A Study on Policy Integration and Anonymous Communication for Protecting Personal Information

個人情報保護のための ポリシー統合と匿名通信に関する研究

Kazuhiro Kono

Division of Electrical, Electronic and Information Engineering

Graduate School of Engineering

Osaka University

Japan

January 2010

Preface

This dissertation presents my research work on policy integration and anonymous communication for protecting personal information. The dissertation is the result of the research during the Ph.D. course at the Division of Electrical, Electronic and Information Engineering, Graduate School of Engineering, Osaka University. The dissertation is organized as follows.

Chapter 1 describes the background, the motivation, and the purpose of this research, and presents the outline of this dissertation.

Chapter 2 presents an overview of two privacy protection technologies, policy integration and anonymous communication systems, which we focus on in the dissertation. We also clarify the characteristic of this dissertation, comparing our approach with related work in the research areas.

Chapter 3 describes a policy integration framework for unifying management of various access rights management systems. Policy integration is a technique which generates integrated policies suitable for all the systems by using inheritance relations of access rights. Since each system has its own inheritance relation, we need to integrate each inheritance relation in order to obtain inheritance relations for policy generation. Focusing on the integration of inheritance relations, we introduce a matrix-based algorithm for integrating inheritance relations of access rights for policy generation. Since inheritance relations of access rights are found in subject, resource, and action categories for existing systems, our algorithm first integrates inheritance relations in each category, and next, integrates inheritance relations of all categories. It is proven that these operations can be carried out by basic matrix operations. This enables us to implement the integration algorithm very easily.

In Chapters 4 and 5, we analyze the performance of an anonymous communication system called 3-Mode Net (3MN). Although 3MN has more advantages than other anonymous communication systems, the performance of 3MN is not analyzed in detail. In Chapter 4, we evaluate the number of relay nodes and the number of encryption required for communication in 3MN. In particular, we give explicit formulas of their probability distributions, expectations, and variances. From the obtained formulas, we investigate the impacts of the probabilities of mode selections and the initial multiplicity of encryption on the behavior of 3MN. We also show several requirements for avoiding the situation where the number of relay nodes becomes extremely large.

In Chapter 5, we investigate sender anonymity against collaborating nodes who collude with each other in order to identify a message sender. We derive the probability that the first immediate predecessor of all the collaborating nodes on the communication path coincides with the message sender. The probability is derived from probability generating functions. Using the result, we consider the influences of the probabilities of mode selections and the initial multiplicity of encryption on sender anonymity against collaborating nodes. From the results obtained in Chapter 4, we also investigate the relationship between the number of relay nodes and sender anonymity, and consider a condition both for providing high sender

anonymity and for reaching a receiver through a small number of relay nodes.

Chapter 6 concludes this dissertation.

Acknowledgments

The research described in this dissertation has been carried out during my tenure of doctoral course at the Graduate School of Engineering, Osaka University under the guidance of Professor Noboru Babaguchi at the Graduate School of Engineering, Osaka University.

First of all, I would like to express my deepest appreciation to my supervisor, Professor Noboru Babaguchi of the Graduate School of Engineering of Osaka University for his instruction, continuing encouragement, and valuable discussions throughout this research. He has taught me a lot of things both in the aspect of my research and the aspect of my life. I have learned many valuable lessons through this research, which have further developed my abilities.

I wish to thank the members of my committee: Professor Noboru Babaguchi, Professor Kyo Inoue, and Professor Tetsuya Takine of the Graduate School of Engineering, Osaka University, who have given me many insightful suggestions for the dissertation.

I also take pleasure in thanking Professor Zen-ichiro Kawasaki, Professor Ken-ichi Kitayama, Professor Shozo Komaki, and Professor Seiichi Sampei of the Department of Information and Communications Technology of the Graduate School of Engineering, Osaka University, and Professor Riichiro Mizoguchi and Professor Takashi Washio of the Institute of Scientific and Industrial Research, Osaka University for providing thoughtful comments on the dissertation.

For the research of policy integration, I am especially grateful to Dr. Hiroaki Kamoda, Dr. Masaki Yamaoka, and other members of the Research and Development Headquarters of NTT DATA Corporation for a number of fruitful discussions and suggestions. I also show my appreciation to Mr. Akihito Aoyama of the Financial Systems Sector, NTT DATA Corporation for helpful comments.

I would like to express my gratitude to all the past and present members in the Media Integrated Communication Laboratory in the Department of Information and Communications Technology (Babaguchi Laboratory). They have always provided encouragement and their friendship have given me strength through the difficult times. I especially thank Lecturer Naoko Nitta, Assistant Professor Yoshimichi Ito, and Postdoctoral Fellow Dr. Minh Son Dao of the Graduate School of Engineering, Osaka University, and Dr. Xiaoyi Yu of the School of Software and Microelectronics, Peking University, who provided me with helpful comments and invaluable advice. I give my thanks to Mr. Naoki Miyake and Mr. Shinnosuke Nakano for their useful discussions about anonymous communication.

Finally, I would like to thank my entire family for their understanding, support, encouragement, and patience during the whole period of my education.

Contents

List of Figures

List of Tables

List of Symbols

Symbol	Meaning
s_i	The i-th subject in the set of subjects
r_l	The l-th resource in the set of resources
a_m	The m-th action in the set of actions
$G(V, E)$	A directed graph
V	A set of vertices
E	A set of edges
v_i	The i-th vertex in a directed graph
e_{ij}	An edge directed from v_i to v_j in a directed graph
$G_h(\mathrm{ACT}_h, \mathrm{IR}_h^{\mathrm{ACT}})$	A graph which represents inheritance relations in action category of the h-th system ($1 \leq h \leq H$)
$G(\mathrm{ACT}, \mathrm{IR}^{\mathrm{ACT}})$	A graph which represents integrated inheritance relations in action category of the integrated system
$G(\mathrm{RES}, \mathrm{IR}^{\mathrm{RES}})$	A graph which represents integrated inheritance relations in resource category of the integrated system
$G(\mathrm{SUB}, \mathrm{IR}^{\mathrm{SUB}})$	A graph which represents integrated inheritance relations in subject category of the integrated system
A_h	An adjacency matrix expressing inheritance relations in action category of system h ($1 \leq h \leq H$)
A	An adjacency matrix expressing integrated inheritance relations in action category
R	An adjacency matrix expressing integrated inheritance relations in resource category
S	An adjacency matrix expressing integrated inheritance relations in subject category
X_{RA}	An adjacency matrix of integrated inheritance relations of resource and action categories
X_{SRA}	An adjacency matrix of integrated inheritance relations of subject, resource, and action categories
I	A unit matrix

p_{D}	The probability to choose D-Mode
p_{T}	The probability to choose T-Mode
p_{E}	The probability to choose E-Mode
d	The number of times when D-Mode is chosen
t	The number of times when T-Mode is chosen
e	The number of times when E-Mode is chosen
k	The initial multiplicity of encryption
N	A random variable representing the number of relay nodes
N_{e}	A random variable representing the number of encryption
K	A random variable representing the number of multiplicity of encryption
τ_k	A random variable representing the number of relay node under the condition that the initial multiplicity of encryption is k
ϵ_k	A random variable representing the number of encryption under the condition that the initial multiplicity of encryption is k
M_{N}	The expectation of the number of relay nodes
V_{N}	The variance of the number of relay nodes
M_{E}	The expectation of the number of encryption
V_{E}	The variance of the number of encryption
L_i	The event where the first collaborating node appears the i-th node on a communication path
L_{i+}	The event where the first collaborating node appears the i-th node or more than the i-th node on a communication path
J	The event where the first immediate predecessor among all collaborating nodes is the message sender
$P(N = r)$	The probability distribution of the number of relay nodes
$P(N_{\mathrm{e}} = r)$	The probability distribution of the number of encryption
$g_{\tau_k}(\lambda)$	The probability generating function for τ_k
$g_{\epsilon_k}(\lambda)$	The probability generating function for ϵ_k
$E(X)$	The expectation of X
$V(X)$	The variance of X
$P(J\|L_{1+})$	The conditional probability that the first immediate predecessor among all collaborating nodes is a message sender under the condition that a collaborating node receives a data set

