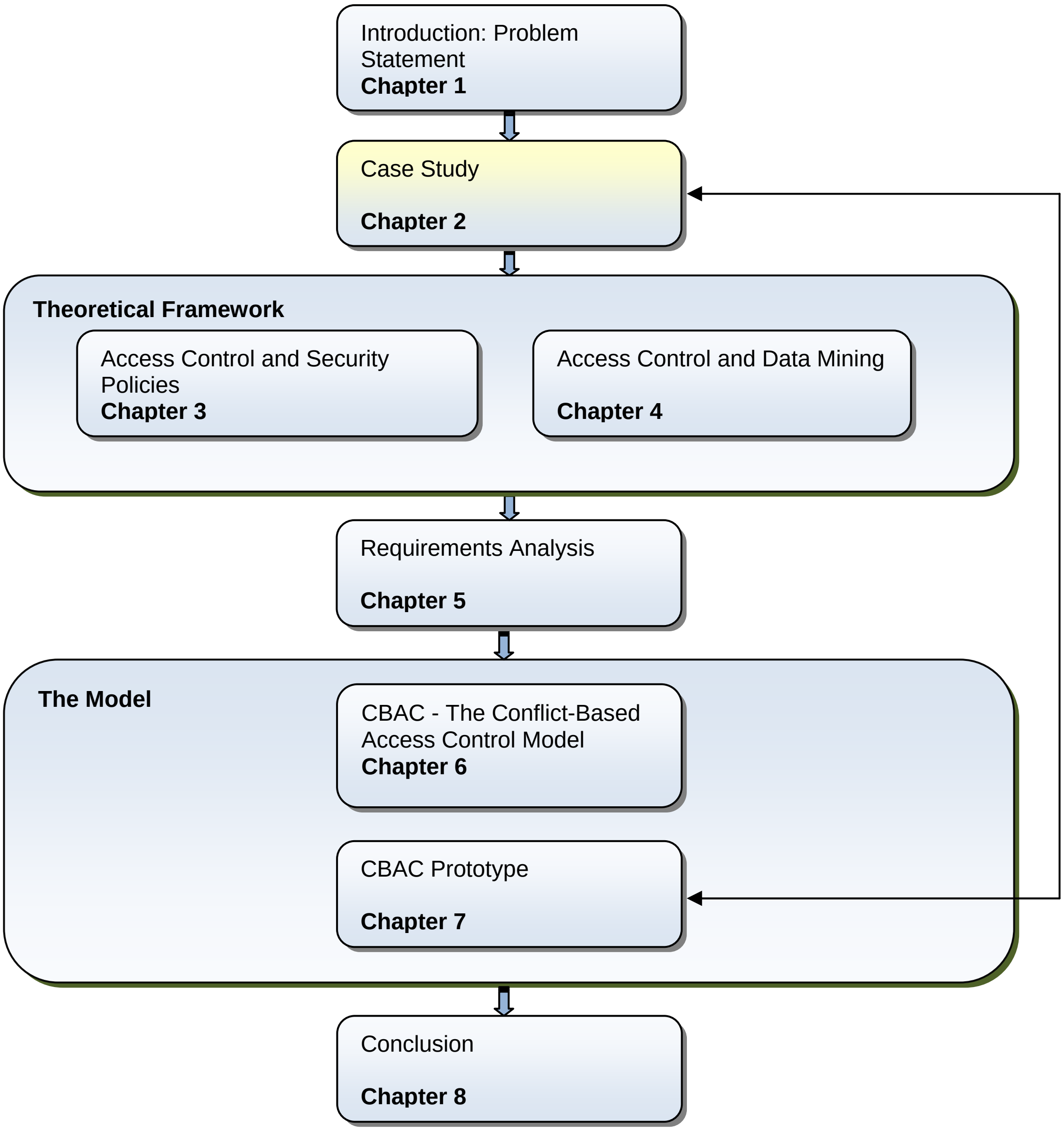


Figure 1.2: Graphical depiction of the thesis layout

CHAPTER 2: CASE STUDY - DM BANK

THESIS LAYOUT



2.1 INTRODUCTION

Data mining allows a mining agent to find information in data and this information can grow to knowledge in the course of the same data-mining activity. If not managed in a responsible manner, the knowledge gained from data can lead to a breach of confidentiality between the owner of the data and the owner of the data-mining activity. The focus of this chapter is on building and explaining the example of ABC Petrol company (a secondary company) and its position in the proposed access control (CBAC) model within a typical data-mining environment controlled by DM Bank (the global company). The discussion emphasises reasons for and the background to a model based on conflict. Chapter 2 concludes with a discussion of high-level conflict-based access control requirements for DM Bank, a data-mining environment, and difficulties faced when data miners are shared between conflicting companies.

2.2 CASE STUDY

The data-mining environment described here consists of a bank (DM Bank) and various clients (companies, private persons and all other legal entities that are usually clients of a bank). The one company that will be concentrated on for the purposes of this case study is the ABC Petrol company.

2.2.1 ABC Petrol’s environment

DM Bank conducts data mining for prediction and estimation purposes for its clients. Some of DM Bank’s clients have specific business relationships with one another and these relationships are described as follows:

ABC Petrol is in conflict with two other companies, Green Petrol and P+P Petrol, because the three of them belong to the same functional business domain, namely the petroleum companies’ functional business domain. Their conflicting relationships are depicted in Figure 2.1. Another important concept depicted in Figure 2.1 is the fact that the lines between ABC Petrol and P+P Petrol, as well as between ABC Petrol and Green Petrol are of equal length because the conflict between ABC Petrol and the other two companies is specified to be the same.

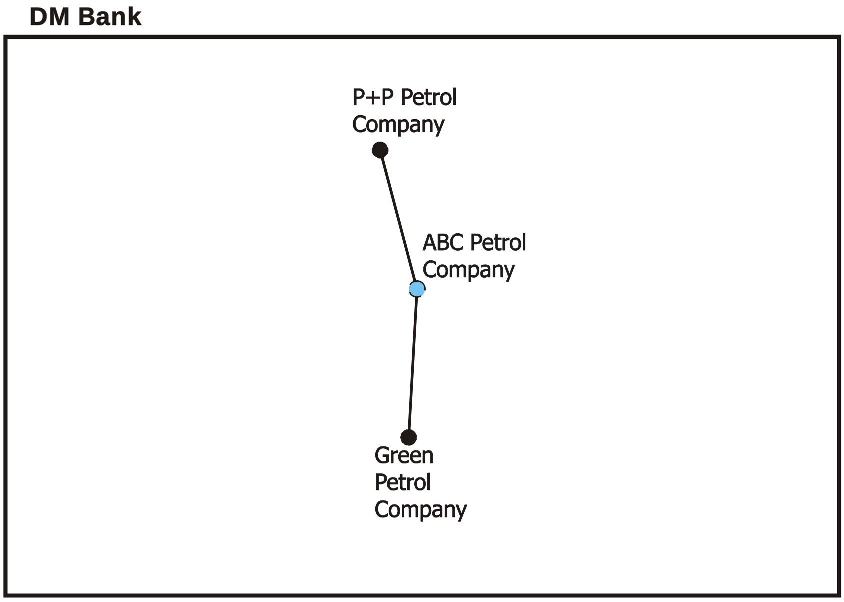


Figure 2.1: ABC Petrol is in conflict with Green Petrol and with P+P Petrol

Furthermore, Figure 2.2 shows that ABC Petrol is in conflict with HighFly Airline, because ABC Petrol holds shares in FlySave Airlines. For this reason, ABC Petrol initially specified HighFly Airline as a conflicting company. ABC Petrol also indicated Pick&Save Food as a conflicting company, because ABC Petrol holds shares in QuickPay Food. Lastly, ABC Petrol specified SelectShoe as a conflicting company for reasons unknown to DM Bank. These extended conflicting relationships are depicted in Figure 2.2. The length of the lines that represent the conflict between the companies also denotes the extent of the conflict between them. For example, the line between ABC Petrol and HighFly Airlines is longer than the line between ABC Petrol and Green Petrol, which implies that the conflict between ABC Petrol and Green Petrol is more intense (a shorter line) than the conflict between ABC Petrol and HighFly Airlines. Hence, companies that are in greater conflict with ABC Petrol are closer placed to ABC Petrol and the further a company is placed from ABC Petrol, the less intense is the conflict between that company and ABC Petrol.

The environment as seen by ABC Petrol comprises five conflicting companies at this stage. The situation as perceived by DM Bank will be explained next.

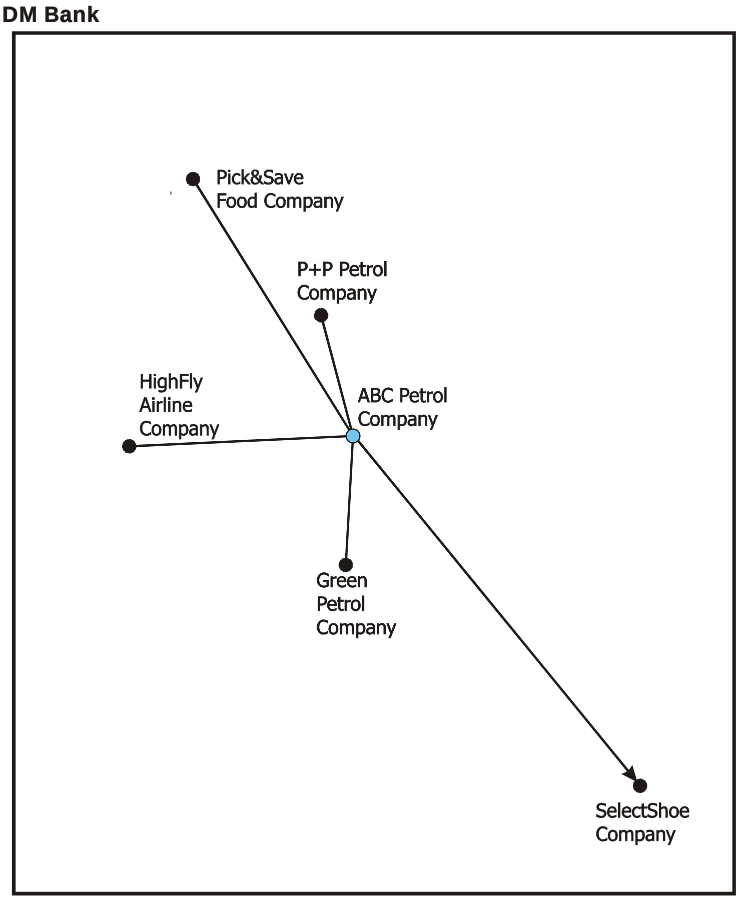


Figure 2.2: ABC Petrol in conflict with Pick&Save Food, HighFly Airline, Green Petrol and SelectShoe (according to ABC Petrol's own indications)

2.2.2 DM Bank's environment

DM Bank is a financial institution. It has clients on whose data DM Bank performs data mining so as to improve its services to these clients by suggesting better individualised and group financial and banking options for them. For purposes of this case study, all of DM Bank's clients will be grouped into one or more functional business domains, depending on each client's main function. ABC Petrol belongs to the *Petroleum* functional business domain because of its interest in petrol, but it also belongs to the *Airline* functional business domain because of its interest in the FlySave Airline company. By the same token it also belongs to the *Food* functional business domain because of its interest in the QuickPay Food company.

Functional business domains such as *General* and *Anonymous* are also present in the DM Bank environment. Table 2.1 illustrates all the functional business domains that exist within DM Bank.

Table 2.1: Functional business domains mapped onto clients of DM Bank

Functional Business Domains	Food	Petroleum	Airline	Shoes	General	Anonymous
Clients of DM Bank	EasyEat	ABC Petrol	HighFly	SelectShoe	A Smith	Client 194
	Pick&Save	Green Petrol	FlySave		BG Harper	Client 382
	QuickPay	P+P Petrol	BlueSky		R Baber	
		Fast Petrol				

The environment from DM Bank's perspective is somewhat more complex than the one perceived by ABC Petrol. DM Bank sees a bigger picture and has more information regarding other clients that ABC Petrol is not even aware of. This extra information is important in a data-mining context as one should concentrate on the conflicts that exist between the different companies when deciding on which data miners should be allocated to which company. The environment as it is known to DM Bank is depicted in Figure 2.3.

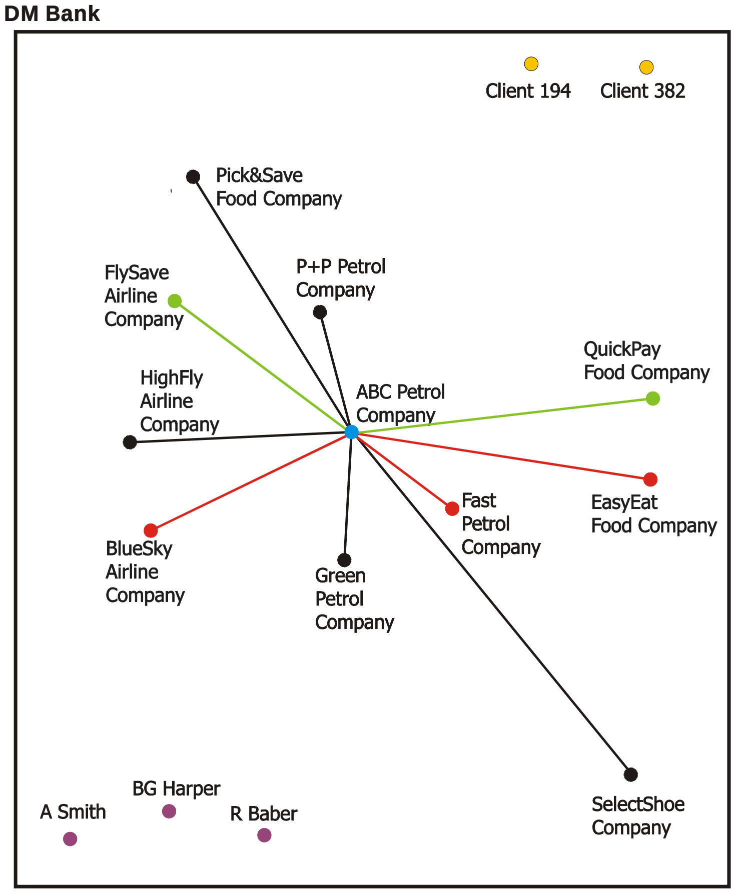


Figure 2.3: The client environment as seen by DM Bank

In Figure 2.3 the already known conflicting relationships are represented by black lines. Where ABC Petrol initially indicated an interest in another company, these relationships are indicated by green lines. All other relationships of ABC Petrol within the functional business domains are indicated by red lines. These red lines imply a conflicting relationship with ABC Petrol that ABC Petrol was not aware of, but that DM Bank could point out. The clients in the other two functional business domains (*General* and *Anonymous)* are also shown but no relationship between them was implied by ABC Petrol or could be detected by DM Bank.

2.3 Access control and the data-mining environment of DM Bank

The following scenario is now possible when performing data mining without considering access control between data miners and the clients of DM Bank.

While conducting a data-mining exercise, DM Bank’s mining agents detect a negative growth in Green Petrol’s cash flow. If one or more of the mining agents that work on the data sets of Green Petrol also work on those of ABC Petrol, Green Petrol’s position may well be jeopardised. Green Petrol could be exposed to the risk of information leakage directly to a conflicting company, namely ABC Petrol. Should information about Green Petrol’s sensitive cash flow – in comparison to ABC Petrol’s better cash flow situation – be leaked, it could easily become known to people who can use this information to their benefit and to the detriment of Green Petrol.

The above example emphasises some important data-mining access control requirements to be considered by DM Bank to minimise any possible reputation risks that may follow because of information leakages coming from data miners that work on conflicting companies’ data sets.

One could, for example, ask the following questions and resolve them with answers from the suggested high-level conflict-based access control policy.

- May the same mining agent who has access to the data sets of ABC Petrol obtain access to those of Green Petrol or, by the same token, to any other company in ABC Petrol’s main functional business domain, namely the Petroleum functional business domain? If one considers the fact that all petroleum companies mainly do business in the petroleum functional business domain, all petroleum companies should be considered to be in conflict.

  For confidentiality purposes, the answer to this question should be: No, the same mining agent who has access to the data sets of ABC Petrol should not be given access to those of Green Petrol, or to any other company in the Petroleum functional business domain.

- May the mining agent who has access to the data sets of Green Petrol obtain access to those of QuickPay Food? The answer to this question lies in the 'derived relationship' between these two companies, which looks as follows: ABC Petrol is in conflict with Green Petrol (as indicated by ABC Petrol) and ABC Petrol has a specified interest in QuickPay Food. This makes QuickPay Food and Green Petrol conflicting companies.

  For confidentiality purposes the answer to this question should be: No, the same mining agent who has access to the data sets of Green Petrol may not obtain access to those of QuickPay Food, because information from ABC Petrol can flow via QuickPay Food to Green Petrol.

- May the mining agent who has access to the data sets of ABC Petrol obtain access to those of BlueSky Airline? Again a derived relationship must be investigated. ABC Petrol indicated an interest in FlySave Airline and because of that interest, ABC Petrol is in conflict with HighFly Airline. By the same token, ABC Petrol is in conflict with BlueSky Airline.

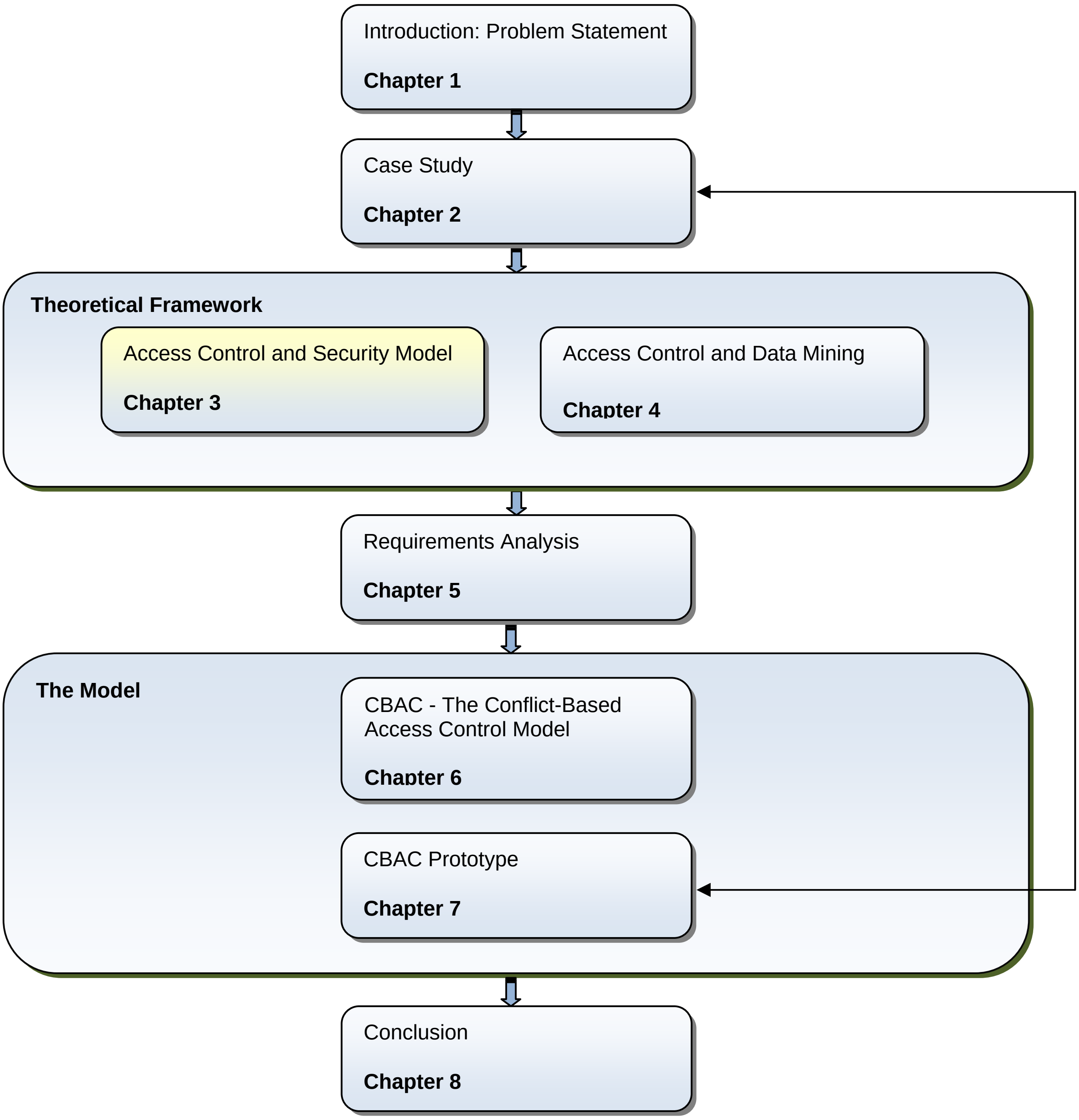
  For confidentiality purposes the answer to this question should be: No, the same mining agent who has access to the data sets of ABC Petrol may not obtain further access to those of BlueSky Airline because ABC Petrol is in conflict with BlueSky Airline, although ABC Petrol was not aware of this situation, DM Bank could deduct this relationship.

When the given case study is analysed and some questions regarding confidentiality are considered critically, it has to be deduced that confidentiality is not guaranteed during data-mining activities. It is also acceptable for the clients of DM Bank to expect from DM Bank to regulate the access control of mining agents in such a manner that confidentiality is preserved during data-mining activities wherever required.

In the remainder of this thesis, DM Bank is used as an example of a global agent that provides a service to secondary agents, namely the companies who are clients of DM Bank. The service to secondary agents entails typical financial services that banks deliver to their clients. Some of the global agent's secondary agents have specific business relationships with one another and the main relationships were described in this chapter.

CHAPTER 3: EMERGING THEORIES AND PERSPECTIVES RELEVANT TO ACCESS CONTROL AND SECURITY POLICIES

THESIS LAYOUT



3.1 INTRODUCTION

Information security can be divided into the following three high-level categories that form the traditional CIA in information security:

- *Confidentiality* – when only authorised parties can access information-related assets. Confidentiality refers to secrecy and privacy.

- *Integrity* – when only authorised parties can modify assets and only in authorised ways. Modification includes writing, changing, changing status, deleting and creating.

- *Availability* – when assets are accessible to authorised parties at appropriate times. The opposite of availability is denial-of-service.

Another important category that was not part of the traditional CIA is *nonrepudiation* – this category guarantees that a message or data can be proved to have originated from a specific user, which means once an action has taken place, a user cannot claim that he did not perform it.

These information security categories form part of managing and ensuring the protection of information. An elementary element in enforcing security is represented by the access control service that has the task to control all access to a system and its resources and to ensure that only authorised accesses can take place. According to Samarati and De Capitani di Vimercati [2001] the development of an access control system is based on security policies and models.

Chapter 3 briefly discusses access control terminology in general and then outlines security policies in more detail. Models related to these policies are then discussed and put in perspective as they relate to this research.

3.2 ACCESS CONTROL TERMINOLOGY

The word ‘access control’ suggests an active *subject* (or a *principal)* accesses a passive *object* with a specific *access operation*, while a *reference monitor* grants or denies access. Figure 2.1 captures this view of access control.

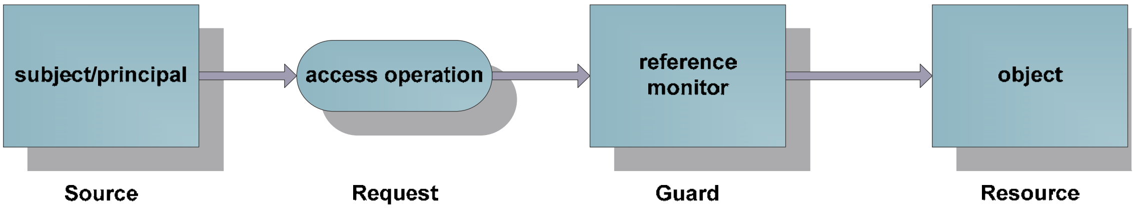


Figure 3.1: The Fundamental Model of Access Control [Gollmann, 2006]

Subject and *principal* are used in the security literature to refer to the entity making the access request. This research will use the following definition:

> Subjects operate on behalf of human users we call principals, and access is based on the principal's name bound to the subject in some unforgeable manner at authentication time. Because access control structures identify principals, it is important that principal names be globally unique, human-readable and memorable, easily and reliably associated with known people. [Gasser, 1990]

Thus, when discussing security policies, this research will refer to *principals*, and when discussing operational systems that should enforce a security policy, this research will refer to *subjects*.

3.3 SECURITY POLICIES AND ACCESS CONTROL POLICIES

A *security policy* defines the security aims and goals of a system. A *policy* is a document that contains a contract of insurance ["Collins Concise Dictionary," 2004]. Habib [2009] refers to a security policy as a notion of secure states. Thus, a security policy is a contract that ensures the security that one expects the system to enforce. According to Samarati and De Capitani di Vimercati [2001] a security policy defines the high-level rules according to which access control must be governed. Chaula, Yngström and Kowalski [2005] also states that a good security policy explicitly states how breaches are to be detected.

Management administers the construction of a security policy for either a military environment or a commercial environment. Security policies for military environments focus primarily on confidentiality, while those for commercial environments focus primarily on integrity.

A security policy may use different types of *access control policies*, either alone or in combination. Access control policies describe the rules for regulating the interactions between subjects and objects, and can be grouped into the following three main classes [De Capitani di Vimercati, Paraboschi, & Samarati, 2006; Samarati & De Capitani di Vimercati, 2001]:

Discretionary (DAC) (authorisation-based) access control policies control all access, based on the identity of the principal and on access rules (authorisations) stating what principals are (or are not) allowed to do [De Capitani di Vimercati, et al., 2006]. This means the access control is left to the discretion of the owner.

Mandatory (MAC) (rule-based) access control policies control all access, based on mandated regulations as determined by a central authority [De Capitani di Vimercati, et al., 2006].

Role-based (RBAC) access control policies control access, depending on the roles that users have within the system and rules stating what accesses are allowed to users in given roles [De Capitani di Vimercati, et al., 2006; Ferraiolo & Kuhn, 1992; Sandhu, Coyne, Feinstein, & Youman, 1996].

The implementation of a security policy adopted by an organisation/enterprise is based on the particular access control model/s in use. These models can be informal, which means they may lack mathematical or logical strictness (although an informal model can be rigorous and well specified); or they may be formal, which means they have mathematical or logical rigidity; or these models may be a mixture of the two. Different access control models will be addressed next and be clarified for purposes of this research.

3.4 Access Control Models

Bishop [2003] states that a *security model* represents a particular security policy and Habib [2009] refers to a model as the implementation of the monitor. It was earlier mentioned that security policies can comprise either of a military security policy – focusing on confidentiality – or a commercial security policy – focusing on integrity.

The question that formulates the problem statement for this research is as follows:

Is it possible to minimise the risk of a breach of confidentiality in a data-mining environment by managing the access control in such an environment based on the conflicts that exist between the different subjects that form part of this environment?

To be able to investigate this problem, the access control models that are important for further research (according to the problem statement), are models that concentrate on *confidentiality,* models that are able to handle *conflict,* and models for a *data-mining environment*. The rest of this paragraph will therefore describe access control models as grouped in the following three categories:

- Access control models that concentrate on confidentiality
- Access control models that can handle types of conflict
- Access control models for a data-mining environment

3.4.1 Access Control Models for Confidentiality

When referring to access control models for confidentiality, the Bell-La Padula Confidentiality Model is always an important model. However, at least three other models also need mentioning, namely those of Lampson, Graham and Denning, as well as Harrison, Ruzzo and Ullman.

3.4.1.1 Bell-La Padula Model (BLP)

The Bell-La Padula model [Bell & LaPadula, 1973, 1976] is a state machine model that focuses on data confidentiality and controls access to classified information. It is a model that is typically used together with a military security policy. In this formal model, the entities in an information system are divided into subjects and objects. In determining the allowance of a specific access mode, a comparison is made between the clearance of a subject and the classification of the object (to be more exact, it is compared to the combination of classification and set of compartments that constitute the *security level*). A lattice expresses such a clearance, or classification scheme. [De Capitani di Vimercati, Foresti, & Samarati, 2007] refer to the security policy that is represented by the Bell-La Padula model as a Secrecy-Based Mandatory Policy. The BLP functions according to two mandatory access control (MAC) rules and one discretionary access control (DAC) rule that must comply with the following three security properties:

In the Bell-La Padula model, information from a high-sensitivity document may be transferred to a lower-sensitivity document via the concept of trusted subjects. Although trusted subjects are not restricted by the ★–property, untrusted subjects are. The BLP security model focuses on access control and is characterised by the phrase: "*no read up, no write down*". One of its main limitations involves the fact that it only addresses confidentiality.

The ability of the Bell-La Padula to handle *conflict* has not been documented. Literature also does not have examples of the Bell-La Padula model being used in a data-mining environment. Due to these limitations the Bell-La Padula model is not considered an appropriate model for addressing the problem statement as formulated in Par. 3.4.

3.4.1.2 Lampson model; Graham-Denning model; Harrison-Ruzzo-Ullman model

Butler W. Lampson [1971] was the first person to formulate an access control model for confidentiality – i.e. the Lampson model. It was developed with 'protection' as a starting point where 'protection' meant to keep the maliciousness or error of one user from harming other users. The Lampson model was later refined by Graham and Denning [1972] who then proposed their Graham-Denning model. The latter has the structure of a state machine where each state is a triple (S; O; M)

where S is a set of subjects,

O is a set of objects (which has S has a subset), and

M is an access matrix that has

one row for each subject,

one column for each object,

and is such that cell M [s; o] contains the access rights that subject s has for object o. These access rights are taken from a finite set of access rights, A. States are changed by requests for altering the access matrix, M.

An individual machine in the model is called a system.

In 1976 a variation on the Graham-Denning model was proposed, called the Harrison-Ruzzo-Ullman model [Harrison, Ruzzo, & Ullman, 1976]. This model is based on commands, where each command involves conditions and primitive operations. Even though this upgraded model was available, the model that was used for confidentiality in most cases remained the Bell-La Padula model [Marciniak, 2001].

3.4.2 Access Control Models for Conflict

The proposed model for access control is specified for a data-mining environment within a commercial environment. The latter is particularly vulnerable for conflict, especially of the sort where companies are in competition with one another for the same market segment. The Chinese Wall Security Policy model has been used with great success in the commercial environment.

3.4.2.1 Chinese Wall Security Policy Model

The Chinese Wall Security Policy (CWSP) model, designed by Brewer and Nash [1989], addresses access control requirements in a commercial environment. The CWSP model can however also address access control requirements in a more military environment, because Bishop [2003] argues that the CWSP model is a type of security policy – called hybrid policies – that refers equally to confidentiality and integrity. These requirements definitely reflect commercial environments and potential military environments where a possible conflict of interest can arise. The concept of a conflict of interest is applicable in a data-mining environment where the same mining agent is exposed to data of different companies.

The CWSP model makes use of actors and objects, and defines the access right to data, on a specific side of a ‘Chinese Wall’ to which the actor already has access. It develops a set of rules aimed at preventing people (actors) from accessing data (objects) on the opposite side of the 'Chinese Wall'. All corporate information is stored in a hierarchy as shown in Figure 2.2.

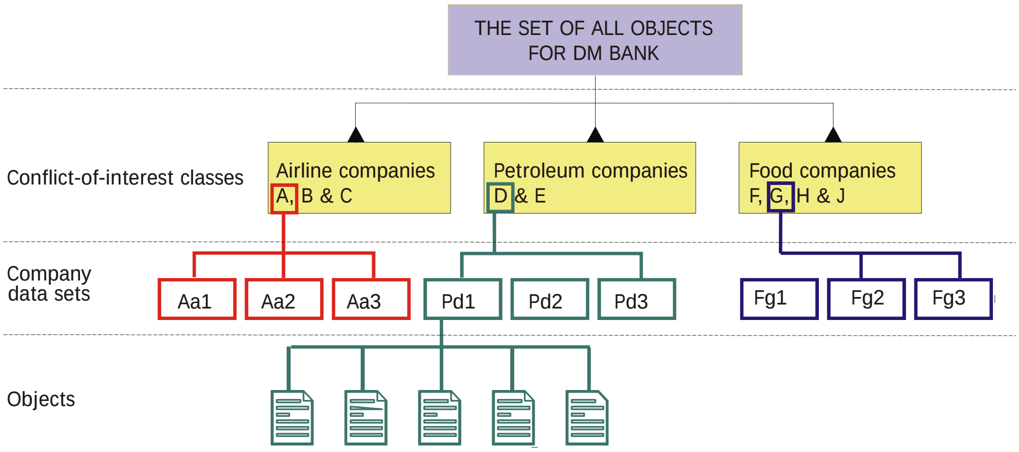

THE SET OF ALL OBJECTS FOR DM BANK
Conflict-of-interest classes
Airline companies A, B & C
Petroleum companies D & E
Food companies F, G, H & J
Company data sets
Aa1
Aa2
Aa3
Pd1
Pd2
Pd3
Fg1
Fg2
Fg3
Objects

Figure 3.2: Chinese Wall Security Policy Model

There are three levels of significance in Brewer and Nash's CWSP model:

- At the lowest level, objects are considered. These objects contain specific data items, each concerning a specific company.
- At the intermediate level, all objects concerning the same company are grouped together. These are referred to as company data sets.
- At the highest level, all data sets of companies that are in conflict are grouped together into classes. These are referred to as conflict-of-interest classes.

The foundation of the Chinese Wall Security Policy model is that actors are allowed access only to information that is not in conflict with any other information to which they already have access. The CWSP model builds a collection of impassable walls, known as Chinese walls, around the data sets of conflicting companies. No data sets that are in conflict can be stored on the same side of the Chinese walls.

The concept of Chinese walls is altered and integrated into the CBAC model as proposed in this thesis.

3.4.2.2 Aggressive Chinese Wall Security Policy Model

According to Lin [1989; 2003], the Brewer-Nash model was based on the incorrect assumption that corporate data can be grouped into separate and 'disjoint' conflict-of-interest classes (CIR classes [Tsau Young Lin, 2003]). In his arguments, Lin explains that CIR classes are seldom disjoint; in fact, they generally overlap. For example, it is an incorrect assumption that all airline companies could be grouped into one, and only one, conflict-of-interest class. The truth is that the contents in CIR classes often intersect and overlap.

Lin subsequently suggested a modified model called the Aggressive Chinese Wall Security Policy model (ACWSP model). This model is based on the development of a methodology known as Granular Analysis and Computing [Tsau Young Lin, 2003] and it allows CIR classes to overlap.

The CBAC model that is introduced in this study considers CIR classes that are not mutually disjoint as an important access requirement for data-mining environments. The ability of the CWSP and the ACWSP to handle *conflict* in classes or in 'not mutually disjoint' classes is proved, but this conflict cannot be handled in ad hoc defined spheres as in the CBAC model.

Since the available literature does not have examples of either the CWSP or the ACWSP models being used in a data-mining environment, neither of these models are appropriate for studying the already mentioned problem statement.

3.4.3 Access Control Models for a Data-Mining Environment

Current security approaches in the data-mining environment concentrate on privacy and not on confidentiality. The following examples serve to prove this statement.

3.4.3.1 Privacy-Preserving Access Control

Privacy-preserving access control means the service provider can neither learn what access rights a customer has, nor link a request for accessing an item to a particular customer for example with electronic voting [Katsikas, Gritzalis, & Balopoulos, 2005]. As a result, the privacy of both customer activity and customer access rights is protected.

When using the CWSP model, data sets and databases may be sanitised by removing private information. In some cases, however, data-mining techniques allow the recovery of the removed information. Privacy preservation is thus an important technique to be used when sensitive data is to be released to a third party, for example a mining agent within a data-mining environment [Byun, Bertino, & Li, 2005]. Although privacy preservation emphasises privacy as opposed to access control, it still needs to be considered within the overall context of addressing access control requirements for data-mining environments. For this reason, privacy-preserving access control will be discussed again in Chapter 4.

3.4.3.2 Privacy-Aware Access Control

Privacy-aware access control combines the concepts of privacy and access control. Current research into privacy-aware access control focuses on two issues. Firstly, it concentrates on the definition and development of access control and privacy languages, such as XACML (eXtensible Access Control Markup Language) [OASIS, 2010]. Secondly, current research in this field also focuses on the definition of architectures to protect and preserve the privacy of either services or clients [Ardagna, Cremonini, Damiani, De Capitani di Vimercati, & Samarati, 2006]. Privacy-aware access control is important for the development of access control models for data-mining environments in general. Privacy-aware access control indeed facilitates controlled access to individual records, which is not the case with privacy-preserving approaches. The privacy-aware approach also protects privacy in the networked society because of the growing amount of personal information being collected by numerous commercial and public services [Ardagna, Cremonini, De Capitani di Vimercati, & Samarati, 2008]. Privacy-aware access control will be discussed again in Chapter 4 because of the exciting effect it has on the definition of architectures to protect and preserve the privacy of both services and clients in a data-mining environment.

3.4.3.3 Privacy-Enhanced Access Control

Research on privacy-enhanced access control [Claudio et al., 2010] focuses on the definition of a privacy-enhanced access control system that includes different models and languages. The infrastructure is then aimed not only at regulating access to resources but also at protecting the privacy of the users. The use of an access control approach that concentrates on privacy rather than on confidentiality (whether used in a data-mining environment or not) is not applicable to the research question of this study.

Although the current research acknowledges the importance of access control mechanisms associated with access control models when securing a computer system or network of any size, access control mechanisms do not form part of this study and will not be discussed here.

3.5 ACCESS CONTROL MODEL SUMMARY

It seems that the models that were discussed so far can be grouped into three main categories, as depicted in Figure 3.3:

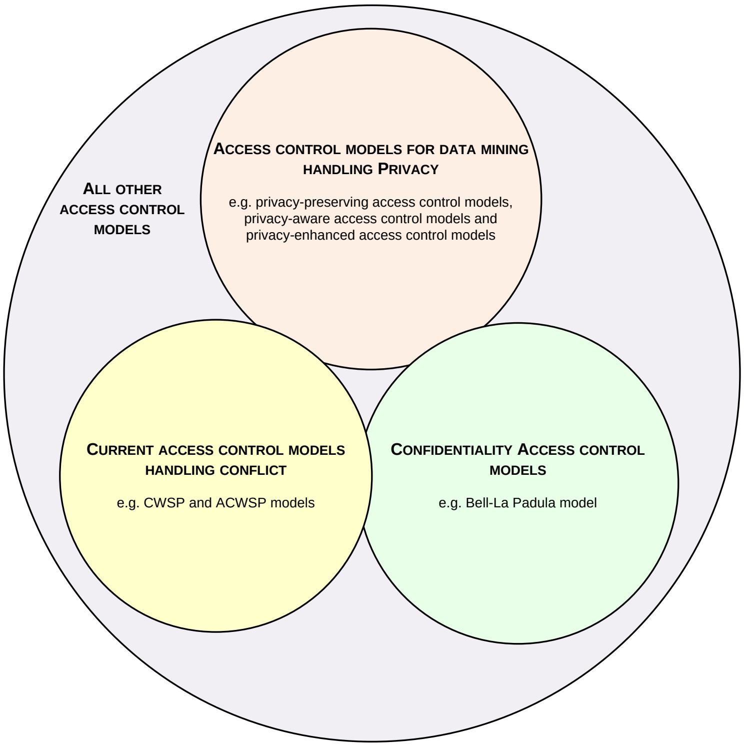

ALL OTHER ACCESS CONTROL MODELS
ACCESS CONTROL MODELS FOR DATA MINING HANDLING PRIVACY
e.g. privacy-preserving access control models, privacy-aware access control models and privacy-enhanced access control models
CURRENT ACCESS CONTROL MODELS HANDLING CONFLICT
e.g. CWSP and ACWSP models
CONFIDENTIALITY ACCESS CONTROL MODELS
e.g. Bell-La Padula model

Figure 3.3: Three main categories for the discussed access control models

- An example of a model for *confidentiality* is the Bell-La Padula model. Unfortunately, this model addresses only confidentiality and the ability of the Bell-La Padula to handle *conflict* has not been documented.

- The models that were found to be able to handle *conflict* to a certain extent are the CWSP and ACWSP models. Both can handle conflict in classes or in 'not-mutually-disjoint' classes, but neither can deal with conflict in ad hoc defined spheres. The CWSP model with its strict conflict-of-interest classes is not fitting for a *data-mining environment*.

- Security approaches in a *data-mining environment* concentrate on privacy rather than on *confidentiality*. In this environment, the following examples of access control are present: privacy-preserving, privacy-aware and privacy-enhanced access control.

3.6 CONCLUSION

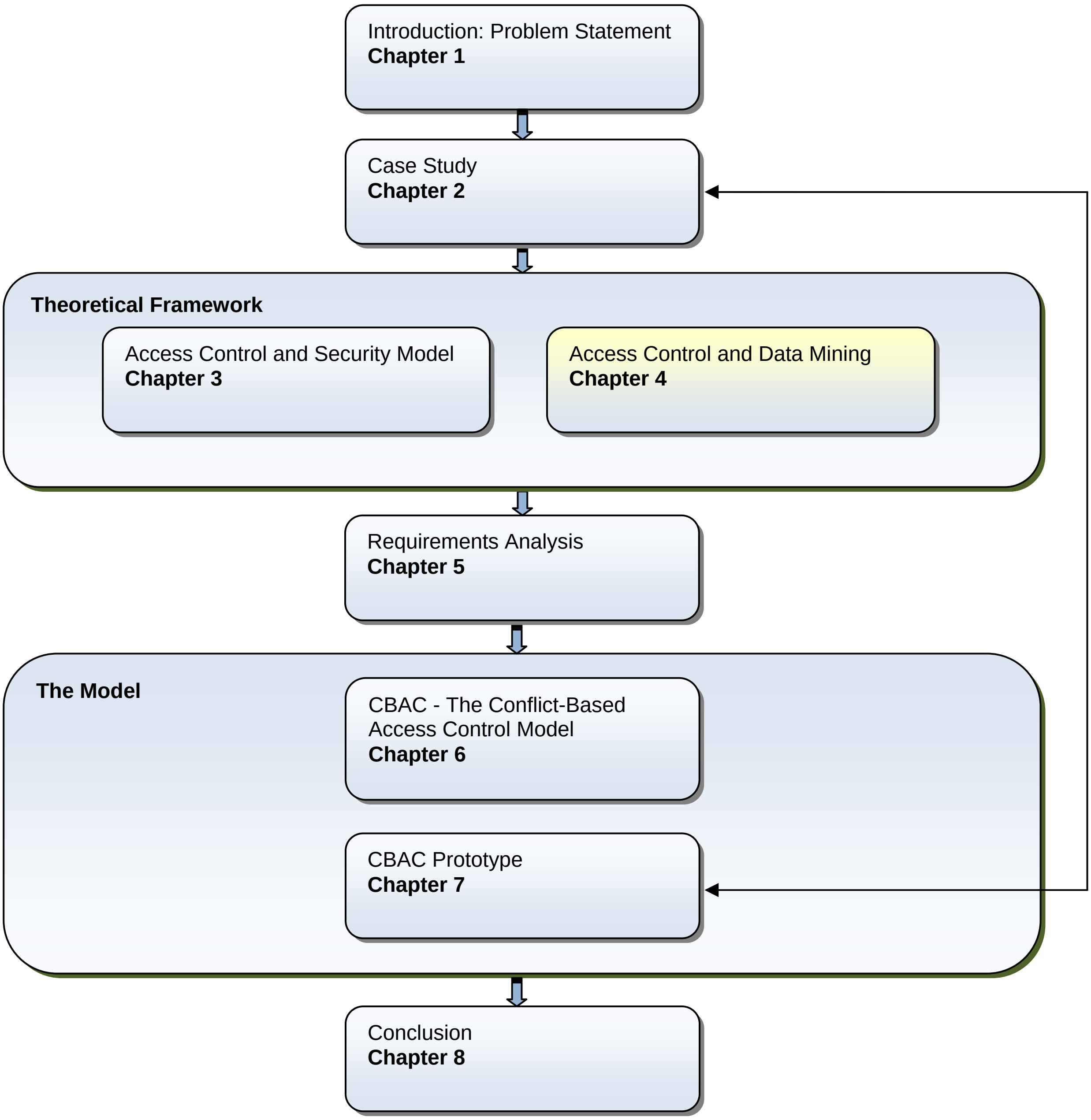
In this chapter, a number of security access control policies, security policies and models were identified. The different models were grouped into three different categories namely models for confidentiality, models that can handle conflict and models for the data-mining environment.

From the literature studied, the researcher was not able to confirm the existence of a single model that handles conflict as well as confidentiality in a data-mining environment.

In the next chapter, the focus will be on the data-mining environment and aspects that are necessary when concentrating on access control within a data-mining environment.

CHAPTER 4: EMERGING THEORIES AND PERSPECTIVES RELEVANT TO ACCESS CONTROL AND DATA MINING

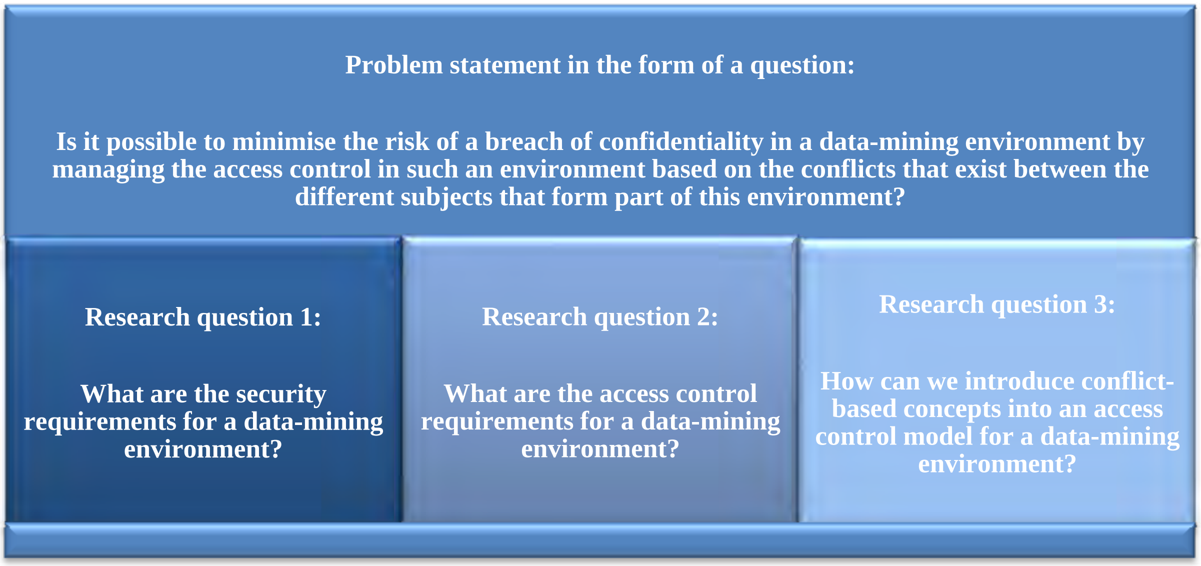
THESIS LAYOUT



4.1 INTRODUCTION

Data mining involves the searching for meaning within enormous masses of data. A formal definition of the concept is formulated as follows: Data mining refers to extracting or ‘mining’ knowledge from large amounts of data [Han & Kamber, 2006]. The information and knowledge acquired from the data-mining process can be used for applications ranging from stock market analysis, fraud detection and science exploration, to medical research and client profile analysis.