Chapter 1

Introduction

As Information Technology (IT) has developed rapidly, various IT systems and services are essential components for the present society. A lot of people have personal computers and use Internet. Thus, they treat various information on their computers, and receive numerous services on the Internet. For example, we can buy almost all goods on the Internet and conduct financial transactions online without going outside. We also create many documents as electrical data with computers without creating by hand.

In the deep penetration of IT into our lives, one of the important issues is how "personal information" is handled. According to the personal information protection law in Japan, personal information means information about an individual which can identify the specific individual. For example, the name and the address of an individual are his personal information. On the Internet world, personal information includes the mail address and the IP address of an individual as well as the information caused by the behavior of an individual such as communication records and purchase records.

We must address our personal information appropriately because the leakage of such personal information might lead to the infringement of our privacy. If the personal information of an individual is exposed on the Internet, the information is known to a number of people. The information might be also used to fraudulent actions. In addition, in recent years, there exists a social trend where any trifling personal information is not disclosed. Therefore, technologies for protecting personal information are strongly required.

The purpose of our research is to develop technologies for protecting personal information. First of all, in order to give a better understanding of technologies necessary for preventing the misuse and the leakage of personal information, we consider the life-cycle, the requirements, and the protection methods of personal information when an individual requests and receives a service from a service provider, which is shown in Table 1.1 [1, 2]. From Table 1.1, the life-cycle of personal information consists of the following four phases: collection, practical use, preservation, and disposition. First, in the collection phase, the service provider receives the personal information of the individual necessary for receiving a service. Second, in the practical use phase, the provider provides a requested service. The

provider also uses the personal information on the intended use. Third, in the preservation phase, the provider manages the given personal information adequately. Finally, in the disposition phase, the provider disposes the given personal information.

In the first and second phases, since it communicates between an individual and a service provider, secure communication technologies based on the confidential level of a requested service are needed. Among secure communication methods, anonymous communication is a communication method where confidentiality is of the highest concern. In the third and final phases, since personal information is managed in a service provider, the provider needs to treat personal information accurately and appropriately in order to prevent information leakage from the inside of the provider. One of the most effective methods for preventing information leakage is to introduce an access control system. Therefore, in this dissertation, we focus on the following two topics: 1) access control; 2) anonymous communication.

Access Control An access control system permits accesses from proper users and rejects accesses from improper users. The use of access control systems enables a provider to treat personal information accurately and appropriately. Access control systems include access rights management systems and network admission control systems.

In order to prevent information leakage, it is very important to establish the rules of access control systems appropriately. As shown in Fig. 1.1, the appropriate establishment of access control systems is effective to prevent the information leakage. The reason why it is effective is that the establishment mistakes such as administration errors and operational errors account for a higher percentage of the causes of information leakage than attacks from the outside of the organization such as worms/viruses and bugs/security holes without physical causes such as losses/misplaces and thefts. Accordingly, it is desirable to develop techniques for managing access control systems, in particular, access rights management systems adequately. We aim at realizing policy integration by which administrators can unify management of access rights management systems.

Anonymous Communication Anonymous communication not only protects the content of a message but also protects the sender and the receiver of the message. Anonymous communication also serves as a foundation in other anonymous technologies such as anonymous authorization and anonymous ID. Anonymous communication is important because it

Table 1.1: Life-cycle, requirements, and protection methods of personal information [1].

Life-cycle	Requirement	Protection method
Collection	Security and anonymity of network	Secure communication
Practical use	Security and anonymity of network	Secure communication
Preservation	Protection of leakage	Access control system
Disposition	Protection of retrieval	Erasion of information

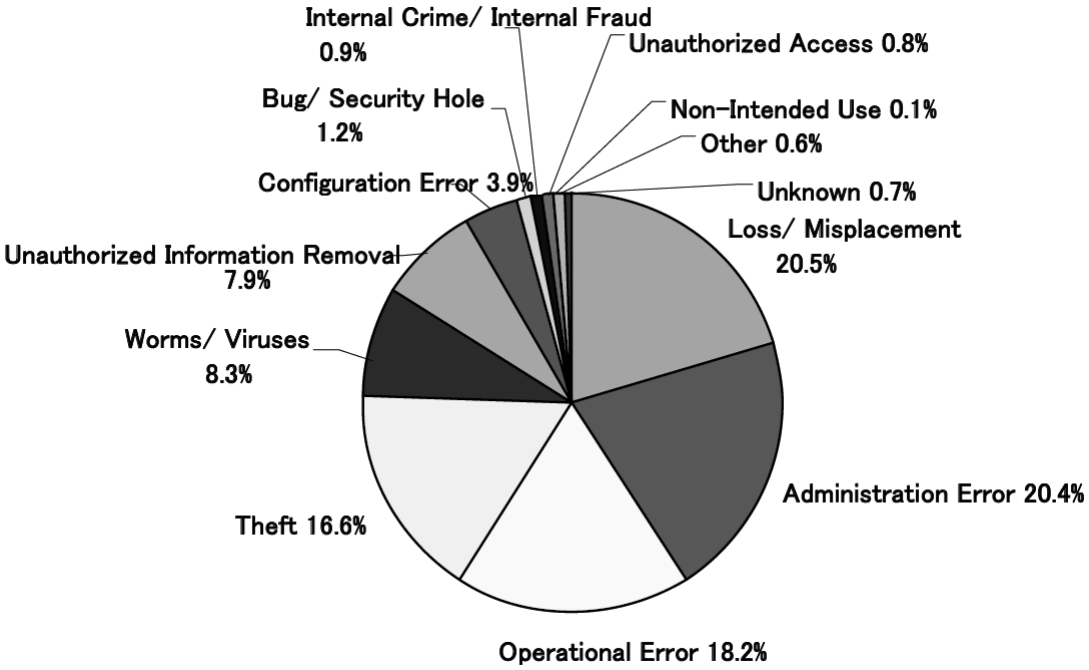


Figure 1.1: Ratio of leaks by cause (Published by Survey Report of Information Security Incident 2007).

provides higher confidentiality than other secure communication technologies such as encryption communication. Although encryption communication represented by TSL/SSL and ID authorization is used commonly all over the world, encryption communication exposes facts that an individual uses a service. Suppose that an individual consults his medical treatment for medical staff by using SSL or E-mail. In this case, the individual cannot consult anonymously because the medical staff can know who communicates by investigating the header of a IP packet or his mail address. Thus technologies for achieving anonymity on the Internet are needed.

In this dissertation, we develop policy integration and anonymous communication for protecting personal information. Administrators can decrease their time and effort because policy integration enables them to unify management of access rights management systems. This technique decreases administration errors and operational errors which remain important causes of information leakage as indicated in Fig 1.1. We also focus on an anonymous communication system called 3-Mode Net (3MN) [3]. Compared to the existing anonymous communication systems Crowds [4] and Onion Routing [5], 3MN provides receiver anonymity unlike Crowds and the number of encryption in 3MN is smaller than Onion Routing. Although 3MN is superior to Onion Routing and Crowds, 3MN has not been analyzed in detail, and thus, it is difficult to decide the parameters of 3MN from required performances.

In our research, we investigate the impacts of the parameters of 3MN on the performance of 3MN.

The outline of the dissertation is as follows. In Chapter 2, we present an overview of two personal protection technologies, policy integration and anonymous communication. We also describe the characteristics of our approaches, comparing them with related work in these research areas. Chapter 3 describes our framework of policy integration for unifying management of various access rights management systems. In particular, we present a method for integrating inheritance relations of access rights for integrated policy generation. In Chapter 4 and Chapter 5, we analyze the performance of an anonymous communication system called 3-Mode Net. In Chapter 4, we investigate the number of relay nodes and the number of encryption required for communication. In Chapter 5, we evaluate sender anonymity against collaborating nodes who collude with each other in order to identify a message sender. Chapter 6 concludes this dissertation, and presents future research directions.

Chapter 2

Technologies for Protecting Personal Information

2.1 Introduction

In this chapter, we describe two technologies for protecting personal information, policy integration and anonymous communication, which we focus on in the dissertation. We describe the framework of policy integration with inheritance relations of access rights and an overview of an anonymous communication system named 3-Mode Net (3MN). We first present the problems of related work in the research area of policy integration, and clarify the characteristics of our approach. We then discuss several issues on 3MN.

2.2 Policy Integration Framework

With the increase and distribution of information, data security and privacy are critical issues. System administrators pay much attention on prevention of information leakage caused by fraudulent actions and mis-operations, and introduce access rights management systems to establish access control policies for various applications and file systems.

Since each access rights management system has its own access control policy, administrators have to configure policies individually in all systems. Therefore, in order to eliminate inappropriate policies, techniques which treat policies in each system appropriately are strongly required.

Policy integration is one of the access rights management/establishment techniques which prevent information leakage caused by fraudulent actions and mis-operations. A classification map about access rights management/establishment techniques is shown in Fig. 2.1. Techniques for preventing information leakage are as follows: policy refinement, policy integration, and policy conflict detection/resolution. Policy refinement is a technique for deriving specific policies from abstract policies. A specific policy indicates a policy estab-

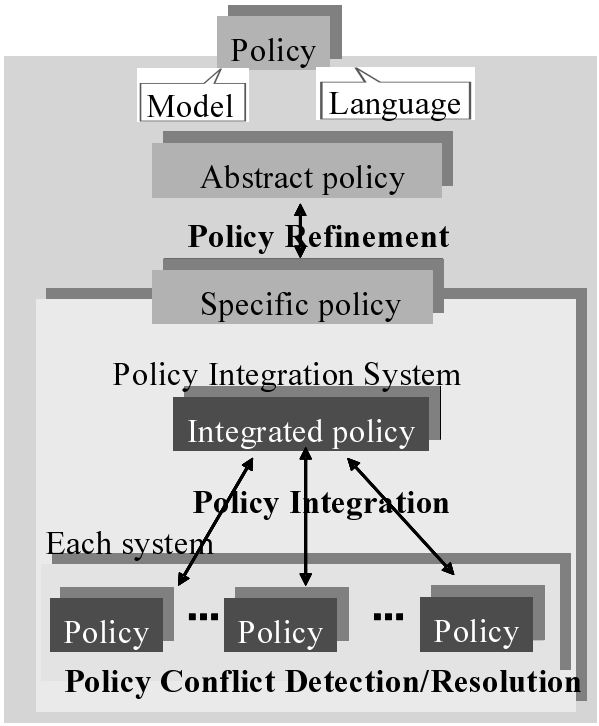


Figure 2.1: Access rights management technologies.

lished for systems in practice, whereas an abstract policy indicates a policy specified in a guideline. The detail of this technique has been discussed in [6, 7, 8]. Policy integration is a technique which unifies management of policies of various access rights management systems. Policy conflict detection/resolution is a technique which detects the conflict and the mis-establishment of policies and changes such inappropriate policies to appropriate policies. The detail of this technique has been discussed in [9, 10, 11, 12, 13, 14].

Policy integration and policy conflict detection/resolution are crucial in order to prevent the establishment of incorrect policies in each system. A couple of methods on policy conflict detection/resolution are developed by Kamoda et al. [11, 12, 13]. In this research, we limit our attention on a policy integration technique.

Policy integration is useful for adequately managing each access rights management system with small effort and time. Suppose that the personal information protection law and the security guideline of organizations are changed. In this case, administrators have to pay enormous attention for updating every policy for each system because each system is independent each other. Thus, they can unify management of each system without excessive effort by introducing a policy integration system.

In what follows, we describe policy integration in detail and consider required conditions by comparing with related work.

2.2.1 Preliminary

We introduce very important notions of our research, that is, an access control policy and inheritance relations of access rights.

Access control policy In policy specification languages, e.g., eXtensible Access Control Markup Language (XACML) [15, 16] and Ponder [17], every access control rule is essentially specified by three elements: *subject*, *resource*, and *action*. Subjects are the entities that can perform actions in systems (e.g., users, computers). Resources are the entities that contain or receive information for which access is requested (e.g., files, devices). Actions are the types of access that is being requested (e.g., edit, copy, read, write).

Using these elements, administrators establish each access control rule called *access control policy* or *policy* for each system. A policy is represented as "a subject can or cannot perform an action for a resource". Examples of policies are as follows: everyone can read public documents; administrators can edit a document.

Inheritance relations of access rights An inheritance relation of access rights is a partial order relation between the attributes and behaviors of objects, whereby the senior attributes and behaviors of objects acquire the permissions of their juniors, and thus, the access rights of their juniors are inherited to the seniors. For example, suppose that there exist the group of general users and the group of system administrators. In this case, since the group of the administrators is superior to the group of general users, the group of the administrators can read a document when the group of the general users can read the document.

The use of inheritance relations of access rights prevents administrators from establishing inappropriate policies. In the above example, when there exists a policy where the group of general users can read a document, administrators need to establish a policy where the group of administrators can read the document from the inheritance relations. That is, it is shown that they must not establish a policy where the group of administrators cannot read the document from the inheritance relations.

Inheritance relations are introduced in various access rights management systems, e.g., Microsoft Windows® Rights Management Services (RMS) [18] for OFFICE documents and Adobe® LiveCycle™ Rights Management ES [19] for PDF documents. In many practical situations, such inheritance relations of access rights are found in subject, resource, and action categories.