The large amounts of data are stored in many different kinds of information repositories and databases. The data warehouse is such a data repository architecture that has emerged in the past four to five decades [Han & Kamber, 2006]. Data warehouse technology includes data cleaning, data integration and on-line analytical processing (OLAP). OLAP is an analysis technique with functionalities such as summarisation, aggregation and consolidation, as well as the ability to view information from different perspectives. Data mining goes beyond the scope of the summarisation- and consolidation-style analytical processing of data by incorporating more superior techniques for data analysis [Han, Kamber, & Pei, 2012].



Problem statement in the form of a question:

Is it possible to minimise the risk of a breach of confidentiality in a data-mining environment by managing the access control in such an environment based on the conflicts that exist between the different subjects that form part of this environment?

Research question 1:

What are the security requirements for a data-mining environment?

Research question 2:

What are the access control requirements for a data-mining environment?

Research question 3:

How can we introduce conflict-based concepts into an access control model for a data-mining environment?

In this chapter, the focus of the study is on answering the first two research questions namely: **What are the security requirements for a data-mining environment?** and **What are the access control requirements for a data-mining environment?**

First of all the data-mining process will be discussed in broad terms, followed by a description of the data-mining environment. The security requirements for data mining will be considered next, and lastly the access control requirements when performing data mining will be stated.

For the sake of transparency and clarity, some terminology needs clarification before we can discuss data mining in general.

4.2 The Data-Mining Process

Data mining can be considered as a synonym for Knowledge Discovery from Data (KDD). Alternatively, data mining can be viewed as an fundamental step in the process of knowledge discovery [Cabena, Hadjinian, Stadler, Verhees, & Zanasi, 1998; Han & Kamber, 2006]. Data mining is the analysis step in the knowledge discovery process, but to adapt to industry, media and the database research environment, this research will use the term 'data-mining process' for the complete knowledge discovery process as discussed in the rest of this section.

In 1996 an initiative was launched by Daimler Chrysler (then Daimler-Benz), SPSS Inc. (Statistical Package for the Social Sciences, then ISL (Integral Solutions Ltd.)) and NCR (National Computer Registry Corporation) to standardise the data-mining process and develop a framework for the data mining tasks. This Cross-Industry Standard Process for Data Mining [Daimler-Benz, et al.] was abbreviated as CRISP-DM, and developed and refined through a series of workshops from 1997 to 1999. Whereas CRISP-DM refers to the agent doing the tasks within this process as an analyst or data analyst, the research in hand refers to this person as a mining agent. CRISP-DM 1.0 was published in 1999 and consists of an iterative sequence of steps. This iterative sequence of steps was adopted in the current research and is depicted in Figure 4.1 as a data-mining process for this research:

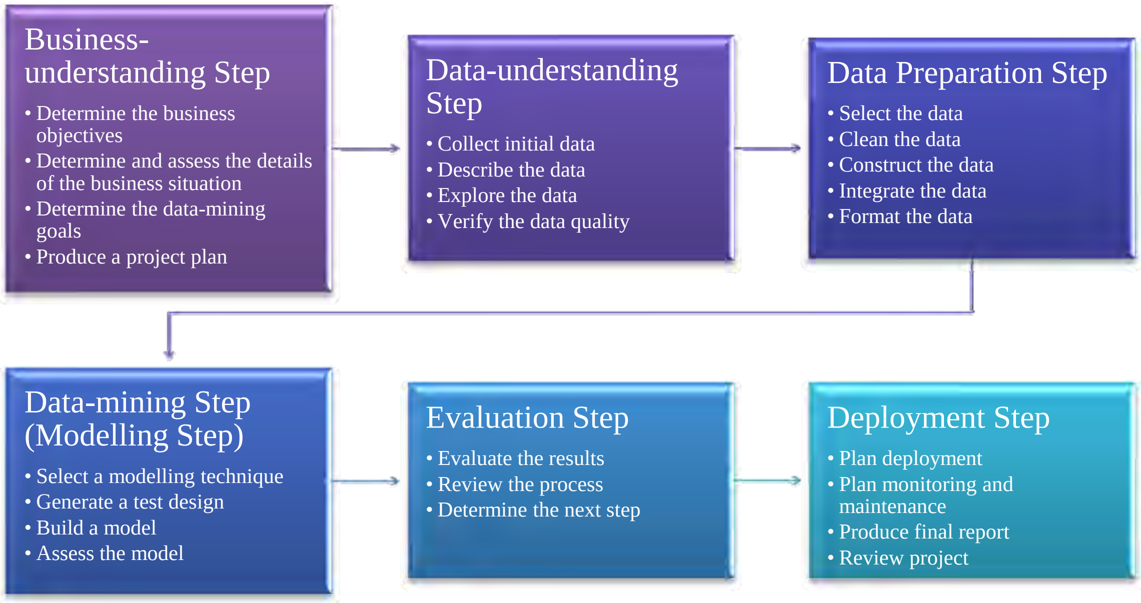


Figure 4.1: The Process of Data Mining (adapted)

The different steps in the data-mining process will now be discussed in the same order as they have been depicted in Figure 4.1.

4.2.1 Business-understanding Step [Daimler-Benz, et al.]

Understanding the project objectives and project requirements as seen from a business perspective constitutes the focus of this step. This knowledge and insight are transformed into a data-mining problem definition and a preliminary plan projected to achieve the objectives and requirements. Business knowledge is the source of the questions that should be asked and can be answered by means of data mining, for example: ‘What is wrong with the *home loan* accounts in the bank?’ The first stage in the business-understanding step is to clearly determine the business objectives concerned.

4.2.1.1 Determine the business objectives

The first objective of the mining agent (data analyst) is to realise and understand from a business perspective what the client wants to achieve. The client might have many competing constraints and objectives and these must be properly balanced. The initial goal of the mining agent (analyst) is to uncover important factors that may sway or control the result of the project. A potential result of neglecting this task in the business-understanding step is to spend a great deal of effort producing the right answers to the wrong questions. For example, the principle business goal is to maintain current customers by forecasting when they may be prone to move to an opponent. An example of a related business question is: ‘Will more easily accessible *home loan* accounts considerably lessen the number of high-value customers who leave?’

4.2.1.2 Determine and assess the details of the business situation

This task involves more thorough fact finding about all of the resources, constraints, assumptions and other factors that should be considered in determining the data analysis goal and project plan. In the previous task, the objective is to rapidly get to the core of the situation. Here, you want to sort out the particulars. For example, list the resources available to the project; list all requirements of the project; list the assumptions made by the project; list the constraints on the project; list the risks or events that might occur to delay the project or cause it to fail; compile a glossary of terminology relevant to the project; and construct a cost-benefit analysis for the project.

4.2.1.3 Determine the data-mining goals

In the context of business terminology, a business goal declares objectives. A data-mining goal declares project objectives in technical terms. For example, the business goal might be: ‘Increase house loan payments of existing customers’. A data-mining goal, on the other hand, might be: ‘Predict the size of a house loan a client will ask for, given his cheque account activities and credit card activities over the past three years, demographic information (age, salary, city, etc.) and the average price of houses in the city where the client stays’.

4.2.1.4 Produce a project plan

The task at issue here is to explain the proposed plan for achieving the data-mining goals and thereby also the business goals. The project plan should specify the probable set of steps to be performed during the rest of the project, including an initial selection of tools and techniques to be used.

4.2.2 Data-understanding Step [Daimler-Benz, et al.]

The data-understanding step starts with the collection of initial data and proceeds with actions or tasks aimed at getting familiar with the data, identifying possible data quality problems, recognising first insights into the data, or discovering interesting subsets to formulate hypotheses for hidden information.

4.2.2.1 Collect initial data

This primary collection includes data loading (if necessary) so as to ultimately arrive at data understanding. For example, if you apply a particular tool for data understanding, it makes great sense to load your data into this tool. This effort possibly leads to primary data preparation steps.

4.2.2.2 Describe the data

Inspect the ‘gross’ or ‘surface’ properties of the attained data and report on the results, for example the number of records and fields in each table and the identities of the fields.

4.2.2.3 Explore the data

This task concentrates on the data-mining questions that can be dealt with by means of querying, visualisation and reporting. Such exploration includes the distribution of key attributes, for example the target attribute of a prediction task; results of simple aggregations; properties of significant sub-populations; relations between pairs or small numbers of attributes; simple statistical analyses. These analyses may address the data-mining goals directly; they may also add to or refine the data description and quality reports and supply the transformation and other data preparation needed for further analysis. If appropriate, one could, as part of this task, include graphs and plots that point out data characteristics or lead to exciting data subsets for further examination.

4.2.2.4 Verify the data quality

Analyse the quality of the data by covering questions such as: Is the data complete (does it cover all the cases required)? Is it correct or does it contain errors? If there are errors, how common are they? Are there missing values in the data? If so, how are they represented, where do they occur and how common are they?

4.2.3 Data Preparation Step [Daimler-Benz, et al.]

The data preparation step involves the activities required to build the final data set (data that will be supplied to the data-mining step (modelling tool) from the original raw data. (During this step in this process, multiple data sources may be combined.) The data preparation tasks within the data preparation step are likely to be performed numerous times and not in any prescribed order. Tasks include tabling, recording and attributing the selection, as well as the transformation and cleaning of data (e.g. removing noise and inconsistent data) for the data-mining step (modelling step).

4.2.3.1 Select the data

Choose the data to be used for data mining. Criteria include relevance to the data-mining goals, quality, and technical constraints such as limits on data volume or data types. It is important to note that data selection covers the selection of attributes (columns) as well as the selection of records (rows) in a table.

4.2.3.2 Clean the data

Elevate the data quality to the level required by the selected data-mining techniques. This may involve the selection of clean subsets of the data, the insertion of suitable defaults or more ambitious techniques such as the estimation of missing data by modelling.

4.2.3.3 *Construct the data*

This task takes into account constructive data preparation operations such as the production of derived attributes, entire new records or transformed values for existing attributes.

4.2.3.4 *Integrate the data*

These are methods whereby information is combined from multiple tables or records to create new records or values.

4.2.3.5 *Format the data*

Formatting transformations refer to primarily *syntactic* modifications made to the data that do not change its meaning, but might be required by the data-mining (modelling) tool.

4.2.4 Data-mining Step (Modelling Step) [Daimler-Benz, et al.]

This is a fundamental step in the data-mining process where intelligent methods (the appropriate combination of data-mining techniques) are applied in order to extract data patterns. In this step, different data-mining techniques are selected and applied. Typically, there are a number of techniques for the same data-mining problem type. Some techniques have specific requirements with regard to the form of the data. Therefore, stepping back to the data preparation step is often required.

4.2.4.1 *Select a modelling technique*

As the first step, the actual modelling technique that is to be used should be selected. While you probably already selected a tool in the business understanding step, this task refers to the specific modelling technique to be used, e.g. decision tree building with C4.5 [Quinlan, 1993] or neural network generation with back propagation [Han, et al., 2012]. If multiple techniques are applied, perform this task for each technique separately.

4.2.4.2 Generate a test design

Before building a model, the quality and validity of such model must be tested by generating a procedure or mechanism for this test. For example, in supervised data-mining tasks such as classification, it is common to use error rates as quality measures for data-mining models. Therefore, we usually separate the data set into a train and test set, build the model on the train set and estimate its quality on the separate test set.

4.2.4.3 Build a model

The next task in the data-mining or modelling step is to run the modelling tool on the prepared data set in order to create one or more models.

4.2.4.4 Assess the model

The mining agent interpret the created models according to his domain knowledge, the data-mining success criteria, and the desired test design. While the mining agent judges the success of the application of modelling and discovery techniques in a more technical context, he contacts business analysts and domain experts later on in order to discuss the data-mining results in a business context. Moreover, this task considers models only, whereas the evaluation phase also takes into account all other results that were produced in the course of the project. The mining agent tries to rank the models and assesses them according to the evaluation criteria. As far as possible, he also takes into account business objectives and business success criteria. In most data-mining projects the mining agent applies a single technique more than once or generates data-mining results by using different alternative techniques. In this task, the mining agent also compares all results in accordance with the evaluation criteria.

4.2.5 Evaluation Step [Daimler-Benz, et al.]

Before proceeding to set the new model into final operation, it is essential to assess this model and review the steps that were implemented to build it so as to be certain that it properly accomplishes the business objectives. The goal of this step is to determine if there is some important business issue that has not been sufficiently considered. On conclusion of this step, a decision on the use of the data-mining results should be reached.

4.2.5.1 Evaluate the results

Previous evaluation steps dealt with factors such as the accuracy and generalisability of the model. This step assesses the extent to which the model meets the business objectives and seeks to determine if there is some business reason why this model is deficient. Another option of evaluation is to assess other data-mining results that have been generated. These results cover models that are inevitably related to the original business objectives, as well as all other findings that are not inevitably related to the original business objectives. They might also uncover additional challenges, information or hints for future advice.

4.2.5.2 Review the process

At this stage the resultant model should satisfy business needs. It is now suitable to do a more systematic review of the data-mining commitment in order to determine if there is any important aspect or task that has been overlooked. This review also covers quality assurance issues, e.g. did we correctly build the model? Did we only use attributes that we are allowed to use and that are available for future analyses?

4.2.5.3 Determine the next steps

Based on the assessment results and the process review, a decision is made at this stage about how the project is to proceed. It needs to be decided whether to conclude the project and move on to deployment (if appropriate), whether to commence further iterations, or whether to start new data-mining projects. This task includes the analyses of remaining resources and budgets that influence the decisions.

4.2.6 Deployment Step [Daimler-Benz, et al.]

The creation of the model is not the end of the project. Interesting patterns are extracted that represent the knowledge gained. This knowledge gained needs to be organised and presented in a way so that the customer can use it (e.g. this may typically involve some visualisation techniques). Depending on the requirements, the deployment step may be as straightforward as generating a report or as difficult as implementing a repeatable data-mining process across the enterprise. In many cases it is the customer, not the mining agent (or the data analyst), who carries out the deployment steps. However, even if the mining agent will not carry out the deployment effort, it is important for the customer to understand in advance what activities need to be carried out in order to truly make use of the created models. The client must be able to incorporate business insights gained from the data-mining step in the organisation's business and information systems.

4.2.6.1 Plan the deployment

During this task a plan for deployment of the data mining results into the business, is formed by basing it on the evaluation results.

4.2.6.2 Plan the monitoring and maintenance

To prevent unnecessary long periods of incorrect usage of data mining results, a careful preparation of a maintenance strategy is important. Also, a detailed plan on the monitoring process of the deployment of the data mining results should be in place.

4.2.6.3 Produce the final report

Depending on the deployment plan, the project leader and his team write up a final report consisting out of either a summary of the project or if the report is a final presentation of the data mining results.

4.2.6.4 Review the project

This is the final assessment task where the following questions are answered: What went right and what went wrong? What was done well and what needs to be improved?

This iterative sequence of steps as mentioned in Sections 4.2.1 through to 4.2.6, known as the data-mining process, takes place in the data-mining environment. For purposes of this research, the data-mining environment will comprise of all physical storage devices needed for data, all data and knowledge stored for possible data manipulation and all programs and humans authorised to work on the data-mining process.

4.3 THE DATA-MINING ENVIRONMENT

The data-mining process has been discussed and a naming convention has been established. The other important role players in this data-mining environment are the mining agents themselves. (When referring to mining agents in this thesis, this may include business analysts, system analysts or statisticians that have access to the data in the data-mining process.)

The mining agents are in direct contact with the data that becomes information during the data-mining process and even more so with the data that becomes knowledge during the pattern evaluation step. It is during these steps that the relevant mining agents may become aware of possible sensitive information and knowledge. If a bank should do data mining on its clients, it will be for the good of its reputation to have an access control model in place that minimises the security situation that arises when mining agents become aware of possible sensitive information. One of the challenges in knowledge management is the maintaining of security [Bertino, et al., 2006].

When discussing the data-mining environment in general, it is true to acknowledge a general aim of knowledge discovery, which implies finding knowledge that is otherwise hidden by large volumes of data. This points to a potential problem in respect of information security risk, namely – if the knowledge is hidden, how do we know that a security risk exists? Unexpected security issues may arise during the data-mining process. For instance, where the security of individual data items is not a concern, there may well be patterns in the mined data that pose an information security risk.

The aim of this research is to minimise the chances of a potential breach of confidentiality made possible by the hidden knowledge revealed through the data-mining process. For example, a mining agent should not be able to leak sensitive (positive or negative) hidden knowledge of one client to a conflicting client and, by doing this, adversely affect the first client and put at risk the reputation of the global company for whom the mining agent works.

Performing data mining is important in obtaining underlying and important information about competitors and clients. Getting hold of unknown information by means of data mining is quite a common phenomenon, but is it equally commonly known that one's own information is accessible at remote sites where competitors are doing data mining on data about you?

The following example depicted in Figure 4.2 illustrates this problem:

Enterprise A bought shares from Enterprise B at R10 a share. Enterprise A and Enterprise B use the same bank, namely DM Bank. DM Bank's data-mining activities show that there is a negative change in Enterprise B's credit rating. A breach of confidentiality is now a possibility for Enterprise B via DM Bank's data-mining activities. Under these circumstances it is important for Enterprise B to minimise its information security risk exposure as revealed by data-mining activities. If Enterprise B allows such activities without security in mind, it will be possible for DM Bank and Enterprise A to know about Enterprise B's credit rating drop without Enterprise B even knowing that anybody else has knowledge of these circumstances. On the other hand, if Enterprise B is 'security aware', it will have policies in place to handle new knowledge discovery by another party, e.g. DM Bank. These policies may include written agreements on how to handle knowledge resulting from the data-mining process.

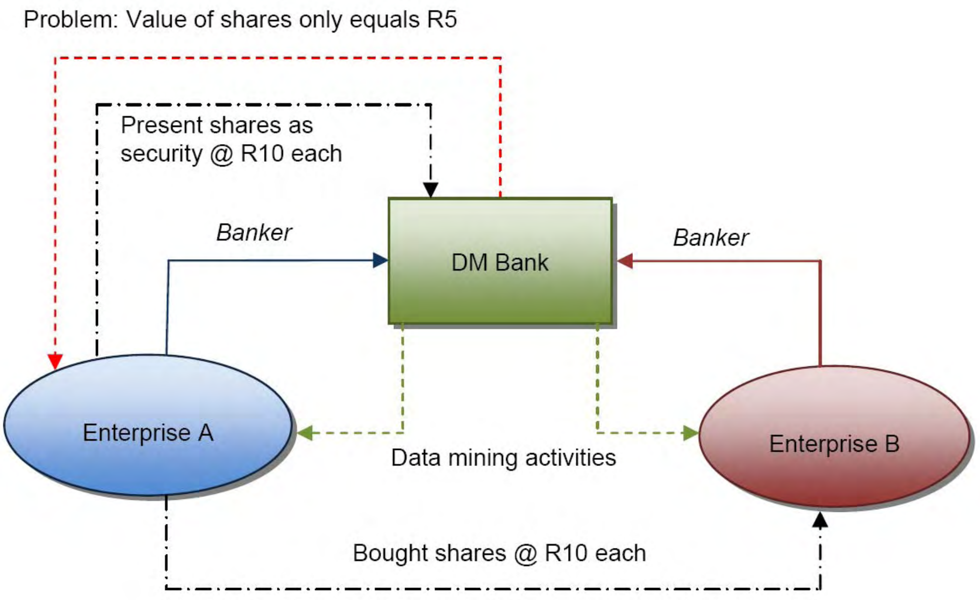


Figure 4.2: Example of a possible breach of confidentiality

The current research will not concentrate on privacy, but rather on the confidentiality of the data of for example Enterprise A and Enterprise B.

4.4 SECURITY REQUIREMENTS FOR DATA MINING

The different steps in the data-mining process have been explained so far. The different security problems within these steps will be highlighted next. However, the main emphasis is on confidentiality and possible ways of solving a breach of confidentiality while performing data mining.

In Table 4.1 the data-mining process is confirmed in the first column (orange) as a series of steps that were observed by the researcher during industry discussions, as well as in experiences from industry and real life. This specific data-mining process was confirmed by literature when the researcher compared her results with the Cross-Industry Standard Process for Data Mining (CRISP-DM) [Daimler-Benz, et al.]. The CRISP-DM steps are thus represented in the first column of Table 4.1. In the second and third columns (green) of Table 4.1 integrity and availability with regard to the data-mining steps in column 1 are explained. The explanations are based on expert input from industry as well as the researcher’s own knowledge and experience in this field. The fourth column (pink) looks specifically at confidentiality. The researcher followed the same steps as for columns 2 and 3 to obtain this information, but because confidentiality constitutes the emphasis of this research, this column has its own colour and will be extended in Table 4.2. Once Table 4.1 was constructed, the information was discussed and verified with different experts in industry.

Table 4.1: The Data-mining Process Mapped onto the Classic CIA Security Triad (Confidentiality, Integrity and Availability)

	Core Information Security Services		
Data-mining process	**Integrity**	**Availability**	**Confidentiality**
	CNSSI-4009 [CNSS, 2010]: “Quality of an IS reflecting the logical correctness and reliability of the operating system; the logical completeness of the hardware and software implementing the protection mechanisms; and the consistency of the data structures and occurrence of the stored data.” *Data cannot be modified undetectably.*	CNSSI-4009 [CNSS, 2010]: “Timely, reliable access to data and information services for authorized users.” *The computing systems used to process and store the information, the security controls used to protect it, and the communication channels used to access it must be functioning correctly, which involves the prevention of denial-of-service attacks.*	CNSSI-4009 [CNSS, 2010]: “Assurance that information is not disclosed to unauthorized individuals, processes, or devices.” *The disclosure of information to unauthorised individuals or systems must be prevented. Confidentiality is necessary (but not sufficient) for maintaining the privacy of the people whose personal information a system holds.*
Business-understanding step Determine business objectives Determine and assess details of	The unauthorised modification or destruction of information during the *business-understanding step* is possible, and results in low **integrity** of the *business-understanding step*. When *determining business objectives* or *data-mining goals*, unauthorised	The **availability** of data in the *business-understanding step* can be compromised by authorised users not having timely and reliable access to the clients so as to be able to establish from a business perspective what the client wants. Users must be able to *determine and assess*	The *business understanding step* cannot be declared **confidential** if the process of *producing a project plan* through *determining business objectives, assessing the situation* and

	Core Information Security Services		
Data-mining process	**Integrity**	**Availability**	**Confidentiality**
the situation Determine data-mining goals Produce project plan	modification is possible. Business objectives as seen by the client can apparently be modified by the mining agent and data-mining goals can on purpose be defined erroneously by the mining agent.	*the details* of the situation as well as the *data-mining goals* in order for the *project plan to be produced*.	*determining data-mining goals* is disclosed to a mining agent that also mines data on a conflicting secondary company's data sets (e.g. two petrol companies). Such a mining agent can now be exposed to the data of two conflicting secondary companies in an unauthorised way.
Data-understanding step Collect initial data Describe data Explore data Verify data quality	The output of the *data-understanding step* is the *initial data collection report*, the *data description report*, the *data exploration report* and the *data quality report*. In all of these reports data can be modified undetectably (e.g. possible solutions for data quality problems must be listed in the data quality report and generally depend heavily on both data and business knowledge). The possible undetected modification of these	The *data-understanding step* starts with an *initial collection of data*. During this task the timely and reliable access to data and information for authorised users is essential. The **availability** of the data that is needed to produce these reports can be compromised if timely and reliable access to this data is not possible.	During the *data-understanding step* the mining agent *collects the initial data*, *describes and explores it*, and *verifies the quality of the data*. If the same mining agent should perform this process on the data sets of two conflicting secondary companies (e.g. two airline companies), the entire process can be seen as unauthorised,

	Core Information Security Services		
Data-mining process	**Integrity**	**Availability**	**Confidentiality**
	reports has serious implications for the **integrity** of all the reports emanating from the *data-understanding step*.		thus risking the **confidentiality** of the output of the *data-understanding step*.
Data preparation step Select data: The data that is used in the analysis task is retrieved from the database during this step. Clean data Construct data Integrate data Format data	The main output of this step is the data set(s) produced for modelling or analysis of the project. The *Select Data task* may compromise the integrity of this step to a major degree because the output for this task is the *rationale for inclusion/exclusion*. This is a list of data that should be included or excluded, as well as the reasons for these decisions. Such a subjective task poses a risk for the integrity of this step because it is possible to change the composition of selected data.	During this step a selection must be made of the data to be used for analysis. The quality of the data must be raised to the level required for the selected analysis technique. *Constructive data* preparation techniques such as the production of new records may follow. The *integration of data* where information is combined from multiple tables or records is followed by the *formatting of data* (only syntactic) as required by the formatting tool. For all these tasks **availability** of data from the database is important. If this availability is not timely, this step is of little value.	The data sets produced for modelling or analysis of the project (though still 'unworked' in the sense that no analysis has been done on it) constitute the final output of this step and contain major information about the project. Possible hidden information might be revealed and at this stage secondary companies may by themselves decide that the same mining agent may not work on the data sets of conflicting secondary companies. This is to ensure **confidentiality** of information and may prescribe

	Core Information Security Services		
Data-mining process	**Integrity**	**Availability**	**Confidentiality**
			that data mining may only be done on some data sets or that specific data sets are not available for data mining at all.
Data-mining (modelling) step This is a fundamental step in the knowledge discovery process where intelligent methods (the appropriate combination of data-mining algorithms) are applied in order to extract data patterns. Select a modelling technique Generate a test design	During this step different data-mining techniques can be used and if the wrong technique is deliberately used for the problem or questions at hand, the wrong answers will come forward. This will be the same as destroying or changing the output of this step and will definitely affect the **integrity** of the data-mining step.	The **availability** of prepared data sets on which to run the modelling tool so as to create one or more models is important. If these prepared data sets are not timely available, the quality of the model that will be built is questionable because it was not tested in a proper manner by building more than one model and assessing these models afterwards.	**Confidentiality** during this step is very important because the results of this step are unpredictable and the agents who should have access to the results are impossible to define beforehand. One of the outputs is a report on the interpretation of the models. During this step it is important to consider each secondary company's preferences of possible specific mining agents or the different secondary companies' different priority levels on the confidentiality of the results of this specific step. Some secondary companies may

	Core Information Security Services		
Data-mining process	**Integrity**	**Availability**	**Confidentiality**
Build the model Assess the model			see these results as workable data and other may see them as sensitive data that should be handled separately from the masses.
Evaluation step Evaluate results Conduct a review process Determine next steps	This step assesses the degree to which the model meets its business objectives. The **integrity** during this step might be compromised because this evaluation is a subjective test (e.g. evaluating results may become a personal opinion).	To evaluate the results by reviewing the process means that the **availability** of all the data and information is crucial. The intentional or accidental missing of data or information makes this step obsolete and the security of the data-mining process will be questioned.	**Confidentiality** is of utmost importance during the evaluation of results. Clients may request that the *evaluation of the results* and the *review process* must be done by specific mining agents or only in specific circumstances, for example if the results of the data-mining (modelling) step were positive.
Deployment step Plan the model deployment Plan monitoring and maintenance Produce the	**Integrity** in the *deployment step* can be compromised by the possibility of following a human method of ignoring all negative aspects of the project and underlining all positive aspects of the project.	The **availability** of data is essential to be able to plan the deployment of the model, plan its monitoring and maintenance, produce the final report and then to review the report.	During deployment, **confidentiality** is important for the global company responsible for the data-mining process so as not to affect any participating secondary companies either

	Core Information Security Services		
Data-mining process	**Integrity**	**Availability**	**Confidentiality**
final report Review the project			positively or negatively. During this step the participating secondary companies may also request to monitor and personally be involved in producing the final report to make sure that they agree with the end result.

Security requirements for different security problems encountered in the data-mining steps (as indicated in Table 4.1) will be described once the access control requirements for doing data mining in general have been discussed.

4.5 ACCESS CONTROL REQUIREMENTS WHEN DOING DATA MINING

As earlier indicated in Section 3.4.3.1, it is possible to sanitise data sets and databases by the removal of private information when the Chinese Wall Security Policy model is used. It is, however, possible for the removed information to be recovered by data-mining techniques.

When sensitive data has to be released to a third party, privacy preservation is a technique to be used for example when a mining agent has to work with sensitive data [Byun, et al., 2005]. It is acknowledged that the emphasis is on privacy in privacy preservation and not on access control but it still needs to be considered within the overall aim of addressing access control requirements for data-mining environments. An important goal of privacy-preserving data mining is the development of new models for inferred or indirect information when access to exact and precise information in the original individual records is not possible [Agrawal & Srikant, 2000]. The study of data mining as far as privacy preservation is concerned has become an active area of research in computer science [Agrawal & Srikant, 2000; Clifton & Marks, 1996; Conrado, Petkovic, & Jonker, 2004; Kantarcioglu & Clifton, 2004]. It is important to emphasise that, for such research, access to the precise information in the original individual records is essential. For this reason privacy preservation is a technique that will not be further investigated here.

Privacy-aware access control merges the concepts of privacy and access control. Current research in privacy-aware access control has a twofold focus: firstly, it is concerned with the definition and explanation as well as development of access control and privacy languages such as XACML (eXtensible Access Control Markup Language)[OASIS, 2010] and secondly, it aims at defining architectures to protect and preserve the privacy of either services or clients [Ardagna, Cremonini, Damiani, Vimercati, & Samarati, 2006]. Privacy-aware access control is central for the development of access control models for data-mining environments in general. Unlike in the case of privacy-preserving approaches, privacy-aware access control facilitates controlled access to individual records. Since the research in hand concentrates on confidentiality, the concept of privacy is too narrow, and hence privacy-aware access control will not be investigated further.

The new access control requirements will be derived from the broader security perspective as depicted in Table 4.1. Table 4.2 depicts the confidentiality problems brought to light in Table 4.1, which can cause a breach of confidentiality while doing data mining. It also explains new requirements for access control to prevent these potential confidentiality problems.

In Table 4.2 the data-mining steps are repeated in column 1 (orange) and the confidentiality issues that arose during data mining (see Table 4.1) are added to this table in column 2 (pink). The third (blue) column of Table 4.2 contains suggestions for maintaining confidentiality during data mining. After Table 4.2 was set up, the information was discussed and verified with different experts in industry.

Table 4.2: New access control requirements to prevent confidentiality problems during data mining

Data-mining process as seen in Table 4.1	**Confidentiality issues during data mining as seen in Table 4.1**	**Suggestions for maintaining confidentiality during data mining**
Business-understanding step Determine business objectives Determine and assess details of the situation Determine data-mining goals Produce project plan	The *business-understanding step* cannot be declared **confidential** if the process of *producing a project plan* through *determining business objectives, assessing the situation* and *determining data mining goals* is disclosed to a mining agent that also mines data on a conflicting secondary company's data sets (e.g. two petrol companies). Such a mining agent can now be exposed to the data of two conflicting secondary companies in an unauthorised way.	**i** No single mining agent or single group of mining agents should work on the data sets of two (or more) conflicting secondary companies. **ii** All secondary companies whose data will be mined should be in a position to declare to which functional business domain/s it belongs (e.g. Does it belong to the petrol functional business domain or is it a food company with a small interest in the petrol functional business domain?)
Data-understanding step Collect initial data Describe data Explore data Verify data quality	During the *data-understanding step*, the mining agent *collects the initial data, describes and explores it*, and *verifies the quality of the data*. If the same mining agent should perform this process on the data sets of two conflicting secondary companies (e.g. two airline companies), the entire process can be seen as unauthorised, thus risking the **confidentiality** of the output of the *data-understanding step*.	**iii** The allocation of mining agents by the global company that is responsible for the data-mining environment should be such that the risk of a breach of confidentiality is minimised. Any breach of confidentiality may cause a reputation risk to the global company that is responsible for the data-mining process.

Data-mining process as seen in Table 4.1	**Confidentiality issues during data mining as seen in Table 4.1**	**Suggestions for maintaining confidentiality during data mining**
Data preparation step Select data: The data used in the analysis task is retrieved from the database during this step. Clean data Construct data Integrate data Format data	The final output of this step is the data set(s) produced for modelling or for the major analysis work of the project. Though still ‘unworked’ in the sense that no analysis has been done on them, the data sets contain major information about the project. Possible hidden information may be revealed and at this stage secondary companies may by themselves decide that the same mining agent may not work on the data sets of conflicting secondary companies to ensure the **confidentiality** of information. They may also prescribe that data mining may be done on some data sets only and that specific data sets are not available for data mining at all.	**iv** Secondary companies whose data sets are exposed to data mining should be in a position to state whether they want to share mining agents with other secondary companies or whether they want a mining agent to be assigned solely to themselves.
Data-mining (modelling) step This is a fundamental step in the knowledge discovery process where intelligent methods (the appropriate combination of data-mining algorithms) are applied in order to extract data	**Confidentiality** during this step is very important because the results of this step are unpredictable and the agents who should have access to the results are impossible to define beforehand. One of the outputs is a report on the interpretation of the models. During this step it is important to consider each secondary company’s preferences of possible specific mining agents or the different secondary companies’ different	**v** Secondary companies whose data sets are exposed to data mining should be in a position to state a sensitivity level for conflict between themselves and any other secondary company. This sensitivity level for conflict should come into play when deciding which mining agent can work on the data sets of the secondary company

Data-mining process as seen in Table 4.1	**Confidentiality issues during data mining as seen in Table 4.1**	**Suggestions for maintaining confidentiality during data mining**
patterns. Select a modelling technique Generate a test design Build the model Assess the model	priority levels on the confidentiality of the results of this specific step. Some secondary companies may see these results as workable data and other may see them as sensitive data that should be handled separately from the masses.	and also whether secondary companies can share mining agents.
Evaluation step Evaluate results Conduct a review process Determine next steps	**Confidentiality** is of utmost importance during the evaluation of results. Clients may request that the *evaluation of the results* and the *review process* must be done by specific mining agents or only in specific circumstances, for example if the results of the data-mining (modelling) step were positive.	
Deployment step Plan the model deployment Plan monitoring and maintenance	During deployment, **confidentiality** is important for the global company responsible for the data-mining process so as not to affect any participating secondary companies either positively or negatively. During this step the participating secondary companies may also request to monitor and personally be involved in producing the final report to make	**vi** The global company responsible for the data-mining process should assist in the process of assigning mining agents by adding information to the decision-making process (e.g. What other secondary companies are present and will be part of the data mining process? What other secondary

Data-mining process as seen in Table 4.1	Confidentiality issues during data mining as seen in Table 4.1	Suggestions for maintaining confidentiality during data mining
Produce the final report Review the project	sure that they agree with the end result.	companies are present in the original secondary company's functional business domain, without this company being aware of it?

The new access control requirements for a data-mining environment should be able to maintain confidentiality while managing data-mining results that are unpredictable in nature. To be able to maintain confidentiality, the suggestions in Table 4.2 will be summarised and used as input requirements for the proposed access control model.

(i) No single mining agent or single group of mining agents should work on the data sets of two (or more) conflicting secondary companies. Every secondary company whose data will be mined should be in a position to declare to which functional business domain/s it belongs, e.g. does it belong to the petrol functional business domain or is it a food company with a small interest in the petrol functional business domain?

Security requirement*: A mining agent must not work on the data sets of conflicting secondary companies.*

(ii) Every secondary company should be in a position to declare to which functional business domain/s it belongs.

Security requirement*: Every secondary company must declare to which functional business domain/s it belongs, e.g. does it belong to the petrol functional business domain? Is it a food company with a small interest in the petrol functional business domain? Does it belong to a functional business domain called 'general'?*

(iii) The global company that is responsible for the data-mining environment should allocate mining agents in such a way that the risk of a breach of confidentiality is minimised. Any breach of confidentiality may cause a reputation risk to the global company that is responsible for the data-mining process.

> Security requirement*: The global company responsible for the overall data-mining environment must manage the allocation of mining agents so as to minimise the risk of a breach of confidentiality. It should adopt an access control model based on the conflicts between the different secondary companies on which the global company wants to perform data mining.*

(iv) Secondary companies whose data sets are exposed to data mining should be in a position to state whether they want to share mining agents with other secondary companies or whether they want a mining agent to be assigned solely to themselves.

> Security requirement*: By definition, a secondary company should be able to insist on being the only secondary company in the 'group' of secondary companies that are served by a specific mining agent.*

(v) Secondary companies whose data sets are exposed to data mining should be in a position to state a sensitivity level for conflict between themselves and any other secondary company. This sensitivity level for conflict should then come into play when deciding which mining agent may work on the data sets of the secondary company, as well as whether secondary companies may share mining agents.

> Security requirement*: Each secondary company must be able to determine on a predefined scale the conflict levels between itself and other secondary companies, and such conflict levels must be moderated by the global company.*

(vi) The global company that is responsible for the data-mining process should assist in the process of assigning mining agents by adding information to the decision-making process, e.g. what other secondary companies are present and will be part of the data-mining process? What other secondary companies, of which the original secondary company was not aware, are present in the original secondary company's functional business domain?

> Security requirement*: The global company that is involved in the data-mining process can, with the help of the proposed access model, add 'intelligence' to the*

process of assigning mining agents to different secondary companies, because it often has more information about these secondary companies than the secondary companies themselves.

Based on the research conducted in the current study, the above are the new access control requirements that are suggested in Table 4.2 in order to maintain confidentiality and solve the broad security problems encountered during data mining (see Table 4.1).

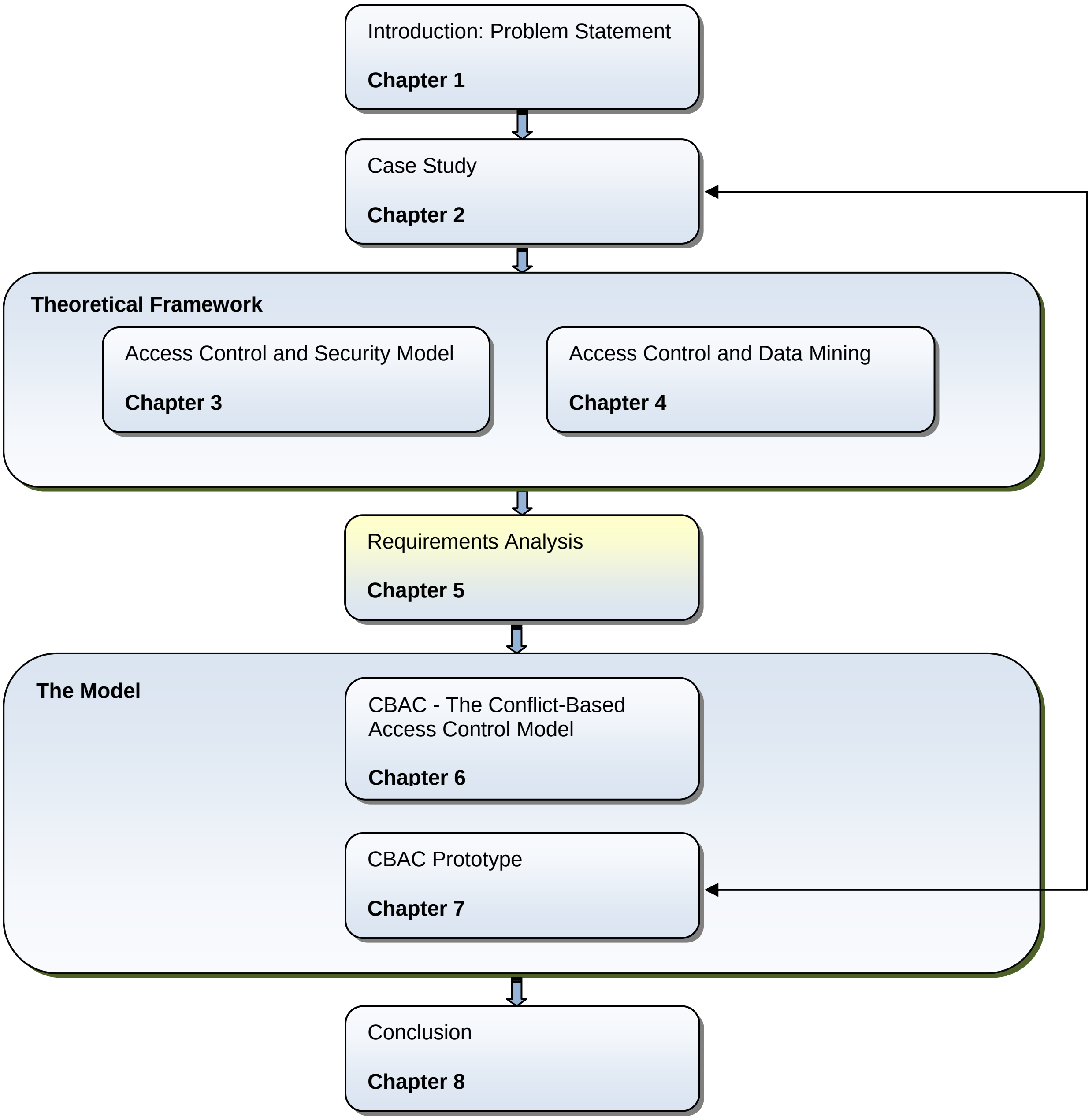
4.6 CONCLUSION

It is essential to consider the access control requirements within the data-mining process when deciding how to apply mining agents to different secondary companies. If a breach of confidentiality is allowed, this may cause a serious reputation risk for the global company that is doing data mining.

A model that concentrates on the (potential) conflict between the different competing clients in a specific functional business environment (e.g. the banking environment) will minimise the risk of a breach of confidentiality during the data-mining process.

CHAPTER 5: REQUIREMENTS ANALYSIS

THESIS LAYOUT



5.1 INTRODUCTION

A number of access control models theories and views on security in general were discussed in Chapter 3 and it was concluded that access control models can, for purposes of this thesis, be grouped into three main categories: models for confidentiality; models that can handle conflict to a certain degree; and models for the data-mining environment that concentrate on privacy. The current research proposes an access control model that captures relevant aspects from these three main categories to the benefit of the data-mining environment, while concentrating on confidentiality. This concept is depicted in Figure 5.1.

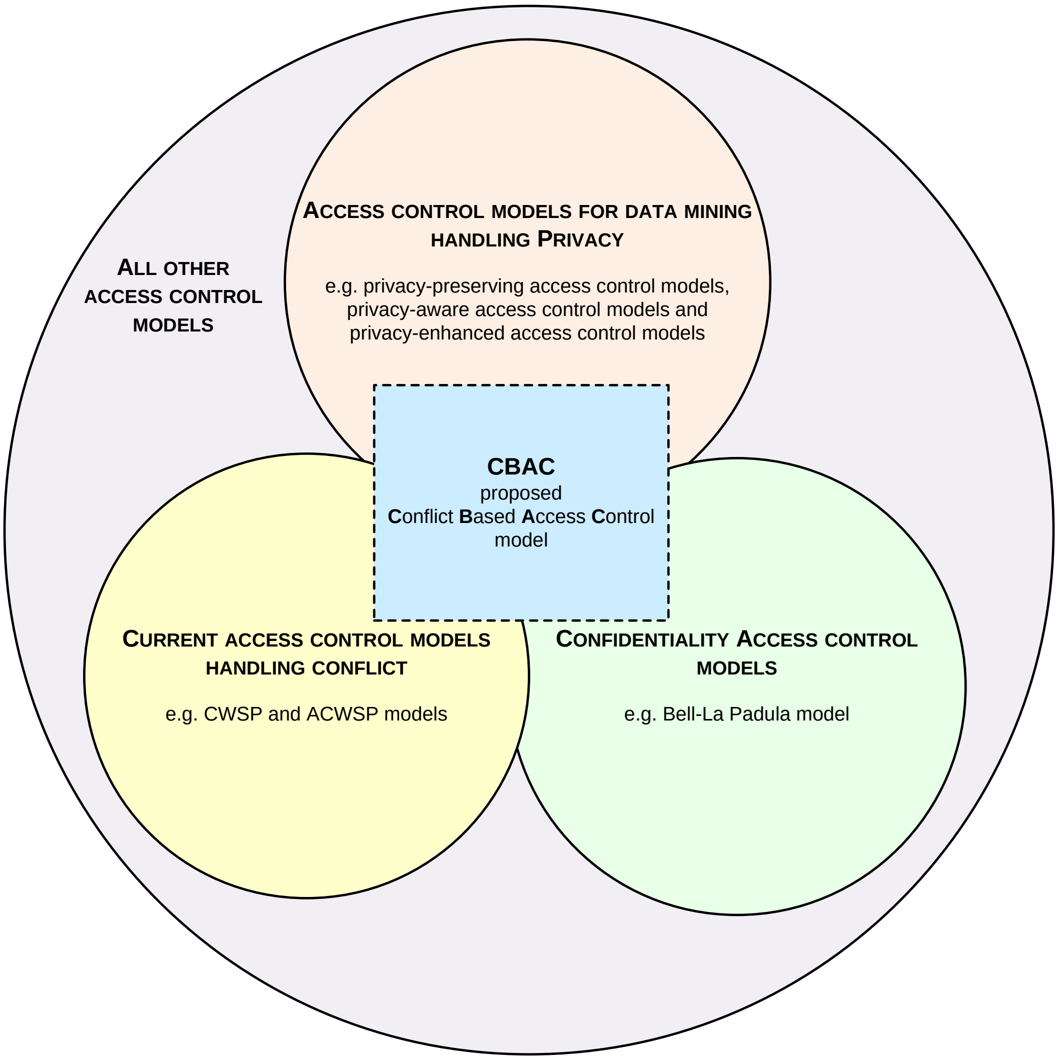
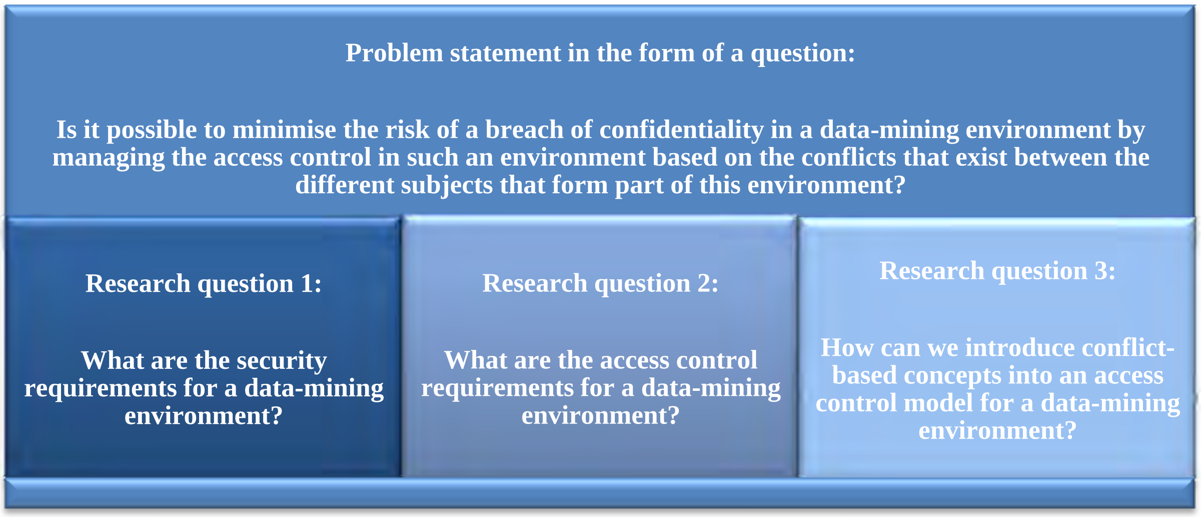

ALL OTHER ACCESS CONTROL MODELS
ACCESS CONTROL MODELS FOR DATA MINING HANDLING PRIVACY
e.g. privacy-preserving access control models, privacy-aware access control models and privacy-enhanced access control models
CBAC
proposed
Conflict Based Access Control model
CURRENT ACCESS CONTROL MODELS HANDLING CONFLICT
e.g. CWSP and ACWSP models
CONFIDENTIALITY ACCESS CONTROL MODELS
e.g. Bell-La Padula model

Figure 5.1: Conceptual framework showing how the application field of the proposed CBAC model overlaps with the application fields of other access control models

In Chapter 4 the data-mining process was analysed. The following requirements proved to be important when doing data mining while keeping security (especially confidentiality) of access control in mind:

- A mining agent must not work on the data sets of conflicting secondary companies.
- Every secondary company must declare to which functional business domain/s it belongs. (E.g. does it belong to the petrol functional business domain? Is it a food company with a small interest in the petrol functional business domain? Does it belong to a functional business domain called 'general'?)
- The global company that is responsible for the overall data-mining environment must manage the allocation of mining agents in such a way that the risk of a breach of confidentiality is minimised. It should adopt an access control model, based on the conflicts between the different secondary companies on which the global company wants to perform data mining.
- By definition, a secondary company should be able to insist on being the only secondary company in the 'group' of secondary companies that are served by a specific mining agent.
- Each secondary company must be able to determine on a predefined scale the conflict levels between itself and other secondary companies, and such conflict levels must be moderated by the global company.
- The global company that is involved in the data-mining process can, with the help of the proposed access control model, add 'intelligence' to the process of assigning mining agents to different secondary companies, because it often has more information about these secondary companies than secondary companies themselves. For example, a potential conflict of interest can occur between two secondary companies that are not in the same functional business domain. They may not even be aware of such a conflict, as it arises from an indirect path that exists between the two secondary companies. However, the global company, which has access to both secondary companies' information, can prevent the possible harmful assignment of the same mining agent to these two specific secondary companies.