2.2.2 Overview of Policy Integration with Inheritance Relations

Policy integration is a technique which unifies management of access rights management systems. Policy integration is divided into two classes: integration of policies in each system; generation of integrated policies by using inheritance relations. We describe only the framework of integrated policy generation because the policy integration in the former case is within the scope of policy conflict detection/resolution.

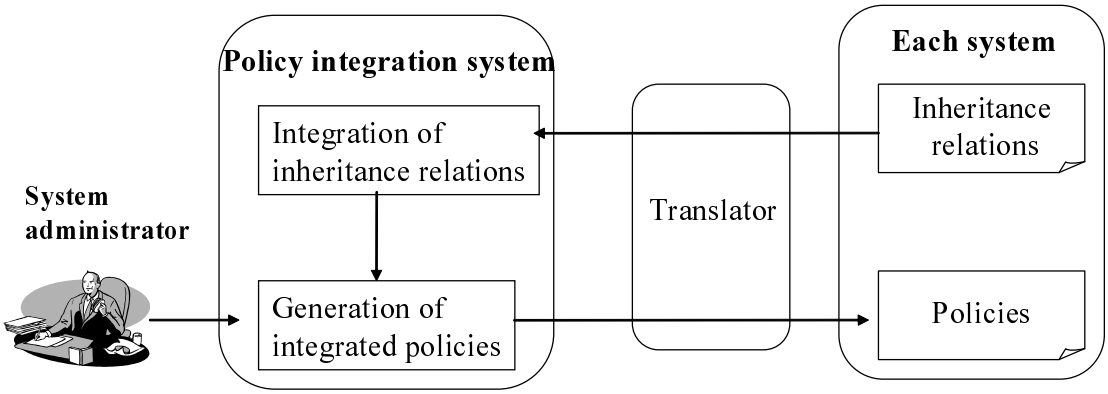


Figure 2.2: Framework of policy integration by using inheritance relations of access rights.

The framework of integrated policy generation with inheritance relations is shown in Fig. 2.2. A policy integration system first extracts inheritance relations of access rights from each system. Note that inheritance relations in different systems may be specified in different languages. Therefore inheritance relations are translated to the language used in the policy integration system through translators. These translators are also used for translating integrated policies to the language used in each system when the integrated policies are generated. A procedure of the generation of integrated policies is as follows:

1. Integrate inheritance relations in each system.

2. Establish some fundamental policies manually and generate related policies automatically by referring to integrated inheritance relations.

In the latter process, policies which relate the fundamental policies can be created easily by using inheritance relations. The important issues in this research are as follows:

1. What kind of inheritance relations is treated?

2. How inheritance relations are integrated?

2.2.3 Issues on Policy Integration with Inheritance Relations

Regarding policy integration frameworks, several methods have been proposed [9, 20, 21, 22, 23, 24, 25, 26, 27]. Among them, in [20], [21], [22], and [26], policy generation frameworks by integrating inheritance relations of access rights are considered.

For the integration of inheritance relations of different systems, Gong et al. discuss complexity, composability, and conditions for integration [21]. In [22], Dawson et al. consider

Table 2.1: Requirements for policy integration.

	P. Bonattie [20]	S. Dawson [22]	M. Sugano [26]
Automatic integration	Yes	No	No
Three categories	No	No	Yes
Easily implementable	Yes	No	No
Versatility	Yes	Yes	No

the conditions for which integrated inheritance relations do not include conflicts of inheritance relations and redundant inheritance relations. Although these two works clarify the requirements for integrating inheritance relations, they offer no specific method for generating integrated inheritance relations. In contrast with these two works, Bonatti et al. propose an algorithm for generating integrated inheritance relations [20]. In their work, they introduce a graph-theoretic approach and derive a logic programming-based algorithm for integrating inheritance relations. Their algorithm, however, seems rather complicated and is not easily implementable. Another drawback of the above three works is that they only deal with inheritance relations in a single category. Unlike the above works, Sugano et al. propose a policy integration framework called *PolicyComputing*^{TM} which treats inheritance relations in three categories [26]. Their method, however, does not create inheritance relations automatically because they integrate inheritance relations manually.

In this dissertation, we develop the framework of policy integration with inheritance relations satisfying the following four requirements:

1. Inheritance relations are integrated automatically.
2. The framework can deal with inheritance relations in three categories.
3. The integration algorithm is easily implementable.
4. The number of systems and the number of inheritance relations are arbitrary (versatility).

The four requirements discussed in the works are listed in Table 2.1. Although there exist works which satisfy requirements partially, no work satisfies all requirements. We point out that the novelty of our research is to develop a policy integration framework which satisfies the above four requirements.

2.3 Anonymous Communication System

Achieving anonymity of communications on the Internet is one of the most important issues in communication engineering. In a communication, encryption protocols protect the content of a message. They do not, however, hide the identities of a sender and a receiver of

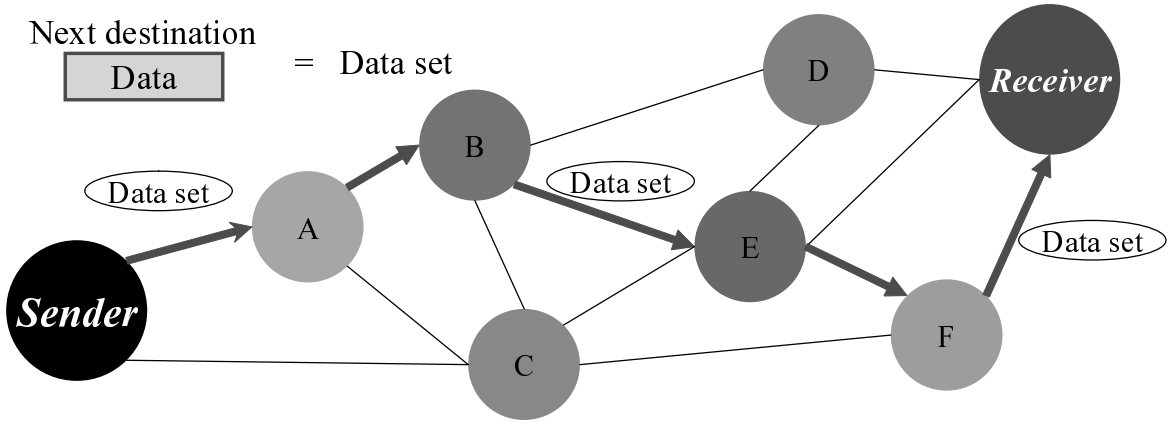


Figure 2.3: General framework of anonymous communications systems.

the message because one can easily get their IP addresses from the header of its IP packet. Once their identities are revealed, one can easily infer sender's preferences and human relationships between the sender and the receiver. Therefore, a communication method for protecting the addresses of senders and receivers is strongly required.

Anonymous communication systems not only protect the content of a message but also protect the addresses of the sender and the receiver of the message. In general, an anonymous communication system is regarded as a communication system as shown in Fig. 2.3, which forwards a data set from the sender to the receiver through several relay nodes, where we refer to the data set as the set of data composed of the address of the next destination and encrypted data. Since sender anonymity and receiver anonymity depend on the ways of forwarding and creating of a data set, several anonymous communication systems have been proposed [4, 5, 28, 29, 30, 31, 32, 33, 34, 35], and they can be applied to several services, such as electronic vote systems [28], web access systems [36, 37], e-mail systems [38], and P2P softwares [39].