Problem statement in the form of a question:

Is it possible to minimise the risk of a breach of confidentiality in a data-mining environment by managing the access control in such an environment based on the conflicts that exist between the different subjects that form part of this environment?

Research question 1:

What are the security requirements for a data-mining environment?

Research question 2:

What are the access control requirements for a data-mining environment?

Research question 3:

How can we introduce conflict-based concepts into an access control model for a data-mining environment?

In this chapter, the focus of the study is on answering the second research question namely:

What are the access control requirements for a data-mining environment?

The requirements as captured in Chapter 3 and Chapter 4 will thus be summarised in the following section.

5.2 REQUIREMENTS FOR THE CBAC MODEL

The current research proposes an access control model based on the conflicts that exist between companies that are competing in a commercial environment for the same market. The model minimises the possibility of a breach of confidentiality in respect of the information of, for example, different competing shoe companies that are clients at the same commercial bank. The name of the proposed model has been derived from the model description, namely a **C**onflict-**B**ased **A**ccess **C**ontrol model – CBAC.

Based on the initial research done during the literature study for this thesis, some preliminary requirements were identified [Loock & Eloff, 2005a] and after subsequent research [Loock, Eloff, & Heidema, 2012], the following requirements were set for the CBAC model:

- The proposed model should be based on the idea of conflict-of-interest classes. This will have the result that a mining agent cannot work on the data sets of conflicting secondary companies.

- All secondary agents must declare to which functional business domain they belong, e.g. do they belong to a petrol functional business domain, do they belong to a food functional business domain, or do they belong to both functional business domains?

- The model must provide for different degrees of conflict to determine whether a potential conflict may influence an access control decision. This should be managed by the global company that is responsible for the overall data-mining environment as well as for the allocation of mining agents.

- Secondary companies can, up to a certain extent, determine the conflict levels between each other on a predefined scale. This decision on conflict levels must be moderated by the global company.

- The model must be able to predict conflict over and above the conflict defined by the secondary agents. The global company that is involved in the data-mining process can, with the help of the proposed access control model, add knowledge and insight to the process of assigning mining agents to different secondary companies because it has more information about secondary companies than the secondary companies themselves.

- The proposed model should minimise the global agent’s reputation risk by enabling the global agent to manage possible confidentiality leaks more effectively.

- In terms of the proposed model it is compulsory for a secondary agent to belong to one or more functional business domain(s). This requirement forces all secondary companies to belong to the ‘general’ functional business domain if they cannot declare a specific functional business domain to which they belong. If secondary company X is accessed by a user (a data-mining agent), then all other secondary companies in all of the functional business domains of the selected secondary company X become unavailable to that specific user, except for the secondary companies in the ‘general’ functional business domain.

- The concept of a history record is important. A record should be kept of all the data sets that a mining agent is currently working on, as well as those accessed in the past.

- A secondary company may have an interest in a functional business domain for a variety of reasons. A secondary company may also have different conflict levels in respect of different functional business domains.

The CBAC model will be described in more detail in Chapter 6.

5.3 INDUSTRY AND THE REQUIREMENTS FOR THE CBAC MODEL

The above-mentioned requirements were discussed with various reviewers in industry and a summary of their comments follows below:

Industry reviewer 1 from the banking environment was of the opinion that the model was relevant and appropriate, but that it was not developed for single banking clients. The model was perhaps more inclined towards wealthy clients within the banking environment who might have a need for defining a conflict area, and it was definitely inclined to a greater extent towards conflicting companies within the banking environment. These companies should then be combined into functional business domains to assist with the classification of the model when conflicting boundaries are defined. Such classification should be mandatory to all companies, even if a company should only belong to a group called *General*. According to industry reviewer 1, the management and allocation of mining agents should definitely be the responsibility of the global agent as prescribed by the model. Reputation risk is an important factor in the banking environment and if the CBAC model could assist in minimising this risk, it would certainly constitute one of the model’s significantly positive features.

Industry reviewer 2, who hailed from an Information Technology company, was very excited about the model and could clearly see how it would have a positive influence on the breach of confidentiality problem that exists at the moment. It was the view of this reviewer that the CBAC model would be better used within an organisation of a substantial size with a large data-mining department, because differentiating between data miners when there were only a few to make use of, would be impractical.

The third industry reviewer also came from the banking environment and was very positive about the model and the conflict-based access control concept in general. This reviewer was also aware of websites that sell information gathered through various data-mining processes. The concept of conflicts as defined in this thesis is known to this industry reviewer as so-called toxic combinations and these combinations are known threats in the data-mining environment in which he works. Also, for any secondary company to insist on being in a specific 'group' that is serviced by a specific mining agent or even being the only secondary company in the 'group' of secondary companies that are serviced by a specific mining agent is perfectly acceptable. To be able to define the conflict levels between different secondary companies on a predefined scale, even though these conflict levels are moderated by the global company, is a practical and useful feature of the CBAC model.

5.4 THE CONCEPTUAL FRAMEWORK FOR THE CBAC MODEL

The CBAC model will now be explained with a set of ordered and known concepts namely the fundamental model of access control [Gollmann, 2006]. The problem domain within which the CBAC model will be explained, is access control in a data-mining environment. The slightly adapted fundamental access control model as originally explained by Gollmann [2006] is depicted in Figure 5.2.

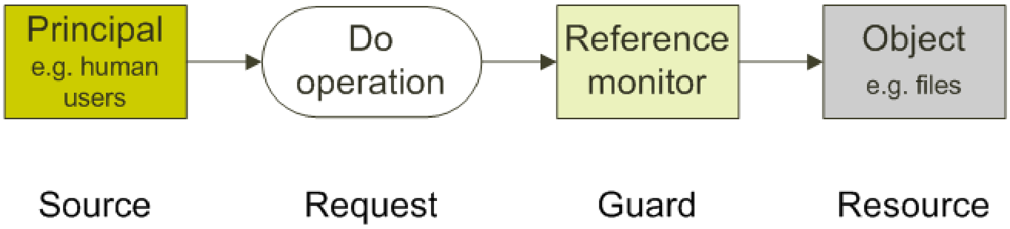


Figure 5.2: The Fundamental Model of Access Control [Gollmann, 2006]

The objective of the CBAC model is to manage the access of a mining agent (principal) in a data-mining environment (protected system consisting of all the data sets of all the secondary agents). The fundamental model of access control can be re-drawn with the proposed CBAC model in mind – this is depicted in Figure 5.3.

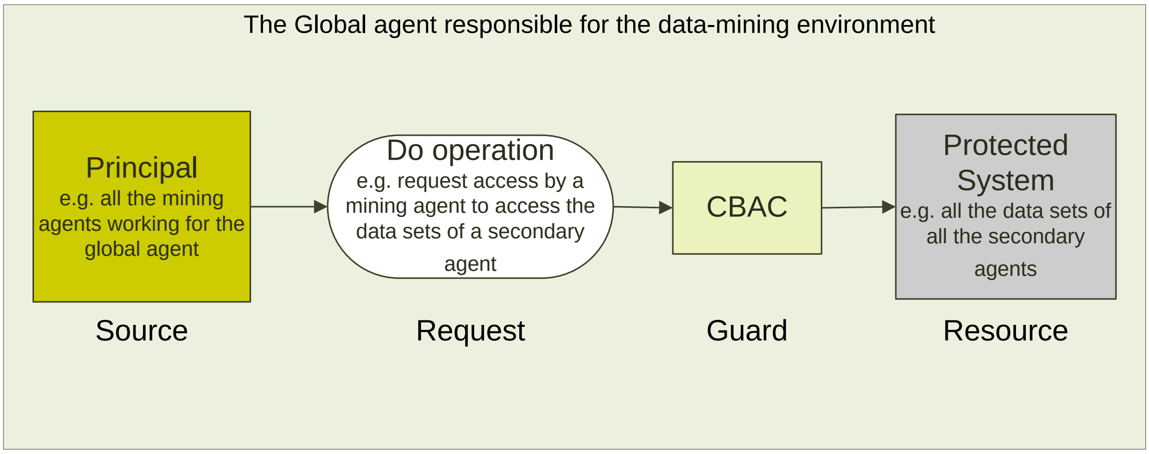

The Global agent responsible for the data-mining environment
Principal
e.g. all the mining agents working for the global agent
Do operation
e.g. request access by a mining agent to access the data sets of a secondary agent
CBAC
Protected System
e.g. all the data sets of all the secondary agents
Source
Request
Guard
Resource

Figure 5.3: CBAC and the Fundamental Model of Access Control

The following is important in the relationships between the mining agents, the secondary agents and the global agent:

- The principal comprises one or more mining agents, e.g. DM_1, DM_2, …, DM_m.
- The ‘do operation’ is a request to access the data sets of a secondary agent. If a mining agent is able to access one data set of a secondary agent, it will be able to access all the data sets of this secondary agent. For this reason, the access request is to access a secondary agent and not a specific data set of a secondary agent. Also, the type of access right required by the data-mining agent may vary and may include for example doing a decision tree analysis on a secondary agent. The CBAC model as proposed in this research focuses on modelling the relationship(s) between data-mining agents and secondary agents (protected system) and, for reasons of complexity, does not differentiate between the types of access.
- The proposed CBAC model functions as the guard between the mining agents (requesting access to the data sets of the secondary agents) and the protected system (consisting of the secondary agents).
- The protected system consists of all the secondary agents (e.g. c_1, c_2, … c_i, …, c_z) on whose data sets the global agent (e.g. DM Bank) is performing data mining.

These relationships are depicted in Figure 5.4.

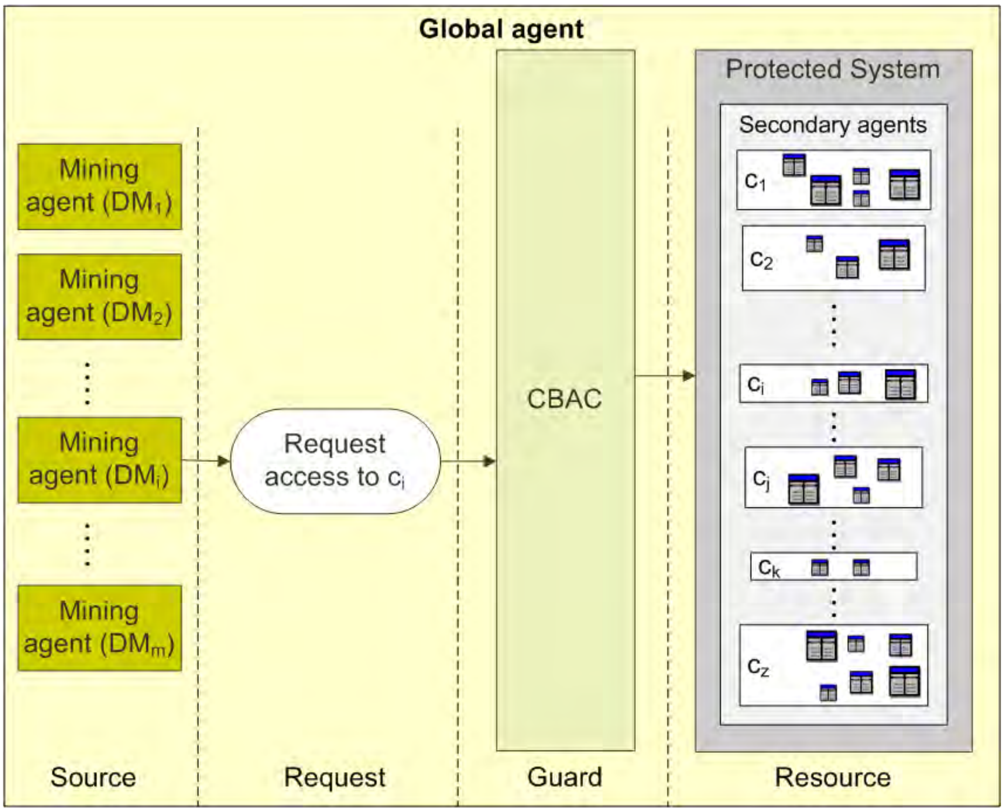


Figure 5.4: CBAC in terms of a general access control model

With the case study in mind (Chapter 2), the relationships in this conceptual framework are as follows:

- The global agent, DM Bank, conducts data mining as part of its business intelligence procedures. The data-mining process results in DM Bank having one or more active mining agents that are doing data mining for DM Bank.
- Mining agents can work on their own or in groups, depending on the stage of the data-mining process and the size of the project.
- A client of DM Bank (the global agent) is a secondary agent. A global agent can thus have many secondary agents and these secondary agents are the resources on whose data sets data mining is done.
- Some secondary agents (e.g. two different petrol companies) may be in conflict with one another because they conduct business in the same functional business domain.

- For purposes of this research, if a mining agent requests access to a specific data set of a specific secondary agent c_i, the mining agent will be granted (or denied) access to the secondary agent c_i with all its data sets.

The CBAC model, as seen in the operational environment of the global agent, is depicted in Figure 5.5.

When a *mining agent* (‘principal’) *requests access* (‘do operation’) to the *data sets of a secondary agent* (‘protected system’), the *CBAC model* (‘reference monitor’) validates whether the request should be granted or denied. This validation is done according to the rules defined by the CBAC model as will be explained in Chapter 6.

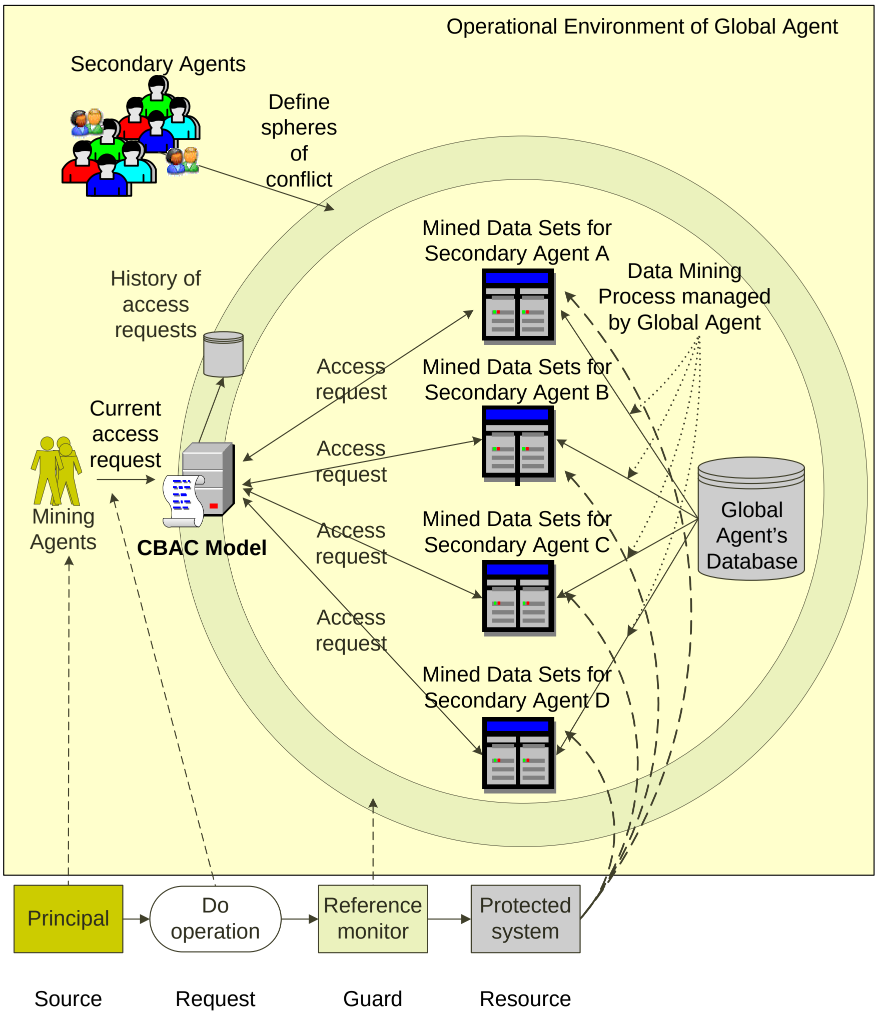

Operational Environment of Global Agent
Secondary Agents
Define spheres of conflict
Mined Data Sets for Secondary Agent A
Data Mining Process managed by Global Agent
History of access requests
Access request
Mined Data Sets for Secondary Agent B
Current access request
Access request
Global Agent's Database
Mining Agents
CBAC Model
Access request
Mined Data Sets for Secondary Agent C
Access request
Mined Data Sets for Secondary Agent D
Principal
Do operation
Reference monitor
Protected system
Source
Request
Guard
Resource

Figure 5.5: CBAC demonstrated in the operational environment of the global agent

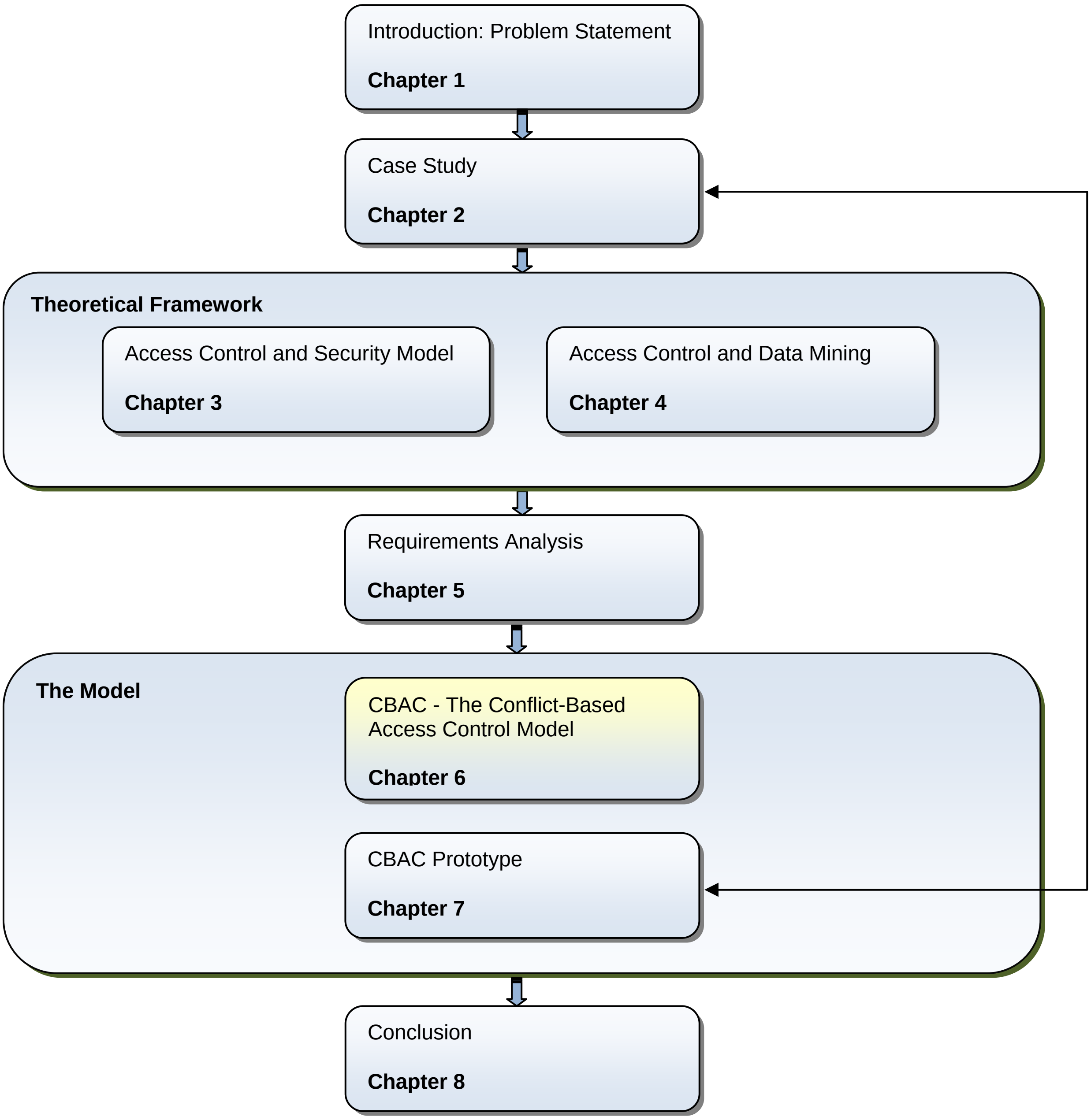
5.5 CONCLUSION

The conclusions reached in Chapters 3 and 4 were combined to obtain a final list of requirements that are needed for the proposed conflict-based access control model. The CBAC model was then depicted in a conceptual framework to explain its relationship with the fundamental access control model and the data-mining environment.

In the next chapter the proposed CBAC model will be explained in detail.

CHAPTER 6: THE CONFLICT-BASED ACCESS CONTROL MODEL

THESIS MAP



6.1 INTRODUCTION

This chapter sets out to explain an access control model that meets the access control requirements of conflicting agents. A set of requirements and a conceptual framework (both of which had already been established) were used as the foundation for constructing the proposed model. This construction process resulted in the requirements being refined until a final set of requirements for the CBAC model was established.

The **C**onflict-**B**ased **A**ccess **C**ontrol model introduced and defined in this chapter, henceforth called the CBAC model, provides an access control model for data-mining environments. The aim of this chapter is to explain the CBAC model in detail by setting out its final set of requirements, explaining its basic elements and explaining all operational elements of the model. A typical access request is discussed subsequently by making use of the principles of the model and the chapter is concluded with a discussion of the advantages of the CBAC model.

6.2 THE CONFLICT-BASED ACCESS CONTROL (CBAC) MODEL

6.2.1 Requirements for the CBAC model

Initial requirements were set for the CBAC model [Loock & Eloff, 2005a] and after subsequent research and discussions with industry, the requirements were adapted to the point where the following were applied in the development of the model:

- The proposed model was based on the idea of conflict-of-interest classes.
- A secondary agent within this proposed model was compelled to belong to one or more functional business domain(s).
- The model had to provide for a degree of conflict to determine whether a potential conflict would influence an access control decision.
- A secondary company was allowed to have a dedicated mining agent assigned to it.
- Secondary companies had to determine the conflict levels between one another on a predefined scale and within limits.

- The model had to be able to predict conflict over and above the conflict defined by the secondary agents.
- The proposed model had to minimise the global agent's reputation risk by enabling the latter to manage potential confidentiality leaks more effectively.
- The concept of a history record is important. A record should be kept of all the data sets that a mining agent is currently working on, as well as those accessed in the past.
- A secondary company may have an interest in a functional business domain for a variety of reasons. A secondary company may also have different levels of conflict with regard to different functional business domains.

The objective of the CBAC model is to determine whether a mining agent (subject) may obtain access to a data-mining object. The type of access right required by the data-mining agent could vary and may for example include doing a decision tree analysis on an object or on a set of objects. The CBAC model as proposed in this research focuses on modelling the relationship(s) between data-mining agents and objects (data sets) and, for reasons of complexity, does not differentiate between the types of access.

6.2.2 Basic elements of the CBAC model

The CBAC model implements an access control policy that regulates data-mining activities between *data-mining agents* working for a *global agent*, *secondary agents* to whom the global agent renders a service, and the '*data sets*' of the secondary agents. Stated differently, a *global agent* is regarded as the company that instructs the data-mining activity to be conducted, while *secondary agents* are those companies whose data sets are exposed to data-mining activities as instructed by the global agent.

The remainder of this paragraph presents the definitions of the elements of the CBAC model. This is followed by (if applicable) a graphical illustration and a brief reference to the case study as discussed in Chapter 2.

6.2.2.1 Secondary agent

We shall use symbol c_i to denote both a secondary agent and the data sets of this particular secondary agent.

6.2.2.2 Global agent

Let D denote the global agent.

The set of secondary agents are in a specific business relationship with the global agent for example they are clients (c_i) of a bank (D).

Figure 6.1 depicts global agent D:

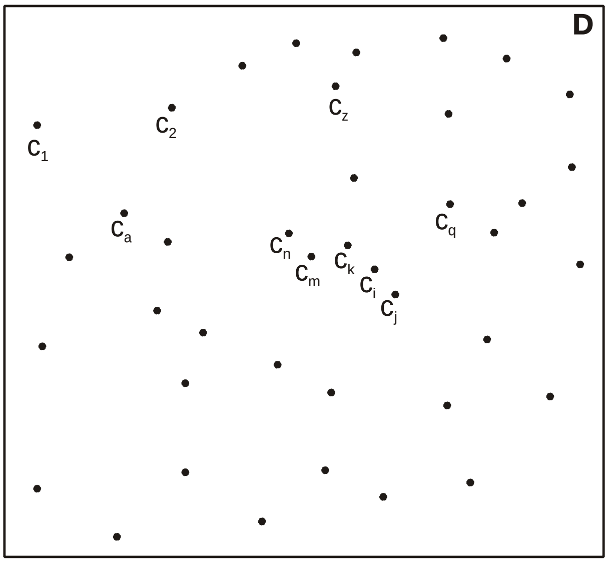


Figure 6.1: All secondary agents (c_i) associated with a global agent (D)

Case study:

The CBAC model focuses on minimising the potential security risks created when a mining agent who works for the global agent, DM Bank, performs data-mining activities on the objects (data sets) of two or more secondary agents who might, or might not be, conflicting clients – such as Green Petrol and ABC Petrol.

6.2.2.3 Conflict-of-interest field

A conflict-of-interest field is the set of secondary agents that are in conflict with one another.

Let CNC denote all possible secondary agents that belong to D.

CNC = $\{c_1, c_2, \ldots, c_z\}$,

where:

c_i is a secondary agent and

z is the total number of secondary agents associated with the global agent (D).

Let C be the set of secondary agents called the *conflict-of-interest field* of all possible secondary agents CNC.

C = $\{c_{p\_1}, c_{p\_2}, c_{p\_3}, \ldots, c_{p\_k}, \ldots, c_{p\_q}\} \subset \{c_1, c_2, \ldots, c_z\}$ = CNC

where:

c_p are the secondary agents belonging to C and

$1 \leq q \leq z$ where:

q is the total number of secondary agents in C ⊂ CNC and claiming to be in conflict with one or more other secondary agents in C, or where any other secondary agent of the global agent claims to be in conflict with these secondary agents in C (in which case this other secondary agent also belongs to C), or where the global agent claims that these secondary agents in C are in conflict with one another.

Case study:

The secondary agents ABC Petrol and Green Petrol are elements of C, because they both claim to be in conflict with another secondary agent – in this case with each other.

6.2.2.4 Non-conflict-of-interest field

Let NC represent the set of secondary agents called the non-conflict-of-interest field of global agent D.

NC = CNC – C = $\{c_i \,|\, c_i \in CNC, c_i \notin C\}$

Note:

The elements of NC are all those secondary agents who are not part of any conflict claim argued by any secondary agent of the global agent or by the global agent D itself.

Figure 6.2 depicts the global agent D with the two subsets C and NC.

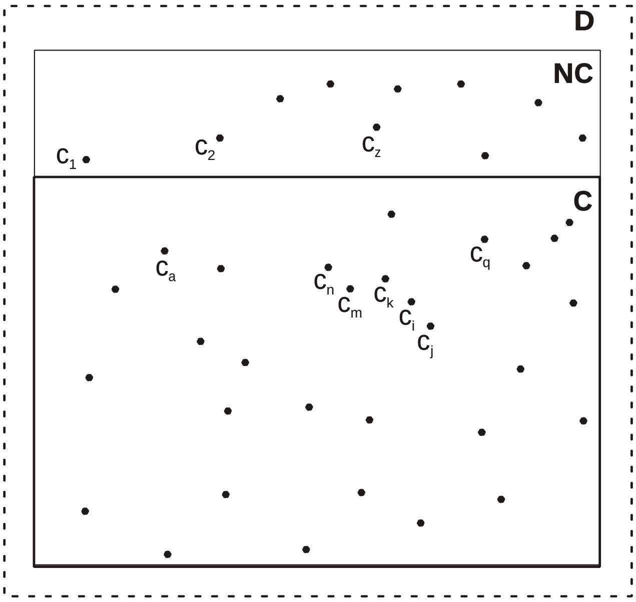


Figure 6.2: The conflict-of-interest set of secondary agents (C) and the set of non-conflicting secondary agents (NC) of global agent D

6.2.2.5 MiningAgent$_x$

Let MiningAgent$_x$ denote a data-mining agent / user / person / group who conducts data-mining activities on behalf of the global agent D.

6.2.2.6 *Session*

A session is defined as the time interval during which no mining agent's history record needs to be deleted. Thus it is the total time of interaction between a mining agent and the data sets on which the agent is working during a specific time interval.

6.2.2.7 *Secondary agents accessed by mining agents*

Let NS be the set of all secondary agents' data sets to which all the mining agents who perform data mining on behalf of the global agent D, was already granted access to.

$NS = \cup_x NS_x$

where:

NS_x is the set of secondary agents to which a specific $MiningAgent_x$ was already granted access during the running session:

$NS_x = \{c_{r\_1}, c_{r\_2}, c_{r\_3}, \ldots, c_{r\_p}\}$

where:

$c_r \in C$ and r is a specific secondary agent's data sets and

p is the number of the secondary agents' data sets to which $MiningAgent_x$ has access to.

x = 1 … n, where n is the total number of mining agents available to the global agent.

NS_x represents a *history record* for $MiningAgent_x$.

6.2.2.8 *Functional business domain*

A functional business domain is a domain in which companies conduct the same type of business, such as airline companies that all belong to the airline domain. This can be formulated as:

Let a_i denote a functional business domain or a particular market segment such as the airline business.

Here $1 \leq i \leq j$ and j is the total number of functional business domains of all the secondary agents serviced by global agent D, constituting the set

A = {a_1, …, a_i, …, a_j}

Case study:

Consider, for example, the global agent D for which A = {Food, Petrol, Shoes, Airlines} and m = 4.

6.2.2.9 Associations between secondary agents and functional business domains

Let the fact that secondary agent c_i has an interest in functional business domain a_h be represented by e(c_i, a_h).

There are two ways to construct the relationship between c_i and a_h namely:

i Each c_i associates itself with an a_h, for example e(c_i, a_h)

ii For each a_h consider all c_i's that fall in that functional business domain, for example e(a_h, {c_i, c_j, ...}), also

For each c_i consider all a_h's that are relevant to that secondary agent, for example e(c_i, {a_h, a_g, ...})

Note:

- For a specific e(c_i, a_h), let w(c_i, a_h) denote the weight of conflict of interest of secondary agent c_i in functional business domain a_h. The weight function gets its value from the distance function (section 6.2.3.4) indicating the degree of conflict of interest between a secondary agent $c_i \in$ C and any other conflicting secondary agent $c_j \in$ C that belongs to the functional business domain a_h, as defined by c_i or c_j.
- w(c_i, a_h) is a number in N ∪ { ∞ } = {1, 2, …, ∞}.
- w(a_h, c_i) = 0 ∀ h and i. The weight function does not indicate a weight of conflict of interest between a functional business domain a_h and a secondary agent c_i. Only secondary agents can initiate conflicts of interest.
- The smaller the number w(c_i, a_h), the stronger the interest of secondary agent c_i in functional business domain a_h; thus w(c_i, a_h) can be seen as a 'distance' from c_i to a_h. When a_h is the main functional business domain of c_i, we define w(c_i, a_h) = 1. A value ∞ indicates no interest at all.

- It is possible for a secondary agent of the global agent ($c_i \in$ CNC) to be associated with one or more functional business domains, $a_h \in$ A.

Case study:

The ABC Petrol company belongs to the Petrol functional business domain because of its main business activity, but also to the Airlines functional business domain because of the majority of shares that it has in the FlySave Airline company.

6.2.3 Operational elements of the CBAC model

The functioning of the CBAC model is explained by means of the following operational elements:

- Determine an *access group list (*see 6.2.3.1). For each secondary agent, a list is determined that records all other secondary agents in which the original secondary agent has a business interest.
- Identify conflict of interest between secondary agents, as identified by themselves. The identification of a conflict of interest by one secondary agent – between himself and all the other secondary agents – is called a *sphere of conflict* (6.2.3.2) for that specific secondary agent.
- Determine a *cut-off point for conflict* (see 6.2.3.3) beyond which conflict has no impact on access control decisions.
- Determine the degree of conflict between secondary agents – this is referred to as the *distance* (6.2.3.4) parameter.
- Determine if there is any potential conflict, i.e. conflict not identified by the secondary agents themselves. This is referred to as the *path* (6.2.3.6) parameter.

6.2.3.1 Access group list for secondary agent c_i

Let AG-List$_i$ be the access group list for c_i.

Note:

AG-List$_i$ $\subset$ D

c_i defines its own AG-List$_i$

AG-List$_i$ = {c_j, c_k, … c_s} is a list of all those secondary agents in which secondary agent c_i has a specific business relationship namely a positive relationship or business interest, for example secondary agent c_i owns 60% of secondary agent c_j. This is a business relationship for which there exists no conflict with c_i. However, this interest causes a conflict between c_i and all the other secondary agents that are in the same functional business domain as any agent in AG-List$_i$.

6.2.3.2 Sphere of conflict – conflict of interest

Let S_i be the *sphere of conflict* for secondary agent c_i

Note:

i. $S_i \subset CNC$: $S_i \subset C \subset CNC$ (S_i is either part of C) or $S_i = \{c_i\}$

and $c_i \notin S_j$ for any $j \neq i$,

in which case $S_i \subset NC \subset CNC$ (S_i is part of NC).

ii. Each c_i defines its own S_i. In the following instance, $S_i = \{c_i, c_j, c_k, c_m, c_n\}$, it is depicted that c_j, c_k, c_m and c_n are identified by c_i as conflicting secondary agents. The *sphere-of-conflict* definition [Loock & Eloff, 2005b] explains the concept of conflict between secondary agents. A secondary agent c_i might be in conflict with another secondary agent c_y because c_i and c_y are in the same functional business domain. Any other reason could also apply, for example secondary agent c_i might be in conflict with another secondary agent c_x because c_i is having a substantial interest in a company c_z, which is in the same functional business domain as c_x.

iii. There may be a c_i ($1 \leq i \leq q$) where $q \leq z$, which chooses not to define an S_i; thus $S_i = \{c_i\}$. However, this does not exclude the fact that c_i can be included in another S_k ($1 \leq k \leq q$) as defined by c_k.

6.2.3.3 Cut-off point for conflict of interest

Let r_i be the cut-off point for conflict, beyond which conflict is to be regarded by secondary agent c_i as insignificant. r_i is defined as the radius of the sphere of conflict S_i defined by c_i :

rad $(S_i) = r_i \geq 0$

where r_i is a natural number ($N \cup \{0, \infty\}$)

Note:

No range of possible values for r_i is pre-specified. During implementation time of the CBAC model these ranges will be specified, for example:

- S_1: $r_1 = (0,3)$, which means all secondary agents with a conflict larger than 3 with secondary agent c_1 will be seen as not in conflict with secondary agent c_1.

This example can be shown for all $S_i \subset NC$.

The *sphere of conflict* and its cut-off point for conflict (radius) are depicted in Figure 6.3:

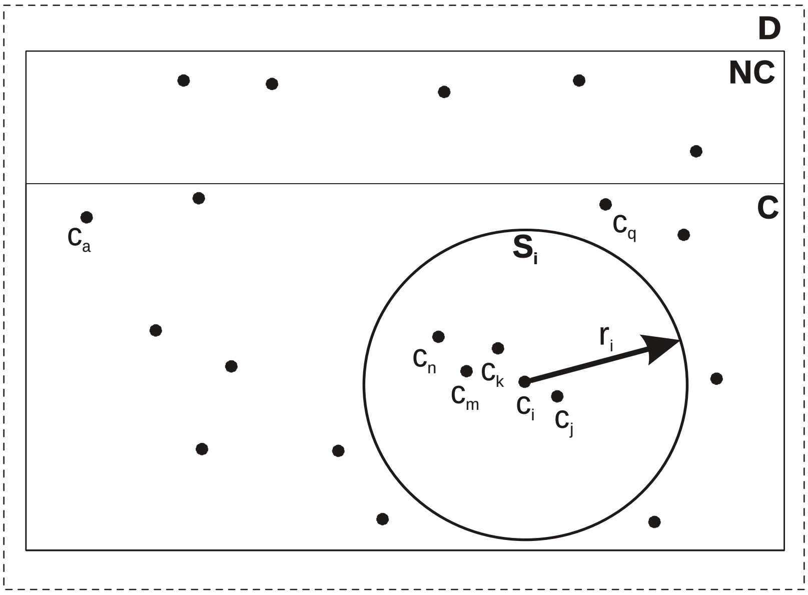


Figure 6.3: r_i is the radius for S_i defined around c_i

6.2.3.4 Distance – degree of conflict of interest

Let $d_{i,j}$ denote the distance between c_i and c_j

where $d_{i,j}$: C x C $\rightarrow$ N $\cup \{0, \infty\}$. Thus d_{ij} is a number in $\{0, 1, 2, \ldots, \infty\}$

$(c_i, c_j) \rightarrow d_{i,j} (c_i, c_j)$

$d_{i,j} (c_i, c_j) \geq 0$ for all $c_i, c_j \in C$

$d_{i,j} (c_i, c_j) = 0$ if and only if $c_i = c_{j.}$

Note:

- $d_{i,j}$ is a function indicating the degree of conflict between a secondary agent $c_i \in C$ and any other conflicting secondary agent $c_j \in C$, as defined by c_i or c_j. Note that this only applies to agents in C and not to all $c_i \in$ CNC.
- The chosen value of $d_{i,j}$ is determined by c_i. c_i determines the value of $d_{i,j}$ based on the degree of conflict between c_i and c_j, as determined by c_i. However, care should be taken in deciding on the value of $d_{i,j}$. If this value is too lowhigh, it can result in unwanted releases of information and if it is too highlow, it can result in a delay in services.
- $d_{i,j} = \infty$ between c_i and c_j implies an infinite distance (c_i did not determine a value for $d_{i,j}$), which means no conflict of interest exists between c_i and c_j.

It is important to highlight the difference between cut-off point for conflict (*radius*) and *distance*. The radius r_i defined by c_i denotes a cut-off point beyond which conflict is considered as insignificant between c_i and any c_j. In case $r_i < d_{ij}(c_i, c_j)$, the degree of conflict between c_i and c_j is considered as insignificant and should not influence any access decision.

Case study:

In the example depicted in Figure 6.4, a $d_{\text{ABC Petrol, SelectShoe}} > 3 = r_{\text{ABC}}$ refers to a non-existent conflict, such as with the SelectShoe company, where $c_{\text{SelectShoe}} \notin S_{\text{ABC Petrol}}$. The radius r_i (in this example $r_{\text{ABC Petrol}}$), which represents the ‘non-existing conflict’ cut-off, is determinable by c_i (in this example, ABC Petrol).

Consider the following example by also referring to Figure 6.4. The ABC Petrol company defines a sphere-of-conflict field, $S_{\text{ABC Petrol}}$, around itself. It starts with secondary agents such as P+P Petrol company and Green Petrol company defined at a distance of 1. This means $d_{\text{ABC Petrol, P+P Petrol}} = d_{\text{ABC Petrol, Green Petrol}} = 1$, which implies a high degree of conflict because they are in direct conflict with ABC Petrol.

Other conflicting secondary agents that ABC Petrol defines in this ‘sphere of conflict’ ($S_{\text{ABC Petrol}}$) are the HighFly Airline company, with $d_{\text{ABC Petrol, HighFly Airline}} = 2$ and the Pick&Save Food company, with $d_{\text{ABC Petrol, Pick\&Save Food}} = 3$. The latter two secondary agents are in the sphere-of-conflict field $S_{\text{ABC Petrol}}$ for the following reasons:

ABC Petrol has the majority of shares in FlySave Airline, which is in conflict with HighFly Airline. Ownership of the majority of shares in FlySave Airline makes the conflict between ABC Petrol and HighFly Airline severe (for this example, $d_{\text{ABC Petrol, HighFly Airline}} = 2$). ABC Petrol further holds some shares in QuickPay Food, which is clearly in conflict with Pick&Save Food. Ownership of some shares in QuickPay Food makes the conflict between ABC Petrol and Pick&Save Food a matter that needs attention (for this example, $d_{\text{ABC Petrol, Pick\&Save Food}} = 3$). ABC Petrol further decided that $r_{\text{ABC Petrol}} = 3$ and any secondary agent that exists at a distance with $d_{\text{ABC,c\_i}} > 3$ is not in conflict with ABC Petrol and need not be part of ABC Petrol’s sphere of conflict, $S_{\text{ABC Petrol.}}$

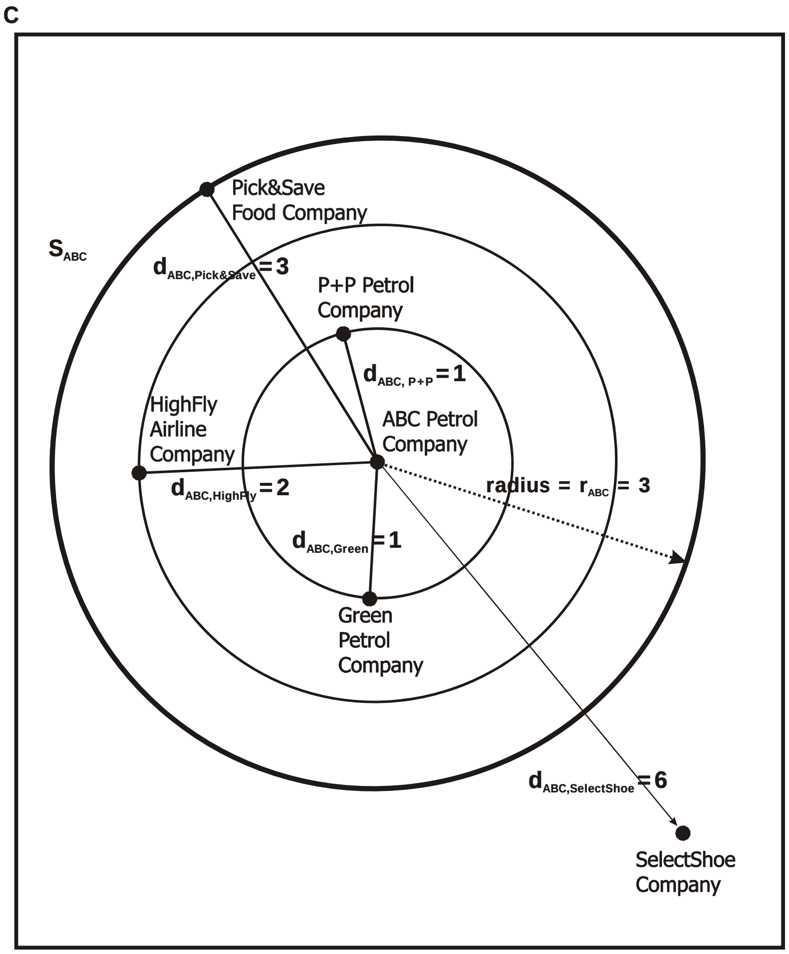


Figure 6.4: Sphere of conflict around ABC Petrol company

Sphere of conflict, cut-off point (radius) and *distance* focus on conflict identified by the secondary agents themselves, i.e. representing the secondary agent’s view of the world. It is also important to assess whether there is any other potential conflict over and above the conflicts as identified by the secondary agents. This potential conflict of interest (PIT) is evaluated by the global agent and is addressed in the next section.

6.2.3.5 Potential Conflict of Interest (PIT)

Let G = (V, E) represent a mechanism to assist in the assessment of potential conflict of interest between secondary agents
where:

(1) G is a bipartite graph consisting of the following two sets:

(i) a set V = V(G) whose elements (finite set of nodes) are called vertices of G

(ii) a set E = E(G) of ordered pairs (may be labelled with 'e' for 'edge') of distinct vertices called edges (lines between nodes) of G

(2) V, the vertices of G, is partitioned into two subsets A and D such that each edge of G connects a vertex of A to a vertex of D, i.e. the lines only exist between a node in A and a node in D.

(3) G is a directed graph where the edges between vertices go in a specific direction, i.e. e(c_i, a_h) is different from e(a_h, c_i). For the sake of simplicity, e(c_i, a_h) refers to the upward direction whereas e(a_h, c_i) is the downward direction. (See Figure 6.5 below.)

(4) G is a weighted graph meaning that each edge e $\in$ E of G is assigned a non-negative number w(e) called the weight of e [Lipschutz & Lipson, 2003].

Furthermore:

(5) The subset A of V represents all the functional business domains a_h of the global agent (see 6.2.3.5 (1, 2) and Figure 6.5).

(6) The subset D of V represents all the secondary agents c_i of the global agent (see 6.2.3.5 (1, 2) and Figure 6.5).

(7) The set of edges E represents the associations between secondary agents and functional business domains. Thus e_i represents the association from c_i to a_h and is represented by e(c_i, a_h) (see 6.2.3.5 (1) and Figure 6.5).

(8) The weight of the edge e(c_i, a_h) is denoted by w(c_i, a_h) and represents the strength of the associations between secondary agents (subset D of V) and functional business domains (subset A of V). (See 6.2.3.5 (2, 4) and Figure 6.5.)

(9) w(c_i, a_h) = $d_{i,j}$ where c_i in subset D is related to c_j in subset D via functional business domain a_h in subset A. (See 6.2.3.5 (2, 4) and Figure 6.5.)

(10) The weight of the edge e(a_h, c_i) is denoted by w(a_h, c_i) and represents the strength of the associations between functional business domains (subset A of V) and secondary agents (subset D of V). This weight will always be 0 because the representing edge, e(a_h ,c_i) is directed from a functional business domain to a secondary agent thus w(a_h, c_i) = 0 is the direction from a_h in subset A to c_i in subset D. (See 6.2.3.5 (3, 4) and Figure 6.5.)

(11) w(c_i, a_h) > 0 where the edge e_i links c_i in subset D to a_h in subset A. (See 6.2.3.5 (3, 4) and Figure 6.5.)

Note:

G is used to detect all possible relationships (paths) between all secondary agents via the functional business domains in A.

6.2.3.6 Potential Conflict of Interest (PIT) – Path

Let ~ denote a path between any two secondary agents from different functional business domains c_a and $c_{x,}$ represented as $c_a \sim c_x$.
where:

(1) $c_a \sim c_x$ means at least one path exists between c_a and c_x

(2) $c_a \sim c_x$ implies a sequence $c_a\, a_x\, c_p\, a_y \ldots a_z\, c_x$, where every pair $c_u a_v$ or $a_v c_w$ is an edge in E(G).

Note:

- There can be multiple paths between any two secondary agents.

- The *path* definition is used to determine potential conflict. This potential conflict can only be noticed by the global agent who is able to consider all paths between all agents, as opposed to direct conflict that can be identified by the secondary agents themselves.

- *Path* establishes whether a potential conflict of interest exists between any two secondary agents at a given point in time and not necessarily in the same $S_{x.}$

- The existence of a path implies that a mining agent can obtain access to the data sets of two secondary agents, namely, $c_a \notin S_x$ and $c_x \notin S_a$. There may be a potential conflict of interest between these two secondary agents at both ends of the path. This potential conflict of interest (which was not foreseen by c_a and c_x themselves) can be determined by the global agent with the help of the *path* parameter.

6.2.3.7 Potential Conflict of Interest (PIT) – The k-th path

All possible paths between c_a and c_x must be established to be able to calculate the shortest path between c_a and c_x. The shortest path between c_a and c_x is the path that best represents the degree of potential conflict between c_a and c_x.

Let P_k (c_a, c_x) denote the k-th path between c_a and c_x
where:
k is any path of all possible paths between c_i and c_j and
where:
P_k (c_a, c_x), between c_a and c_x is a sequence $c_a\, a_x\, c_p\, a_y \dots a_z\, c_x$, where every pair $c_u a_v$ or $a_v c_w$ is an edge in E(G).

6.2.3.8 Potential Conflict of Interest (PIT) – The weight of a path

Let wP_k (c_a, c_x) denote the weight of path P_k (c_a, c_x)
wP_k $(c_a, c_x) = \sum w(c_s, a_t)$, where $\sum$ goes over all pairs (c_s, a_t) and (a_t, c_s) along the path P_k (c_a, c_x) thus wP_k (c_a, c_x) is a number in $\{0, 1, 2, \dots, \infty\}$
where:

(1) w is explained in section 6.2.3.5

(2) k is any path of all possible paths between c_a and c_x

(3) c_s is any secondary agent on any path between c_a and c_x.

c_s, c_a and $c_x \in$ D and s, a and x are numbers in {0, 1, 2, ..., ∞}

(4) a_t is any functional business domain on any path between c_a and c_x.

$a_t \in$ A and t is a number in {0, 1, 2, ..., ∞}

(5) $wP_k (c_a, c_x) = \infty$ when the calculated path becomes so long that the assumption can be made that there is no path between two secondary agents.

Note:

- In the same manner that r_a is used as the cut-off point for conflict for secondary agent c_a, a cut-off point of the path weight will be established by the global agent. The smaller the cut-off point of path weight, the more conflict will exist between the two secondary agents c_a and c_x. The efficiency of this cut-off point will be established by the global agent over time, based on internal business risk analysis by the global agent himself. The path weight is based on the distance parameters that the secondary agents established over time, in other words on the business risk that the secondary agents were prepared to take when establishing the conflicts between themselves and the other secondary agents. The global agent should calculate its business risk on the different mining agents that is doing data mining on behalf of the global agent, when establishing a cut-off point for the path weight.

- For illustration purposes, the cut-off point of the path weight can be defined as 5.

$wP_k (c_a, c_x)$ > 5 : weight of the path too high to be considered as a conflicting path

≤ 5 : weight of the path indicates a conflict

$= \infty$: indicates the calculated path became so long that the assumption can be made that there exists no path, thus no conflict

6.2.3.9 Potential Conflict of Interest (PIT) – Shortest path

Because there is a possibility of multiple paths between any two secondary agents, the shortest path between c_a and c_x is the path that best represents the degree of potential conflict between c_a and c_x.

For illustration purposes, consider path P_k (c_a, c_x) (k=1,2) as depicted in Figure 6.5. Figure 6.5 shows two different paths between two secondary agents:

P_1 (c_a, c_x) = c_a a_k c_x and P_2 (c_a, c_x) = c_a a_m c_e a_n c_x

Note:

It is important to introduce a mechanism for determining which path, out of multiple possible paths, should be regarded as the shortest path, because the shortest path between c_a and c_x is the path that best represents the degree of potential conflict between c_a and c_x.

Calculation of Shortest path between c_a and c_x.

(1) Calculate all possible P's that exist between c_a and c_x

P_k (c_a, c_x) (k=1, 2, ...)

(2) Calculate the weight, wP of all possible P's between c_a and c_x

wP_k (c_a, c_x) (k=1, 2, ...)

(3) Determine the smallest wP, which is the shortest path between c_a and c_x

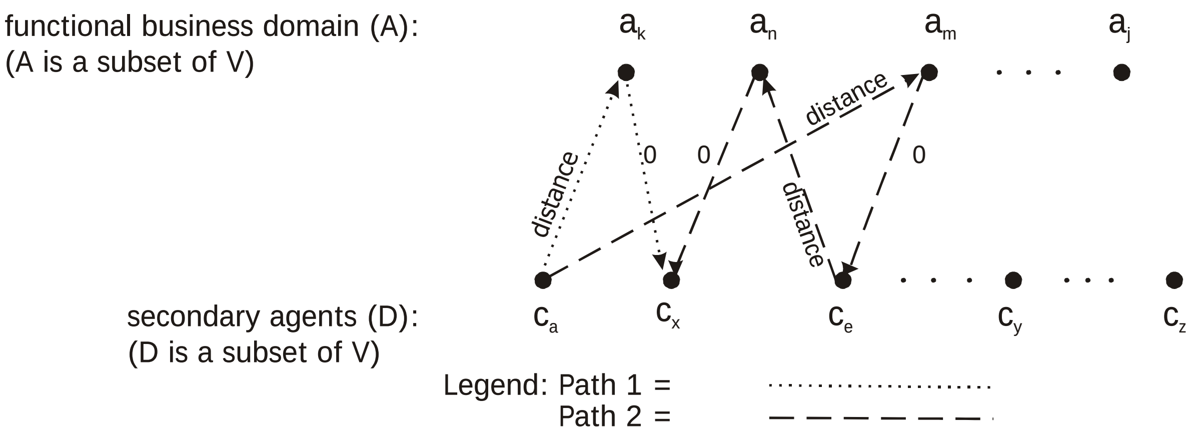


Figure 6.5: Two different paths from c_a to c_x

Case study:

Consider the example as depicted in Figure 6.6. ABC Petrol has a strong interest in the Petrol functional business domain, because it is a petrol company itself. Thus, w(ABC Petrol, Petrol) = 1. ABC Petrol is in conflict with Green Petrol and P+P Petrol according to its own definition of its sphere of conflict $S_{ABC\ Petrol}$. It is also clear from Figure 6.6 that ABC Petrol is in potential conflict with Fast Petrol, although the latter was not identified by ABC Petrol as being in its sphere of conflict. ABC Petrol furthermore has a strong interest in the Airlines functional business domain, because it holds many shares in FlySave Airline company. Thus, w(ABC Petrol, Airlines) = 2 = $d_{ABC,HighFly}$ (Figure 6.4). ABC Petrol frankly stated that it is in conflict with HighFly Airline company, but from Figure 6.6 it is clear that ABC Petrol is also in potential conflict with BlueSky Airline, which was not included in ABC Petrol's initial definition of its sphere of conflict. In addition, Figure 6.6 shows that ABC Petrol has an interest in the Food functional business domain, because it has shares in QuickPay Food. Thus, w(ABC Petrol, Food) = 3 = $d_{ABC,Pick\&Save}$ (Figure 6.4). ABC Petrol defined a conflict with Pick&Save Food, but Figure 6.6 shows that ABC Petrol is also in potential conflict with EasyEat Food, which was not defined by ABC Petrol to be in S_{ABC}.

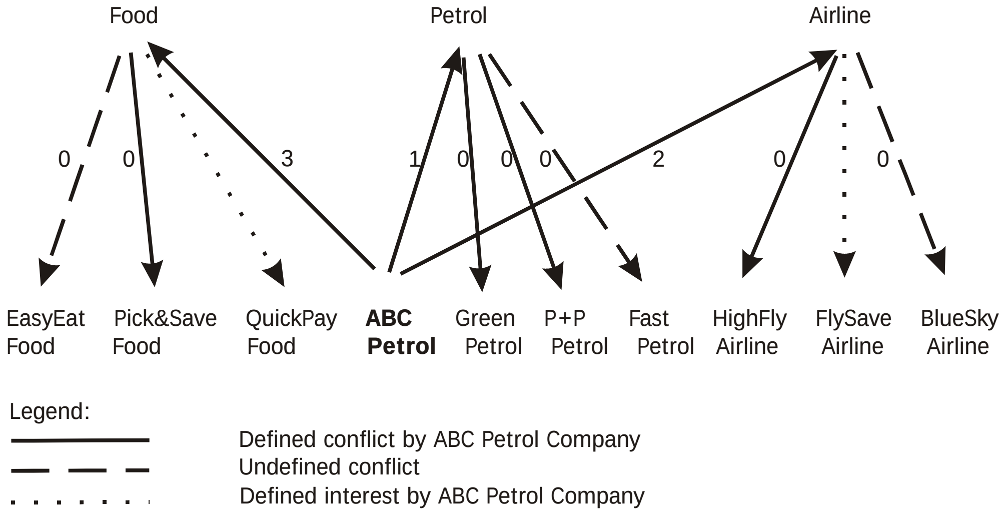


Figure 6.6: Possible paths from ABC Petrol company to three undefined companies

6.3 ACCESS REQUEST

Let MiningAgent_x (NS_x, c_p) represent an access request
where:

(1) MiningAgent_x is the data miner issuing the access request.

(2) NS_x is a set of non-conflicting secondary agents representing all data sets of secondary agents to which MiningAgent_x already had access prior to the current request:

$NS_x = \{c_{\text{MiningAgent\_x\_1}}, c_{\text{MiningAgent\_x\_2}}, c_{\text{MiningAgent\_x\_3}}, \ldots, c_{\text{MiningAgent\_x\_p}}\} \subset CNC$

(3) c_p is the data set to which MiningAgent_x is requesting access.

6.3.1 Grant / reject an access request

Figure 6.7 depicts the decision-making process regarding access requests. All access requests are between mining agents and data sets of secondary agents.

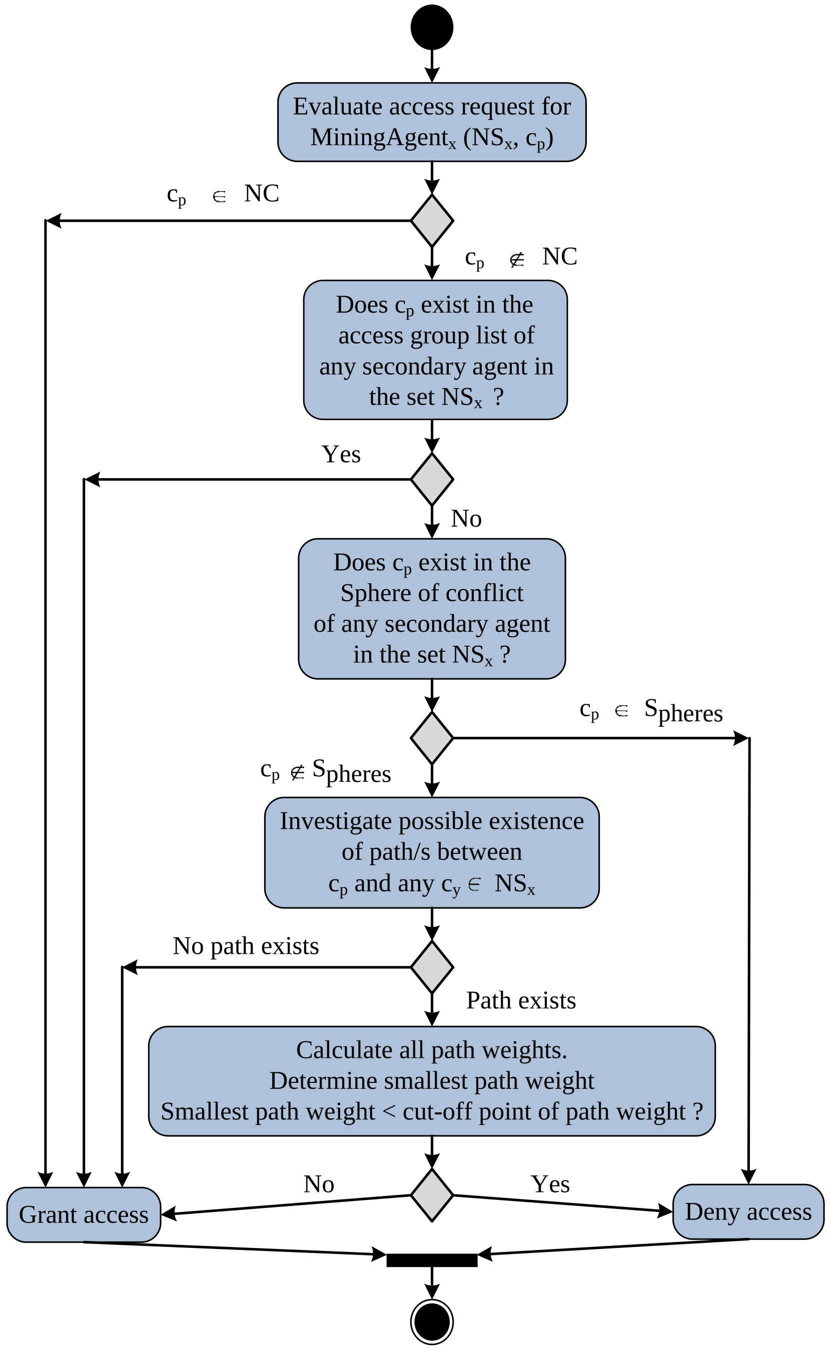

Evaluate access request for MiningAgent$_x$ (NS$_x$, c$_p$)
c$_p$ ∈ NC
c$_p$ ∉ NC
Does c$_p$ exist in the access group list of any secondary agent in the set NS$_x$?
Yes
No
Does c$_p$ exist in the Sphere of conflict of any secondary agent in the set NS$_x$?
c$_p$ ∈ Spheres
c$_p$ ∉ Spheres
Investigate possible existence of path/s between c$_p$ and any c$_y$ ∈ NS$_x$
No path exists
Path exists
Calculate all path weights. Determine smallest path weight Smallest path weight < cut-off point of path weight ?
No
Yes
Grant access
Deny access

Figure 6.7: High-level logic of an access request – MiningAgent$_x$ wants access to data sets of secondary agent c$_p$

The first step is to test whether c_p is an element of all the non-conflicting secondary agents.

Is $c_p \in NC$?

If the answer is yes,

then access is granted.

If the answer is no,

the second step is to test whether c_p exists in the access group list of any secondary agent in the set NS_x.

If the answer is yes,

then access is granted.

If the answer is no,

the third step is to test whether c_p is in the sphere of conflict of any secondary agent in the set NS_x.

If the answer is yes, which means that

$c_p \in S_{pheres} = \{S_{MiningAgent\_x\_1}, S_{MiningAgent\_x\_2}, S_{MiningAgent\_x\_3}, \ldots, S_{MiningAgent\_x\_p}\}$,

then access is rejected.

If the answer is no, which means that

$c_p \notin S_{pheres} = \{S_{MiningAgent\_x\_1}, S_{MiningAgent\_x\_2}, S_{MiningAgent\_x\_3}, \ldots, S_{MiningAgent\_x\_p}\}$,

then

the third step is to test whether a path exists or paths exit between c_p and any secondary agent c_y in the set NS_x.

If the answer is no,

then $MiningAgent_x$ is granted access, but

if the answer is yes,

the fourth step is to calculate the weight of all possible paths and determine the shortest path which is the smallest number.

If the weight of the shortest path is small enough to be defined as *'a conflict does exist'*,

then no access can be granted to $MiningAgent_x$.

If, however, the weight of the shortest path is big enough to be defined as *'no conflict exists'*,

then access is granted to $MiningAgent_x$.

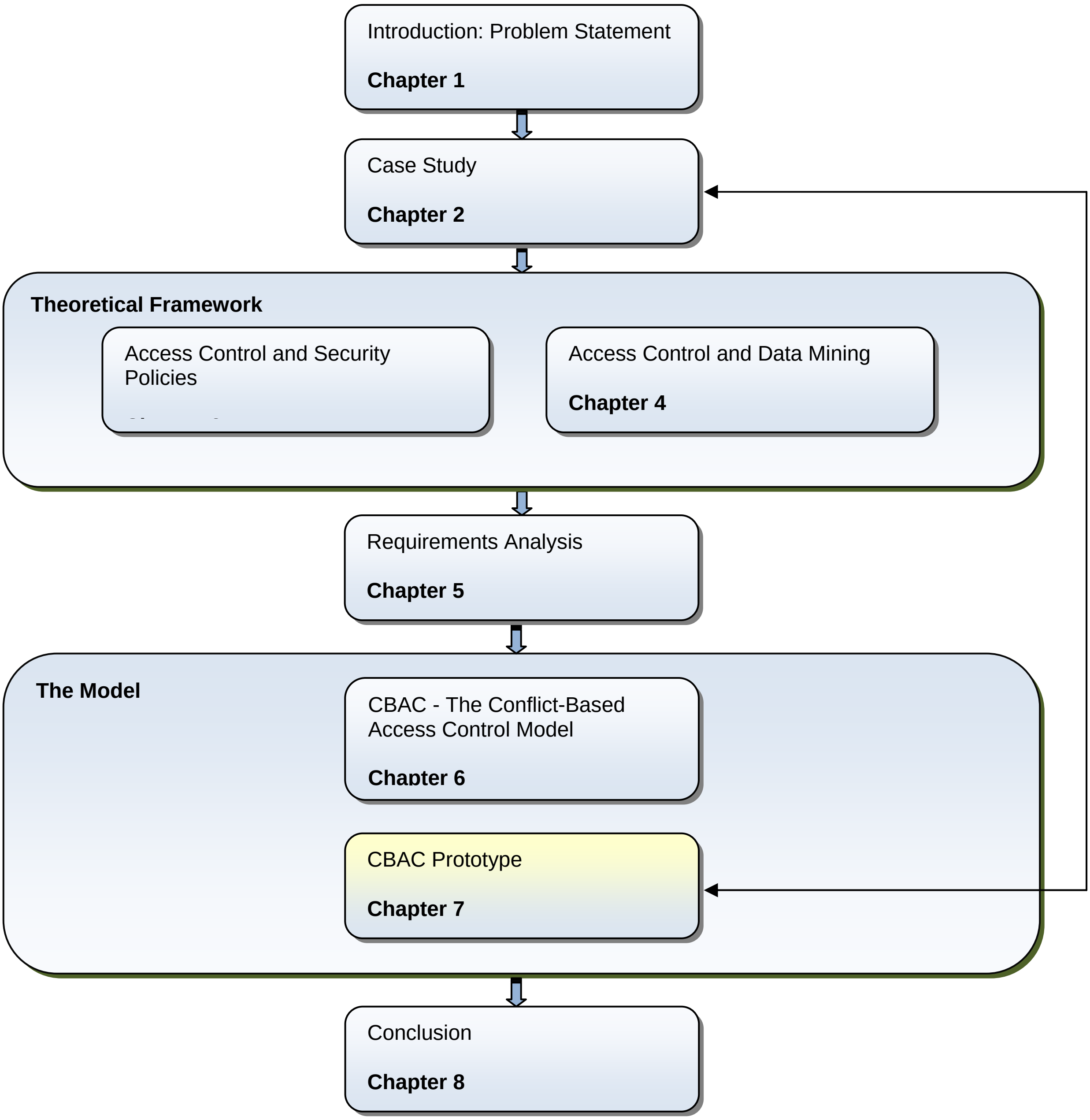
The implementation of this logic is explained in greater detail and enhanced with more figures in the chapter that follows.

6.4 CONCLUSION

This thesis proposes a new access control model called the Conflict-Based Access Control (CBAC) model, which deals with a changing business environment, i.e. a data-mining environment [Loock & Eloff, 2005a]. The CBAC model supports the definition of *conflict-of-interest classes* and adds a *sphere of conflict* definable by the secondary agents themselves, rather than by functional business domains. The model works on the assumption that conflict-of-interest classes are not 'disjoint' or separate. The proposed model keeps a history of access requests for each mining agent and allows the *degree of conflict* between secondary agents to be quantifiable. The CBAC model is *dynamic* and thus able to cope with a rapidly changing commercial environment. Furthermore, it is a commercial security model that is typically applicable in a data-mining environment. The private information of banking clients can and must be used in data-mining activities (to the benefit of the banking clients themselves and the banks in general) without sacrificing client confidentiality or bank credibility. The CBAC model makes this possible.

CHAPTER 7: A CBAC PROTOTYPE IMPLEMENTATION

THESIS LAYOUT



7.1 INTRODUCTION

The literature survey presented in chapters 3, 4 and 5 has led to the development of a Conflict-Based Access Control (CBAC) model (Chapter 6). The current chapter presents the design of a proof-of-concept prototype to demonstrate a subset of the model. The majority of requirements as stated in Chapter 5 were implemented in the proof-of-concept prototype. Some requirements were left to ensure a small and relatively simple proof-of-concept model so as to prove the concept of a conflict-based access control. More involved issues such as the performance abilities of the model are seen as topics for future research.

The chapter starts with the aims of the proof-of-concept prototype. An overview of the implementation process is then provided and the operation of the prototype is subsequently illustrated by means of the case study example.

7.2 THE AIMS OF THE PROOF-OF-CONCEPT PROTOTYPE

The aim of the CBAC prototype is to demonstrate the following features of the CBAC model:

- The use of conflict-based access control by a global agent for confidentiality purposes.
- The specification of spheres of conflict by secondary agents, as well as the identification of functional business domains to which each secondary agent belongs.
- The identification of potential conflict-of-interest (PIT) paths between any two secondary agents from different functional business domains.
- The restriction of mining agents to work in a ‘group’ of secondary companies that are not in conflict with one another so as to prevent any possible breach of confidentiality.

The focus of the prototype is on calculating

- the different spheres of conflict for each secondary agent based on each sphere’s degree of conflict,

- the bi-partite graph showing all the relationships between all secondary agents and all functional business domains, with all 'conflict strengths' added to it, and finally

- all the potential conflict-of-interest paths (PITs) as seen by the global agent.

- Access groups are groups of companies that are accessed by only one user at any given point in time. Hence, not all companies in an access group are conflicting companies. For purposes of the prototype these access groups are generated in the groups window and in the browser window it is possible to see how all companies in the same group are granted access simultaneously.

7.3 PROOF-OF-CONCEPT PROTOTYPE IMPLEMENTATION

7.3.1 Technical platform used for developing the prototype

This prototype was developed in Microsoft C#, using Microsoft Visual Studio 2008. It requires the .NET Framework v3.5 to run, since the prototype uses a number of features provided by this framework. Three of the most significant features used are discussed in Appendix A.

To make the prototype more user friendly, some variable names were changed from the names that were used in the model discussion up to this point. The following table serves as guidance for clarity purposes.

Table 7.1: CBAC model terminology mapped onto CBAC prototype terminology

CBAC *model* terminology	**CBAC *prototype* terminology**
functional business domain	domain (can be called an item)

CBAC *model* terminology	**CBAC *prototype* terminology**
secondary agent	company (can be called an item)
data sets of a secondary agent (model does not concentrate on this level)	entity (can be called an item)
contents of data sets (model does not concentrate on this level)	contents (can be called an item)
radius = d	conflict radius
secondary agent or data sets of a secondary agent or contents of data sets of a secondary agent	item
list of all the names of the secondary agents that may be accessed at the same time	access group
data-mining agent	user or a data-mining agent or agent

CBAC *model* terminology	**CBAC *prototype* terminology**
secondary agent k from the non-conflicting set of secondary agents: $k \in NC$	special node k
distance	distance

7.3.2 Structure of prototype

The prototype can be divided into two subsystems, namely the database-and-graph subsystem, and the presentation subsystem. Figure 7.1 depicts the structure of the database-and-graph subsystem by means of a class model that represents this subsystem.

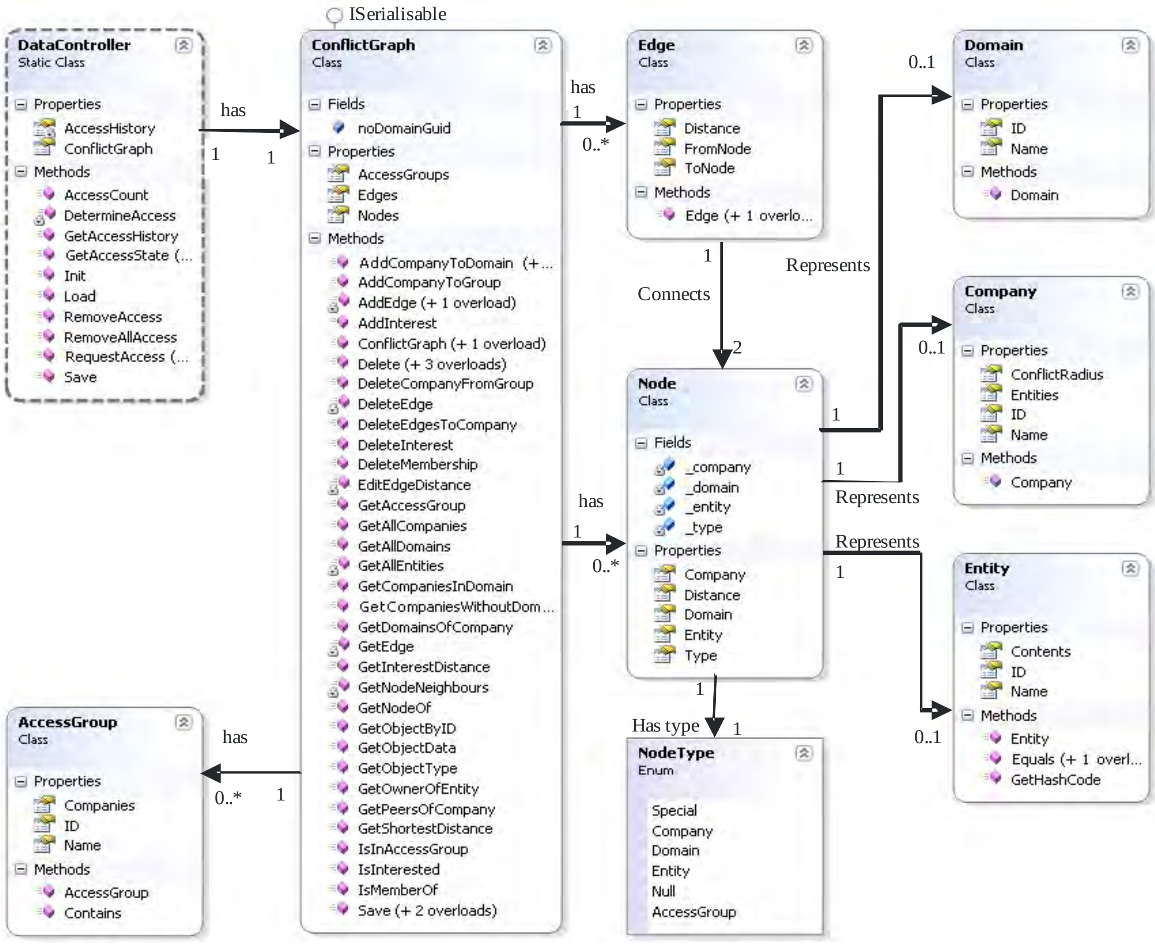


Figure 7.1: The Database-and-Graph Subsystem

The basic structure of the graph is implemented using node and edge classes. Nodes each contain either a domain, a company, or an entity in a company. Each node stores the type of item it contains as well. Each domain, company and entity class stores the respective details of a single object, and is uniquely identified using a **G**lobally **U**nique **Id**entifier (GUID). All of the above are stored in a ConflictGraph class, which contains all functionality related to Create, Read, Update and Delete (CRUD) on the graph, as well as conflict detection. The conflict graph furthermore contains the list of access groups. Finally, the DataController static class instantiates the ConflictGraph and maintains the access history of a mining agent. Note that the prototype only considers a single mining agent, so only a single access history is kept. The DataController class also handles serialisation of the ConflictGraph.

All operations by the presentation subsystem (depicted in Figure 7.2) on the database-and-graph subsystem are performed using the properties and methods provided by the static *DataController* class or its instantiated *ConflictGraph* object.

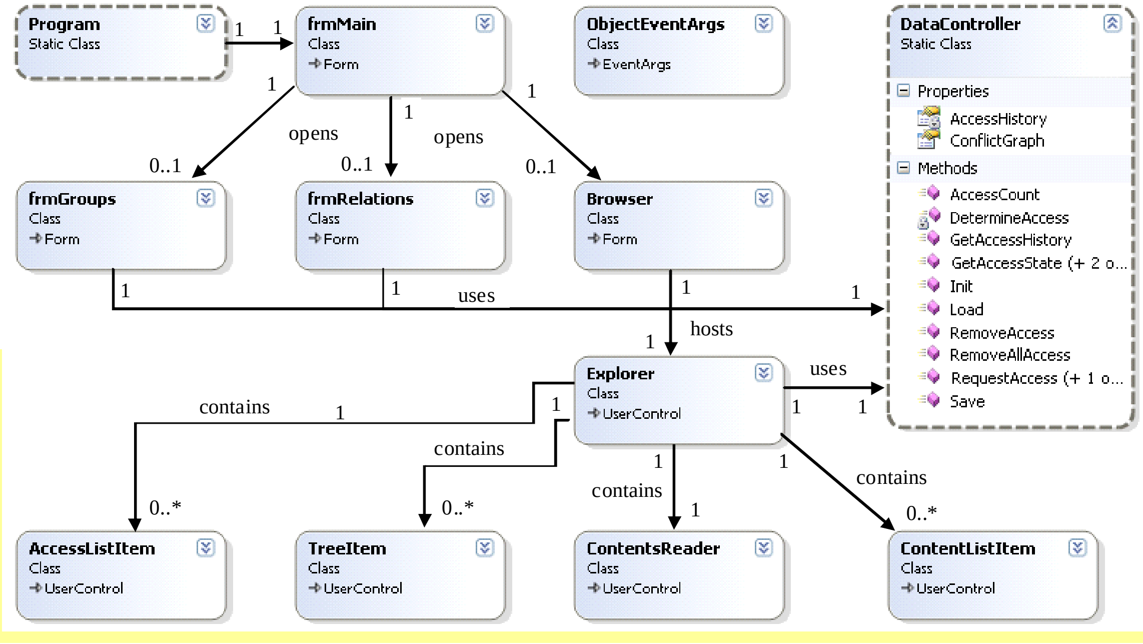


Figure 7.2: The Presentation Subsystem

Although the prototype provides features for adding, editing or removing items in the database, the primary algorithm of note is used to determine whether a user with a given access history is allowed to access a specific company by checking for conflicts on the conflict graph. This algorithm is depicted in Figure 7.3.

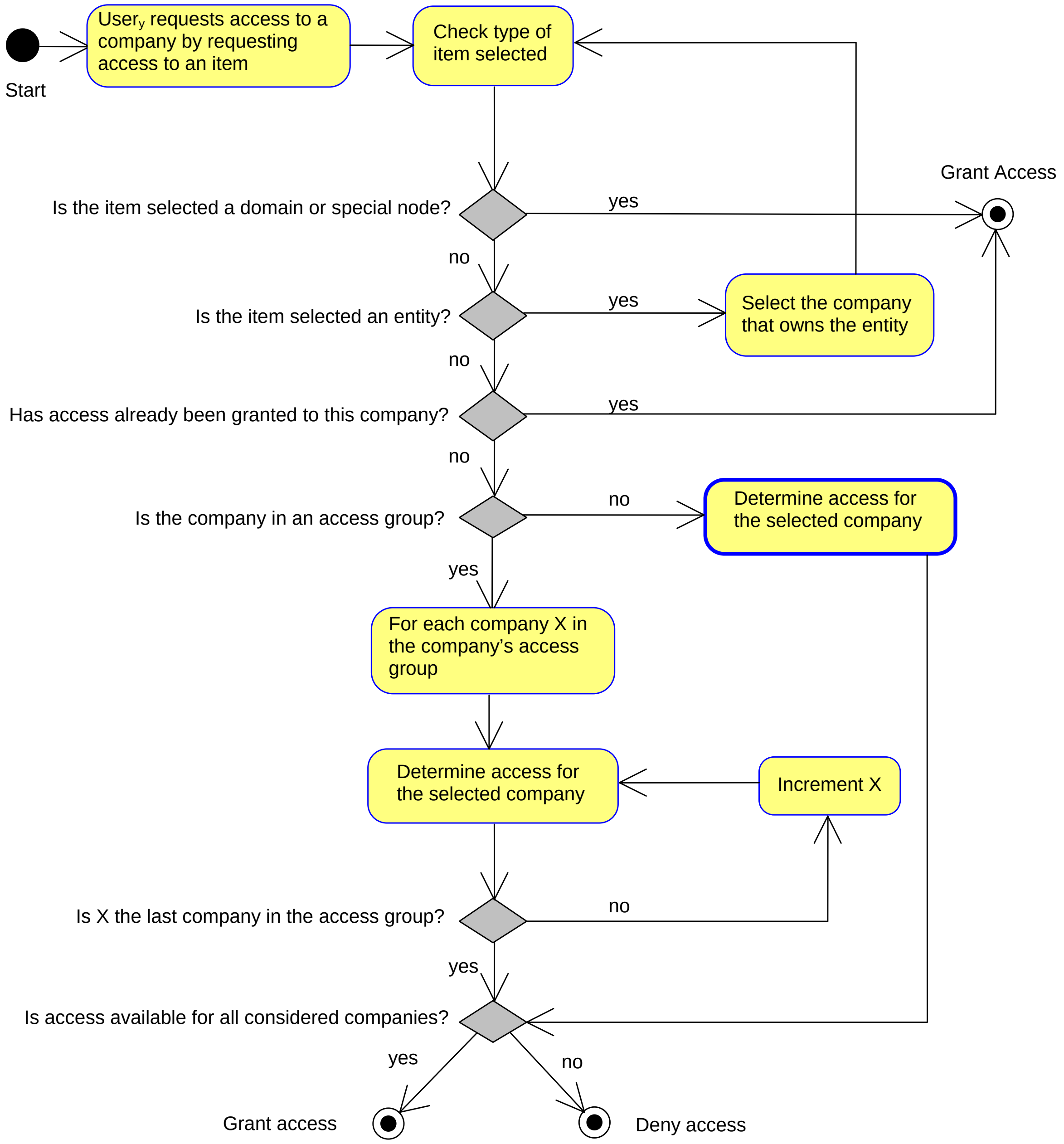

Start
User$_y$ requests access to a company by requesting access to an item
Check type of item selected
Is the item selected a domain or special node?
yes
Grant Access
no
Is the item selected an entity?
yes
Select the company that owns the entity
no
Has access already been granted to this company?
yes
no
Is the company in an access group?
no
Determine access for the selected company
yes
For each company X in the company's access group
Determine access for the selected company
Increment X
Is X the last company in the access group?
no
yes
Is access available for all considered companies?
yes
no
Grant access
Deny access

Figure 7.3: High-level activity diagram to determine access control for a user with a given access history

Due to the complexity of the ‘Determine Access for the selected Company’ process, it is depicted in more detail in Figure 7.4.

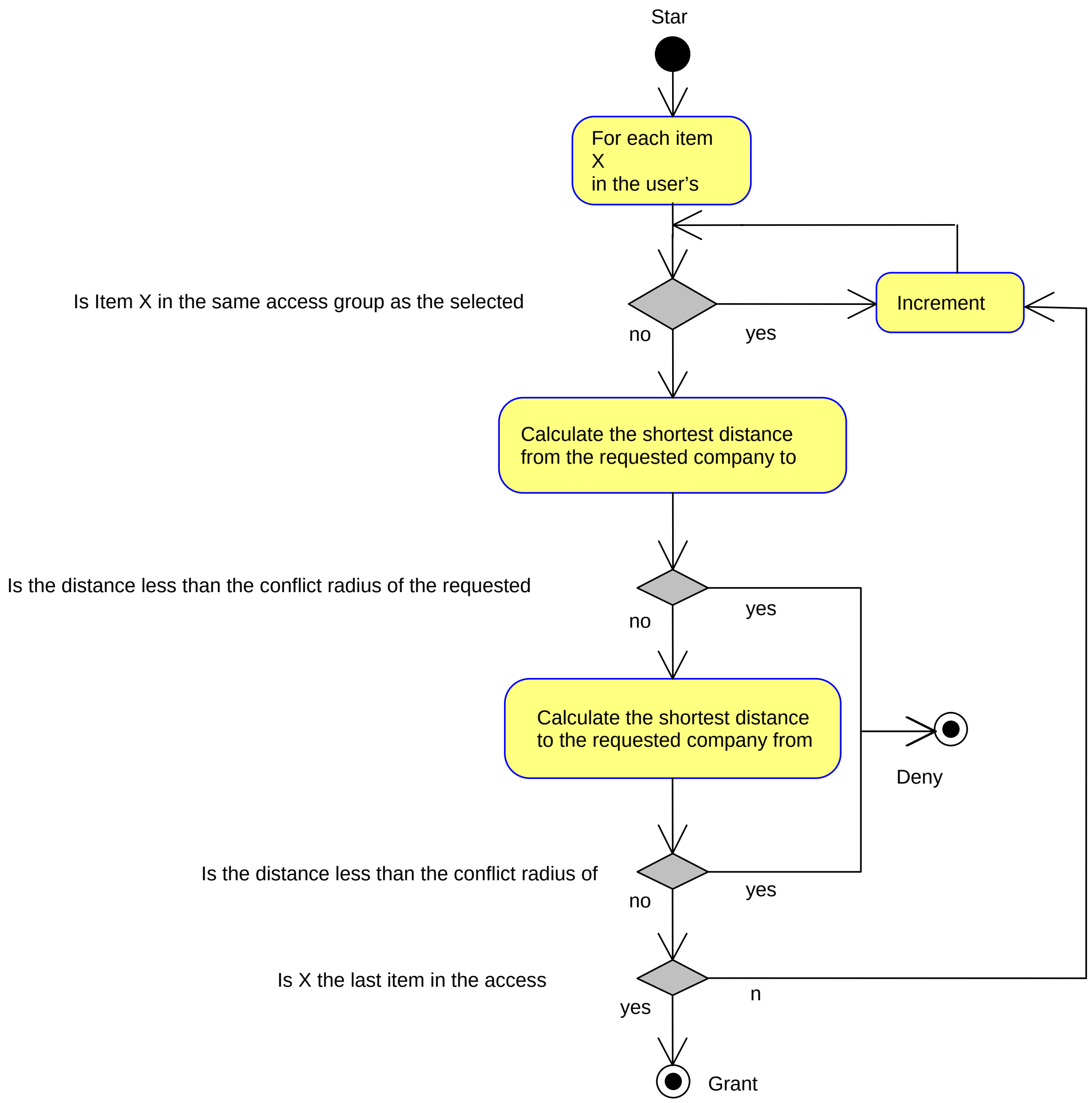

Star
For each item
X
in the user’s
Is Item X in the same access group as the selected
no
yes
Increment
Calculate the shortest distance
from the requested company to
Is the distance less than the conflict radius of the requested
no
yes
Calculate the shortest distance
to the requested company from
Deny
Is the distance less than the conflict radius of
no
yes
Is X the last item in the access
yes
n
Grant

Figure 7.4: High-level activity diagram – ‘Determine Access for the selected Company’s process

7.3.3 A statechart diagram for the prototype

A statechart diagram for the prototype is depicted in Figure 7.5. Entities of a company can be in one of four access states with respect to a specific user. The first access state, ‘available’, indicates that the company’s entities are not in conflict with any other company’s entities in the user’s access history. The second access state, ‘denied’, indicates that a conflict exists between the entities of this company and those of another in the access history of the specific user. The third access state, ‘granted’, indicates that the company’s entities are already in the specific user’s access history, and hence access has already been granted to this user. The fourth access state, ‘group conflict’, indicates that the company’s entities are part of a group. This means that access must be granted to all items in that group simultaneously. Because this is not possible for at least one item in that group, a situation of ‘group conflict’ arises. The access state of a particular company’s entities can change whenever the access history contents of a particular user change.

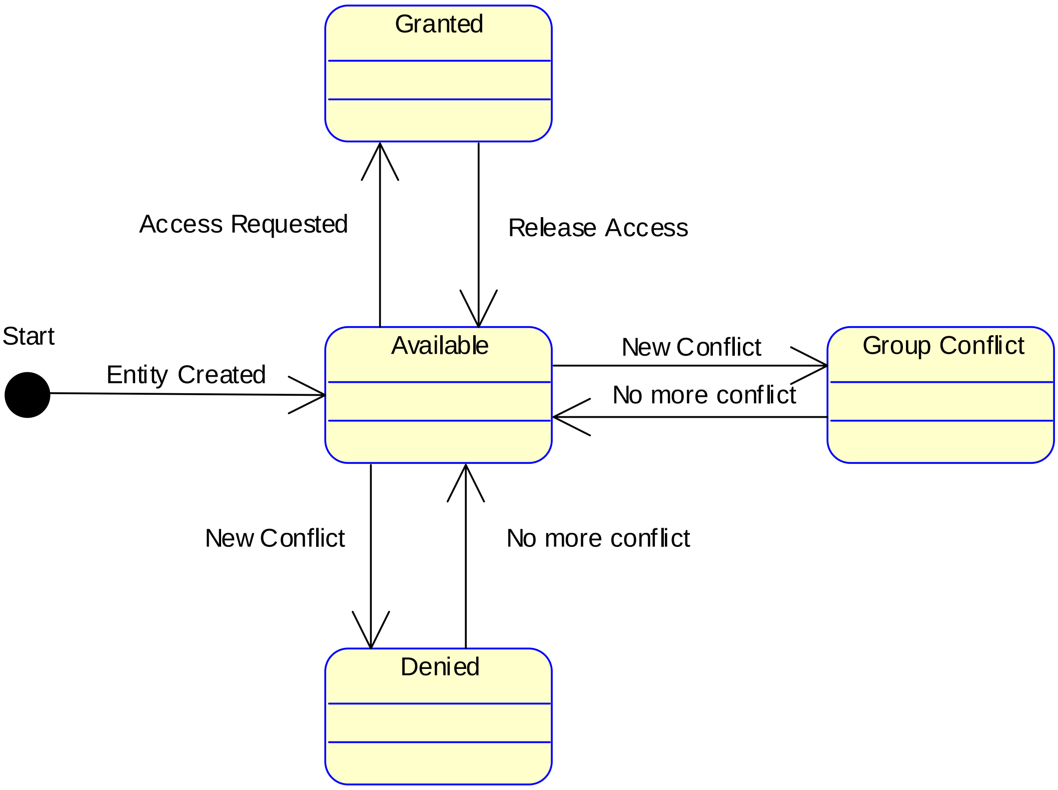


Figure 7.5: The statechart diagram

7.4 PROOF-OF-CONCEPT PROTOTYPE OPERATION

Start the prototype by running the provided executable file.

(1) Download the program and data files at

http://osprey.unisa.ac.za/temp/SEC780_Project1.zip

(2) Extract the files to your PC. The extracted folder name will be SEC780_Project1

(3) Navigate to the subfolder bin/Debug

(4) Click on SEC780_Project1.exe

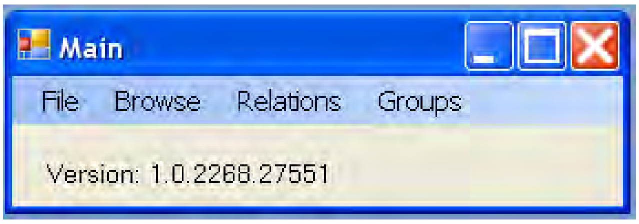
The main window will then be displayed (see Figure 7.6).

(5) On the program menu, click on File -> Load database

(6) Select *DMBank-Prototype Operation*

The main window will again be displayed (see Figure 7.6)

(7) On the program menu, click on Browse



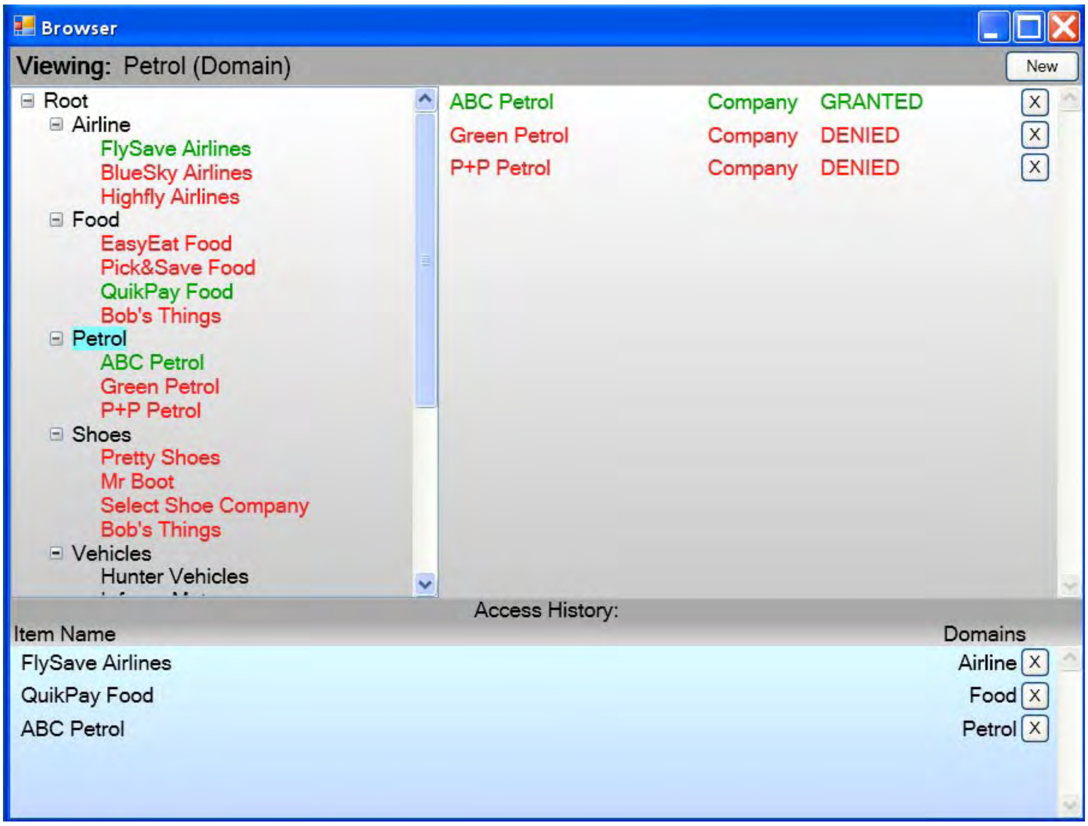
Version: 1.0.2268.27551

Figure 7.6: Main Window of CBAC

When the prototype is executed, its database is empty by default. The database contains all the contents of the Conflict Graph, including nodes, edges, groups and their contents. Therefore, it is recommended to follow the next steps in sequence to be able to value the demonstration. To later save the contents of the database, select *Save Database* from the *File* menu. The remaining menu items, *Browse*, *Relations* and *Groups* all open new windows. Their use will be described in sequence.