Recently, a new anonymous communication system called 3-Mode Net (3MN) has been proposed [3, 40]. 3MN can be regarded as an extension of Crowds-based anonymous communication systems [4, 41]. As shown in Table 2.2, 3MN has two advantages compared to the two famous anonymous communication systems Crowds and Onion Routing: first, 3MN

Table 2.2: Characteristics among Onion Routing, Crowds, and 3-Mode Net.

	Onion Routing [5]	Crowds [4]	3-Mode Net [3]
Sender anonymity	Yes	Yes	Yes
Receiver anonymity	Yes	No	Yes
Number of encryption	Large	None	Small

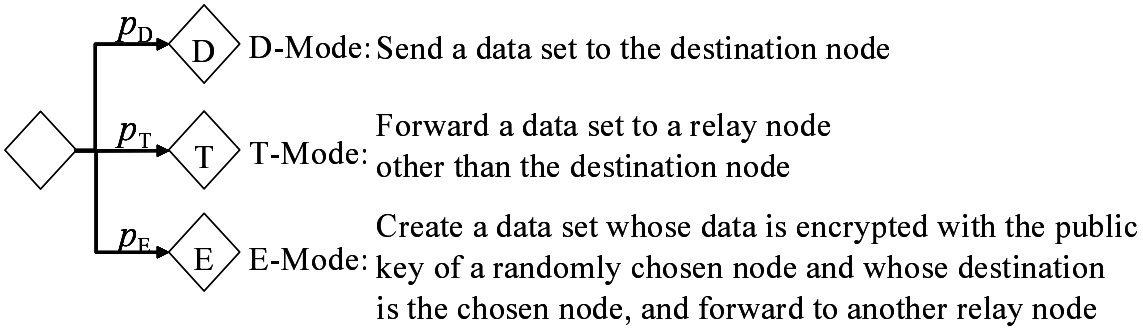


Figure 2.4: Actions of a node in 3-Mode Net.

provides anonymity to the proper receiver unlike Crowds; second, as shown in [3], 3MN has an advantage of the smaller number of encryption than Onion Routing-based anonymous communication systems [5, 33, 42, 43].

2.3.1 Overview of 3-Mode Net

We proceed to describe an overview of 3-Mode Net which we focus on in this dissertation.

a) Three modes in 3-Mode Net

3MN has three modes as shown in Fig. 2.4, i.e., Decryption Mode (D-Mode), Transmission Mode (T-Mode), and Encryption Mode (E-Mode). Each relay node chooses one of the three modes randomly with predefined probabilities.

In Fig. 2.4, the first mode is the mode where a node transmits a received data set to its destination directly. In this case, the destination node that receives the data set decrypts it with the decryption key of the destination node, and produces a new data set, which is similar to the case of Onion Routing [5]. This mode is called Decryption Mode (D-Mode).

The second mode is the mode where a node forwards a received data set to a node other than the destination node. This mode is called Transmission Mode (T-Mode).

The third mode consists of the following two processes: 1) create a new data set whose destination is a newly-chosen node except for the destination of a received data set and whose data is created by encryption of the received data set with the public key of the newly-chosen node; 2) forward the new data set to another node except for the destination of the new data set. This mode is called Encryption Mode (E-Mode).

Because of the existence of E-Mode, the destination of a data set does not always indicate the proper receiver of a message, and thus, 3MN guarantees receiver anonymity. This makes sharp contrast with the case of Crowds [4, 41]. In addition, sender anonymity is provided because each node cannot understand whether the immediate predecessor of the node is the

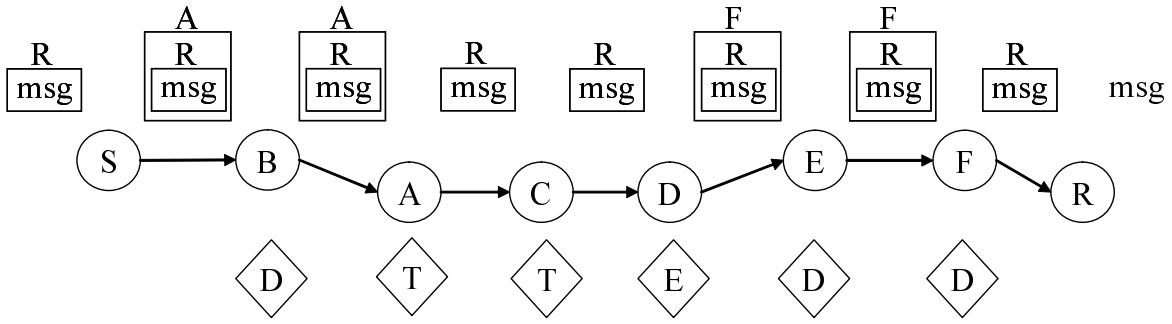


Figure 2.5: An example of the behavior of 3-Mode Net.

proper sender of a message or one of the relay nodes. This is similar to the cases of Onion Routing and Crowds.

Each relay node chooses one of the three modes randomly with predefined probabilities. Let p_{D}, p_{T}, and p_{E} denote the probabilities to choose D-Mode, T-Mode, and E-Mode, respectively, where $p_{\mathrm{D}} + p_{\mathrm{T}} + p_{\mathrm{E}} = 1$.

b) Behavior of 3-Mode Net

We describe the behavior of 3MN with Fig. 2.5, where each square frame indicates a data set composed of a multiple-encrypted message and the addresses of the next destinations. The letters on square frames indicate the next destinations.

Sender S first prepares a data set $\mathrm{R}||K_{\mathrm{R}}(\mathrm{msg})$ which consists of the address of a proper receiver R and an encrypted message $K_{\mathrm{R}}(\mathrm{msg})$ with R's public key K_{R} (|| represents the combination of data). Next, S creates a data set $\mathrm{A}||K_{\mathrm{A}}(\mathrm{R}||K_{\mathrm{R}}(\mathrm{msg}))$ (A and K_{A} represent the destination of a node chosen randomly and A's public key, respectively). After that, S forwards the data set $\mathrm{A}||K_{\mathrm{A}}(\mathrm{R}||K_{\mathrm{R}}(\mathrm{msg}))$ to another node B. In this case, the initial number of the multiplicity of encryption is equal to two. We denote the initial multiplicity of encryption by k.

When a relay node has received a data set, the node first checks its destination. If the destination corresponds to the node, the node decrypts the multiple-encrypted data set, produces a new data set, and chooses one mode randomly with predefined probabilities. Otherwise, the node only chooses one mode randomly with predefined probabilities. In this example, B chooses D-Mode. Thus B forwards the received data set to node A.

When node A receives the data set, A can obtain a new data set $\mathrm{R}||K_{\mathrm{R}}(\mathrm{msg})$ by decrypting it. After that, A chooses T-Mode and forwards $\mathrm{R}||K_{\mathrm{R}}(\mathrm{msg})$ to another node C.

In a similar fashion, node C and the following nodes forward a data set with encryption and decryption by choosing one of three modes. Finally, the proper receiver R receives a data set $\mathrm{R}||K_{\mathrm{R}}(\mathrm{msg})$. Then, R acquires the message msg by decrypting the received data set,

Table 2.3: Relationships among Onion Routing, Crowds, and 3-Mode Net.