7.4.1 The Browser

Selecting *Browse* from the main menu will open the browser window. A similar window as depicted in Figure 7.7 will open. This window allows the user to view the contents of any domain (its companies), company (its entities), or entity (its contents). This demonstration works with ABC Petrol as the active secondary agent and for this reason, click on *ABC Petrol*, and then on *Petrol* in the left column.



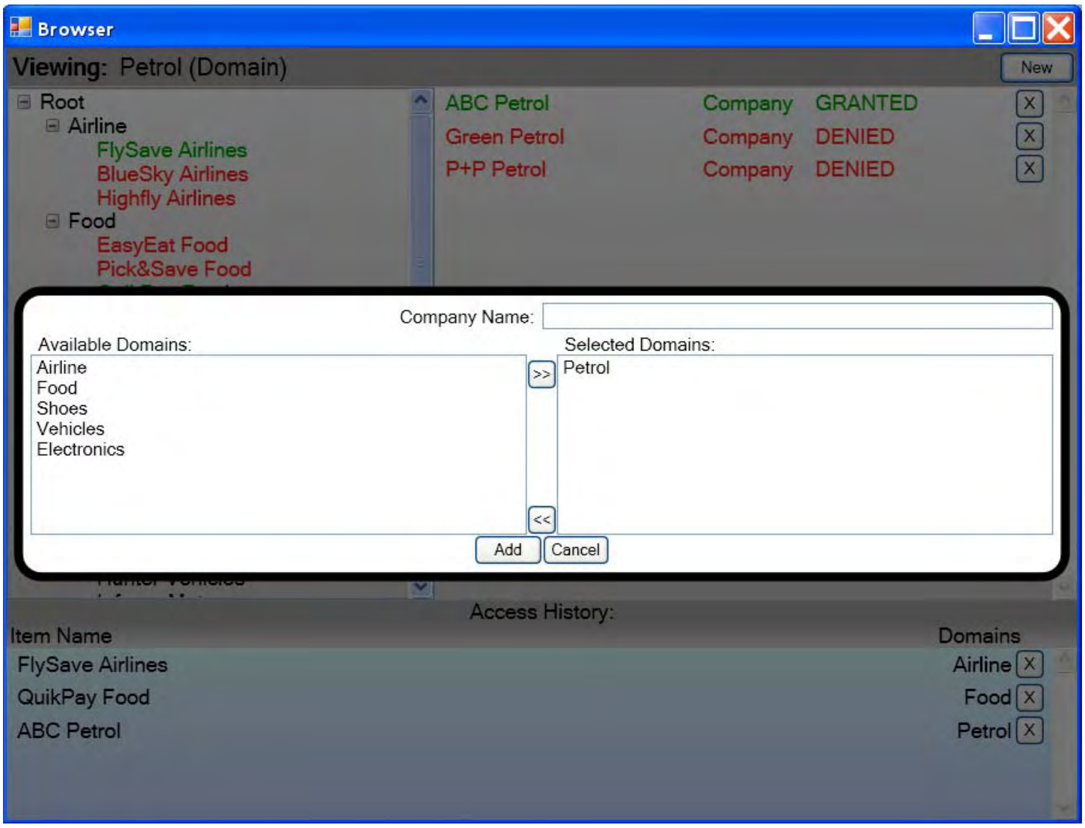
ABC Petrol	Company	GRANTED	X
Green Petrol	Company	DENIED	X
P+P Petrol	Company	DENIED	X

Item Name	Domains	
FlySave Airlines	Airline	X
QuikPay Food	Food	X
ABC Petrol	Petrol	X

Figure 7.7: The Browser

1) The bar at the top of the window indicates what is currently displayed in the contents part of the window.

2) The *New* button to the right of the top bar is used to add a new domain, company or entity to the database. The type of item created depends on what is currently selected. For example, if a domain is selected, a company will be created in the domain. To facilitate this, an overlay will appear over the browser window to allow the user to provide the details of the new item, as depicted in Figure 7.8.



Viewing: Petrol (Domain)

Root
Airline
FlySave Airlines
BlueSky Airlines
Highfly Airlines
Food
EasyEat Food
Pick&Save Food

ABC Petrol	Company	GRANTED
Green Petrol	Company	DENIED
P+P Petrol	Company	DENIED

Company Name:

Available Domains:
Airline
Food
Shoes
Vehicles
Electronics

Selected Domains:
Petrol

Access History:

Item Name	Domains
FlySave Airlines	Airline
QuikPay Food	Food
ABC Petrol	Petrol

Figure 7.8: Details of new item

3) The panel on the left (below the top bar) displays the structure of the database as a tree. The database root is shown at the top, with the domains as its children. For each domain, its companies are listed as children. Since companies can be in multiple domains, they can be listed in multiple parts of the tree. The colour of each item on the tree indicates its access state. Black indicates the item is available. Green indicates that access has already been granted to this item, and red or orange indicates that access has been denied due to a conflict of interest. The current item selected is highlighted with a cyan background.

4) The panel on the right (below the top bar) displays the contents of the selected item. This can be domains, companies or entities, depending on the selected item. Again, the colour of the text indicates its access state.

5) At the bottom of the window appears a list of the current companies (secondary agents) to which the user has been granted access.

6) To select an item, either click on it in the database tree or in the contents list. If necessary, the system will check if access can be granted and will refuse access if necessary. If granted, the database tree, contents list and access history are updated accordingly. Selecting an entity (only possible from the contents list) will cause an overlay to be displayed with the entity's contents.

7) To delete an item, select its parent in the database tree, and then click on the 'X' button to the right of the item in the contents list. An overlay will appear asking for confirmation.

8) To release access to a company's contents, click on the 'X' button to the right of the company listed in the access list. The database tree and contents list will update the access state of each displayed item accordingly.

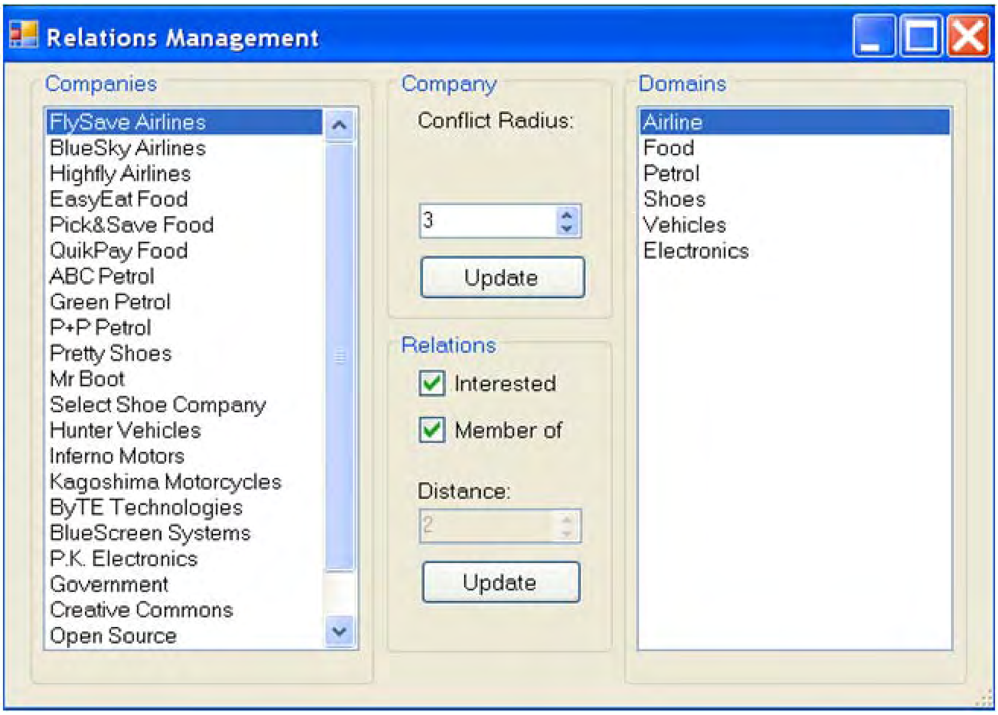
9) Note that any additions or removals will only be saved to the database file when the *Save Database* command is used from the main menu. Access history is not saved.

7.4.2 The Relations Manager

Clicking on *Relations* from the main menu will cause the Relations Management window to be displayed (see Figure 7.9). From this window, the user can add, edit or remove relations between companies and their domains, as well as each company's conflict radius,

as defined by the model. This system allows a company to be a member of a domain (or multiple domains) or to merely have an interest in a domain (the strength of which can be specified).

1) To add a company to a domain, select the company from the list on the left, and the domain from the list on the right. Then, add a checkmark to the *Interested* and *Member Of* check boxes, and click on *Update*.



Relations Management

Companies

FlySave Airlines
BlueSky Airlines
Highfly Airlines
EasyEat Food
Pick&Save Food
QuikPay Food
ABC Petrol
Green Petrol
P+P Petrol
Pretty Shoes
Mr Boot
Select Shoe Company
Hunter Vehicles
Inferno Motors
Kagoshima Motorcycles
ByTE Technologies
BlueScreen Systems
P.K. Electronics
Government
Creative Commons
Open Source

Company

Conflict Radius: 3

Relations

☑ Interested

☑ Member of

Distance: 2

Domains

Airline
Food
Petrol
Shoes
Vehicles
Electronics

Figure 7.9: Relations Management Window

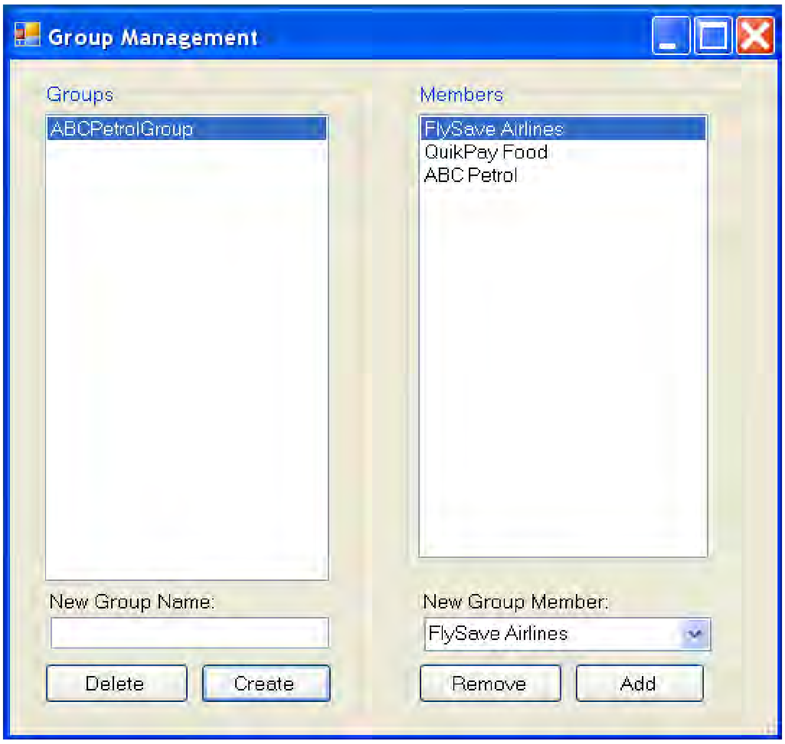
2) To remove a company from a domain, apply the same steps as above, but remove the checkmark from the *Interested* and *Member Of* domains. Note that removing the *Interested* checkmark will automatically remove the *Member of* checkmark.

3) Then add an interest in a domain to a company without adding it as a member of the domain, apply the same steps as in 1) above, but do not check the *Member of* box. Also, specify the distance in the *Distance* box before clicking *Update*. The lower the number, the higher the interest. Also note that making a company a member of a domain implies a very strong interest.

4) To remove an interest in a domain to a company, apply the same steps as in 2) above.

5) To change a company's conflict radius, select the company from the list on the left, change the value of the *Conflict radius* box, and click *Update*.

6) Note that these changes will only be saved to the database file when the *Save Database* command is used from the main menu.



Groups: ABCPetrolGroup

Members: FlySave Airlines, QuikPay Food, ABC Petrol

New Group Name:

New Group Member: FlySave Airlines

Figure 7.10: Groups Management Window

7.4.3 The Groups Manager

Selecting *Groups* from the main menu will cause the *Groups Management* window depicted in Figure 7.10 to be displayed. This window allows the user to add and remove groups, as well as assign members to these access groups. Unlike domains, companies can only be part of zero or one group at a time. In the browser window, all companies in the same group are always granted access simultaneously.

1) To add a group to the database, enter the new group's name in the *New Group Name* text box, and click on *Create*.

2) To remove a group from the database, select the group from the list on the left, and click on *Delete.*

3) To add a company to a specific group, select the group from the list on the left. Next, select the company from the *New Group Member* combo box, and click on *Add*. If the company is already part of another group, the user will be prompted to indicate whether the company should be transferred to the new group.

4) To remove a company from a specific group, select the group from the list on the left. Then, select the company from the list on the right, and click on *Remove*.

5) Note that these changes will only be saved to the database file when the *Save Database* command is used from the main menu.

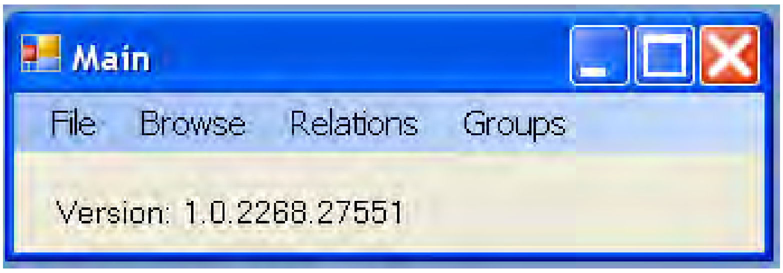
7.5 EXAMPLE OF THE PROOF-OF-CONCEPT PROTOTYPE

The case study is now revisited to illustrate the operation of the CBAC model. The discussion focuses on the implementation of a sphere of conflict for the ABC Petrol company (see Chapter 2), one specific company of DM Bank; the acknowledgement of its conflicting companies as indicated by ABC Petrol; and the effect on access control when assigning a data miner to the ABC Petrol company. The discussion is structured to include the following aspects:

- Specifying the sphere of conflict for ABC Petrol
- Setting the severity level of the conflict for ABC Petrol
- Specifying the conflict radius for ABC Petrol

7.5.1 Specifying the sphere of conflict for the ABC Petrol company

The first component to be determined is the sphere of conflict of the ABC Petrol company and this must be specified by ABC Petrol itself. To do this, the Main window of the CBAC proof-of-concept prototype must be activated by double clicking on the icon. The Main window will open. See Figure 7.11.



Version: 1.0.2268.27551

Figure 7.11: Main Menu Window

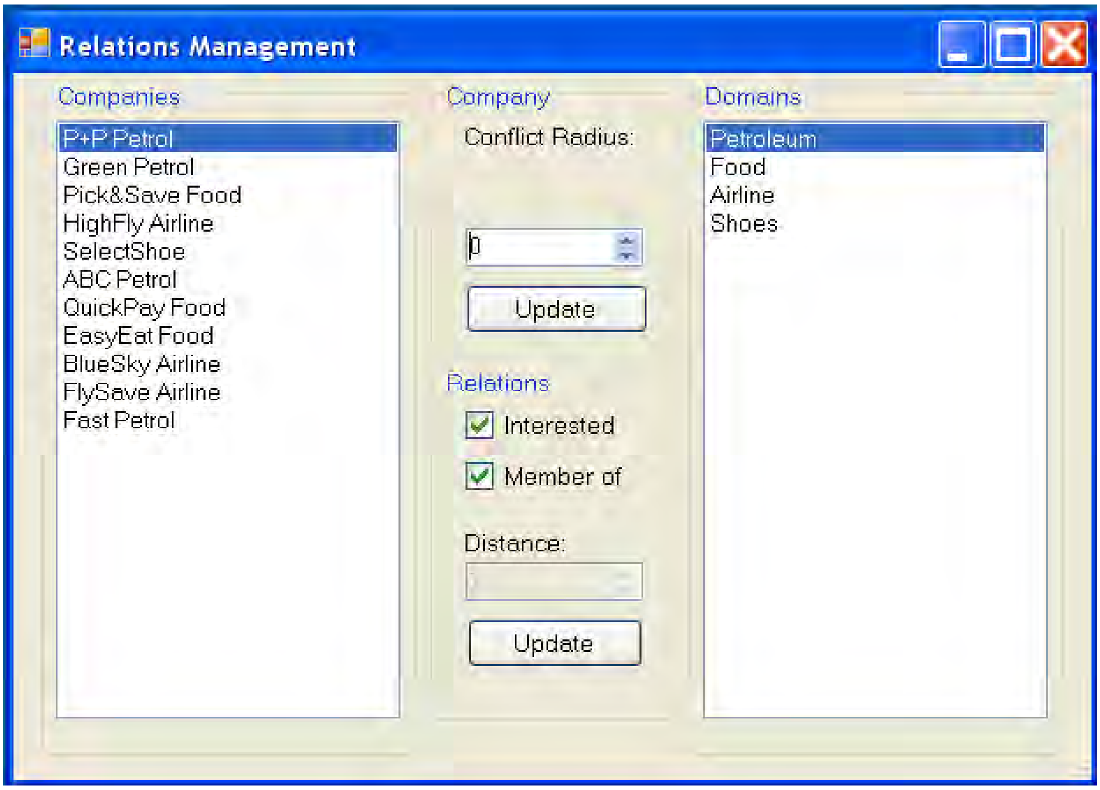
Click on *Browse* to open an empty file. Start entering the secondary agents with which ABC Petrol is in conflict, namely the five secondary agents as identified in Figure 2.2. Add each of them to a functional business domain (see the four functional business domains as depicted in Table 2.1). Then add all the other secondary agents known to DM Bank that belong to the four identified functional business domains.

7.5.2 Setting the severity of the conflict for ABC Petrol

Open the *Relations Manager* and add all the known relations.

Relations between ABC Petrol and all other petrol companies in the Petroleum functional business domain:

All Petroleum companies must have a "✓" in the *Interested* block, as well as in the *Member of* block. *Distance* should be "1" indicating that they are members of the petroleum functional business domain and that they have a strong interest in the petroleum functional business domain. This also means that they are in strong conflict with any other company belonging to the petroleum functional business domain.



Relations Management

Companies

P+P Petrol
Green Petrol
Pick&Save Food
HighFly Airline
SelectShoe
ABC Petrol
QuickPay Food
EasyEat Food
BlueSky Airline
FlySave Airline
Fast Petrol

Company

Conflict Radius: 0

Relations

☑ Interested
☑ Member of

Distance:

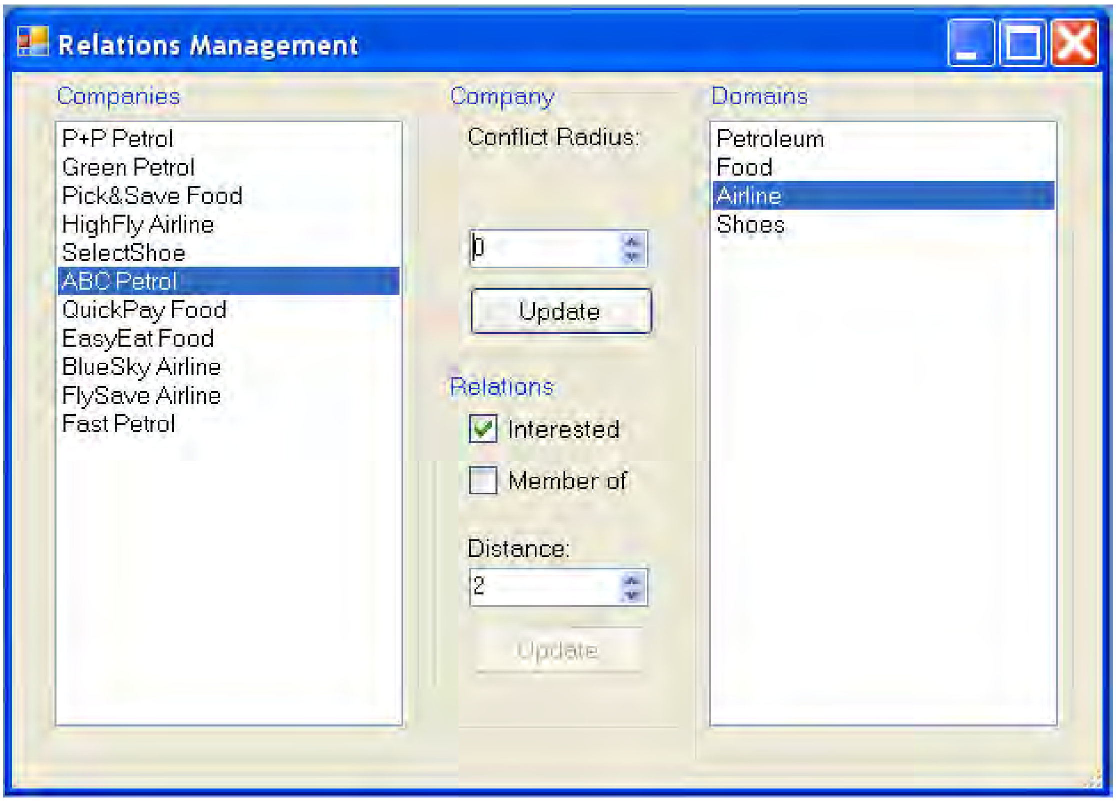
Domains

Petroleum
Food
Airline
Shoes

Figure 7.12: Relations Management Window for ABC Petrol

Relations between ABC Petrol and all the other specified interested companies:

ABC Petrol must have a "✓" in the *Interested* and a "□" in the *Member of* block with a *Distance* of "2" from the airline functional business domain, indicating the strength of the relationship between ABC Petrol and the airline functional business domain.



Relations Management

Companies
P+P Petrol
Green Petrol
Pick&Save Food
HighFly Airline
SelectShoe
ABC Petrol
QuickPay Food
EasyEat Food
BlueSky Airline
FlySave Airline
Fast Petrol

Company
Conflict Radius: 0
Update

Relations
☑ Interested
□ Member of
Distance: 2
Update

Domains
Petroleum
Food
Airline
Shoes

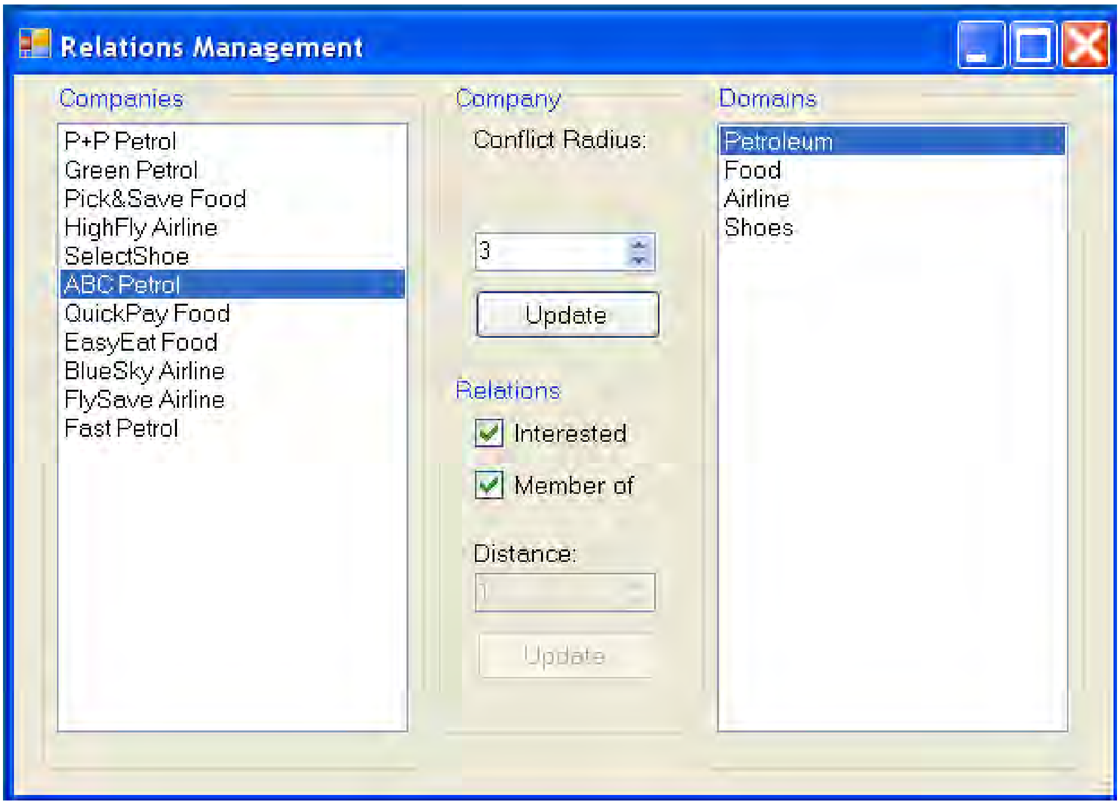
Figure 7.13: The relationship between ABC Petrol and the airline functional business domain

ABC Petrol must have a "✓" in the *Interested* and a "□" in the *Member of* block with a *Distance* of "3" from the food functional business domain, indicating the strength of the relationship between ABC Petrol and the food functional business domain.

ABC Petrol must have a "✓" in the *Interested* and a "□" in the *Member of* block with a *Distance* of "6" from the shoes functional business domain, indicating the strength of the relationship between ABC Petrol and the shoes functional business domain.

7.5.3 Specifying the conflict radius for ABC Petrol

For all the established relationships, ABC Petrol can at this stage specify a conflict radius that allows all other conflicting companies on a radius longer than the conflicting radius to be seen as non-conflicting companies. In the example, this conflict radius is set to 3 in the Conflict Radius box at the top of the Relations Management window.



Companies:
- P+P Petrol
- Green Petrol
- Pick&Save Food
- HighFly Airline
- SelectShoe
- ABC Petrol
- QuickPay Food
- EasyEat Food
- BlueSky Airline
- FlySave Airline
- Fast Petrol

Conflict Radius: 3

☑ Interested

☑ Member of

Distance: 1

Domains:
- Petroleum
- Food
- Airline
- Shoes

Figure 7.14: Specifying the conflict radius for the ABC Petrol company

This concludes the example for the specifications as seen by the ABC Petrol company.

7.6 EXAMPLE: HOW ACCESS PERMISSION FOR A USER WITH A GIVEN ACCESS HISTORY IS DETERMINED

The following example is applicable when demonstrating the process that is followed for a user with an existing access history to request access to a given company. Mining agent$_1$ has an access history in which there is one company, namely Pick&Save Food. Mining agent$_1$ requests access to ABC Petrol.

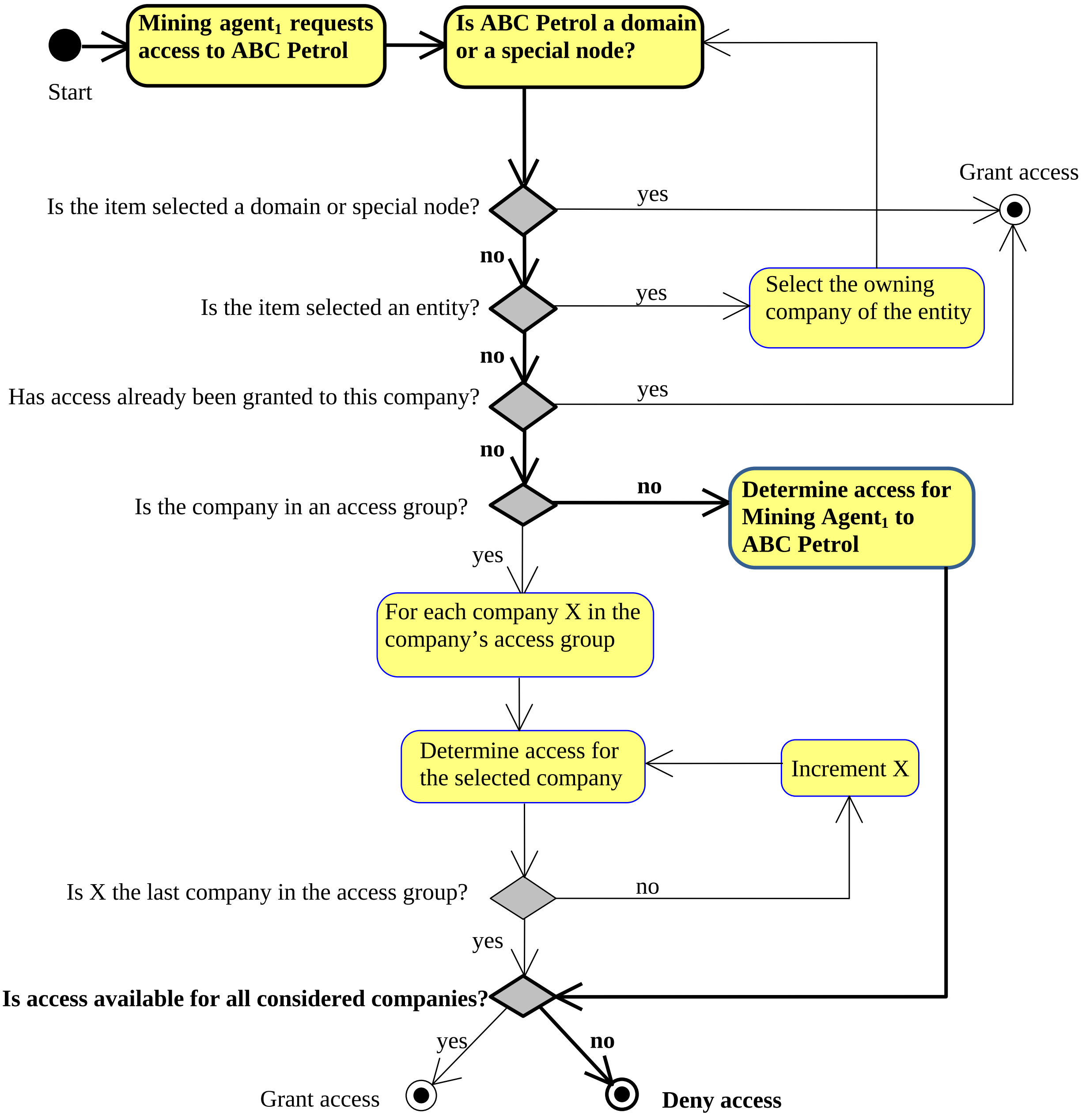

Start
Mining agent₁ requests access to ABC Petrol
Is ABC Petrol a domain or a special node?
Is the item selected a domain or special node?
yes
Grant access
no
Is the item selected an entity?
yes
Select the owning company of the entity
no
Has access already been granted to this company?
yes
no
Is the company in an access group?
no
Determine access for Mining Agent₁ to ABC Petrol
yes
For each company X in the company's access group
Determine access for the selected company
Increment X
Is X the last company in the access group?
no
yes
Is access available for all considered companies?
yes
no
Grant access
Deny access

Figure 7.15: Determine access to ABC Petrol for Mining Agent$_1$

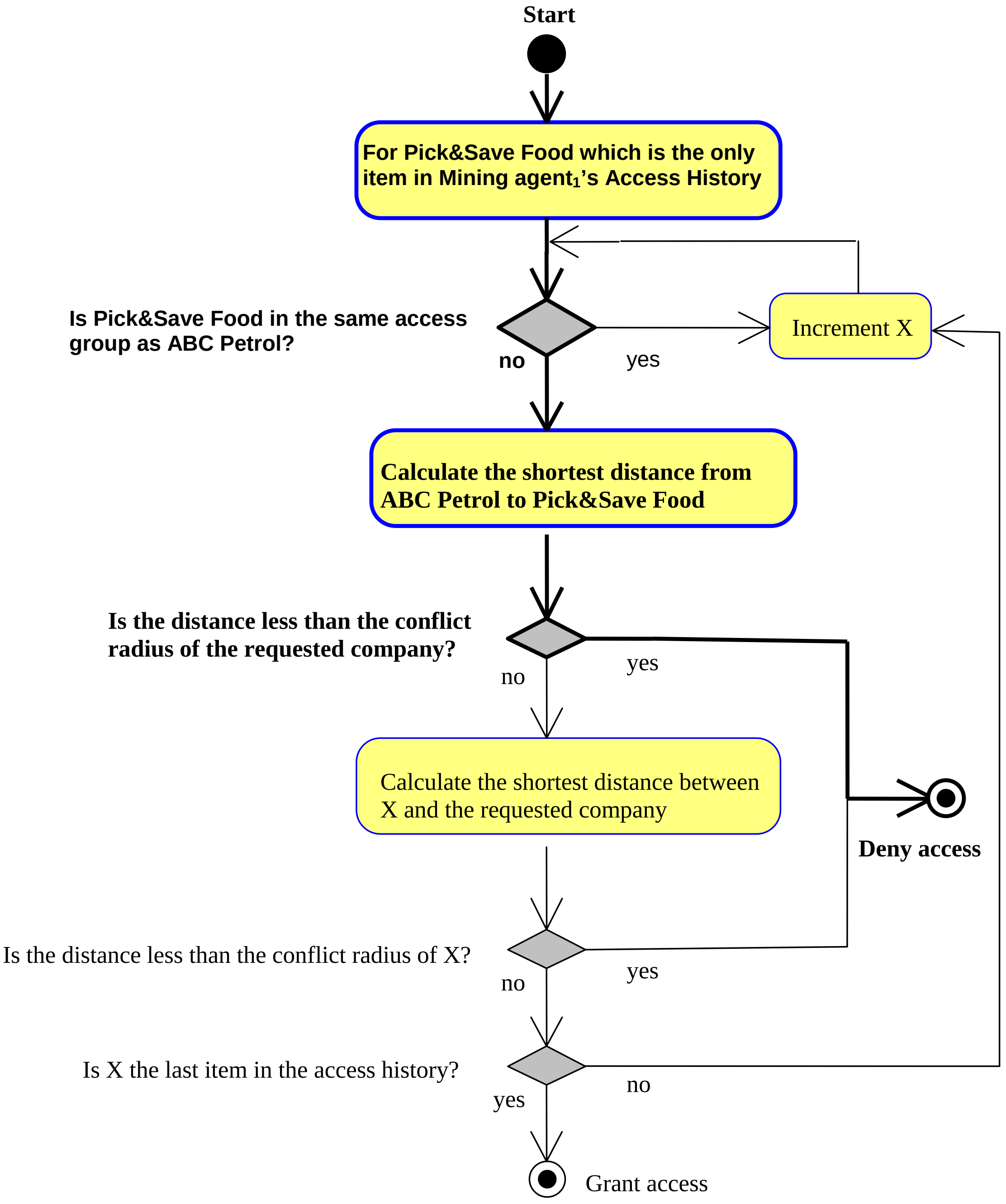

Start
For Pick&Save Food which is the only item in Mining agent₁'s Access History
Is Pick&Save Food in the same access group as ABC Petrol?
no
yes
Increment X
Calculate the shortest distance from ABC Petrol to Pick&Save Food
Is the distance less than the conflict radius of the requested company?
no
yes
Calculate the shortest distance between X and the requested company
Deny access
Is the distance less than the conflict radius of X?
no
yes
Is X the last item in the access history?
yes
no
Grant access

Figure 7.16: Determine possible conflict between ABC Petrol and Mining agent$_1$'s access history

In conclusion, Mining agent$_1$ is not allowed to work on the data sets of ABC Petrol because it is already working on the data sets of Pick&Save Food. Also, ABC Petrol and Pick&Save Food is in conflict with one another.

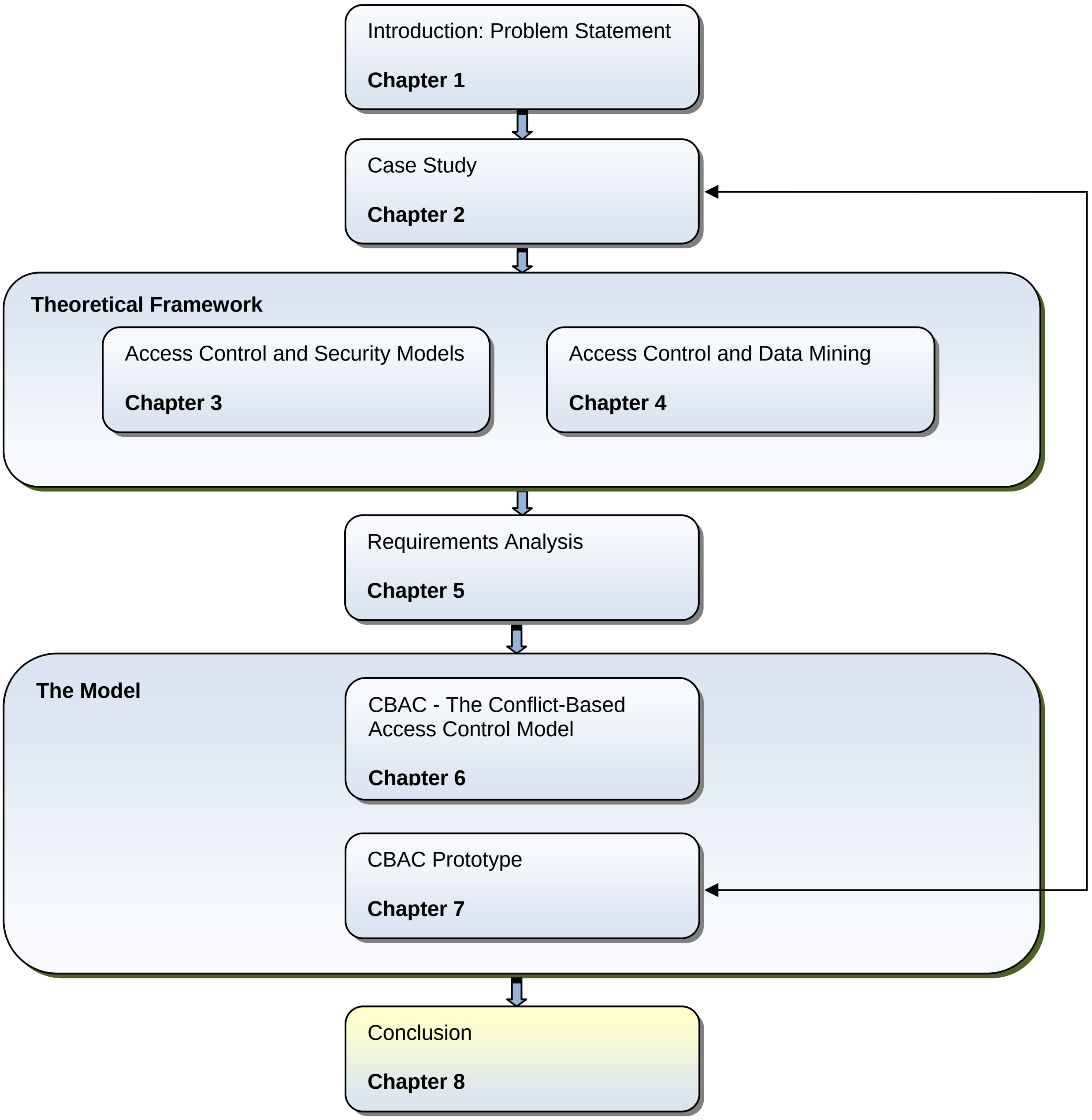
7.7 CONCLUSION

The access control model based on conflict was tested in terms of a proof-of-concept prototype and is not a working product in itself. It was found that it is possible to design a conflict-based access control model for a data-mining environment according to minimum requirements. Having done that, it is now also possible to minimise the risk of a breach of confidentiality in a data-mining environment by managing the access control in such an environment based on the conflicts that exist between the different subjects that form part of it.

This prototype can be developed into a working product in a further study.

CHAPTER 8: CONCLUSION

THESIS LAYOUT



8.1 INTRODUCTION

The study in hand focused on a model for conflict-based access control in a data-mining environment, i.e. the CBAC model. To increase the pertinence of the model, it was presented within a banking environment, an active part of the greater commercial environment. The pragmatic issue of developing such a model within the wider context of access control models constituted the topic of this thesis. The motivation for the study was postulated in Chapter 1 and required a number of research goals to be addressed. In this closing chapter the researcher revisits the research questions and evaluates the extent to which the objectives of the research goals have been met. Finally, the chapter is concluded with a discussion of the main contribution of the research and suggestions on further research forthcoming from this work.

8.2 REVISITING THE PROBLEM STATEMENT

The main focus of this research was to specify an access control model based on conflict. The access control decision was given new focus by not asking "Who may access this resource?*"* or *(subject*$_i$*, object*$_i$*, read)*, but rather "Who may access this resource while also honouring the resource's conflicting business relationships*?"* or *(subject*$_i$*, object*$_i$*, read, severity of conflict)*. The study confirmed that over and above the attributes of users, the business relationships of users play an important role when access control decisions are made.

The aim of this research was to minimise the chances of a potential breach of confidentiality made likely by the hidden knowledge revealed through the data-mining process.

This thesis therefore endeavoured to answer the following research questions:

What are the security requirements for a data-mining environment?

The data-mining process was examined and each step of the process received due emphasis. Based on this examination, an ideal list was compiled of those security requirements for data mining that had to be present to ensure sufficient confidentiality during the course of the data-mining process.

What are the access control requirements for a data-mining environment?

After investigating security requirements in a data-mining environment, a study was made of those access control requirements for data mining that had to be present to ensure confidentiality during the course of the data-mining process. A list of relevant access control requirements was also compiled.

How can we introduce conflict-based concepts into an access control model?

Based on the security and access control requirements already established, a model was developed that minimises the risk of a breach of confidentiality in a data-mining environment. This is done by managing the access control in such an environment, based on the conflicts that exist between the different subjects that form part of this environment. Once the CBAC model was ready, the concept was demonstrated in a proof-of-concept prototype.

8.3 MAIN CONTRIBUTION

The main contribution of this thesis is its proposition of a new access control model called the Conflict-Based Access Control (CBAC) model.

Although data-mining technology can be seen as having many advantages, there are also some clear disadvantages that need to be addressed. As mentioned earlier, a problem associated with data mining is the lack or loss of confidentiality pertaining to knowledge management, in other words the problem of how to maintain security [Bertino, et al., 2006].

While data mining is a term that is well understood in some circles, it has not really entered the vocabulary of the general population. The interesting cases of Jérôme Kerviel from the French bank Société Générale and Raj Rajaratnam, founder of the Galleon Group, emphasise the fact that the secure management of data and information so as to minimise a breach of confidentiality forces one to concentrate on conflicting relationships in general, and data mining in particular. In this thesis a novel access control model that concentrates on these conflicting relationships was built. The CBAC model implements the notion of "conflict-of-interest" classes by introducing a new concept – the so-called sphere-of-conflict. Graph theory is used to model a Potential Conflict-of-Interest (PIT) path between different agents. The CBAC provides a dynamic approach for addressing the access control requirements of rapidly changing business environments. Furthermore, the model can be successfully deployed in a complex use case to investigate data-mining activities at financial institutions without sacrificing client confidentiality or jeopardising the credibility of the service provider.

The CBAC deals with a changing business environment, i.e. a data-mining environment [Loock & Eloff, 2005a] and the model supports the following:

- The definition of conflict-of-interest classes as well as the assumption that such classes are not ‘disjoint’ or separate.

  The original concept of conflict-of-interest classes as defined for the CWSP involved strict ‘walls’ based on business definitions [Brewer & Nash, 1989]. This means that these conflict-of-interest classes were based on functional business domains that compartmentalised organisations with no overlaps between the different compartments. The ACWSP model pointed out that these conflict-of-interest classes are ‘seldom disjoint’ [T Y Lin, 1989] and that they ‘overlap’ [T Y Lin, 1989], which is equally true for functional business domains.

- A sphere of conflict for a secondary agent specified by that secondary agent itself rather than by functional business domains.

  The different *degrees of conflict of interest* between the secondary agent and all its known conflicting secondary agents are quantifiable, as well as the *cut-off point for conflict of interest* for the secondary agent. This is the first step towards an access control model that takes into account the different conflicting relationships that exist between competing secondary agents within the same global environment.

- The addition of a bi-partite graph that assists the global agent to add knowledge to the model by giving information regarding a *potential conflict of interest (PIT)* between two secondary agents.

  This potential conflict of interest is calculated with the assistance of a *potential conflict-of-interest path*. The existence of more than one *potential conflict-of-*

interest path is acknowledged and a possible *k potential conflict-of-interest paths* are calculated and added to the model. The method of weighting these *k* different *potential conflict-of-interest paths* so as to select the shortest *potential conflict-of-interest paths* is then part of the model description. The global agent may specify a *cut-off point of path weight* to ensure that the potential conflict-of-interest path does not get too long.

The CBAC model also introduces the concept of Potential Conflict of Interest, which can be determined by the global agent. In the CBAC model, the functional business domain definition is a dynamic concept that plays a role in the *conflict-of-interest classes*. The CBAC model allows for the implementation of a potential conflict of interest in the same or another functional business domain as determined by a secondary agent or by the global agent at a specific point in time. For example: if secondary agent c_i has developed a new interest in another secondary agent c_j, then agent c_i may ask to be treated as such. From then on, the same mining agent cannot work on the data sets of both secondary agents again. Secondary agents can change this *conflict of interest* during execution time.

- The current research finds the CBAC model to be *dynamic* and thus able to cope with a rapidly changing commercial environment. Today’s business environment is much more dynamic compared to that of a decade ago. Companies do not only diversify, but also (in some cases) change their business strategy, which might entail entering a new functional business domain. For example, a secondary agent develops a ‘new’ major interest in another secondary agent, which creates potential conflict of interest in the future. It is argued that such changes have to be reflected in the decision-making process of granting or rejecting access requests, especially so in a data-mining environment. This dynamic adaptation to business requirements, in conjunction with the preservation of separation of duties, is the main contribution of the CBAC model. The CBAC model bases the definition of conflict on conflict between secondary agents as defined by a specific secondary agent, in other words, it incorporates the latter agent’s ‘view of the world’.

- The current research makes an important contribution towards commercial security models in general in that the CBAC model is typically applicable in a data-mining environment. The private information of banking clients can and must be used in data-mining activities (for the benefit of the banking clients themselves and the banks in general), yet without sacrificing client confidentiality or bank credibility. The CBAC model makes this possible.