	Initial multiplicity	Probability of mode selections		
		D-Mode	E-Mode	T-Mode
3-Mode Net	k	p_{D}	p_{E}	p_{T}
Onion Routing	k	1	0	0
Crowds	1	$1 - p_f$	0	p_f

and the transmission of the message finishes.

Note that 3MN provides a unified framework which can deal with Onion Routing and Crowds as special cases by selecting the probabilities of three modes and the initial multiplicity of encryption appropriately. The relationships among Onion Routing, Crowds, and 3MN are shown in Table 2.3, where p_f represents the probability of forwarding a received encrypted message to another node in Crowds [4, 41].

2.3.2 Issues on 3-Mode Net

As stated in [3], 3MN is an anonymous communication system which provides sender anonymity and receiver anonymity with a small number of encryption; however, there exist several issues on 3MN, which we discuss in this subsection.

3MN has not been analyzed from the viewpoint of performance. Although the expectations of the number of relay nodes and the number of encryption are derived in [3], the other important factors such as their probability distributions and their variances have not been derived. As a result, we do not understand the impacts of the probabilities of mode selections and the initial multiplicity of encryption on the behavior of 3MN. In this dissertation, we analyze the number of relay nodes and the number of encryption.

Sender anonymity and receiver anonymity in 3MN are discussed qualitatively in [3]. Unfortunately, it is not shown that to what extent anonymity is guaranteed for senders and receivers when the probabilities of mode selections and the initial multiplicity of encryption are given. Attackers may reveal senders and receivers with high probability if the probabilities of mode selections are chosen inappropriately. Intuitively, when both the number of relay nodes and the number of encryption are quite small, anonymity would be low. In order to evaluate the performance of 3-Mode Net, it is a crucial issue to clarify the effects of the probabilities of mode selections on sender anonymity and receiver anonymity. In this dissertation, we evaluate anonymity in 3MN, in particular, sender anonymity in 3MN.

2.4 Conclusion

In this chapter, we have presented policy integration and anonymous communication as technologies for protecting personal information. In policy integration, we have defined a policy and inheritance relations of access rights, and have described a framework for generating integrated policies by using inheritance relations. Furthermore, we have clarified the characteristics of our approach by comparing with related work in the research area of policy integration. Our detailed framework of policy integration with inheritance relations will be described in Chapter 3.

We have also discussed anonymous communication, focusing on a system named 3-Mode Net. We have described the framework, the characteristics of 3MN, and the performance analysis of 3MN in terms of the number of relay nodes, the number of encryption, and sender anonymity.

Chapter 3

Policy Integration Algorithm with Inheritance Relations of Access Rights

3.1 Introduction

In this chapter, we show a policy integration framework for unifying management of various access rights management systems. A procedure for generating integrated policies is as follows; first, integrate inheritance relations of access rights in each system; second, establish fundamental policies by administrators and generate related policies automatically by referring to the integrated inheritance relations.

As discussed in Chapter 2, the method for generating integrated inheritance relations is crucial. The main purpose of this chapter is to provide an algorithm for integrating inheritance relations of access rights, which satisfies the four requirements presented in Chapter 2: Inheritance relations are integrated automatically; the framework can deal with inheritance relations in three categories; the integration algorithm is easily implementable; the number of systems and the number of inheritance relations are arbitrary.

A procedure for generating integrated policies of access rights management systems is as follows: first, integrate inheritance relations in a single category; second, integrate inheritance relations of all categories; third, establish fundamental policies by administrators, and generate related policies automatically by referring to the integrated inheritance relations. The first and the second steps generate integrated inheritance relations, and we show that these operations can be accomplished by matrix operations. The derivation of the algorithm is based on a graph-theoretic approach, where a useful property of adjacency matrices is exploited. Further, it is shown that the elimination of conflicts and the removal of redundant inheritance relations are carried out by matrix operations. This chapter is related to the work published in [44, 45].

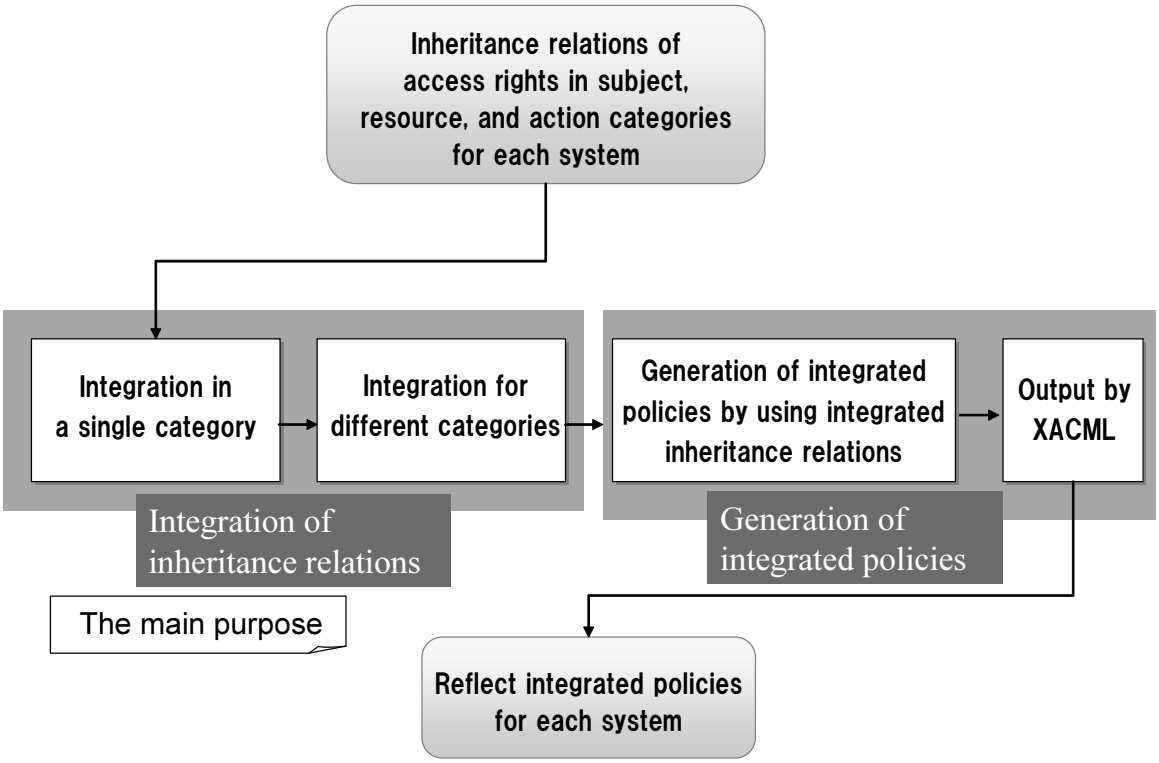


Figure 3.1: Procedure for policy integration.

3.2 Procedure for Generating Integrated Policies

This section presents our policy integration framework briefly, and clarifies the purpose of this chapter. We illustrate our procedure for policy integration, which is shown in Fig 3.1. The procedure of our policy integration is as follows:

1. Integrate inheritance relations of all systems for each category.

2. Integrate three integrated inheritance relations of each category into one integrated inheritance relation.

3. Establish some fundamental policies manually.

4. Generate related policies automatically by referring to the integrated inheritance relation.

5. Output these policies by using a policy specification language such as XACML [15].

As discussed in Chapter 2, what kind of inheritance relations is treated and how inheritance relations are integrated are very important. We focus on the first and the second processes for integrating inheritance relations of access rights.

The first step integrates inheritance relations in a single category. The method for this step is shown in Section 3.5.2, where some basic results of graph theory and some matrix operations are applied. This step allows us to obtain three integrated inheritance relations for action, subject, and resource categories.

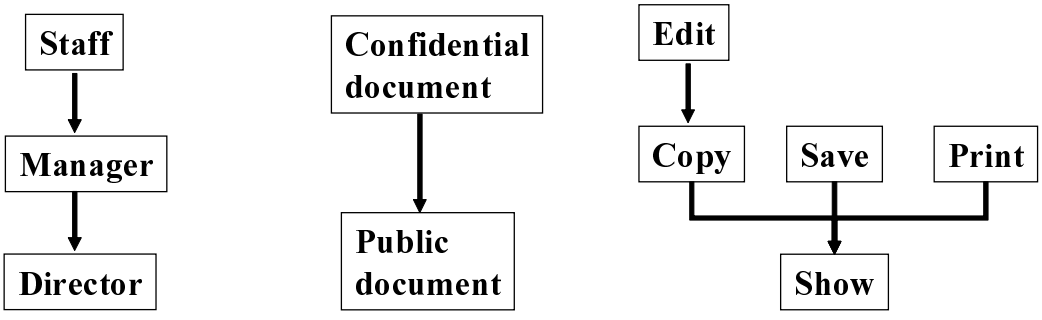
The second step integrates the above three integrated inheritance relations into one integrated inheritance relation. The method for this step is shown in Section 3.5.3, where we show that the inheritance relations in different categories can be integrated with Kronecker sum.

Note that when different names are assigned to the same action in different access rights management systems, we need to unify them. For example, in system A, “Read” is assigned to a certain action, and in system B, “Show” is assigned to the same action. In this case, we have to use the same name (“Read” or “Show”) for the same action in both systems. This unification is accomplished by translators, which is shown in Fig. 2.2 of Chapter 2.

In what follows, we discuss our algorithm for integrating inheritance relations of access rights. First, we discuss inheritance relations of access rights in detail.

3.3 Inheritance Relations of Access Rights

As discussed in Chapter 2, in many practical situations, we can find inheritance relations of access rights in each category, that is, subject category, resource category, and action category. Examples of inheritance relations in subject, resource, and action categories are shown in Fig. 3.2 (a), Fig. 3.2 (b), and Fig. 3.2 (c), respectively.



(a) Subject category. (b) Resource category. (c) Action category.

Figure 3.2: Examples of inheritance relations of access rights.

In Fig. 3.2 (a), the arrow directed from "Staff" to "Manager" represents that if staff can perform an action on a resource, managers can perform the same action on the resource. Similarly, the arrow from "Confidential document" to "Public document" in Fig. 3.2 (b) indicates that when a subject can perform an action on confidential documents, the subject can perform the same action on public documents. In the same way, the arrow from "Edit" to "Copy" in Fig. 3.2 (c) implies that if a subject can edit a resource, the subject can also copy the resource.

In general, an inheritance relation is described by the following statement: if a subject s_i can perform an action a_m on a resource r_h, a subject s_j can perform an action a_n on a resource r_l. We express this statement as $(s_i, r_h, a_m) \to (s_j, r_l, a_n)$. As shown in Fig. 3.2, we focus on inheritance relations described in a single category. Therefore, we only deal with inheritance relations expressed by the following forms:
inheritance relations in subject category:

$$(s_i, r_h, a_m) \to (s_j, r_h, a_m) \qquad (\forall r_h \in \text{res}, \forall a_m \in \text{act}),$$

inheritance relations in resource category:

$$(s_i, r_h, a_m) \to (s_i, r_l, a_m) \qquad (\forall s_i \in \text{sub}, \forall a_m \in \text{act}),$$

inheritance relations in action category:

$$(s_i, r_h, a_m) \to (s_i, r_h, a_n) \qquad (\forall s_i \in \text{sub}, \forall r_h \in \text{res}),$$

where res, act, and sub represent the sets of resources, actions, and subjects, respectively.

In the first expression, access rights for subject s_i are inherited to subject s_j regardless of resources and actions. In this case, we simply denote this inheritance relation by $s_i \to s_j$, and we say that the inheritance relation is *independent* from resource and action categories. The independence of inheritance relation in resource category and that in action category are similarly defined, and we use notations $r_h \to r_l$ and $a_m \to a_n$ for those relations. Throughout this chapter, we assume that all inheritance relations in each category are independent from other categories.

3.4 Overview of the Integration of Inheritance Relations

As shown in Fig. 3.2, inheritance relations are expressed by directed graphs. First, we introduce some notions in graph theory [46, 47] to express inheritance relations.

A directed graph $G(V, E)$ consists of a set of vertices denoted by V, and a set of edges denoted by E. An edge $e = (v, u)$ in E is an ordered pair of two different vertices v and u in V, where v and u are called *initial* and *terminal* vertices of the edge e, respectively. An edge is called a loop if its initial vertex and terminal vertex are the same. We only deal with directed graphs without any loops throughout the chapter. A path in a directed graph is defined as a finite sequence $e_1 e_2 \dots e_n$ of edges where the terminal vertex of each intermediate

edge coincides with the initial vertex of the succeeding edge, and we refer to the number of edges as the length of the path. A path $e_1 e_2 \dots e_n$ is called a circuit if the initial vertex of e_1 coincides with the terminal vertex of e_n.

3.4.1 Integration of Inheritance Relations in a Single Category

We only integrate inheritance relations in action category for all access rights management systems because the integration of inheritance relations in subject and resource categories is performed in a similar manner. As stated in Chapter 2, the integration of inheritance relations is accomplished as follows: first, integrate inheritance relations for each category; second, integrate inheritance relations for three categories. We assume that all inheritance relations in each category are independent from other categories. In this subsection, we only deal with inheritance relations such as $a_m \rightarrow a_n$, and do not consider general inheritance relations such as $(s_i, r_h, a_m) \rightarrow (s_j, r_l, a_n)$.

a) Basic Operation for Integration

We provide a method for integrating inheritance relations in action category by using graphs. The integration in subject and resource categories is carried out in a similar manner.

Let $G_h(\mathrm{ACT}_h, \mathrm{IR}_h^{\mathrm{ACT}})$ denote a graph which represents inheritance relations in action category of system h ($h \in \{1, \dots, H\}$), where ACT_h denotes a set of vertices representing actions of system h, and $\mathrm{IR}_h^{\mathrm{ACT}}$ denotes a set of edges representing inheritance relations between pairs of actions of system h. A basic operation for integrating inheritance relations is an operation taking the union of these inheritance relations:

$$\mathrm{ACT} = \mathrm{ACT}_1 \cup \mathrm{ACT}_2 \cup \cdots \cup \mathrm{ACT}_H, \tag{3.1}$$

$$\mathrm{IR}^{\mathrm{ACT}} = \mathrm{IR}_1^{\mathrm{ACT}} \cup \mathrm{IR}_2^{\mathrm{ACT}} \cup \cdots \cup \mathrm{IR}_H^{\mathrm{ACT}}, \tag{3.2}$$

where ACT denotes a set of actions of the integrated system, and $\mathrm{IR}^{\mathrm{ACT}}$ denotes a set of inheritance relations between pairs of actions of the integrated system. This allows us to obtain the graph which represents integrated inheritance relations of the integrated system as $G(\mathrm{ACT}, \mathrm{IR}^{\mathrm{ACT}})$.