8.4 FUTURE RESEARCH

The proposed model achieved the set of objectives to the extent described in the section above, but it suffers some limitations. These limitations provide opportunities to extend and support the work described in this thesis by a number of future research projects:

- The implementation of a bi-partite graph is important for the model as it is described in this research. However, further research is necessary to establish if this is the most cost effective way to implement the model.

- The element of trust to be added to the different mining agents warrants an investigation into human behaviour and should have an interesting effect on the performance of the model.

- The performance abilities of the model should be investigated further.

- The model was developed for a data-mining environment. Future research could well involve using the CBAC model in other environments such as in the offices of financial advisors of an organisation servicing different wealthy clients, or in the offices of big attorney practices that represent conflicting parties.

- It is also suggested that future studies should answer the question: Are access rights important when you discuss a conflict-based access control model? Does this model need refinement to add access rights? Will access rights help to avoid conflict on the level of business addressed by the CBAC model?

BIBLIOGRAPHY

Agrawal, R., & Srikant, R. (2000, 14 - 19 May). *Privacy-Preserving Data Mining.* Paper presented at the ACM SIGMOD Conference on Management of Data, Dallas, Texas.

Ardagna, C. A., Cremonini, M., Damiani, E., De Capitani di Vimercati, S., & Samarati, P. (2006). The Architecture of a Privacy-aware Access Control Decision Component *Lecture Notes in Computer Science* (Vol. 3956/2006, pp. 1 - 15). Heidelberg: Springer Berlin.

Ardagna, C. A., Cremonini, M., Damiani, E., Vimercati, S. D. C. d., & Samarati, P. (2006). The Architecture of a Privacy-Aware Access Control Decision Component *Lecture Notes in Computer Science* (Vol. Volume 3956/2006, pp. 1 - 15). Heidelberg: Springer Berlin.

Ardagna, C. A., Cremonini, M., De Capitani di Vimercati, S., & Samarati, P. (2008). A Privacy-Aware Access Control System. *Journal of Computer Security, 16*(4).

Bell, D., & LaPadula, L. (1973). Secure Computer Systems: Mathematical Foundations and Model (pp. 42). Bedford, MA: National Technical Information Service.

Bell, D., & LaPadula, L. (1976). Secure Computer System: Unified Exposition and Multics Interpretation. Bedford, MA: MITRE Corporation.

Bertino, E., Khan, L. R., Sandhu, R., & Thuraisingham, B. (2006). Secure Knowledge Management: Confidentiality, Trust, and Privacy. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 36*(No. 3), 10.

Bishop, M. (2003). *Computer Security: art and science*: Pearson Education Inc.

Brewer, D. F. C., & Nash, M. J. (1989, 01-03 May 1989). *The Chinese Wall Security Policy.* Paper presented at the IEEE Symposium on Research in Security and Privacy, Oakland, California.

Byun, J.-W., Bertino, E., & Li, N. (2005). *Purpose based access control of complex data for privacy protection*. Paper presented at the Proceedings of the tenth ACM symposium on Access control models and technologies, Stockholm, Sweden.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering Data Mining from concept to implementation*. New York: Prentice Hall PTR.

Chaula, J. A., Yngstrom, L., & Kowalski, S. (2005). *A Framework for Evaluation of Information Systems Security*. Paper presented at the Information Security South Africa Conference (ISSA 2005), Johannesburg, South Africa.

Claudio, A. A., Jan, C., Markulf, K., Ronald, L., Gregory, N., Bart, P., . . . Mario Verdicchio, B. (2010). Exploiting Cryptography for Privacy-Enhanced Access Control A result of the PRIME Project.

Clifton, C., & Marks, D. (1996). *Security and privacy implications of data mining*. Paper presented at the ACM SIGMOD International Conference on Management of Data.

CNSS. (2010). Committee on National Security Systems Instruction No. 4009 Retrieved 21 June, 2012, from http://www.cnss.gov/Assets/pdf/cnssi_4009.pdf

Collins Concise Dictionary. (2004) (21st Century Edition ed.). Glasgow: HarperCollins Publishers Ltd.

Conrado, C., Petkovic, M., & Jonker, W. (2004). Privacy-Preserving Digital Rights Management *Lecture Notes in Computer Science* (Vol. 3178, pp. 83-99). Heidelberg: Springer Berlin.

Daimler-Benz, ISL, & NCR. CRoss Industry Standard Process for Data Mining Retrieved 14 June, 2011, from http://www.crisp-dm.org/

Davies, L. (1997). *Psychology of Risk, Speculation and Fraud*. Paper presented at the Financial Panel of the European Research Center, Will the EMU Pay Off? Anticipating the Effects on the Market., Amsterdam.

De Capitani di Vimercati, S., Foresti, S., & Samarati, P. (2007). Authorization and Access Control. In M. Petkovic & W. Jonker (Eds.), *Security, Privacy, and Trust in Modern Data Management*. Berlin Heidelberg: Springer-Verlag.

De Capitani di Vimercati, S., Paraboschi, S., & Samarati, P. (2006). Access Control: Principles and Solutions. In H. Bidgoli (Ed.), *Handbook of Information Security* (Vol. 3). Hoboken, New Jersey: John Wiley & Sons.

Ferraiolo, D. F., & Kuhn, D. R. (1992, 13-16 October). *Role-Based Access Controls.* Paper presented at the 15th National Computer Security Conference, Baltimore.

Gasser, M. (1990). *The role of naming in secure distributed systems.* Paper presented at the CS'90 Symposium on Computer Security, pages 97-109, Rome, Italy.

Gauthier-Villars, D. (2010, 06 October). Rogue French Trader Sentenced to 3 Years, *The Wall Street Journal*.

Gollmann, D. (2006). *Computer Security* (2nd ed.). West Sussex: John Wiley & Sons Ltd.

Graham, G., & Denning, P. (1972). *Protection - Principles and Practice.* Paper presented at the Spring Joint Computer Conference.

Habib, L., Jaume, M., & Morisset, C. (2009). Formal definition and comparison of access control models. *Journal of Information Assurance and Security, 4*, 372-381.

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (Second ed.): Morgan Kaufmann Publishers.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (Third ed.). MA: Morgan Kaufmann Publishers.

Harrison, M. A., Ruzzo, W. L., & Ullman, J. D. (1976). Protection in Operating Systems. *Communication of the ACM, 19*(8), 461-471.

Kandias, M., Mylonas, A., Virvilis, N., Theoharidou, M., & Gritzalis, D. (2010). An Insider Threat Prediction Model *Lecture Notes in Computer Science* (Vol. 6264/2010, pp. 26-37).

Kantarcioglu, M., & Clifton, C. (2004). Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data. *IEEE Transactions on Knowledge and Data Engineering, 16*(9), 1026 - 1037.

Katsikas, S. K., Gritzalis, S., & Balopoulos, T. (2005). *Specifying electronic voting protocols in typed MSR.* Paper presented at the 2005 ACM Workshop on Privacy in the Electronic Society (WPES '05), Alexandria, VA, USA.

Lampson, B. W. (1971). Protection. *5th Princeton Symposium in Information Sciences and Systems*.

Lin, T. Y. (1989, 1990). *Chinese Wall Security Policy - An Aggressive Model.* Paper presented at the Conference Proceedings of the Fifth Annual Computer Security Applications Conference, Tucson, Arizona, USA.

Lin, T. Y. (2003, August 4-6, 2003). *Chinese Wall Security Policy Models: Information Flows and Confining Trojan Horses.* Paper presented at the Conference Proceedings of the Seventeenth Annual IFIP WG 11.3 Working Conference on Data and Applications Security, Estes Park, Colorado, U.S.A.

Lipschutz, S., & Lipson, M. L. (2003). *Discrete Mathematics*. New York: McGraw-Hill.

Loock, M., & Eloff, J. H. P. (2005a). *Investigating the usage of the Chinese Wall Security Policy Model for Data Mining.* Paper presented at the The 4th International Symposium on Information and Communication Technologies (Session: WISICT05), Cape Town, South Africa.

Loock, M., & Eloff, J. H. P. (2005b, 29 June 2005 - 01 July 2005). *A new Access Control model based on the Chinese Wall Security Policy Model.* Paper presented at the ISSA2005 New Knowledge Today Conference, Balalaika Hotel, Sandton, South Africa.

Loock, M., Eloff, J. H. P., & Heidema, J. (2012). CBAC: Conflict-Based Access Control. Paper submitted for publication.

Marciniak, J. J. (2001). *Encyclopedia of Software Engineering* (2 ed.).

Mouton, J. (2001). *How to succeed in your Master's & Doctoral Studies*. Pretoria: Van Schaik Publishers.

OASIS. (2010). eXtensible Access Control Markup Language (XACML) TC Retrieved 21 June, 2012, from http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml

Pfleeger, C. P., & Pfleeger, S. L. (2003). *Security in Computing* (3rd ed.). Upper Saddle River, New Jersey 07458: Prentice Hall.

Popper, N. (2011, 08 March). Galleon trial showcases a classic - and dwindling - type of hedge fund, *Los Angeles Times*.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*: Morgan Kaufmann.

Samarati, P., & De Capitani di Vimercati, S. (2001). Access Control: Policies, Models, and Mechanisms. *Lecture Notes in Computer Science*(2171), 137-196

Sandhu, R. S., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based access control models. *Computer*, *29*(2), 38-47.

Treanor, J. (2011, 15 September). Trading tactics, Soc Gen's Jérôme Kerviel and UBS's Kweku Adoboli, *The Guardian*.

APPENDIX A: PROTOTYPE TECHNICAL PLATFORM

A.1 THE TECHNICAL PLATFORM USED FOR CONSTRUCTING THE PROTOTYPE

This prototype was developed in Microsoft C#, using Microsoft Visual Studio 2008. It requires the .NET Framework v3.5 to run, since the program uses a number of features provided by it. Three of the most significant features used are listed below.

A.1.1 LINQ (Language Integrated Queries)

LINQ is a technology that attempts to provide enhanced retrieval features, similar to the abilities of modern relational DBMS queries, to any enumerable data structure (such as Lists, Collections, Queues or any class implementing the IEnumerable interface). This allows the programmer to request any data using syntax similar to SQL. Many of the retrieval methods in the ConflictGraph class use LINQ. An example is provided below, which is used to find all the companies in a specific domain.

```
 public List<Company> GetCompaniesInDomain(Domain d)
{
   var result =
      (from e in Edges where (e.FromNode.Type == NodeType.Domain) && (e.FromNode.Domain == d)
select e.ToNode.Company)
         .ToList();
   return result;
}
```

The above code, which is essentially a single line, is an equivalent of a Foreach loop, which would require several lines of code.

A.1.2 WPF (Windows Presentation Foundation)

The prototype uses WPF for the layout and presentation aspects. It uses a number of objects in WPF format (they have a .XAML extension). The WPF format is an approach to describe the design and layout of a form, which can be used in both Windows Forms and web pages using Silverlight. WPF is an XML application, meaning that it represents a form in a hierarchical manner, using text and markup. The primary WPF object in this program is called “*Explorer.xaml*”, and contains a look and feel not possible with normal windows forms. Note that WPF objects can contain other WPF objects inside them. The following is an extract of one of the WPF objects:

```
<Grid Margin="5,1,5,1" x:Name="LayoutRoot" Height="20" >
        <Grid.Background>
                <LinearGradientBrush EndPoint="0.5,1" StartPoint="0.5,0">
                        <GradientStop Color="#00FFF882" Offset="0"/>
                        <GradientStop Color="#0090270B" Offset="1"/>
                </LinearGradientBrush>
        </Grid.Background>
    <Grid.ColumnDefinitions>
      <ColumnDefinition Width="0"/>
      <ColumnDefinition Width="*"/>
      <ColumnDefinition Width="Auto"/>
      <ColumnDefinition Width="20"/>
    </Grid.ColumnDefinitions>
    <TextBlock x:Name="TxtID" Background="Transparent" Text="ID" Grid.Column="0" Margin="2,0,2,0"
FontSize="14" VerticalAlignment="Center"/>
    <TextBlock x:Name="TxtName" Background="Transparent" Text="Item Name" Grid.Column="1"
Margin="2,0,2,0" FontSize="14" VerticalAlignment="Center"/>
    <TextBlock x:Name="TxtDomain" Background="Transparent" Text="Domain" Grid.Column="2"
Margin="2,0,2,0" FontSize="14" VerticalAlignment="Center"/>
    <Button x:Name="BtnDelete" Content="X" Grid.Column="4" Click="BtnDelete_Click"/>
  </Grid>
```

Notice that many of the items in this object do not have explicitly assigned positions and sizes, as they are automatically scaled to the size of their cell in the grid. This is similar to tables in HTML. This allows the contents of a window to be scaled to any arbitrary size (this feature can be tested by resizing the Browser window).

Another advantage of WPF is that it allows defining of animations through the use of Storyboard and Key Frames, similar to Flash. This feature is used in the program to provide the animated message overlays (such as the "Cannot display this item" message) and Mouseover highlighting.

Finally, each XAML file has a code behind file associated with it, used similarly to the code behind files of ASP web pages. For example, the event handler for *BtnDelete.Click* in the above example is located there.

A.1.3 Object Serialization

The program can load and save the entire *ConflictGraph*, including *Nodes*, *Edges*, *Access Groups*, and their contents, to a single file, by using the Object Serialization provided in the .NET Framework. To achieve this, all classes that need to be saved and loaded are marked with the *Serializable* attribute.

The *ConflictGraph*, which is considered the "root" object to be saved, implements the *ISerializable* interface, and hence must include a special Constructor, and the *GetObjectData* method.

With the above in place, saving and loading is a trivial task, requiring only standard file IO (A *FileStream*, with a *BinaryWriter*). The serialization process is started with the "*BinaryWriter.Serialize(Stream, object)*" method call. Likewise, deserialization is started with the "*BinaryFormatter.Deserialize(Stream)*" method call.

APPENDIX B: PUBLISHED PAPERS

B.1 INTRODUCTION

The following sections describe the conference and journal papers that were published from this research, reflecting the publication title, the citation, the abstract and content of the paper.

B.2 MINIMIZING SECURITY RISK AREAS REVEALED BY DATA MINING

Marianne Loock, Jan H P Eloff: Minimizing Security Risk Areas revealed by Data mining, Proceedings of the Information Security South Africa Conference (ISSA) 2002, Misty Hills, Muldersdrift, Gauteng, edited by Jan Eloff, Les Labuschagne, Mariki Eloff and Hein Venter.

***Abstract**: The aim and intend of Data Mining is knowledge discovery which means Data Mining finds 'knowledge' that is otherwise hidden by large volumes of data. This now points to a potential security risk - if the knowledge is hidden, how do we know that a security risk exists. Unexpected security issues may arise during the data mining process. Interesting is where the security of individual data items is not a concern, but there may be patterns in the mined data that pose a security risk. The aim of this research is to map the data mining process to the five information security services and to minimize information security risk by concentrating on each mapped area.*

The draft article is reflected below:

MINIMIZING SECURITY RISK AREAS REVEALED BY DATA MINING

Introduction

This article will present an example where data mining causes an information security risk. The question will then be asked: What can one do about information security risk problems that surfaces while doing 'normal and legal' data mining? Concentrating on this identified information security risk problem, the proposed answer is mapping the data mining process to the five information security services.

Data Warehouses and Data Mining

Turning the large volumes of data, data about everything around us, into information and knowledge is a necessity. The information and knowledge now gained can be used for applications such as business management, market analysis, and science exploration.

Some time ago, during the late 1980's, an architecture for databases started to grow namely data warehouses. A data warehouse refers to a database that is maintained separately from an organization's operational databases. It collects information about subjects across an entire organization, which means its scope is enterprise-wide. It allows for the integration of a mixture of application systems and it supports information processing by providing a reliable platform of consolidated historical data for analysis. A short and comprehensive definition for a data warehouse is: 'A data warehouse is a *subject-oriented, integrated, time-variant,* and *non-volatile* collection of data in support of management's decision making process' [Inmon, 1996]. For the purposes of this definition *subject-oriented* means organized around major subjects such as customer or sales; *integrated* means combining multiple heterogeneous sources, such as relational databases, flat files, and on-line transaction records; *time-variant* means data are stored to provide information from a historical perspective (e.g., the past 6-12 years); *non-volatile* means a data warehouse is a physically separate store of data deducted from the application data found in the operational environment and for this reason no transaction processing, recovery or concurrency control mechanisms are required except for two operations in data accessing namely initial loading of data and access of data.

To understand data warehouses better, the following is also important. Data warehouse technology includes, amongst others, data cleaning, data integration, and On-Line Analytical Processing (OLAP). OLAP is analysis techniques with functionalities such as summarization, consolidation, and aggregation, as well as the ability to view information from different angles. Although OLAP tools support multidimensional analysis and decision making, additional data analysis tools are required for in-depth analysis, such as data classification, clustering, and the characterization of data changes over time.

Data cleaning and data integration is done when constructing a data warehouse, which can be viewed as an important pre-processing step for data mining.

This now brings us to the next question namely: What is data mining?

Data mining fundamentals

The concept of data mining and what it entails now needs some explanation. Data mining is a logical concept built on already existing fields, techniques, and tools and on the lowest level it is applied to data. Data mining refers to extracting or 'mining' knowledge from large amounts of data. This mining process uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. Sometimes data mining is treated as a synonym for another popular term, Knowledge Discovery in Databases, or KDD [Han J & Kamber M, 2001]. Another view is that data mining is simply an essential step in the process of knowledge discovery in databases.

Knowledge discovery as a process is pictured in Figure 1 and consists of an interactive sequence of the following steps:

1. Define the business problem (Make a clear statement of your objectives, which includes a way of measuring the results of the knowledge discovery process. The business problem may also include a cost justification.).
2. Build a mining database (1st out of 3 Data Preparation steps).
    a. Data collection (identify possible multiple sources of data that will be mined)[1]
    b. Data description (describe contents of each file or database table).

[1] Be sure to make note of special security and privacy issues that your mining database will inherit from the source data.

c. Data selection (this is a gross elimination of irrelevant or unneeded data).

d. Data cleaning and data quality assessment (to remove noise and inconsistent data).

e. Data integration and consolidation (combining data from different sources into a single mining database).

f. Metadata construction (this is a database about the mining database).

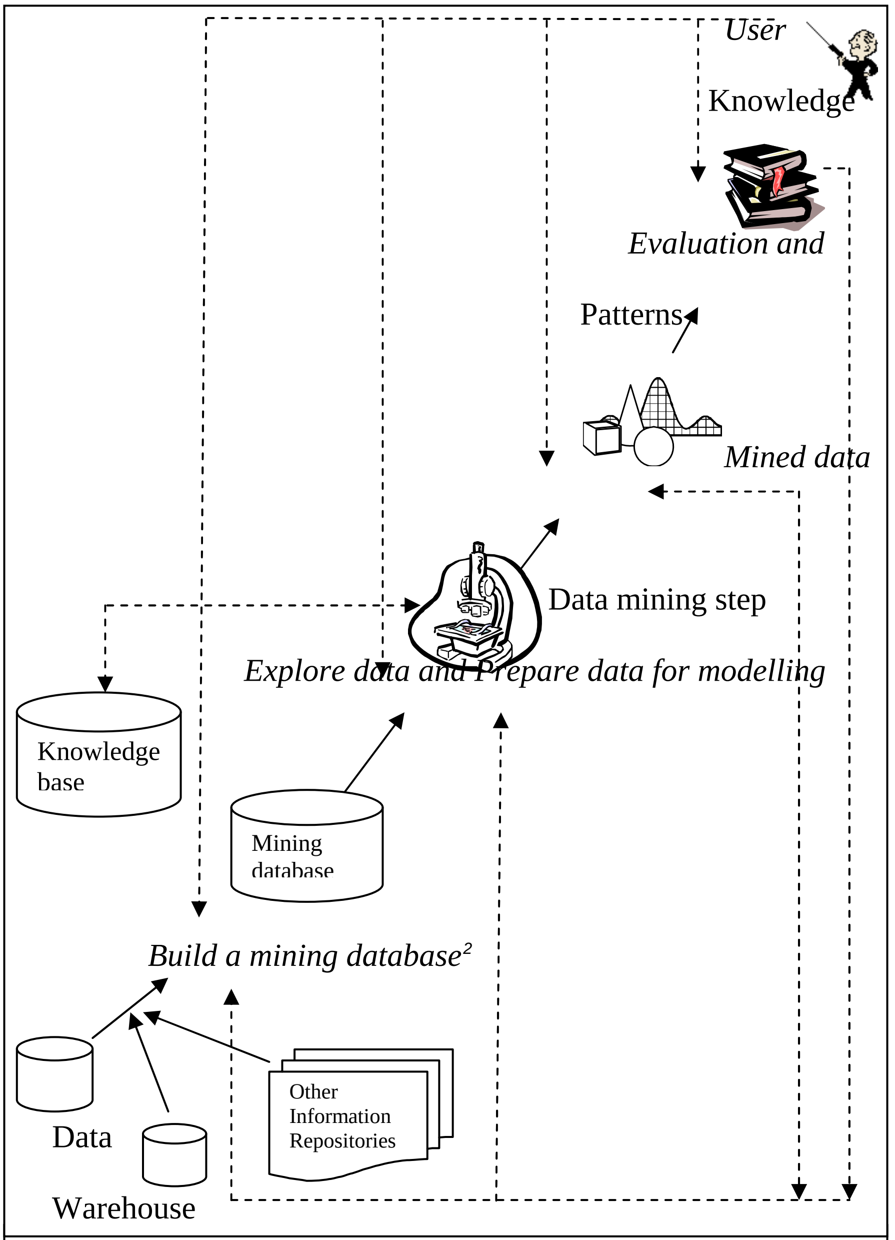


Figure 1: Data mining as a step in the process of knowledge discovery

[2] Build a mining database = data collection, data description, data selection, data cleaning, data integration, metadata construction, load mining database and maintain mining database

   g. Load the mining database.
   h. Maintain the mining database (needs to be backed up, performance monitored, and reorganized).
3. Explore the data (identify most important fields in predicting an outcome, and determine which derived values may be useful). (2nd out of three Data Preparation steps).
4. Prepare data for modelling. (3rd out of three Data Preparation steps).
   a. Select variables (delete irrelevant variables to increase time building a model).
   b. Select rows (delete irrelevant variables to increase time building a model).
   c. Construct new variables (some variables that extend over a wide range may be modified to construct a better predictor, such as using the log of income instead of income).
   d. Transform variables (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)
5. Data mining step (an essential process where intelligent methods are applied in order to extract data patterns and build models in solving business problems). This step may interact with the user or a knowledge base.
6. Pattern evaluation and interpretation (to identify the truly interesting patterns representing knowledge, based on some interestingness measures).
7. Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user).

From this one can see the truth in data mining being a step in the knowledge discovery process but for the purposes of this research the term data mining will be used for the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories.

Information Security Issues

It is now appropriate to ask the question: ‘Where does Information Security fits into all of this?’ When following the Data Mining process as described earlier, there is a noticeable collection of Information Security issues that cannot be ignored. Some of these issues will now be demonstrated by means of an example based on work done by Chris Clifton [Clifton, C. 1998].

Example: Risks posed by patterns in the data of top business enterprises, as mined by an external travel agent - Prediction of sensitive information.

Suppose, for the top business enterprises in South Africa, major corporate announcements required a face-to-face meeting of top and senior management from various locations throughout the country. In addition, negative announcements required participation of top and senior management as well as senior public relations staff throughout the country. Travel records (likely made available to an external travel agent) could then be used to predict the occurrence (‘When will the next major corporate announcements be made?’) and type (‘positive or negative announcements’) of corporate announcements. Here the individual travel records are not a concern, but their correlation with past announcements poses a risk.

The external travel agent will be able to answer the following questions with great ease:

‘When will the next major corporate announcements be made?’

After all the top, senior management, and senior public relations staff has travelled to head office.

‘Will the next corporate announcement be positive or negative?’

It will be positive if only the top and senior management was brought to head office but it will be negative if the senior public relations staff was also brought to head office.

This emphasizes the statement that one should look at the Data mining process and establish the stages that poses an information exposure risk. To be able to establish this risk we must first look at Information Security, as defined in the ISO 7498-2 standard, produced by the International Standards Organization (ISO) [ISO, 2002] and more specific at the Information Security Services.

Information Security Services

If we want to enforce security on information, we must be able to 'measure' all actions on the information. This 'measurement' can be done in five steps [Von Solms & Eloff, 1999], also called the five pillars of Information Security. The five pillars are: Identification and Authentication, Authorisation, Confidentiality, Integrity, and Non-denial.

1. Identification and Authentication is the first step towards enforcing information security. First a person that wants to perform a transaction must be identified and during this identification process, this person must also be authenticated, to ensure that somebody else did not provide the claimed identification.
2. Authorisation (also called logical Access Control) is the second step towards enforcing information security. After a person has been identified and authenticated, one must check whether this person has the access right to the requested resource for example, a transaction, a program, a file, or a database.
3. Confidentiality is the third step towards enforcing information security. The assurance that only authorized people may view the contents of the data or software is called protecting the confidentiality of the data or software.
4. Integrity is the fourth step towards enforcing information security. The assurance that only authorized people may change the contents of the data or software is called protecting the integrity of the data or software.
5. Non-denial (or non-repudiability) is the fifth step towards enforcing information security. After changing the contents of the data or software a person must be forced to 'sign' this change so that he cannot at a later stage deny the fact that he made the change.

The next table maps the data mining process onto the five Information Security Services and reveals important security risk areas.

How to read the table:

- The first column of this table explains the data mining process step-by-step by making use of the above-mentioned example.
- The next five columns represent the five Information Security Services.
- Each row, starting with the number one (1), through to seven (7), concentrates on a specific step in the data mining process.
- Each row indicates all the information security risk areas that are present per data mining step by mapping each data mining step to all five Information Security Services. The indication is done by using a tick, for example - √.
- Some of these information security risk areas that are indicated by a number next to the tick is covered in the explanation that follows. The number used in the table, for example - √(5) refers to the fifth explanation of these information security risk areas.

The Data Mining process mapped onto the five Information Security Services

Data mining process (steps) from the external travel agent's offices and *examples specific to the illustration.*	**Information Security Services**				
	Identification and Authentication	**Authorisation (Logical Access control)**	**Confidentiality**	**Integrity**	**Non-denial**
1. Define the business problem. Express this in business terms. ***Example****: Structure a*	√		**√(1)**		

Data mining process (steps) from the external travel agent's offices and *examples specific to the illustration.*	Information Security Services				
	Identification and Authentication	Authorisation (Logical Access control)	Confidentiality	Integrity	Non-denial
Business Class flight-package for the Management of the top business enterprises.					
2. Build a mining database in other words: Identify a dataset that answers the business question. ***Example****: The names of the top business enterprises, the names of their top and senior management as well as the senior public relations staff, each name's current location, nearest airport and travel-record.*		√(2)	√	√	√
3. Explore the data, in other			√	√(3)	

Data mining process (steps) from the external travel agent's offices and *examples specific to the illustration.*	**Information Security Services**				
	Identification and Authentication	**Authorisation (Logical Access control)**	**Confidentiality**	**Integrity**	**Non-denial**
words: Make some preliminary investigations whether the dataset answers the business question. ***Example****: Will it be possible to build up a history of a specific business enterprise's flight habits and routines with the above-mentioned facts?*					
4. Prepare the data in a format as required by the data mining technique. ***Example****: De normalization or shifting to another platform.*			√	√	**√(4)**
5. Mine the data. ***Example****: Use various data mining techniques to explain the*	**√(5)**	√	√	√	

Data mining process (steps) from the external travel agent's offices and *examples specific to the illustration.*	**Information Security Services**				
	Identification and Authentication	**Authorisation (Logical Access control)**	**Confidentiality**	**Integrity**	**Non-denial**
trends in the dataset.					
6. Do the pattern evaluation in other words: Check whether the business question is answered. ***Example***: *Do we have the evidence for an answer as raised? Do we have statistical convincing evidence that a specific business enterprise functions in a specific manner?*			**√(6)**	√	√
7. Knowledge presentation in other words: Explain the question and answer in simple terms. ***Example***: *Business enterprise X's management tends to travel*			**√(7)**	√	√

Data mining process (steps) from the external travel agent’s offices and *examples specific to the illustration.*	**Information Security Services**				
	Identification and Authentication	**Authorisation (Logical Access control)**	**Confidentiality**	**Integrity**	**Non-denial**
with Business Class tickets twice a year, once during August and once during March, with a tendency to add their senior public relations staff, also Business Class tickets, every now and then.					

Some security risk areas, which are revealed by data mining, will now be explained for the above-mentioned example:

1. The confidentiality of the fact that a data mining exercise is going to take place as well as the outcome of such an exercise must be protected. The assurance that only authorized third party people may view the contents of the data is important when talking about information exposure risks.
2. When building the mining database one must follow a few ‘data changing and maintenance’ steps (for example data collection, data description, data selection, data cleaning, data integration, metadata construction, load mining database and maintain mining database). The people that have to change and maintain this data must be authorized to do so. They must have the access right to the requested data.

3. When working on this step one will not change the data but one might decide to change the collection of facts to be able to answer the correct business question. This also needs assurance that only authorized people will change the contents of the facts collection and by doing so still protect the integrity of the data.
4. After preparing this data in the required format, it must be ‘signed’ so that it is possible to trace the person who has done this preparation. This will eliminate any false accusations and statements with the preparation phase.
5. When starting with the different model building techniques and pattern evaluation techniques, one needs to know who is working with the data.
6. At this stage interesting patterns are being evaluated and compared. Are the right people (only authorized people) doing the evaluation?
7. The question and answer now needs some explanation. Are the right people (only authorized people) looking at the results?

Conclusion

During the data mining process one works intensively with the data involved. The data mining process also expects and allows certain data changes to be made. These initial steps of the data mining process (namely seeing and changing data) cause an information exposure risk. The data mining process as a whole also causes an information exposure risk by making hidden knowledge known.

By agreeing to the fact that a third party may do data mining on an enterprise’s data is a step that must be taken with great care and only if one knows that all the data mining steps in the data mining process, that poses an information exposure risk when cross referenced with the five information security services, has been taken care of.

This research mapped the seven-step data mining process against the five information security services to be able to concentrate on all the information exposure risk areas that are revealed by doing data mining on an enterprise’s data. The rest of this research will concentrate on how to minimize the information exposure risk when doing data mining.

References

Berry Michael J and Linoff Gordon with Linoff Gordon S, Data Mining Techniques: For Marketing, Sales, and Customer Support; Wiley, John & Sons, Incorporated, 1997.

Clifton Chris, Security issues in data warehousing and data mining: panel discussion, in Database Security XI: Status and prospects edited by T Y Lin and Shelly Qian; Chapman & Hall, 1998.

Han Jiawei and Kamber Micheline, Data Mining: Concepts and Techniques; Academic Press, 2001.

Inmon W.H., Building the Data Warehouse; New York: John Wiley & Sons, 1996.

ISO 2002, International Standards Organization, http://www.iso.ch, May 2002

Mattison Rob, Data Warehousing and Data Mining for Telecommunications, Artech House, Incorporated, 1997.

Pazzani Michael J, “Knowledge discovery from data?” article in IEEE Intelligent Systems and their applications: Data Mining II, March/April 2000, pp. 10-13.

Shi Yong, “Data mining” article in The IEBM Handbook of Information Technology in Business edited by Milan Zeleny, 2000, pp. 490–495.

Van Maanen Tom, www.van-maanen.com, April 2002

Von Solms Sebastiaan H. and Eloff Jan H. P., Information Security; Von Solms & Eloff, 1999.

B.3 INVESTIGATING THE USAGE OF THE CHINESE WALL SECURITY POLICY MODEL FOR DATA MINING

Marianne Loock, Jan HP Eloff: Investigating the usage of the Chinese wall security policy model for data mining, ACM International Conference Proceeding Series; Vol.92. Proceedings of the 4th international symposium on Information and communication technologies, Cape Town, South Africa. SESSION: Invited workshop on Information technology and its applications: software development, disaster engineering, and security. Pages 243-248. Year of publication: 2005. ISBN: 1-59593-169-4

***Abstract**: Access control requirements for the protection of information/data objects in an information system environment are well defined and successfully implemented by means of access control policies and models. Whenever these objects are used in a data-mining environment, a change in the access control requirements becomes necessary. During a data-mining activity the data miner may expose unexpected results or trends. It is important for all companies involved in data-mining activities to be aware of these potential access control problems.Brewer and Nash (1989) first defined the Chinese Wall Security Policy model (CWSP model). It provides access control for the commercial environment based on conflict of interest classes. Shortly after the introduction of this model, Lin (1989) reported an error in it and presented a modified version called the Aggressive Chinese Wall Security Policy model (ACWSP model). This model introduced the concept of an overlap between conflict of interest classes. It became evident that these two models had the potential to be used in a data-mining environment.*

The draft article is reflected below:

INVESTIGATING THE USAGE OF THE CHINESE WALL SECURITY POLICY MODEL FOR DATA MINING

1 Introduction

Data mining is important in obtaining underlying and essential information on competitors and clients. This activity involves the exploration and analysis of large masses of data in order to discover meaningful patterns and rules, which means data mining provides the company with intelligence [Berry & Linoff, 1997]. The information and knowledge gained during this activity can be used for applications such as business management, market analysis and science exploration. The data-mining process does not stop the moment that the results become available. The results, in whatever format they are presented, e.g. types of graphs or text, may create an information security risk problem.

The following example illustrates this problem: Airline Company A bought shares from Petroleum Company D for $10 per share. Airline Company A and Petroleum Company D use the same financial institution, namely DM Bank. DM Bank does data mining for prediction and estimation purposes [Berry & Linoff, 1997] using a 'decision tree' technique [Shi, 2000] & [Berry & Linoff, 1997]. From this data-mining exercise DM Bank detects a negative growth in Petroleum Company D's cash flow. Petroleum Company D is exposed to an unwanted information security risk problem, through a possible information leakage – DM Bank could give this sensitive information to Airline Company A. It is possible that this information leakage could damage Petroleum Company D if the parties involved (DM Bank and Petroleum Company D) do not adhere to mutual confidentiality agreements.

DM Bank should consider access control as a means of solving this information leakage problem. There are different data-mining issues to be considered by DB Bank: for example, 'What is DM Bank allowed to do after the results of the data-mining process are revealed?' and 'Who can gain access to the mined data?' To manage and control solutions to these questions, DM Bank needs a Security Policy. This Security Policy must be implemented by security models, for instance file access control models, that can deal with the identified threat. For the purposes of this paper, existing access control models for the commercial environment will be considered with the ultimate aim of determining whether or not they are suitable for solving the data-mining information leakage problem. The Chinese Wall Security Policy model (CWSP model) and the Aggressive Chinese Wall Security Policy model (ACWSP model) are discussed.

The remainder of this paper is structured as follows. In section 2 security policies in general are discussed, with emphasis on the CWSP model in section 2.1 and the ACWSP model in section 2.2. Requirements for a Security Policy model used in a data-mining environment are proposed in section 3.

2 Security Policies

Security policy research reveals that the first such policy to be formally defined was the Military security policy, succeeded by the Bell-LaPadula [Bell & LaPadula, 1973, 1975]. In 1987, Clark and Wilson [Clark & Wilson, 1987] highlighted the importance of commercial security policy models. They claimed that the needs of the commercial community are just as important as the needs of the military community. Furthermore they emphasized that the problems of the commercial community are diverse and therefore require their own security policy models. All of these models (Military security policy, Bell-LaPadula and Clark and Wilson) were designed to operate in a well-defined environment, ranging from a strict military environment to a commercial environment.

In a commercial environment, access control models provide access to employees, based on variables such as the job definitions. In a data-mining environment this approach becomes problematic as most activities take place in an abstract world as opposed to the real world where access to data is controlled. This research project focuses on data-mining activities in a commercial environment. A literature study of access control models designed for commercial environments indicated that the CWSP and ACWSP models showed potential for addressing the access control requirements. The remainder of this paragraph gives an overview of the CWSP model and the ACWSP model.

2.1 Chinese Wall Security Policy Model

Brewer and Nash [Brewer & Nash, 1989] introduced the Chinese Wall Security Policy model. It defines the access right to data to which the data miner (*actor*) already holds title. This model makes use of *actors* and *objects*. It develops a set of rules aimed at preventing people from accessing data (objects) on the wrong side of a wall.

All corporate information, DM Bank’s information, is stored in a hierarchy as shown in Figure 1. There are three levels of significance:

1. At the lowest level *individual objects* are considered, containing specific data items, each concerning a specific company.
2. At the intermediate level all *objects concerning the same company* are grouped together. These are referred to as ‘company data sets’.

3. At the highest level all company data sets whose corporations are in competition (conflict), are grouped together. They are referred to as 'conflict of interest classes' *(*CIR*)*.

Each object is associated with the name of the data set of the company to which it belongs and the name of the conflict of interest class to which that company belongs. For example, the conflict of interest class names could be the business sector headings found in stock exchange listings (Petroleum Companies, Airline Companies and Food Companies) while the individual companies could be the names of companies listed under those headings. Thus, if the data-mining environment of DM Bank, as an example, contains information on Airline Company A, Petroleum Company D and Food Company G:

1. All objects would belong to one of three company datasets – company dataset Airline Company A or company dataset Petroleum Company D or company dataset Food Company G and
2. there would be three conflict of interest classes, one for Airlines (containing Airline Company A, B and C's data sets), one for Petroleum Companies (containing Petroleum Company D and E's data sets), and one for Food Companies (containing Food Company F, G, H and J's data sets).

The basis of the Chinese Wall Security Policy model is that people are only allowed access to information that is not in conflict with any other information that they already possess. When considering the DM Bank scenario and assuming that data-mining activities have already been done, a *new* user (data miner) may freely choose to access whatever data sets he/she chooses. As far as the CWSP model is concerned, a new user does not possess *any* information at this stage and therefore no conflict of interest can exist. However, such a conflict may arise later on.

Suppose a data miner accesses the Airline Company A data set first; at this stage he/she possesses information concerning the Airline Company A data sets. Since Airline Company B is in conflict with Airline Company A, he/she will therefore not be granted access to Airline Company B's data sets. However he/she will be permitted to request access to Petroleum Company D's data sets, because the data sets of these two companies belong to different conflict of interest classes. According to the CWSP model this is permissible because the Airline Companies and the Petroleum Companies are not in competition and therefore they are in different conflict of interest classes. However, when working on a *higher level* called the *data-mined level* as in the DM Bank example, this should not be allowed. Knowing information pertaining to Petroleum Company D (namely the negative change in Petroleum Company D's cash flow situation mentioned in a previous paragraph) and making it known to Airline Company A (which has shares in Petroleum Company D) puts Petroleum Company D at risk. It is therefore inappropriate for DM Bank to use the CWSP model because the information leakage problem will not be solved.

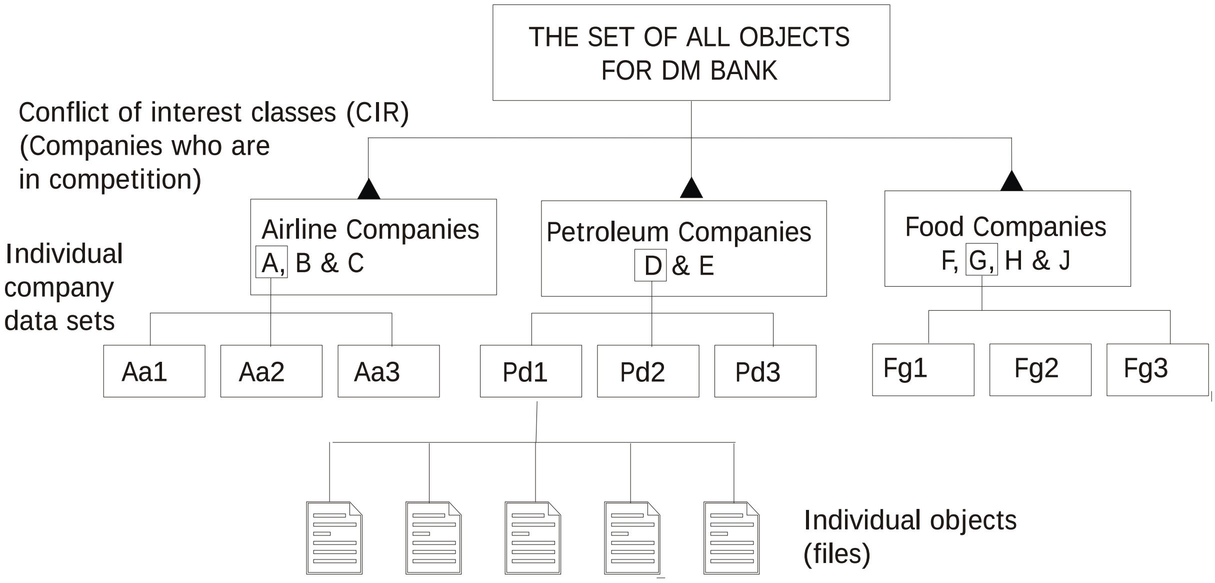

THE SET OF ALL OBJECTS FOR DM BANK
Conflict of interest classes (CIR) (Companies who are in competition)
Individual company data sets
Airline Companies A, B & C
Petroleum Companies D & E
Food Companies F, G, H & J
Aa1
Aa2
Aa3
Pd1
Pd2
Pd3
Fg1
Fg2
Fg3
Individual objects (files)

Figure 1: The composition of objects

The following issues are important:

- *CIR classes can and may overlap* - this is to accommodate the conflict of interest between Airline Company A and Petroleum Company D.

- The *severity* of the *conflict* that exists between two companies must be definable. There may be a *conflict of interest* between Airline Company A and Food Company G but it may be so small that it needs no extra mention and should not inhibit access capabilities.

- Another requirement that a Security Policy model must adhere to is to be *dynamic*. If Airline Company A sells all its shares in Petroleum Company D they are no longer in conflict and must be treated as such.

The problem caused by the fact that the *CIR classes can and may overlap* in the CWSP model was addressed by Lin [T Y Lin, 1989] in the ACWSP model. An overview of the ACWSP model and an example addressing this problem are given in section 2.2.

2.2 The Aggressive Chinese Wall Security Policy Model

The CWSP model builds a collection of impassable walls, called ‘Chinese walls’, around the data sets of competing companies. No data that are in conflict can be stored on the same side of the Chinese walls. According to Lin [T Y Lin, 1989; Tsau Young Lin, 2003] the Brewer-Nash model was based, amongst other factors, on the *incorrect assumption* that corporate data can be grouped into separate and disjoint conflict of interest classes (CIR classes). For example, an incorrect assumption would be that all Airline Companies must be grouped into one, and only one, conflict of interest class. CIR classes are seldom disjoint - they do overlap.

Lin suggested a modified model called an Aggressive Chinese Wall Security Policy model (ACWSP model). This theory is based on the development of a methodology called ‘Granular Analysis and Computing’ [Tsau Young Lin, 2003].

To illustrate the error in the CWSP model’s assumption that the set of all objects of DM Bank could be partitioned into mutually disjoint CIR classes, let us revisit the example in Figure 1. Airline Company A is in the conflict of interest class for all Food Companies. Petroleum Company D is in the conflict of interest class for all Petroleum Companies. This means Airline Company A and Petroleum Company D have no conflicting interests but this is not true because of the *shares* between the two companies. This also means that Airline Company B and Petroleum Company D have no conflicting interest, which is true in this case. These two CIR classes are distinct but they overlap, as depicted in Figure 2.

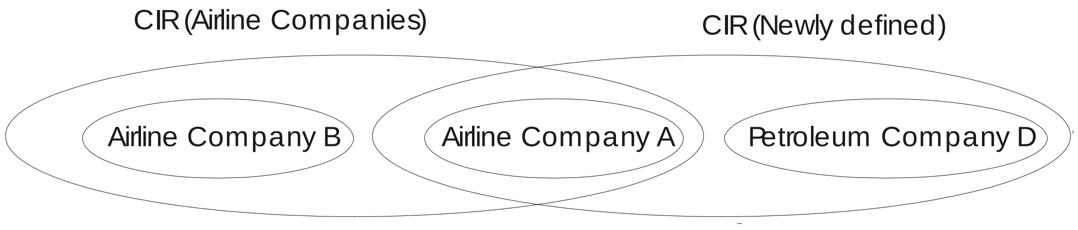


Figure 2: CIR classes are distinct but they overlap

The ACWSP model puts DM Bank in the position where friends and enemies can be defined. This is facilitated by means of a discretionary access control (DAC) approach [Carroll, 1996]. It implies that Airline Company A and Petroleum Company D can be defined as companies in the same CIR class. Furthermore, it also implies that the ACWSP model will satisfy the problem of the mutually disjoint CIR classes that existed in the DM Bank environment, when using the standard CWSP model.

The primary problems are as follows:

- The *severity* of the *conflict* that exists between two companies must be definable. There may be a *conflict of interest* between Airline Company A and Food Company G but it may be so small that it needs no extra mention and should not inhibit access control.
- Another requirement that a Security Policy model must adhere to is to be *dynamic*. If Airline Company A sells all its shares in Petroleum Company D they are no longer in conflict and must be treated as such.

It is important that the Security Policy model used when doing data mining in a commercial environment must be *dynamic* to accommodate an unpredictable environment that can change rapidly.

3 Requirements for a Security Policy Model when used in a Data-Mining Environment

A Security Policy model for data mining in a commercial environment must be able to deal with a changing environment, and also take into consideration that the results of a data-mining activity can be unpredictable. Looking at the example discussed, a Security Policy model for the data-mining environment should take the following requirements into consideration:

1. Facilitating the definition of *conflict of interest* classes, as defined by the CWSP model, should be a fundamental part of such a policy. In the example discussed this implies that all Petroleum Companies (or Airline Companies) are in conflict.
2. Facilitating the definition of CIR-classes not being mutually disjoint, as defined by the ACWSP model, is important. In the example discussed this

implies that Airline Company A and Petroleum Company D can be in the same CIR class because of their business relationship.

3. The *severity* of *conflict of interest* must be quantifiable on a pre-defined scale. When looking at the example, the Security Policy model must be able to define the *severity* of the *conflict of interest* between Airline Company A and Petroleum Company D as a specific number that high, according to a definition, and needs special attention. There may also be a *conflict of interest* between Airline Company A and Food Company G, but it may be so small that it needs no extra mention and should not inhibit access capabilities.

4. A Security Policy model must be *dynamic*. It must be able to cope with a rapidly changing organizational environment. In particular, data mining must be able to reflect organizational changes. If Airline Company A sells all its shares in Petroleum Company D they are no longer in conflict and may ask to be treated as such and request that this be reflected in determining the access capabilities of data miners. Mutual areas of interest may develop into a new *conflict of interest* class. The Security Policy model must be able to handle this change the moment it occurs. In our example it is true that Airline Company A and Food Company G are not in conflict, but a possible mutual interest in Petroleum Company E may develop into a new CIR class. CIR classes similar to this example must immediately be definable.

5. A Security Policy model must preserve confidentiality, integrity and availability in a data-mining environment. Each company (or site) has its own expectations for levels of confidentiality, integrity and availability. These expectations must stay the same before and after a data-mining activity has been performed.

As seen from the example, the CWSP and ACWSP models do not adhere to all of these requirements. For further research this project will focus on the implementation of these requirements in a Security Policy model for commercial environments where data-mining activities take place.

Conclusion

There is a difference between working with data in the real world and working with mined data, especially when it comes to defining security policies. When working with mined data, it may be impossible to predict whether or not a security risk area will emerge. It is thus important to have procedures in place for when such a security risk area suddenly materializes so as to minimize the risk as far as possible.

In this article a discussion of the CWSP model was followed by a summary of the ACWSP model. Both these models proved to have some limitations when applied to mined data. A list of requirements was proposed, which should be used when defining a Security Policy model for commercial environments where data-mining activities take place.

Rerefences

Bell, D. and LaPadula, L. (1973) MITRE Corporation, Bedford, MA.

Bell, D. and LaPadula, L. (1975) MITRE Corporation, Bedford, MA.

Berry, M. J. A. and Linoff, G. (1997) *Data Mining Techniques: For Marketing, Sales, and Customer Support,* John Wiley & Sons, Incorporated.

Brewer, D. F. C. and Nash, M. J. (1989) In *IEEE Symposium on Security and Privacy*Oakland, pp. 206-214.

Carroll, J. M. (1996) *Computer Security,* Butterworth-Heinemann, Newton.

Clark, D. and Wilson, D. (1987) In *IEEE Symposium on Security and Privacy*, pp. 184-194.

Lin, T. Y. (1989) In *Fifth Annual Computer Security Applications Conference*, pp. 282-289.

Lin, T. Y. (2003) In *Seventeenth Annual IFIP WG 11.3 Working Conference on Data and Applications Security*Estes Park, Colorado, U.S.A.

Shi, Y. (2000) In *The IEBM Handbook of Information Technology in Business*(Ed, Zeleny, M.), pp. 490-495.

B.4 A New Access Control Model based on the Chinese Wall Security Policy Model

Marianne Loock, Jan H P Eloff: A New Access Control Model based on the Chinese Wall Security Policy Model, Proceedings of the Information Security South Africa Conference (ISSA) 2005, Balalaika Hotel, Sandton, South Africa, edited by HS Venter, L Labuschagne, MM Eloff. ISBN: 1-86854-625-X

***Abstract**: Access control policies and models successfully implement well defined access control requirements, which are used for the protection of information or data objects in an information system environment.*

Whenever these objects are used in a data mining environment, a change in the access control requirements becomes necessary. During a data mining activity the data miner may expose unexpected results or trends. It is important for all companies involved in data mining activities to be aware of these potential access control problems. Security policies however, can help to resolve this problem. Brewer and Nash (1989) first defined the Chinese Wall Security Policy model (CWSP model). It provides access control for the commercial environment based on conflict of interest classes. Shortly after the introduction of this model, Lin (1989) reported an error in it and presented a modified version called the Aggressive Chinese Wall Security Policy model (ACWSP model). This model introduced the concept of an overlap between conflict of interest classes. When investigating the access control requirements necessary for a data mining environment, it became evident that these two models had the potential, but not fully comply with the requirements.

The purpose of this article is to discuss a new access control model, based on the Chinese Wall Security Policy model, for a data mining environment. This new access control model will address the access control requirements not addressed in the current existing models.

By making use of this new access control model, it will be possible for data miners to work on different company information or data objects without causing access control problems. For example, information leakage problems, after being exposed to unexpected results or trends. All companies involved in data mining activities are in a position of controlling their own level of exposure amongst competitive peer companies, when using the proposed

access control model. This new access control model is dynamic. It copes with a rapidly changing business environment.

The draft article is reflected below:

A NEW ACCESS CONTROL MODEL BASED ON THE CHINESE WALL SECURITY POLICY MODEL

1 Introduction

In general, access control policies and models successfully implement well defined access control requirements, which are used for the protection of real-world information or data objects in an information system environment. When data mining is done on such protected real-world information or data objects, the implementation of the current access control policies and models might not be enough to protect the results of such data mining activities.

Data mining is a well defined and structured activity, which obtains derived information from large masses of basic or core data. It discovers meaningful patterns and rules, which means data mining provides an organisation with intelligence (Berry and Linoff, 1997). This derived information or intelligence is essential for organisations to be competitive. It can be used in areas such as business management, market analysis and science exploration. The data mining process does not stop the moment that the derived information becomes available. This derived information, in whatever format they are presented, for example types of graphs or text, may create an information security risk problem by revealing new and unexpected information. Access control policies and models, for example security policy models, should now be able to improve the secure state of this unexpected information.

Security policy and model research reveals that the first security policy model to be formally defined was the Bell-LaPadula [Bell & LaPadula, 1973], describing the properties of the Military security policy model (DoD, 1983). Clark and Wilson (1987) highlighted the importance of commercial security policy models and two years later Brewer and Nash (1989) defined the Chinese Wall Security Policy (CWSP) model. In the same year the Aggressive Chinese Wall Security Policy (ACWSP) model (Lin, 1989) for a commercial environment was defined. All of these models were designed to operate in a well-defined environment, ranging from a strict military environment to a well-defined commercial environment, unfortunately not addressing access control problems related to derived information such as in a data mining environment.

This article suggests new definitions to be added to an existing access control model, the CWSP model. The new access control model will be functional in a commercial environment where data mining activities take place and it will solve the data mining information leakage problem. The remainder of the paper is structured as follows. The next section describes the access control problem by means of a banking example. The subsequent section shows why the CWSP model does not solve the access control requirements of DM Bank. Then the following section explains a new access control model for a data mining environment. Final section concludes the paper.

2 Understanding the problem by means of an example

The following example illustrates the access control problem that exists when working in a commercial environment where data mining activities takes place.

DM Bank (Figure 1) is a financial institution with hundreds of clients all within the business sector. DM Bank also has a section with data miners, doing data mining on all DM Bank databases for prediction, marketing and risk management reasons (Berry and Linoff, 1997) amongst others using a ‘decision tree’ technique (Berry and Linoff, 1997, Shi, 2000). DM Bank’s data miners that are busy with data mining activities in the set of companies (clients of DM Bank) are called data miners for DM Bank. Airline Companies A, B and C; Petroleum Companies D and E; and Food Companies F, G, H and J are all clients of DM Bank.

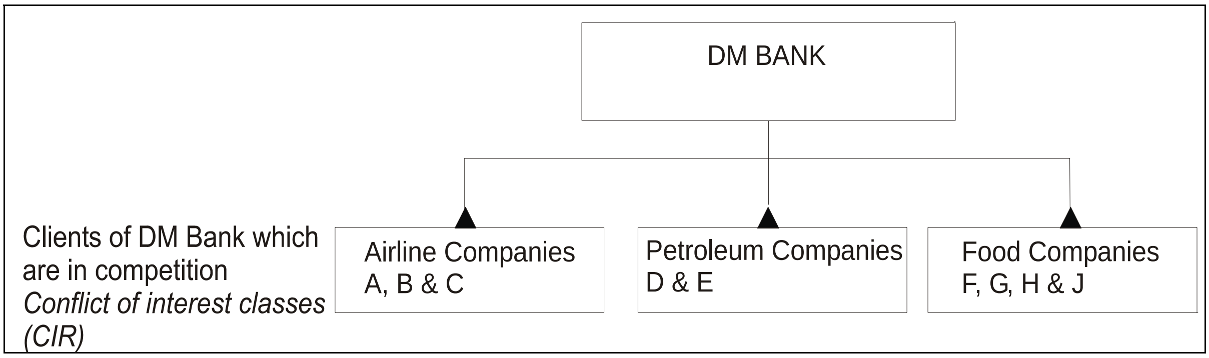

DM BANK
Clients of DM Bank which are in competition
Conflict of interest classes (CIR)
Airline Companies A, B & C
Petroleum Companies D & E
Food Companies F, G, H & J

Figure 1: DM Bank

Assume now that Airline Company A bought shares from Petroleum Company D for $10 per share. From its data mining exercise, DM Bank detects a negative growth in Petroleum Company D's cash flow. Petroleum Company D is exposed to an unwanted information security risk problem, through a possible information leakage – DM Bank could 'leak', by accident this sensitive information to Airline Company A. It is possible that this information leakage could damage Petroleum Company D if the parties involved (DM Bank and Petroleum Company D) do not adhere to mutual confidentiality agreements.

DM Bank should re-consider access control as a means of solving this information leakage. There are also different data mining issues to be considered by DB Bank for example, 'What is DM Bank allowed to do after the results of the data mining process are revealed?' and 'Who can gain access to the mined information?' To manage and control solutions to these questions, DM Bank needs a security policy that can be applied on an environment, which consists of already mined information. This implies that a security policy must be implemented using security models, for instance, file access control models that can deal with the identified threat.

The current CWSP model, as well as the ACWSP model, are two access control models that may be considered for this problem.

3 Why the CWSP model does not solve the access control requirements of DM Bank

The CWSP model makes use of actors (data miners) and objects (information or data objects). It develops a set of rules aimed at preventing people from accessing information or data objects on the wrong side of a wall. This model defines the access right to information or data objects to which the actor (data miner) already holds title. All corporate information, DM Bank's information, is stored in a hierarchy as shown in Figure 2.

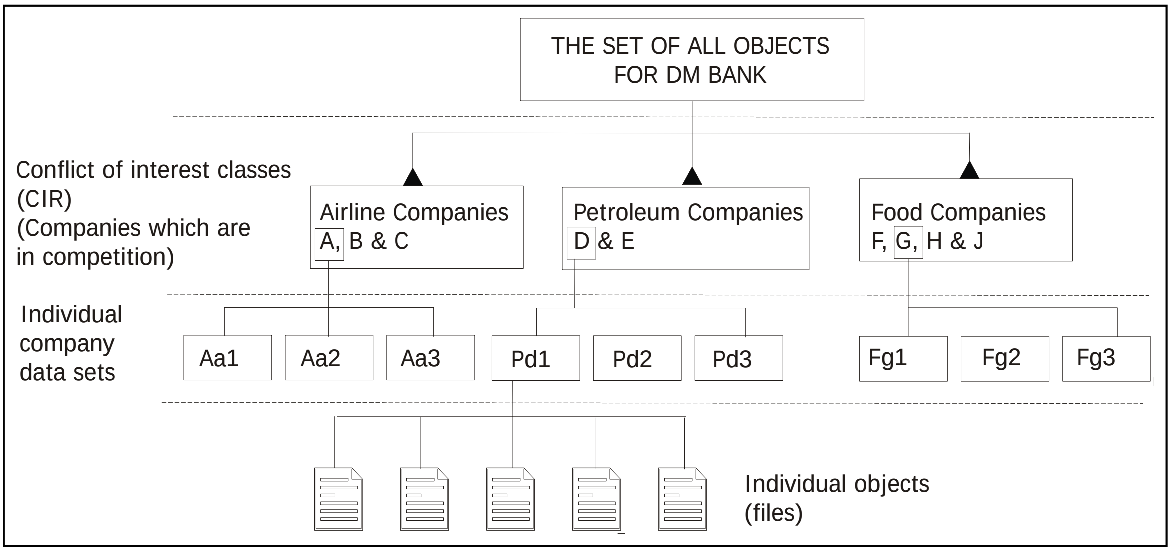

THE SET OF ALL OBJECTS FOR DM BANK
Conflict of interest classes (CIR) (Companies which are in competition)
Airline Companies A, B & C
Petroleum Companies D & E
Food Companies F, G, H & J
Individual company data sets
Aa1
Aa2
Aa3
Pd1
Pd2
Pd3
Fg1
Fg2
Fg3
Individual objects (files)

Figure 2: DM Bank and the Chinese Wall Security Policy model

There are three levels of significance: At the lowest level individual objects are considered, containing specific data items, each concerning a specific company.

At the intermediate level, all objects concerning the same company are grouped together. These are referred to as ‘company data sets’.

At the highest level all company data sets whose companies are in competition (conflict), are grouped together. They are referred to as ‘conflict of interest classes’ (CIR).

Each object is associated with the name of the data set of the company to which it belongs and the name of the conflict of interest class to which that company belongs. For example, the conflict of interest class names could be the business sector headings found in stock exchange listings (Petroleum Companies, Airline Companies and Food Companies), while the individual companies could be the names of companies listed under those headings.

If the data mining environment of DM Bank contains information on Airline Company A, Petroleum Company D and Food Company G:

1. All objects would belong to one of three company datasets – company dataset Airline Company A or company dataset Petroleum Company D or company dataset Food Company G and

2. There would be three conflict of interest classes, one for Airlines (containing Airline Companies A, B and C's data sets), one for Petroleum Companies (containing Petroleum Companies D and E's data sets), and one for Food Companies (containing Food Companies F, G, H and J's data sets).

The basis of the Chinese Wall Security Policy model is that people are only allowed access to information that is not in conflict with any other information that they already possess. When considering the DM Bank scenario and if data mining activities have already been done, a new data miner may be assigned to whatever data sets DM Bank chooses. As far as the CWSP model is concerned, a new data miner does not possess any information at this stage and therefore no conflict of interest can exist. However, such a conflict may arise later on.

Suppose a data miner accesses Airline Company A's data sets first; at this stage he/she possesses information concerning Airline Company A data sets. Since Airline Company B is in conflict with Airline Company A, the data miner will therefore not be granted access to Airline Company B's data sets. However he/she will be permitted to request access to Petroleum Company D's data sets, because the data sets of these two companies belong to different conflict of interest classes. According to the CWSP model, this is permissible because the Airline Companies and the Petroleum Companies are not in competition and therefore they are in different conflict of interest classes. However, when working on a higher level called the data mined level as in the DM Bank example, this should not be allowed. Knowing information pertaining to Petroleum Company D (namely the negative change in Petroleum Company D's cash flow situation mentioned in the previous paragraph) and making it known to Airline Company A (which has shares in Petroleum Company D) puts Petroleum Company D at risk. It is therefore inappropriate for DM Bank to use the CWSP model because the information leakage problem will not be solved.

A new access control model will be suggested. This model will prevent the abovementioned information leakage problem and the following important issues will also be addressed:

- CIR classes can and may overlap - this is to accommodate the conflict of interest between Airline Company A and Petroleum Company D. Lin (1989) addressed the CIR classes that can and may overlap problem in the ACWSP model.

- The severity of the conflict that exists between two companies must be definable. There may be a conflict of interest between Airline Company A and Food Company G but it may be so small that it needs no extra mentioning and should not inhibit access capabilities.
- A Security Policy model must also be dynamic. If Airline Company A sells all its shares in Petroleum Company D they are no longer in conflict and must be treated as such.

4 A new Access Control model for a data mining environment

By adding to the CWSP model's definition, it will be possible to secure also the environment which consists of already mined information.

This article suggests three new definitions, two definitions to be added to the current CWSP model and one that changes the original 'conflict of interest class'-definition of the CWSP model. The first definition to be discussed is the 'conflict of interest class' definition that will be changed slightly to a 'Sphere of conflict' definition.

4.1 Definition 1: Sphere of conflict

A set of companies (all DM Bank's clients) exist where each company has a 'sphere of conflict' around it. This 'sphere of conflict' has a radius r. Consider the following example. ABC Petrol Company has a 'sphere of conflict' defined around it with companies like P+P Petrol Company and Green Petrol Company. (Figure 3).

Other companies that can also exist in this 'sphere of conflict' around ABC Petrol Company are Pick&Save Food Company and HighFly Airline Company. The last two companies exist in this 'sphere of conflict' because ABC Petrol Company has shares in QuickPay Food Company, who is in competition with Pick&Save Food Company and ABC Petrol Company also has shares in FlySave Airline Company which is in competition with HighFly Airline Company.

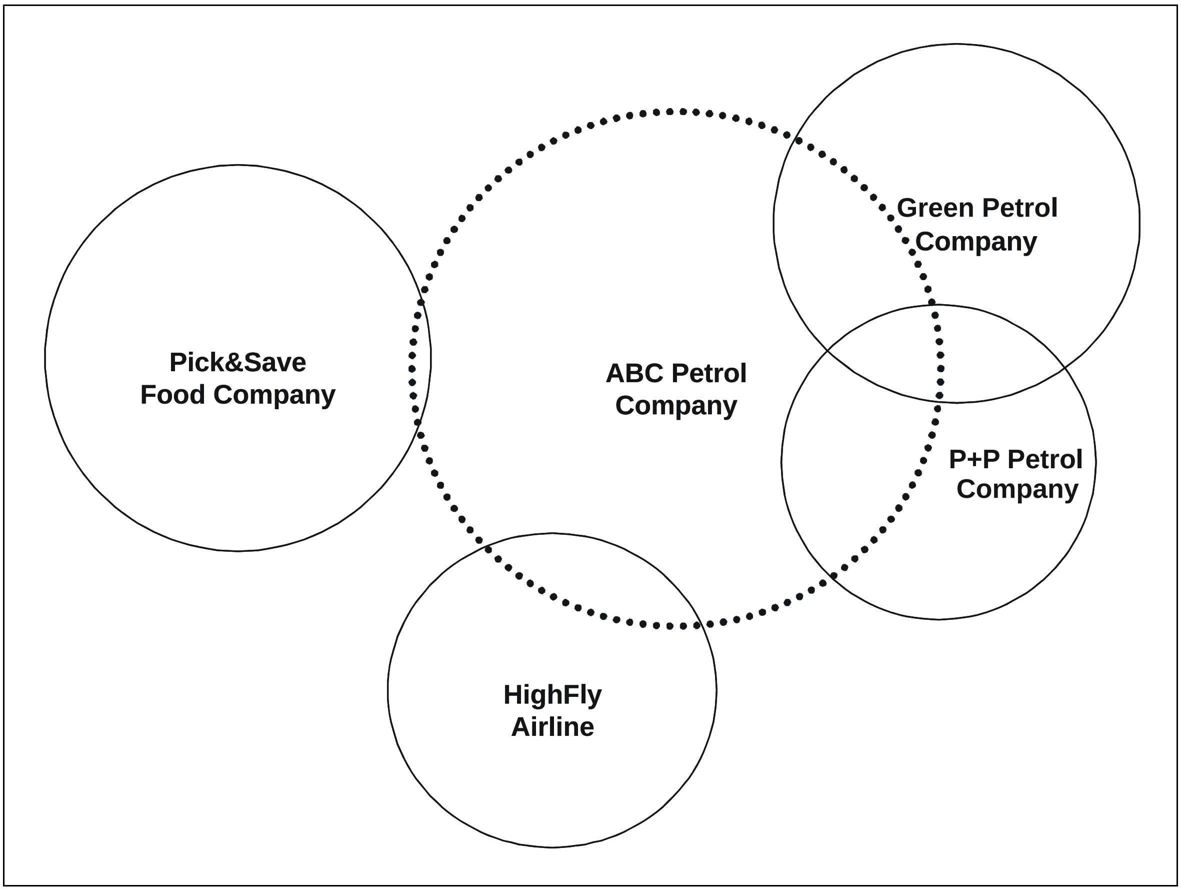


Figure 3: Sphere of conflict

The second definition is called a 'Distance' definition and it defines the severity of conflict between any two companies at a given time.

4.2 Definition 2: Distance

The distance definition works with sets of companies with a pre-defined distance between them. The distance definition defines the severity of the conflict between two companies and the value of distance will always be positive. A distance = 0 implies that the model is working with the same company for example the distance between ABC Petrol Company and itself = 0. A distance = ∞ implies an infinite distance which means no conflict of interest exists between the specified two companies for example, ABC Petrol Company and Lovely Shoes Company.

Drawing a circle around ABC Petrol Company and placing all the companies that are in conflict with ABC Petrol Company on a pre-defined distance from ABC Petrol Company, will result in a 'sphere of conflict'. The pre-defined distance between ABC Petrol Company and for example, P&P Petrol Company is the radius that defines the severity of the conflict between these two companies. This radius can for example, be defined as 1 for all Petrol Companies, radius = 2 for HighFly Airlines and radius = 3 for Pick&Save Food. This is graphically illustrated in Figure 4.

To be able to answer the question whether there is a conflict of interest between any two companies at a given time, the ‘Path’ definition is used.

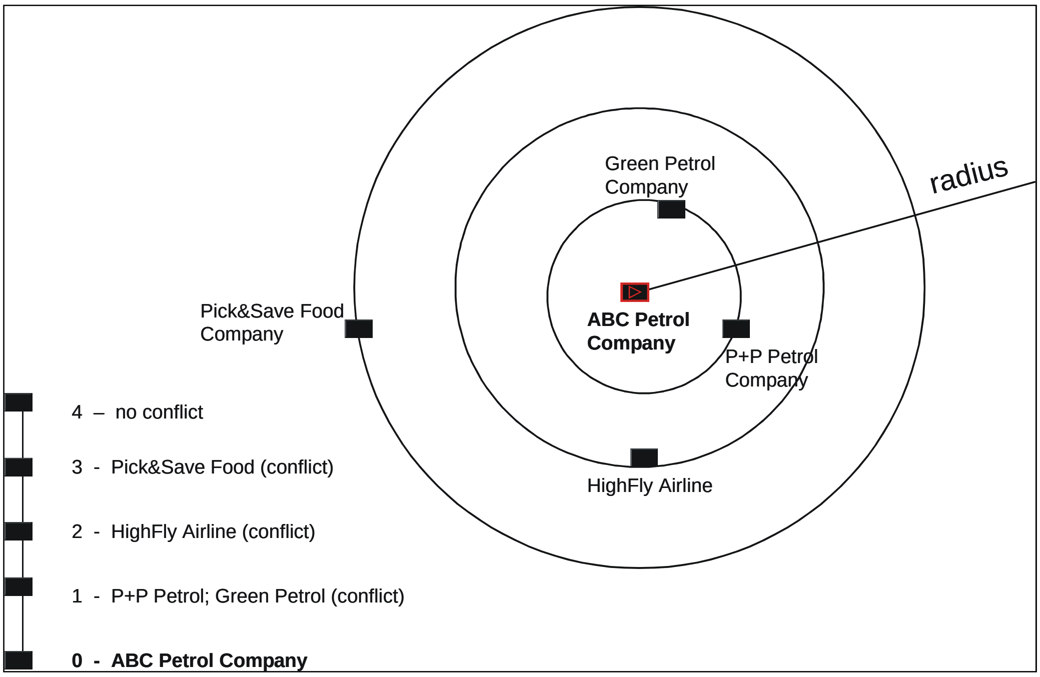


Figure 4: Distance = radius – defines severity of conflict between companies

4.3 Definition 3: Path

The distance definition enables the definition of a path. The existence of a path implicates that there is a conflict of interest between the two companies at both ends of the path. The distance of the path is a positive number. When no path between two companies exist, the distance = ∞. For example, as depicted in Figure 5, if data miner #01 is working with ABC Petrol Company’s information, the following two questions should be answered:

Question 1: Is this data miner allowed to work on Pick&Save Food Company?

Question 2: Is this data miner allowed to work on SoftShoes Company?

In other words, is the distance between ABC Petrol Company and any of the two companies in question equal to ∞? One possible way to represent the relationships in question is illustrated using the graphical representation in Figure 5.

For Question 1, in Figure 5, the path from ABC Petrol Company to Pick&Save Food Company runs through ABC Petrol Company area and Pcik& Save Food Company area. The 'distance' or length of this path = 3. This relationship indicates a conflict of interest between ABC Petrol Company and Pick&Save Food Company and the same data miner can not work on both companies.

For Question 2, in Figure 5, the data miner that works on ABC Petrol Company, will be allowed to work on SoftShoe Company, because no path exists between these two companies. This results in a path with a value of ∞.

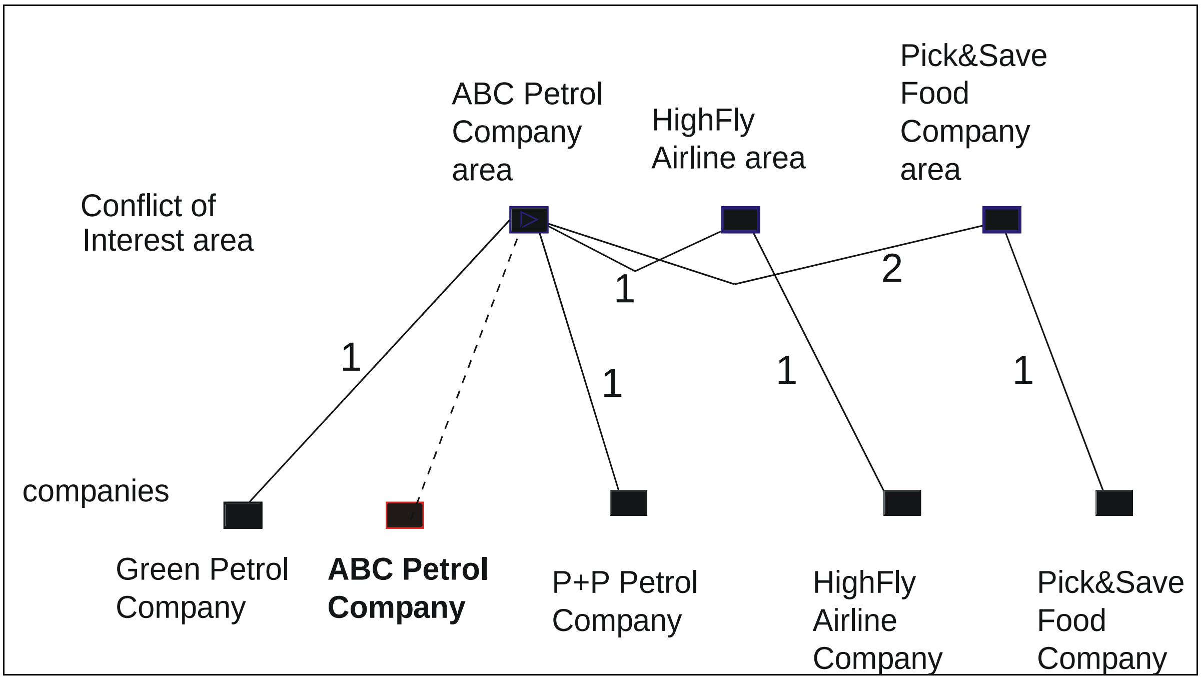


Figure 5: Path between companies

5 Conclusion

There is a difference between working with information or data objects in the real world and working with mined information, especially in the definition of security risk areas and the management thereof. When working with mined information, it is impossible to predict whether or not a security risk area will emerge. It is thus important to have procedures in place for when such a security risk area suddenly materializes, to minimize the risk as far as possible.

In this article an example was discussed to show where the CWSP model and the ACWSP model have some limitations when applied to mined information. The article suggests three new definitions to the CWSP model to sustain a secure state in a data mining environment and thereby minimizing risks of access control compromises.

For future work, it is necessary to investigate the transitive relationships between ABC Petrol Company and BrightLight Company through P+P Petrol Company. Each relationship will need an investigation to determine if it is a conflict or not and if it is possible to represent on the current structure suggested in Figure 5.

References

Bell D. E. and LaPadula, L. J. (1976) Secure computer system: unified exposition and Multics interpretation, MITRE MTR-2997, Available as NTIS AD-A023 588.

Berry, M. J. A. and Linoff, G. (1997) Data Mining Techniques: For Marketing, Sales, and Customer Support, John Wiley & Sons, Incorporated.

Bishop, M. (2003) Computer security: art and science, Pearson Education, Inc.

Brewer, D. F. C. and Nash, M. J. (1989) In IEEE Symposium on Security and PrivacyOakland, pp. 206-214.

Clark, D. and Wilson, D. (1987) In IEEE Symposium on Security and Privacy, pp. 184-194.

DoD (1983) Department of Defense Trusted computer System Evaluation Criteria, CSC-STD-011-83, Department of Defense Computer Security Center Fort Meade, MD.

Lin, T. Y. (1989) In Fifth Annual Computer Security Applications Conference, pp. 282-289.

Shi, Y. (2000) In The IEBM Handbook of Information Technology in Business(Ed, Zeleny, M.), pp. 490-495.

B.5 CBAC: CONFLICT-BASED ACCESS CONTROL

Marianne Loock, Jan HP Eloff, Johannes Heidema: CBAC: Conflict-Based Access Control, Paper submitted for review to an ISI accredited journal, 2012.

Abstract: *Most existing implementations of access control models are based on policies for mandatory access control (MAC) and discretionary access control (DAC) or a variation thereof. In data-mining environments where the behaviour patterns of multiple companies can be exposed, implementations of MAC and DAC policies are not compliant with the conflict management requirements between conflicting companies. The research question dealt with in this paper is whether it is possible to minimise the risk of a breach of confidentiality in a data-mining environment by managing the access control in such an environment based on the conflicts that exist between the different subjects that form part of this environment. The CBAC model presented here is based on a company's own "view of the world" and provides for a degree of conflict and a potential conflict-of-interest path between conflicting companies. A subject defines a degree of conflict or a potential conflict-of-interest path in order to minimise the risk of a breach of confidentiality in the environment of which it is a part.*

The draft article is reflected below:

1 Introduction

While engaged in a data-mining activity, a mining agent may unexpectedly expose trends related to the information of a company. If this mining agent should also be working on the information of another company, that is, a company that is in *conflict* with the first company, attention should be given to access control requirements specific to the conflict scenario [Loock & Eloff, 2002]. The potential conflict between companies may in some cases be so severe that access is completely prohibited, while in other cases it may be insignificant and therefore allow users to be granted access. For purposes of this research, *conflicting* companies are companies that are in competition with one another for any possible business reason and not only because they operate in the same functional business domain. A functional business domain is typically an area in which companies conduct the same type of business, such as airline companies that all belong to the airline business domain. Since subjects and objects in data-mining environments may change rapidly, the access control service of the access control policy should be easily adaptable. It is evident that a set of complex access control requirements is essential in environments such as data mining.

This paper will explain how it is possible to minimise the risk of a breach of confidentiality in a data-mining environment by managing the access control in this environment based on the conflicts that exist between the different subjects that form part of this environment.

The three main concerns in addressing this problem are confidentiality, conflict, and data mining. Since the current literature does not present an access control model based on these three concerns, we propose a new model called the CBAC (or Conflict-Based Access Control) model. The model makes it possible to measure conflict on different levels of severity among the clients of an organisation – not only as specified by the clients, but also as calculated by the organisation. Both types of conflict have their own cut-off points when the conflicts are considered to be of no value any longer.

The remainder of the article is structured as follows: Section 2 explains the rationale for building the new model. Section 3 discusses the related case study and Section 4 gives a definition of the CBAC model and briefly discusses a proof of concept. The article is concluded by a brief summary of the contribution made by the CBAC model.

2 Background and related work

2.1 Confidentiality

When discussing confidentiality policies (also called information flow policies)[Bishop, 2003], an important security model that warrants mentioning is the Bell-La Padula model [Bell & LaPadula, 1973]; [Bell & LaPadula, 1976].

Since the current research takes place in a data-mining environment where the commercial aspect also comes into play, confidentiality is essential and some integrity must therefore be added to the proposed new CBAC model. Access should not be granted to a data miner to the extent that he/she can read the data sets of conflicting companies, even if they exist in different conflict-of-interest classes.

The Bell-La Padula model has for instance been criticised [Gollmann, 2006] for dealing with confidentiality but not with integrity, and also for not addressing the management of access control.

However, because this security model mainly concentrates on confidentiality, it cannot be seen as a possible solution that will address access control problems when managing conflicting entities such as those encountered in data-mining environments.

2.2 Conflict

2.2.1 Chinese Wall Security Policy model

The most important access control approach in existing literature that addresses conflict management between companies is the Chinese Wall Security Policy (CWSP) model. The CWSP model is based on managing potential conflict between companies that belong to the same functional business domain. This model, designed by Brewer and Nash [Brewer & Nash, 1989], addresses access control requirements in a commercial environment where a possible conflict of interest may arise. The concept of conflict of interest is applicable in a data-mining environment where the same mining agent is exposed to the data of different companies.

The CWSP model makes use of actors and objects and defines the access rights to data on a specific side of a “Chinese Wall” to which the actor already has access. It develops a set of rules aimed at preventing people (actors) from accessing data (objects) on the opposite side of the “Chinese Wall”.

There are three levels of significance in the CWSP model:

1. At the lowest level, objects are considered. These objects contain specific data items, each concerning a specific company.
2. At the intermediate level, all objects concerning the same company are grouped together. These are referred to as company data sets.
3. At the highest level, all data sets of companies that are in conflict are grouped together in classes. These are referred to as conflict-of-interest classes.

The founding principle of the Chinese Wall Security Policy model is that actors are allowed access only to information that is not in conflict with any other information to which they already have access. The CWSP model builds a collection of impassable walls called Chinese Walls around the data sets of conflicting companies. No data sets that are in conflict can be stored on the same side of the Chinese Wall.

2.2.2 Aggressive Chinese Wall Security Policy model

According to Lin in [1989]; [2003], the Brewer-Nash model was based on the incorrect assumption that corporate data can be grouped in separate and “disjoint” conflict-of-interest (CIR) classes [Tsau Young Lin, 2003]. In his arguments, Lin explains that CIR classes are seldom disjoint; in fact, they generally overlap. He subsequently suggested a modified model called the Aggressive Chinese Wall Security Policy (ACWSP) model that is based on the development of a methodology known as Granular Analysis and Computing [Tsau Young Lin, 2003]. The ACWSP model allows CIR classes to overlap.

2.3 Data mining

The following research contributions were investigated but proved to be of little value:

2.3.1 Privacy-preservation models

When using the CWSP model, data sets and databases may be sanitised by the removal of private information. In some cases, however, data-mining techniques allow the recovery of such removed information. Privacy preservation is thus an important technique to be used when sensitive data is to be released to a third party, in this case to a mining agent [Bertino, Byun, & Li, 2005]. Although privacy preservation emphasises privacy as opposed to access control, it still needs to be considered within the overall aim of addressing access control requirements for data-mining environments. An important goal of privacy-preserving data mining is the development of new models for inferred or indirect information without access to precise information in the original individual records [Agrawal & Srikant, 2000]. The study of data mining, with regard to privacy preservation, has become an active area of research in computer science [Agrawal & Srikant, 2000]; [Clifton & Marks, 1996]; [Conrado, et al., 2004]; [Kantarcioglu & Clifton, 2004]; [Thuraisingham, 1996]].

2.3.2 Privacy-aware models

Privacy-aware access control combines the concepts of privacy and access control. Current research in privacy-aware access control focuses on two issues. Firstly, on the definition and development of access control and privacy languages such as XACML (eXtensible Access Control Markup Language) [OASIS, 2010], and secondly, on the definition of architectures to protect and preserve the privacy of either services or clients [Ardagna, Cremonini, Damiani, Vimercati, et al., 2006].

Privacy-aware access control is important for the development of access control models for data-mining environments in general. Note that privacy-aware access control indeed facilitates controlled access to individual records, which is not the case with privacy-preserving approaches.

From an analysis of the literature it is clear that the three main concerns, namely confidentiality, conflict and the complications of the data-mining environment, are not solved and described in any single security model. It is also known that the integrity aspect of knowledge management still needs investigation [Bertino, et al., 2006] and in this paper knowledge management is seen as the data-mining process and environment. The next section discusses the case study that is used to explain the proposed Conflict-Based Access Control (CBAC) model.

3 Case study

This section contains a brief overview of the case study used in further discussions.

DM Bank is a financial institution that conducts data mining for prediction and estimation purposes for its clients. In the remainder of this paper, DM Bank is used as an example of a global agent that provides a service to secondary agents (clients). In this case study this service entails typical financial services that banks deliver to their clients. DM Bank (the global agent) has a finite number of clients (secondary agents, e.g. ABC Petrol Company) on whose data sets the mining agents (e.g. DataMiner 3) are performing data-mining activities on behalf of DM Bank (the global agent).

Some of the global agent’s secondary agents have specific business relationships with one another and these relationships are depicted in Figure 1:

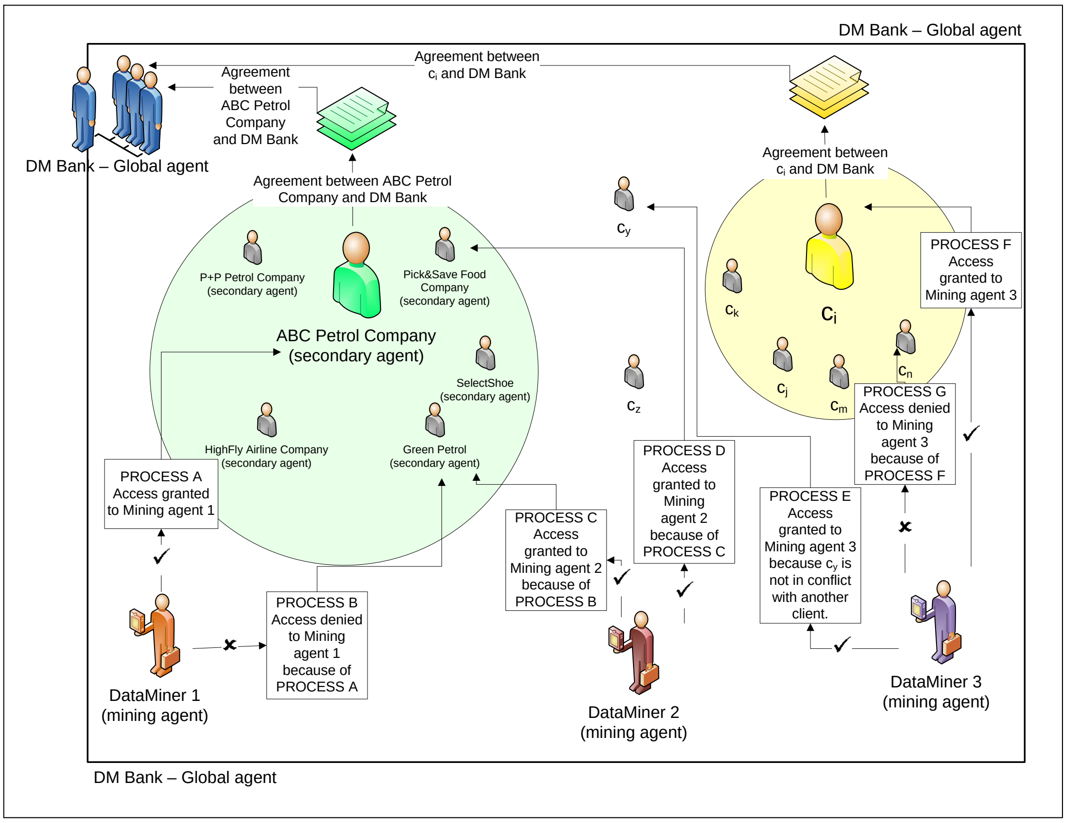

DM Bank – Global agent
Agreement between c_i and DM Bank
Agreement between ABC Petrol Company and DM Bank
DM Bank – Global agent
Agreement between ABC Petrol Company and DM Bank
Agreement between c_i and DM Bank
c_y
P+P Petrol Company (secondary agent)
Pick&Save Food Company (secondary agent)
ABC Petrol Company (secondary agent)
SelectShoe (secondary agent)
HighFly Airline Company (secondary agent)
Green Petrol (secondary agent)
c_k
C_i
c_j
c_m
c_n
c_z
PROCESS F Access granted to Mining agent 3
PROCESS G Access denied to Mining agent 3 because of PROCESS F
PROCESS A Access granted to Mining agent 1
PROCESS B Access denied to Mining agent 1 because of PROCESS A
PROCESS C Access granted to Mining agent 2 because of PROCESS B
PROCESS D Access granted to Mining agent 2 because of PROCESS C
PROCESS E Access granted to Mining agent 3 because c_y is not in conflict with another client.
✓
×
DataMiner 1 (mining agent)
DataMiner 2 (mining agent)
DataMiner 3 (mining agent)
DM Bank – Global agent

Figure 1: Global, secondary and mining agents

ABC Petrol, a secondary agent, is in conflict with another secondary agent, Green Petrol, because both belong to the same functional business domain (petroleum companies). While conducting a data-mining exercise, DM Bank's mining agents (DataMiner 1) detect a negative growth in Green Petrol's cash flow. If one or more of the mining agents who work on the data sets of Green Petrol also work on those of ABC Petrol, Green Petrol is exposed to the risk of information leakage. The latter's sensitive cash flow information could well become known to shareholders and result in their selling their Green Petrol shares and buying into ABC Petrol.

Furthermore, ABC Petrol is in conflict with the HighFly Airline Company, because ABC Petrol holds shares in FlySave Airlines. For this reason, ABC Petrol defined HighFly Airline as a conflicting company. Likewise, ABC Petrol defined Pick&Save Food as a conflicting company, because ABC Petrol holds shares in QuickPay Food. Lastly, ABC Petrol defined the SelectShoe Company as a conflicting company for reasons unknown to DM Bank.

A typical conflict-of-interest scenario for Figure 1 will be a set of conditions in which professional judgement concerning a principal interest (such as keeping information and knowledge regarding conflicting companies undisclosed) is unduly influenced by a less significant interest (such as financial gain) and this will be when e.g. DataMiner 1 has access to ABC Petrol Company and to Green Petrol Company. DataMiner 1 can, for his own financial gain, leak information from one conflicting company to the other.

Figure 1 depicts a secondary agent (ABC Petrol Company) and all the secondary agents that ABC Petrol Company knows to be in conflict with itself. These “conflicting” secondary agents are grouped in a circle around ABC Petrol Company. The same argument goes for secondary agent c_i.

This case study emphasises some important data-mining access requirements to be considered by the global agent, since secondary agents expect the global agent to regulate access control in such a manner that confidentiality during data-mining activities is preserved where required. One could, for example, ask the following questions:

“Can the same mining agent who has access to the data sets of ABC Petrol obtain further access to those of Green Petrol?” (e.g. DataMiner 1). The answer will be ‘No’ because ABC Petrol and Green Petrol are in conflict. Another question might be, “Can the mining agent who has access to the data sets of Green Petrol obtain access to those of Pick&Save Food?” (e.g. DataMiner 2). In this example the answer will be ‘Yes’ because there exists no data stating that Green Petrol and Pick&Save Food are in conflict. Figure 1 also illustrates that if DataMiner 3 has access to c_y, it is possible for DataMiner 3 to get access to c_i, because c_y is in conflict with no other client, therefore also not in conflict with c_i. If, however, DataMiner 3 now wants access to c_n, this request will be denied because DataMiner 3 already has access to c_i, and c_i and c_n are in conflict with one another.

4 The Conflict-Based Access Control (CBAC) model

The following requirements were set for the proposed CBAC model:

1. The idea of conflict-of-interest classes should be part of the proposed model because the CBAC model should be able to group specific classes.
2. The concept of a history record is important. A record should be kept of all the data sets that a subject (e.g. a mining agent) is currently working on, as well as those

accessed in the past. For purposes of the history record a session is defined as the time interval during which no mining agent's history record needs to be deleted. Thus it is the total time of interaction between a mining agent and the data sets in which the agent is working during a specific time interval. Throughout the discussion of the CBAC model, the data-mining environment is used as an environment to demonstrate the principles of the CBAC model. Hence, we will refer to a mining agent as representing the subjects in a normal access control model.

3. The model must provide for a degree of conflict to determine whether a potential conflict would influence an access control decision.
4. The proposed model should focus on confidentiality and be able to enforce access control in a rapidly changing and primarily business environment.

The objective of the CBAC model is to determine if a mining agent (subject) can obtain access to an object. The type of access right required by the data-mining agent can vary and may for example include doing a decision tree analysis of an object or a set of objects. The CBAC model as proposed in this paper focuses primarily on confidentiality. It models the relationship(s) between data-mining agents and objects (data sets) and, for reasons of complexity, does not differentiate between the types of access.

4.1 Elements of the CBAC model

The CBAC model implements an access control policy that regulates data-mining activities between for example data-mining agents working for a global agent, secondary agents to whom the global agent renders a service, and the data sets of the secondary agents. Stated differently, a global agent is regarded as the company that instructs the data-mining activity to be conducted, while secondary agents are those companies whose data sets are exposed to data-mining activities as instructed by the global agent.

The remainder of this paragraph presents the definitions of the different elements of the CBAC model. This is followed in each case by a graphical illustration (if applicable) and a brief discussion with reference to the case study.

4.1.1 Secondary agent

We shall use symbol c_i to denote both a secondary agent and the set of data sets of this particular secondary agent.

4.1.2 Global agent

Let D denote the global agent.

The set of secondary agents are in a specific business relationship with the global agent, for example they are clients (c_i) of a bank (D).

Case study: The CBAC model focuses on minimising the potential security risks that are created when a mining agent who works for the global agent, DM Bank, performs data-mining activities on the objects (data sets) of two or more secondary agents who might or might not be conflicting clients – such as Green Petrol and ABC Petrol.

4.1.3 Conflict-of-interest field

A conflict-of-interest field is the set of secondary agents that are in conflict with one another.

Let CNC denote all possible secondary agents that belong to D.

$$\text{CNC} = \{c_1, c_2, \dots , c_z\}$$

where:

c_i is a secondary agent and

z is the total number of secondary agents associated with the global agent (D).

Let C be the set of secondary agents called the conflict-of-interest field of all possible secondary agents CNC.

$$C = \{c_{p\_1}, c_{p\_2}, c_{p\_3}, \dots , c_{p\_k}, \dots , c_{p\_q}\} \subset \{ c_1, c_2, \dots , c_z\} = \text{CNC}$$

where:

c_p are the secondary agents belonging to C and

$1 \leq q \leq z$ where:

q is the total number of secondary agents (C) where the following is true:

1. a secondary agent in C claims to be in conflict with one or more other secondary agents in C, or
2. any other secondary agent of the global agent claims to be in conflict with a secondary agent or agents in C (in which case the other secondary agent also belongs to C), or
3. the global agent claims that the secondary agents (in C) are in conflict with one another.

Case study: The secondary agents ABC Petrol and Green Petrol are elements of C, because they both claim to be in conflict with another secondary agent – in this case with each other.

4.1.4 Non-conflict-of-interest field

Let NC represent the set of secondary agents called the non-conflict-of-interest field of global agent D.

$$NC = CNC – C = \{c_i \mid c_i \in CNC, c_i \notin C\}$$

4.1.5 MiningAgent$_x$

Let MiningAgent$_x$ denote a data-mining agent / user / person / group conducting data-mining activities on behalf of the global agent D.

4.1.6 Data sets of secondary agents accessed by mining agents

Let NS be the set of all secondary agents' data sets to which all the mining agents who perform data mining on behalf of the global agent D was already granted access.

$NS = \cup_x NS_x$

where:

NS_x is the set of secondary agents' data sets to which a specific $MiningAgent_x$ was already granted access:

$NS_x = \{c_{r\_1}, c_{r\_2}, c_{r\_3}, \ldots, c_{r\_p}\}$

where:

$c_r \in C$ and r are a specific secondary agent's data sets and

p is the number of secondary agents' data sets to which $MiningAgent_x$ has access.

x = 1 … n, where n is the total number of mining agents available to the global agent.

NS_x represents a history record for $MiningAgent_x$.

4.1.7 Functional business domain

A functional business domain is a domain in which companies conduct the same type of business, such as airline companies that all belong to the airline business domain. This can be formulated as follows:

Let a_i denote a functional business domain or a particular market segment such as the Airline business.

Here $1 \leq i \leq m$ and m is the total number of functional business domains of all the secondary agents serviced by global agent D, which constitute the set

$A = \{a_1, \ldots, a_i, \ldots, a_m\}$.

4.1.8 Associations between secondary agents and functional business domains

Let the fact that secondary agent c_i has an interest in functional business domain a_h be represented by $e(c_i, a_h)$.

There are two ways to construct the relationship between c_i and a_h, namely:

1. Each c_i associates itself with an a_h, for example $e(c_i, a_h)$
2. For each a_h consider all c_i's that fall in that functional business domain, for example $e(a_h, \{c_i, c_j, ...\})$.

Also: For each c_i consider all a_h's that are relevant to that secondary agent, for example $e(c_i, \{a_h, a_g, ...\})$.

Furthermore:

For a specific $e(c_i, a_h)$, let $w(c_i, a_h)$ denote the weight of conflict of interest of secondary agent c_i in functional business domain a_h. The weight function gets its value from the distance function (section 4.2.4) indicating the degree of conflict of interest between a secondary agent $c_i \in C$ and any other conflicting secondary agent $c_j \in C$ that belongs to the functional business domain a_h, as defined by c_i or $c_{j.}$

$w(c_i, a_h)$ is a number in $N \cup \{ \infty \} = \{1, 2, \ldots, \infty\}$.

$w(a_{h,} c_i) = 0 \; \forall$ h and i. The weight function does not indicate a weight of conflict of interest between a functional business domain a_h and a secondary agent c_i. Only secondary agents can initiate conflicts of interest.

The smaller the number $w(c_i, a_h)$, the stronger the interest of secondary agent c_i in functional business domain a_h; thus $w(c_i, a_h)$ can be seen as a "distance" from c_i to a_h. When a_h is the main functional business domain of c_i, we define $w(c_i, a_h) = 1$. A value ∞ indicates no interest at all.

It is possible for a secondary agent of the global agent ($c_i \in CNC$) to be associated with one or more functional business domains, $a_h \in A$.

Case study: The ABC Petrol Company belongs to the Petrol functional business domain because of its main business activity, but also to the Airlines functional business domain because it has a majority of shares in the FlySave Airline Company.

In summary, the following table depicts the elements of the CBAC model:

Table 1: Elements of the CBAC model

Name of element	**Description of element**
Global agent	D represents DM Bank with the secondary agents as clients and the mining agents as workers.
Set of all possible secondary agents (clients) of DM Bank	CNC = { c_1, c_2, … , c_z}
Secondary agent	An element of CNC and the set of data sets of a particular secondary agent.
Conflict-of-Interest field	$C = \{c_{p\_1}, c_{p\_2}, c_{p\_3}, \ldots, c_{p\_k}, \ldots, c_{p\_q}\} \subset \{c_1, c_2, \ldots, c_z\} = CNC$
Non-Conflict-of-Interest field	$NC = CNC - C = \{c_i \mid c_i \in CNC, c_i \notin C\}$
MiningAgent_x	MiningAgent_x is a data-mining agent / user / person / group.
The data sets of secondary agents accessed by mining agents	$NS = \cup_x NS_x$ $NS_x = \{c_{r\ 1}, c_{r\ 2}, c_{r\ 3}, \ldots, c_{r\ p}\}$
Functional business domain	$A = \{a_1, \ldots, a_i, \ldots, a_m\}$
Associations between secondary agents and functional business domains	$e(c_i, a_h)$
The weight or “power” of interest of secondary agent c_i in functional business domain a_h	$w(c_i, a_h)$

4.2 Operational elements of the CBAC model

The functioning of the CBAC model is explained by the following operational elements:

4.2.1 Access group list for secondary agent c_i

Let AG-List_i be the access group list for c_i.

Note:

1. $\text{AG-List}_i \subset D$
2. c_i defines its own AG-List_i

AG-List$_i$ = {c_j, c_k, … c_s} is a list of all those secondary agents with whom secondary agent c_i has a specific business relationship (i.e. a positive relationship or business interest), for example secondary agent c_i owns 60% of secondary agent c_j. This is a business interest for which there exists no conflict with c_i. However, this interest causes a conflict between c_i and all the other secondary agents that are in the same functional business domain as any secondary agent in AG-List$_i$.

4.2.2 Sphere of Conflict – conflict of interest

Let S_i be the *Sphere of conflict* for secondary agent c_i

Note:

$S_i \subset$ CNC: $S_i \subset C \subset$ CNC (S_i is either part of C) or $S_i = \{c_i\}$

and $c_i \not\in S_j$ for any $j \neq i$,

in which case $S_i \subset$ NC $\subset$ CNC (S_i is part of NC).

Each c_i defines its own S_i. In the instance $S_i = \{c_i, c_j, c_k, c_m, c_n\}$, it is depicted that c_j, c_k, c_m and c_n are identified by c_i as conflicting secondary agents. The *sphere-of-conflict* definition [Loock & Eloff, 2005b] explains the concept of conflict between secondary agents. A secondary agent c_i may be in conflict with any other secondary agent due to being in the same functional business domain, or for any other reason.

There can be a c_i ($1 \leq i \leq q$) where $q \leq z$, which chooses not to define an S_i; thus $S_i = \{c_i\}$. However, this does not exclude the fact that c_i can be included in another S_k ($1 \leq k \leq q$) as defined by c_k.

4.2.3 Cut-off point for conflict of interest

Let r_i be the cut-off point for conflict beyond which conflict is to be regarded by secondary agent c_i as insignificant. r_i is defined as the radius of the sphere of conflict S_i defined by c_i:

rad (S_i) = $r_i \geq 0$

where r_i is a natural number ($N \cup \{0, \infty\}$)

Note:

No range of possible values for r_i is pre-specified. During implementation time of the CBAC model these ranges will be specified.

This sphere of conflict and its cut-off point for conflict (radius) are depicted in Figure 2:

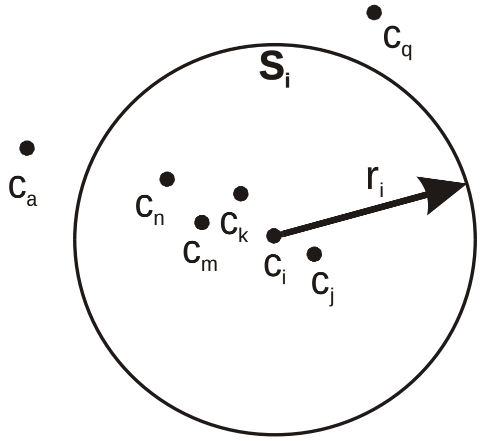


Figure 2: r_i is the radius for S_i (sphere of conflict) defined around c_i (secondary agent)

4.2.4 Distance – degree of conflict of interest

Let $d_{i,j}$ denote the distance between c_i and c_j

where:
$d_{i,j}$: C x C $\rightarrow$ N $\cup$ {0, ∞}. Thus d_{ij} is a number in {0, 1, 2, …, ∞}
$(c_i, c_j) \rightarrow d_{i,j} (c_i, c_j)$
$d_{i,j} (c_i, c_j) \geq 0$ for all $c_i, c_j \in C$
$d_{i,j} (c_i, c_j) = 0$ if and only if $c_i = c_{j.}$

Note:

1. $d_{i,j}$ is a function indicating the degree of conflict between a secondary agent $c_i \in C$ and any other conflicting secondary agent $c_j \in C$, as defined by c_i or $c_{j.}$ Note that this only applies to agents in C and not to all $c_i \in$ CNC.
2. The chosen value of $d_{i,j}$ is determined by c_i. The latter (c_i) determines the value of $d_{i,j}$ based on the degree of conflict between c_i and c_j, as determined by c_i. However, care should be taken in deciding on the value of $d_{i,j}$. If this value is too low, it can result in

unwanted releases of information and if it is too high, it can result in a delay in services.

3. $d_{i,j} = \infty$ between c_i and c_j implies an infinite distance (c_i did not determine a value for $d_{i,j}$), which means no conflict of interest exists between c_i and c_j.

It is important to highlight the difference between cut-off point for conflict (*radius*) and *distance*. The radius r_i defined by c_i denotes a cut-off point beyond which conflict is considered as insignificant between c_i and any c_j. In case $r_i < d_{ij}(c_i, c_j)$ the degree of conflict between c_i and c_j is considered as insignificant and should not influence any access decision.

Case study:

1. In the example depicted in Figure 3, $d_{\text{ABC Petrol Company, SelectShoe}} > 3 = r_{\text{ABC}}$ refers to a non-existent conflict, such as with the SelectShoe Company, where $c_{\text{SelectShoe Company}} \notin S_{\text{ABC Petrol Company}}$. The radius r_i (in this example $r_{\text{ABC Petrol Company}}$) that represents the "non-existing conflict" cut-off can be determined by c_i (in this example, ABC Petrol Company).
2. Consider the following example by also referring to Figure 3. The ABC Petrol Company defines a sphere of conflict, $S_{\text{ABC Petrol Company}}$, around itself. It starts with secondary agents such as P+P Petrol Company and Green Petrol Company defined at a distance of 1. This means $d_{\text{ABC Petrol Company, P+P Petrol Company}} = d_{\text{ABC Petrol Company, Green Petrol Company}} = 1$, which implies a high degree of conflict because they are in direct conflict with ABC Petrol Company.

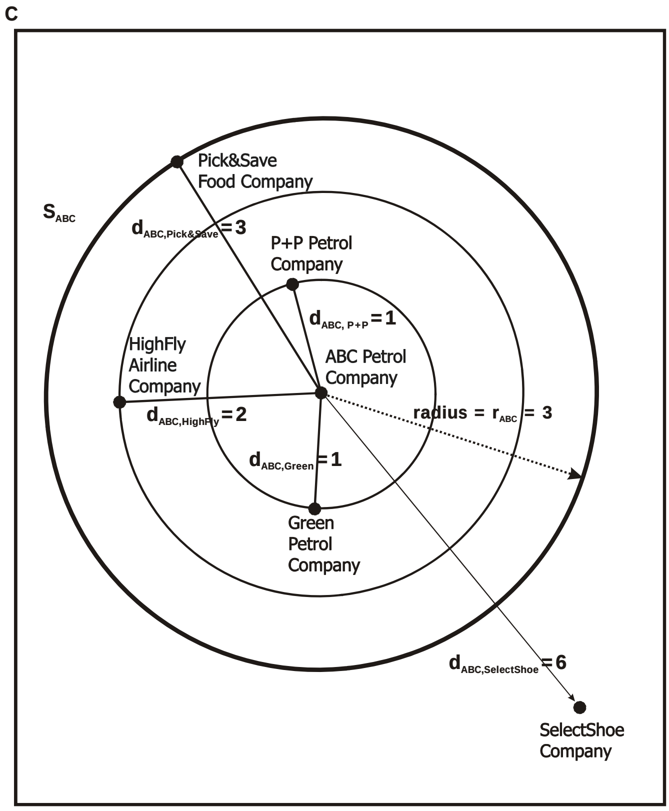


Figure 3: Sphere of conflict around ABC Petrol Company

Sphere of conflict, cut-off point (radius) and distance focus on conflict identified by the secondary agents themselves, i.e. they represent the secondary agent's view of the world. It is also important to assess whether there is any other potential conflict over and above the conflicts as identified by the secondary agents themselves. This Potential Conflict of Interest (PIT) is evaluated by the global agent and is addressed in the next section.

4.2.5 Potential Conflict of Interest (PIT)

Let G = (V, E) represent a mechanism to assist in the assessment of Potential Conflict of Interest between secondary agents

where:

1. G is a bipartite graph consisting of the following two sets:
   (i) A set V = V(G) whose elements (finite set of nodes) are called vertices of G
   (ii) A set E = E(G) of ordered pairs (may be labelled with 'e' for 'edge') of distinct vertices called edges (lines between nodes) of G
2. V, the vertices of G, is partitioned into two subsets A and D such that each edge of G

connects a vertex of A to a vertex of D, i.e. the lines only exist between a node in A and a node in D.

3. G is a directed graph where the edges between vertices go in a specific direction, i.e. $e(c_i, a_h)$ is different from $e(a_h, c_i)$. For the sake of simplicity, $e(c_i, a_h)$ refers to the upward direction whereas $e(a_h, c_i)$ is the downward direction. (See Figure 4 further on.)
4. G is a weighted graph, meaning that each edge $e \in E$ of G is assigned a non-negative number w(e), called the weight of e [Lipschutz & Lipson, 2003].

Note:

All values that are used in the bipartite graph come from already known information between secondary agents or from the different business relationships between the different role players within the global agent.

(1)The subset A of V represents all the functional business domains a_h of the global agent (see 4.1.7).

(2)The subset D of V represents all the secondary agents c_i of the global agent (see 4.1.1).

(3)The set of edges E represents the associations between secondary agents and functional business domains. Thus e_i represents the association from c_i to a_h and is represented by $e(c_i, a_h)$ (see 4.1.8).

(4)The weight of the edge $e(c_i, a_h)$ is denoted by $w(c_i, a_h)$ and represents the strength of the associations between secondary agents (subset D of V) and functional business domains (subset A of V).

(5)$w(c_i, a_h) = d_{i,j}$ where c_i in subset D is related to c_j in subset D via functional business domain a_h in subset A. $d_{i,j}$ was already established and that value is now carried over to the bipartite graph.

(6)The weight of the edge $e(a_h, c_i)$ is denoted by $w(a_h, c_i)$ and represents the strength of the associations between functional business domains (subset A of V) and secondary agents (subset D of V). This weight will always be 0 because the representing edge $e(a_h, c_i)$ is directed from a functional business domain to a secondary agent – thus $w(a_h, c_i) = 0$ is the direction from a_h in subset A to c_i in subset D.

(7)$w(c_i, a_h) > 0$ where the edge e_i links c_i in subset D to a_h in subset A.

Potential Conflict-of-Interest (PIT) path

Let ~ denote a path between any two secondary agents from different functional business domains c_i and c_j, represented as $c_i \sim c_j$

where:

(1) $c_i \sim c_j$ means that at least one path exists between c_i and c_j

(2) $c_i \sim c_j$ implies a sequence $c_i\ a_x\ c_p\ a_y \ldots a_z\ c_j$, where every pair $c_u a_v$ or $a_v c_w$ is an edge in E(G).

Note:

- There can be multiple paths between any two secondary agents.
- The *path* definition is used to determine potential conflict, that is, conflict only noticeable by the global agent, as opposed to direct conflict identifiable by all the secondary agents themselves.
- *Path* establishes whether a potential conflict of interest exists between any two secondary agents at a given point in time and not necessarily in the same $S_{i.}$
- The existence of a path implies that a mining agent can obtain access to the data sets of two secondary agents, namely, $c_i \notin S_j$ and $c_j \notin S_i$. There may be a potential conflict of interest between these two secondary agents at both ends of the path. This potential conflict of interest (which was not foreseen by c_i and c_j themselves) can be determined by the global agent with the help of the *path* parameter.

Potential Conflict of Interest (PIT) – The k^th path

All possible paths between c_i and c_j must be established to be able to calculate the shortest path between c_i and c_j. The shortest path between c_i and c_j is the path that best represents the degree of potential conflict of interest between c_i and c_j.

Let $P_k\ (c_i, c_j)$ denote the k^th path between c_i and c_j

where:

k is any path of all the possible paths between c_i and c_j and

where:

P_k (c_i, c_j), between c_i and c_j is a sequence c_i a_x c_p a_y … a_z c_j, where every pair $c_u a_v$ or $a_v c_w$ is an edge in E(G).

Potential Conflict of Interest (PIT) – The weight of a path

Let wP_k (c_i, c_j) denote the weight of path P_k (c_i, c_j)

where:

(1)wP_k (c_i, c_j) = $\sum w(c_s, a_t)$, where $\sum$ goes over all pairs (c_s, a_t) and (a_t, c_s) along the path P_k (c_i, c_j).

(2)wP_k (c_i, c_j) $\in$ N ∪ {0, ∞}

Thus wP_k (c_i, c_j) is a number in {0, 1, 2, …, ∞}.

(3)wP_k (c_i, c_j) = ∞ when the calculated path becomes so long that the assumption can be made that there is no path between two secondary agents.

(4)Cut-off point of path weight: In the same manner as r_i is used as the cut-off point for conflict for secondary agent c_i, a cut-off point of the path weight will be established by the global agent. The cut-off point of path weight will be used in the following manner: the smaller the cut-off point of path weight, the more conflict will exist between the two secondary agents c_i and c_j. The efficiency of this cut-off point will be established by the global agent over time, based on internal business risk analysis by the global agent himself. The path weight is based on the distance parameters that the secondary agents established over time, in other words on the business risk that the secondary agents were prepared to take when establishing the conflicts between themselves and the other secondary agents. The global agent should calculate its business risk on the different mining agents that are doing data mining on behalf of the global agent when establishing a cut-off point for the path rate.

(5)For illustration purposes, the cut-off point of the path weight can be defined as 5.

wP_k (c_i, c_j) > 5 : weight of the path is too high to be considered as a conflicting path

≤ 5 : weight of the path indicates a conflict

= ∞ : indicates the calculated path became so long that the deduction can be made that there exists no path, which also implies no conflict

The Potential Conflict-of-Interest (PIT) path (shortest path)

Because there is a possibility of multiple paths between any two secondary agents, the shortest path between c_i and c_j is the path that best represents the degree of potential conflict of interest between c_i and c_j. For illustration purposes, consider path P_k (c_i, c_j) (k=1,2) as depicted in Figure 4.

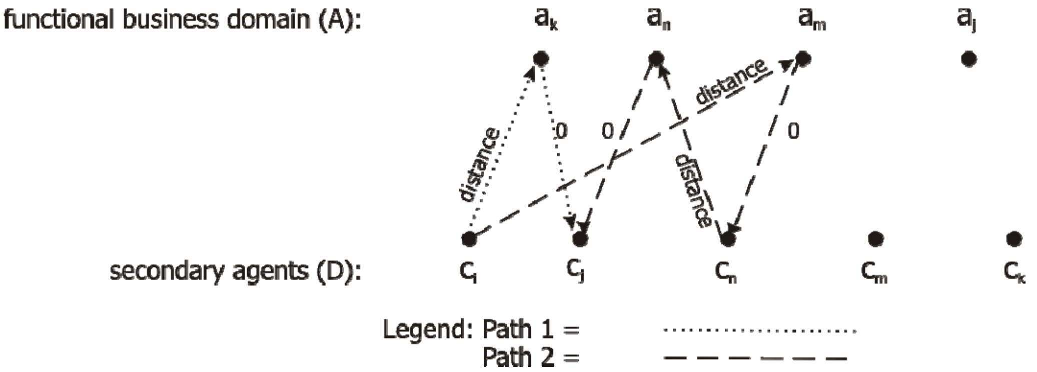


Figure 4: Two different paths from c_i to c_j

The figure above illustrates two different paths between two secondary agents:

P_1 (c_i, c_j) = c_i a_k c_j and P_2 (c_i, c_j) = c_i a_m c_n a_n c_j

It is important to introduce a mechanism to determine which path, out of multiple possible paths, should be regarded as the shortest, because the shortest path between c_i and c_j is the path that best represents the degree of potential conflict of interest between c_i and c_j, which is minP_k.

Case study: Consider the example as depicted in Figure 5 below. ABC Petrol has a strong interest in the Petrol functional business domain, because it is a petrol company itself. Thus, w(ABC Petrol, Petrol) = 1. ABC Petrol is in conflict with Green Petrol as well as with P+P Petrol, according to its own definition of its sphere of conflict $S_{\text{ABC Petrol}}$. Figure 5 shows that ABC Petrol is in potential conflict with Fast Petrol, although the latter was not identified by ABC Petrol as being in its sphere of conflict.

ABC Petrol furthermore has a strong interest in the Airlines functional business domain, because it holds many shares in FlySave Airline Company. Thus, w(ABC Petrol, Airlines) = 2 = $d_{\text{ABC,HighFly}}$ (Figure 3). ABC Petrol stated frankly that it is in conflict with HighFly

Airline Company, but from Figure 5 it is clear that ABC Petrol is also in potential conflict with BlueSky Airlines, which was not included in ABC Petrol's initial definition of its sphere of conflict.

Figure 5 also shows that ABC Petrol has an interest in the Food functional business domain, because it has shares in QuickPay Food. Thus, w(ABC Petrol, Food) = 3 = $d_{ABC,Pick\&Save}$ (Figure 3). ABC Petrol defined a conflict with Pick&Save Food, but Figure 5 shows that ABC Petrol is also in potential conflict with EasyEat Food, which was not defined by ABC Petrol to be part of $S_{ABC.}$

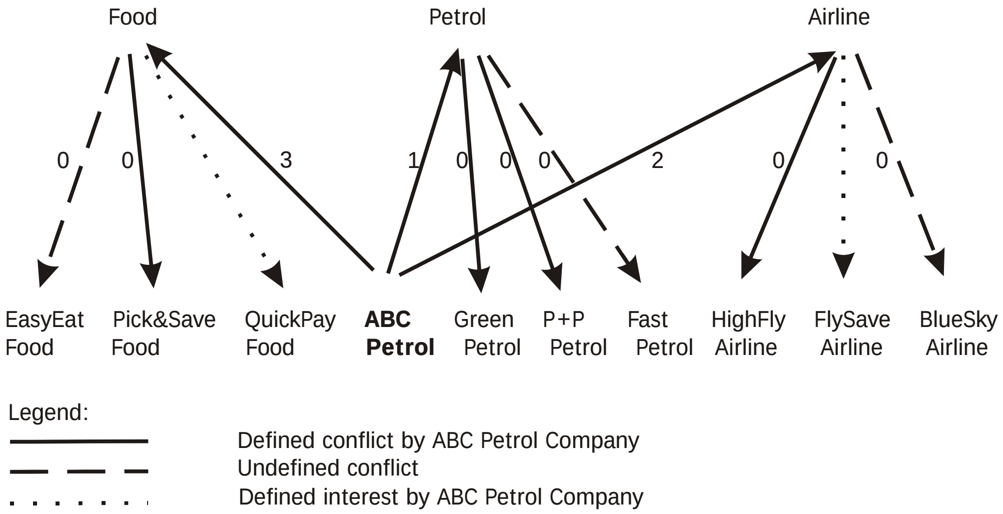


Figure 5: Possible paths from ABC Petrol Company to three undefined companies

4.3 Dealing with an ealing with an access request

Let $MiningAgent_x$ (NS_x, c_p) represent an access request

where:

(1)$MiningAgent_x$ is the data miner issuing the access request.

(2)NS_x is a set of non-conflicting secondary agents (NS_x) that represent all data sets of secondary agents to which $MiningAgent_x$ already had access prior to the current request:

$NS_x = \{c_{MiningAgent\_x\_1}, c_{MiningAgent\_x\_2}, c_{MiningAgent\_x\_3}, \ldots, c_{MiningAgent\_x\_p}\} \subset CNC$

(3)c_p is the data set to which $MiningAgent_x$ is requesting access.

4.3.1 Grant / reject an access request

With regard to access requests, the first step in the decision-making process is to test whether c_p is an element of all the non-conflicting secondary agents. Is $c_p \in$ NC? If the answer is yes, then access is granted. If the answer is no, the second step is to test whether c_p exists in the access group list of any secondary agent in the set NS_x. If the answer is yes, then access is granted. If the answer is no, the third step is to test whether c_p is in the sphere of conflict of any secondary agent in the set NS_x.

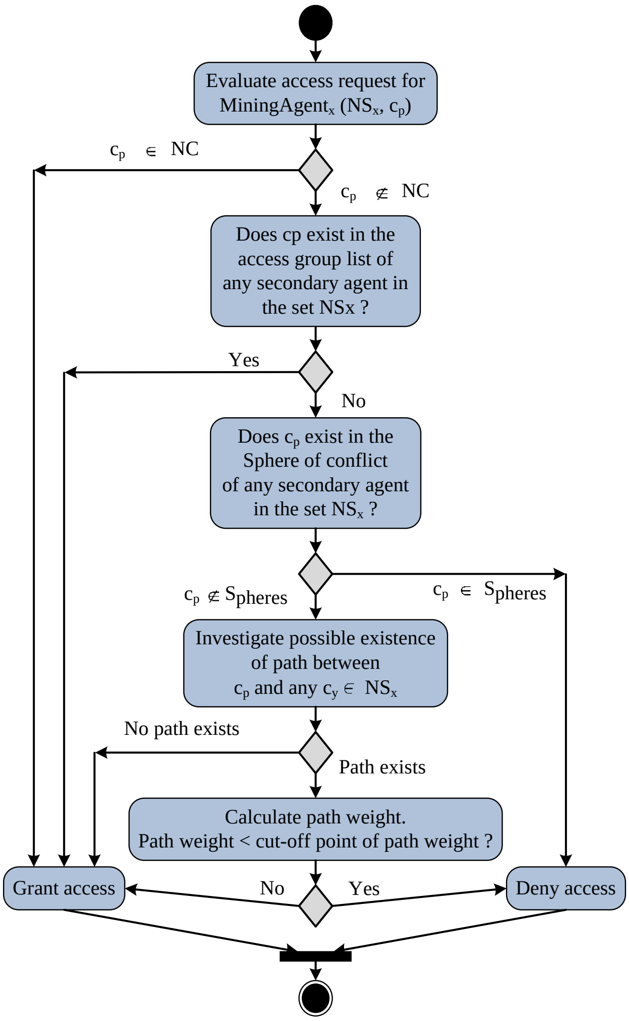

Evaluate access request for MiningAgent$_x$ (NS_x, c_p)
$c_p \in$ NC
$c_p \not\in$ NC
Does cp exist in the access group list of any secondary agent in the set NSx ?
Yes
No
Does c_p exist in the Sphere of conflict of any secondary agent in the set NS_x ?
$c_p \not\in$ Spheres
$c_p \in$ Spheres
Investigate possible existence of path between c_p and any $c_y \in NS_x$
No path exists
Path exists
Calculate path weight. Path weight < cut-off point of path weight ?
Grant access
No
Yes
Deny access

Figure 6: High-level logic of an access request – MiningAgent$_x$ wants access to data sets of secondary agent c_p

If the answer is yes, which means that

$$c_p \in S_{pheres} = \{S_{MiningAgent\_x\_1}, S_{MiningAgent\_x\_2}, S_{MiningAgent\_x\_3}, \ldots, S_{MiningAgent\_x\_p}\},$$

then access is rejected.

If the answer is no, which means that

$$c_p \notin S_{pheres} = \{S_{MiningAgent\_x\_1}, S_{MiningAgent\_x\_2}, S_{MiningAgent\_x\_3}, \ldots, S_{MiningAgent\_x\_p}\},$$

then the fourth step is to test whether a path exists between c_p and any secondary agent c_y in the set NS_x.

If there is no path between the two, $MiningAgent_x$ is granted access, but if a path exists between them, the fifth step is to calculate the weight of the path. If the weight of the path is small enough to be definable as *"a conflict does exist"*, then no access can be granted to $MiningAgent_x$. If, however, the weight of the path is big enough to be definable as *"no conflict exists"*, then access is granted to $MiningAgent_x$.

4.4 Access request exemplified in the case study

Below is an example to explain an access request in the case study. The following data is known information at the time of the access request:

- The set of all possible secondary agents (clients of DM Bank) is CNC = $\{c_1, c_2, c_3, c_7\}$.
- The Conflict-of-Interest field is C = $\{c_1, c_2, c_7\}$.
- The Non-Conflict-of-Interest field is NC= $\{c_3\}$.
- c_2 is in the sphere-of-conflict field of c_1 and c_7 is outside of the sphere-of-conflict field of c_1.
- The cut-off point of the path weight is known and for purposes of this example equals 5.
- Thus (wPk (ci, cj)=5).
- The cut-off point for conflict of interest (r_i) for each c_i is known.
- The distance function, $d_{i,j}$, indicating the degree of conflict between a secondary agent $c_i \in$ C and any other conflicting $c_j \in$ C is known.
- The NS_1 of $MiningAgent_1$ is $\{c_1\}$, which means Mining agent 1 currently has access

only to Secondary agent 1's data sets – thus the history record of MiningAgent_1 is $NS_1 = \{c_1\}$. ….. 1

Consider access request A:

- MiningAgent_1 requests access to the data sets of c_2.

More known facts:

$c_2 \in C$; $c_2 \in$ of the sphere of conflict of c_1; $NS_1 = \{c_1\}$

This request is evaluated as follows:

Answer to access request: *No access granted*

Reason: c_1 and c_2 are in conflict with one another.

Result of history record: $NS_1 = \{c_1\}$ (history record stays the same as in 1)

….. 2

Consider access request B:

- MiningAgent_1 requests access to the data sets of c_3.

More known facts:

$c_3 \in NC$.

This request is evaluated as follows:

Answer to access request: *Access granted*

Reason: c_3 is in conflict with no secondary agent. It is safe to grant access to c_3.

Result of history record: $NS_1 = \{c_1, c_3\}$ (add c_3 to history record in 2) ….. 3

Consider access request C:

- MiningAgent_1 requests access to the data sets of c_7.

More known facts:

$c_7 \in C$; $c_7 \notin$ of the sphere of conflict of c_1; $NS_1 = \{c_1, c_3\}$

This request is evaluated as follows:

A potential conflict-of-interest path between c_3 and c_7 will not exist, because $c_3 \in$ NC.

Establish if a potential conflict-of-interest path exists between c_1 and c_7.

If a path does exist, calculate the weight of the path.

If the smallest possible weight of the path is bigger than the given cut-off point of the path weight (which is 5 for this example), then:

Answer to access request: *Access granted*

Result of history record: $NS_1 = \{c_1, c_3, c_7\}$ (add c_7 to history record in 3) ….. 4a

If the smallest possible weight of the path is smaller than the given cut-off point of the path weight (which is 5 for this example), then:

Answer to access request: *Access denied*

Result of history record: $NS_1 = \{c_1, c_3\}$ (history record stays the same as in 3) ….. 4b

This logic has been implemented in a prototype that will be briefly summarised in the next section.

4.5 Brief overview of the prototype

This prototype was developed to serve as proof of concept of the proposed CBAC model as described in this paper.

The prototype was developed in Microsoft C#, using Microsoft Visual Studio 2008. It requires the .NET Framework v3.5 to run, for the following reasons: Firstly, the prototype uses WPF (Windows Presentation Foundation) for the layout and presentation aspects, and secondly, the prototype utilises LINQ (Language Integrated Queries).

The directed bipartite graph is implemented using two lists of objects. The first list contains all nodes of the graph. Each node contains a single object (such as a functional business domain or secondary agent). The second list contains all edges of the graph. Each edge object references two nodes (referred to as “from” and “to”), as well as the weight of the edge. To retrieve information effectively, LINQ is used to iterate and return specific information from these two lists (such as all edges that lead from a specific node).

A class model for the prototype is depicted in Figure 7.

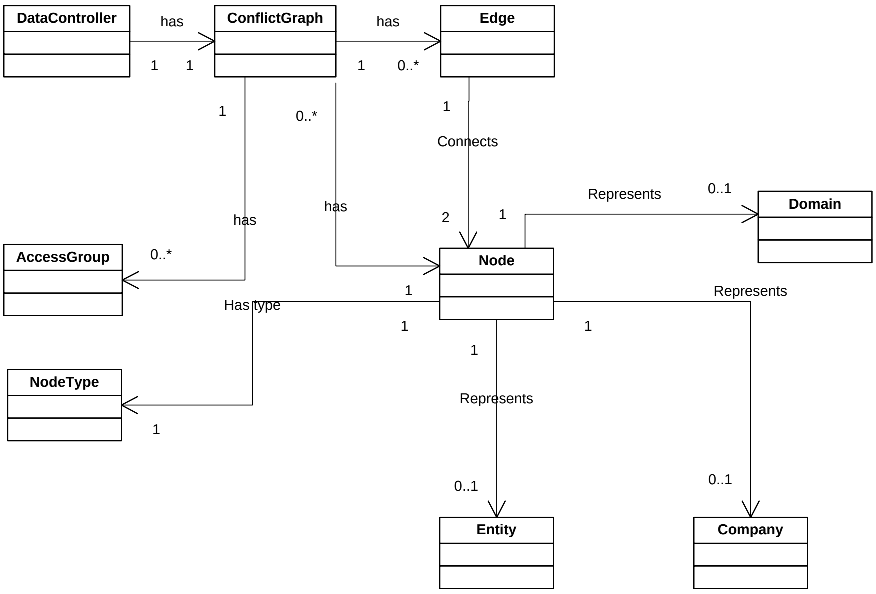


Figure 7: A class model for the prototype

The prototype can be divided into two subsystems, namely the database-and-graph subsystem, and the presentation subsystem. Figure 8 depicts the structure of the database-and-graph subsystem by means of a class model that represents this subsystem.

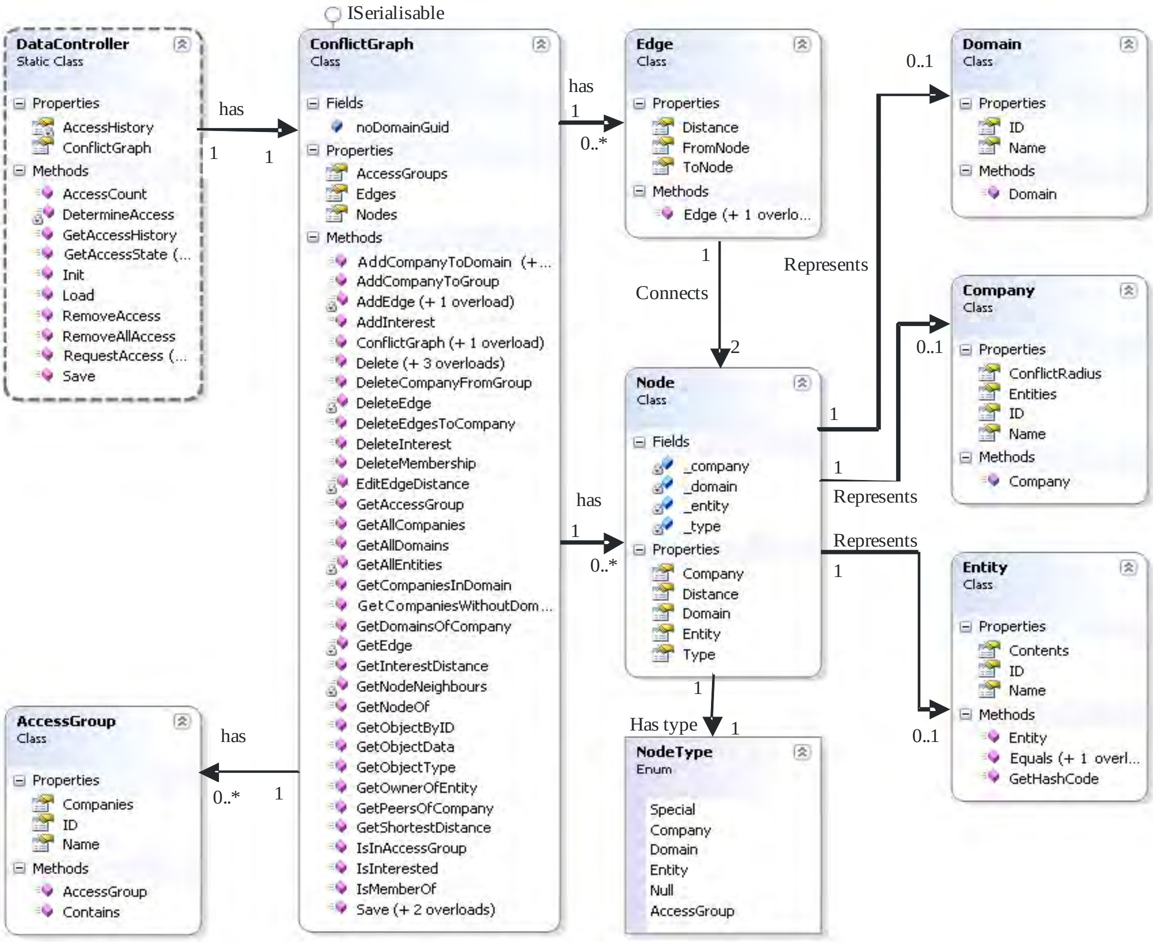


Figure 8: The Database-and-Graph Subsystem

The basic structure of the graph is implemented using the node and edge classes. Each node contains either a functional business domain, a secondary agent (a *company*), or an object in a secondary agent (an *entity*). Each node also stores the type of object it contains. Furthermore, each Domain, Company and Entity object stores the respective details of a single object and is uniquely identified by means of a Globally Unique Identifier (GUID). All of the above are stored in a ConflictGraph object, which contains all functionality related to Create, Read, Update and Delete (CRUD) on the graph and also makes provision for conflict detection. The conflict graph furthermore contains the access group lists.

Finally, the DataController static class instantiates the ConflictGraph and maintains the access history of a mining agent. Note that the prototype only considers a single mining agent, so only a single access history is kept. The DataController class also handles serialisation of the ConflictGraph.

All operations by the presentation subsystem (depicted in Figure 9) on the database-and-graph subsystem are performed using the properties and methods provided by the static *DataController* class or its instantiated *ConflictGraph* object.

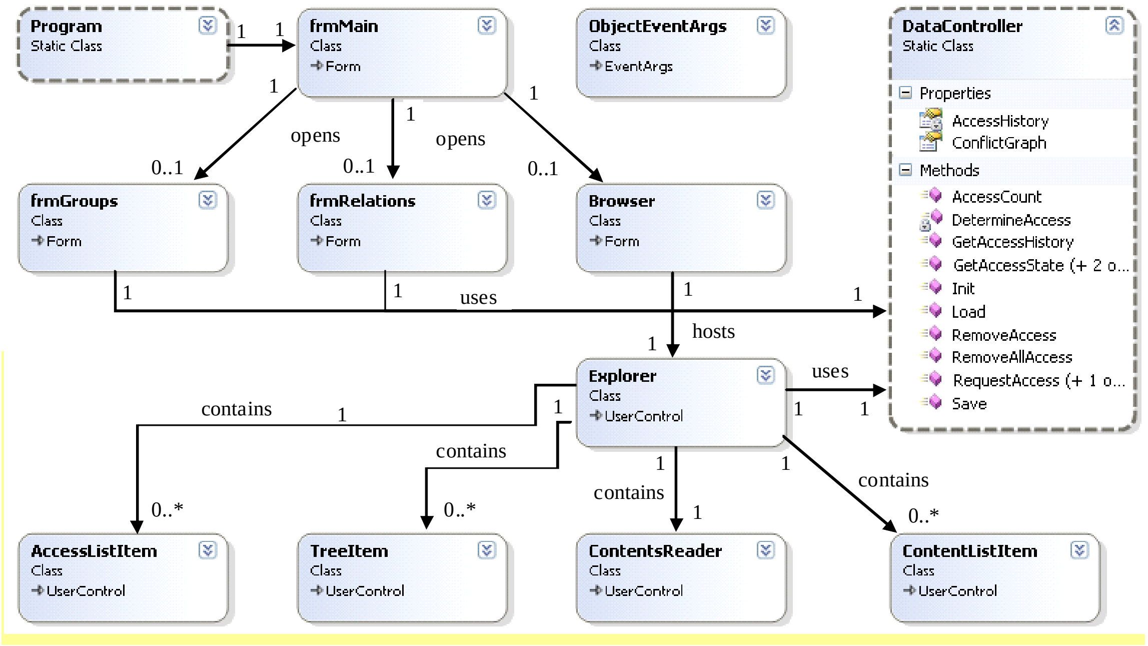


Figure 9: The Presentation Subsystem

Although the prototype provides features for adding, editing or removing items in the database, the primary algorithm of note is used to determine whether a user with a given access history is allowed to access a specific company by checking for conflicts on the conflict graph.

An extensive explanation of this prototype will follow in another paper.

5 Conclusion

The current research proposes a new access control model called the Conflict-Based Access Control (CBAC) model. The CBAC deals with a changing business environment, i.e. a data-mining environment [18] and the model supports the following:

- The definition of conflict-of-interest classes as well as the assumption that such classes are not 'disjoint' or separate.
  The original concept of conflict-of-interest classes as defined for the CWSP involved strict 'walls' based on business definitions [6]. This means that these conflict-of-interest classes were based on functional business domains that compartmentalised organisations with no overlaps between the different compartments. The ACWSP model pointed out that these conflict-of-interest classes are 'seldom disjoint' [8] and that they 'overlap' [8], which is equally true for functional business domains.

- A sphere of conflict for a secondary agent specified by that secondary agent itself rather than by functional business domains.
  The different degrees of conflict of interest between the secondary agent and all its known conflicting secondary agents are quantifiable, as is the cut-off point for conflict of interest for the secondary agent. This is the first step towards an access control model that takes into account the different conflicting relationships that exist between competing secondary agents within the same global environment.

- The addition of a bipartite graph that assists the global agent to add knowledge to the model by giving information regarding a Potential Conflict of Interest (PIT) between two secondary agents.

This potential conflict of interest is calculated with the assistance of a potential conflict-of-interest path. The existence of more than one potential conflict-of-interest path is acknowledged and a possible k potential conflict-of-interest paths are calculated and added to the model. The method of weighting these k different potential conflict-of-interest paths so as to select the shortest potential conflict-of-interest path is then part of the model description. The global agent may specify a cut-off point of path weight to ensure that the potential conflict-of-interest path does not get too long.

- The concept of Potential Conflict of Interest to be determined by the global agent.
  In the CBAC model, the functional business domain definition is a dynamic concept that plays a role in the conflict-of-interest classes. The model allows for the implementation of a potential conflict of interest in the same or another functional business domain as determined by a secondary agent or by the global agent at a specific point in time. For example: if secondary agent c_i has developed a new interest in another secondary agent c_j, then agent c_i may ask to be treated as such. From then on, the same mining agent cannot work on the data sets of both secondary agents again. Secondary agents can change this conflict of interest during execution time.

The current research finds the CBAC model to be dynamic and thus able to cope with a rapidly changing commercial environment. Today's business environment is much more dynamic compared to that of a decade ago. Companies do not only diversify, but (in some cases) also change their business strategy, which might entail entering a new functional business domain. For example, a secondary agent develops a 'new' major interest in another secondary agent, which creates potential conflict of interest in the future. It is argued that such changes have to be reflected in the decision-making process that precedes the granting or rejecting of access requests, especially so in a data-mining environment. This dynamic adaptation to business requirements, in conjunction with the preservation of separated duties, is the main contribution of the CBAC model. The CBAC model bases its definition of conflict on conflict between secondary agents as defined by a specific secondary agent, in other words, it incorporates the latter agent's 'view of the world'.

The current research also makes an important contribution towards commercial security models in general, in that the CBAC model is typically applicable in a data-mining environment. The private information of banking clients can and must be used in data-mining activities – for the benefit of the banking clients themselves and the banks in general – yet without sacrificing client confidentiality or bank credibility. The CBAC model makes this possible.

REFERENCES

[1] M. Loock and J. H. P. Eloff, "Minimizing Security Risk Areas revealed by Data mining," in *ISSA2002 Information Security for South Africa*, Misty Hills, Muldersdrift, Gauteng, 2002.

[2] M. Bishop, *Computer Security: art and science*: Pearson Education Inc., 2003.

[3] D. Bell and L. LaPadula, "Secure Computer Systems: Mathematical Foundations and Model," National Technical Information Service, Bedford, MA 1(M74-244), March 1973.

[4] D. Bell and L. LaPadula, "Secure Computer System: Unified Exposition and Multics Interpretation," MITRE Corporation, Bedford, MA Technical Report MTR-2997 Rev. 1, March 1976 1976.

[5] D. Gollmann, *Computer Security*, 2nd ed. West Sussex: John Wiley & Sons Ltd., 2006.

[6] D. F. C. Brewer and M. J. Nash, "The Chinese Wall Security Policy," in *IEEE Symposium on Research in Security and Privacy*, Oakland, California, 1989, pp. 206-214.

[7] E. Bertino, *et al.*, "Privacy-Preserving Database Systems," in *Lecture Notes in Computer Science*. vol. Volume 3655/2005, ed Heidelberg: Springer Berlin, 2005, pp. 178-206.

[8] R. Agrawal and R. Srikant, "Privacy-Preserving Data Mining," in *ACM SIGMOD Conference on Management of Data*, Dallas, Texas, 2000.

[9] C. Clifton and D. Marks, "Security and privacy implications of data mining," in *ACM SIGMOD International Conference on Management of Data*, 1996, pp. 15 - 19.

[10] M. Kantarcioglu and C. Clifton, "Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, pp. 1026 - 1037, September 2004.

[11] C. Conrado, *et al.*, "Privacy-Preserving Digital Rights Management " in *Lecture Notes in Computer Science*. vol. 3178, ed Heidelberg: Springer Berlin, 2004, pp. 83-99.

[12] B. Thuraisingham, "Data Warehousing, Data Mining and Security," in *IFIP Database Security Conference*, Como, Italy, 1996.

[13] T. Y. Lin, "Chinese Wall Security Policy - An Aggressive Model," in *Conference Proceedings of the Fifth Annual Computer Security Applications Conference*, Tucson, Arizona, USA, 1989, pp. 282-289.

[14] T. Y. Lin, "Chinese Wall Security Policy Models: Information Flows and Confining Trojan Horses," in *Conference Proceedings of the Seventeenth Annual IFIP WG 11.3 Working Conference on Data and Applications Security*, Estes Park, Colorado, U.S.A., 2003.

[15] OASIS. (2010, 21 June). *eXtensible Access Control Markup Language (XACML) TC*. Available: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml

[16] C. A. Ardagna, *et al.*, "The Architecture of a Privacy-Aware Access Control Decision Component " in *Lecture Notes in Computer Science*. vol. Volume 3956/2006, ed Heidelberg: Springer Berlin, 2006, pp. 1 - 15.

[17] E. Bertino, *et al.*, "Secure Knowledge Management: Confidentiality, Trust, and Privacy," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 36, p. 10, May 2006.

[18] M. Loock and J. H. P. Eloff, "A new Access Control model based on the Chinese Wall Security Policy Model," in *ISSA2005 New Knowledge Today Conference*, Balalaika Hotel, Sandton, South Africa, 2005.

[19] S. Lipschutz and M. L. Lipson, *Discrete Mathematics*. New York: McGraw-Hill, 2003.

Marianne Loock received the BSc degree in Computer Science and Mathematics at the University of Pretoria, South Africa, and the BSc (Honours) and MSc degrees in Information Systems at the University of South Africa. She is a PhD student in the Department of Computer Science, University of Pretoria, South Africa, and a Senior Lecturer in the School of Computing at the University of South Africa. Her current research interests include data mining, access control and information security.

Jan Eloff is appointed as the Research Director of SAP Research Pretoria specialising in Mobile Empowerment. He holds a PhD in Computer Science and is appointed as an Extraordinary Professor in Computer Science at the University of Pretoria. Read more about Jan at: http://www.cs.up.ac.za/~eloff and http://za.linked.com/in/janeloff. He can be contacted at: jan.eloff@sap.com.

Johannes Heidema obtained the M.Sc. in Mathematics in 1963 at the PU for CHE, the doctoral exam in Mathematics and Logic in 1964 at the University of Amsterdam, and the D.Sc. in Mathematics in 1966 at the PU for CHE. He lectured Mathematics at the PU for CHE, at UPE, and then (1970-91) as professor at the RAU. He has been a professor in Mathematics at UNISA since 1992, emeritus from 2007. His main academic interests are the foundations of mathematics and logic, and in particular the implications of a model-theoretic approach in theoretical physics (quantum computation and logic), in computer science (nonmonotonic logic for artificial intelligence), and in the philosophy of science (epistemology and realism).