Remark 3.1 *Inheritance relations generated by taking the union of inheritance relations have a major role in the prevention of the establishment of improper policies in each system. As discussed in Section 2.2 and Section 3.2, inheritance relations are used for the generation of appropriate policies. Inheritance relations generated by this basic operation also contain all elements of inheritance relations of each system, and thus, they are applicable for each system. Consequently, we can establish appropriate policies for each system by using inheritance relations which are generated by taking the union of inheritance relations in each system.*

Remark 3.2 *One may consider an operation taking the intersection of inheritance relations in each system as a basic operation. This operation, however, does not fit our purpose which is presented in Section 2.2. The reason is that inheritance relations obtained by this operation consist of common parts of inheritance relations of each system and we only generate policies about the common parts even if we use them.*

However, the above simple operation may cause problems due to the existence of circuits and redundant edges [22]. In the following, we discuss those problems, and provide a way to resolve these problems.

b) Circuits

An example of a circuit in action category is shown in Fig. 3.3 (a). In this example, the actions "Edit", "Save", and "Print" are cyclically dependent. Thus, if the "Edit" action is permitted to a subject on a resource, the "Save" and "Print" actions are also permitted to the subject on the resource.

The problem due to circuits caused by the integration of inheritance relations of all access rights management systems is that the integration yields conflicts between the integrated inheritance relations and the inheritance relations of some access rights management systems. When circuits are generated, there are two different ways to cope with this problem:

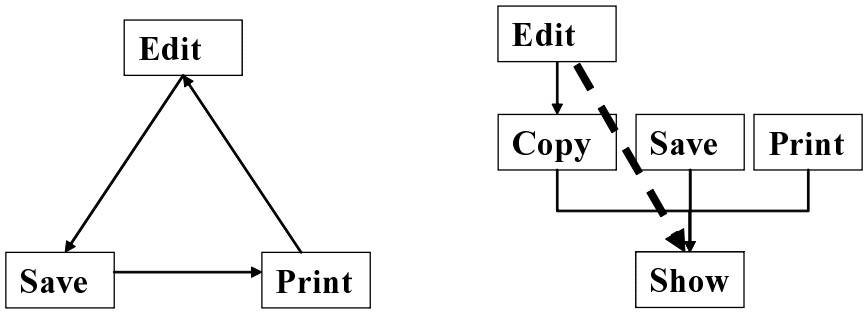
- If administrators think that the conflicts will cause serious problems, stop the integration.
- If administrators think that the problems caused by the conflicts are not serious, continue the integration.

In the latter case, we unify cyclically dependent actions as a single set of actions, and eliminate the circuit. The methods for the detection and the elimination of circuits will be presented in Section 3.5.2, where it is shown that they can be accomplished by some basic matrix operations.

Remark 3.3 *In general, to determine whether the above conflicts are serious or not is a difficult task because it depends on the situation. One of way of this determination is to refer to the security guidelines of the organization. If the integrated inheritance relations contradict their security guideline, we must stop the integration. Otherwise, we unify cyclically dependent inheritance relations.*

c) Redundancy

First, we give the definition of a redundant edge.



(a) A circuit. (b) A redundant edge.

Figure 3.3: Examples of a circuit and a redundant edge.

Definition 3.1 *Let $V = \{v_1, v_2, \ldots, v_n\}$ denote a set of vertices in a directed graph without circuits. We denote an edge directed from v_i to v_j by e_{ij}. The edge e_{ij} is redundant if there exists another path directed from v_i to v_j.*

An example of a redundant edge is shown in Fig. 3.3 (b), where a redundant edge is denoted by a dashed arrow.

Inheritance relations represented by redundant edges are not necessary to be integrated because such relationships can be obtained by tracing the corresponding paths. Therefore we detect and remove all redundant edges. The procedures for the detection and the removal of redundant edges using basic matrix operations are given in Section 3.5.2.

Summarizing this subsection, the integration of inheritance relations in a single category can be accomplished by the following steps:

1. Take a union of inheritance relations of all systems.
2. Detect and eliminate circuits by unifying cyclically dependent inheritance relations.
3. Detect and remove redundant edges.

3.4.2 Integration of Inheritance Relations of Different Categories

In this subsection, we present a basic idea for integrating inheritance relations of different categories through a simple example. A general method using adjacency matrices will be presented in Section 3.5.3.

We describe the reason why inheritance relations of difference categories are integrated. For example, there exist the inheritance relations shown in Fig. 3.2 and a policy where staff can edit confidential documents. In this case, we easily obtain two policies where directors can edit confidential documents and staff can copy confidential documents. It is, however,

difficult to judge whether a policy where managers can show public documents is appropriate or not because the inheritance relations are not related to each other. Therefore, in order to associate inheritance relations of different categories, we need to integrate inheritance relations of different categories.

We consider the integration of inheritance relations of subject, resource, and action categories given by graphs $G_{\mathrm{S}}(V_{\mathrm{S}}, E_{\mathrm{S}})$, $G_{\mathrm{R}}(V_{\mathrm{R}}, E_{\mathrm{R}})$, and $G_{\mathrm{A}}(V_{\mathrm{A}}, E_{\mathrm{A}})$, respectively, where

$$\begin{aligned} V_{\mathrm{S}} &= \{s_1, s_2\},\ V_{\mathrm{R}} = \{r_1, r_2\},\ V_{\mathrm{A}} = \{a_1, a_2\}, \\ E_{\mathrm{S}} &= \{s_1 \to s_2\},\ E_{\mathrm{R}} = \{r_1 \to r_2\},\ E_{\mathrm{A}} = \{a_1 \to a_2\}. \end{aligned}$$

First, we integrate inheritance relations of resource and action categories. By the inheritance relations $r_1 \to r_2$ and $a_1 \to a_2$, and from the assumption that inheritance relations are independent from other categories, the integrated inheritance relations are obtained as follows:

$$\begin{aligned} (s_i, r_1, a_1) &\to (s_i, r_1, a_2),\quad (s_i, r_2, a_1) \to (s_i, r_2, a_2), \\ (s_i, r_1, a_1) &\to (s_i, r_2, a_1),\quad (s_i, r_1, a_2) \to (s_i, r_2, a_2) \qquad (i = 1, 2). \end{aligned} \tag{3.3}$$

For simplicity, we rewrite the above relations as $r_1a_1 \to r_1a_2$, $r_2a_1 \to r_2a_2$, $r_1a_1 \to r_2a_1$, $r_1a_2 \to r_2a_2$, respectively. It then follows that these inheritance relations can be regarded as the edges of the graph defined on the vertex set $\{r_1a_1, r_1a_2, r_2a_1, r_2a_2\}$.

Next, we integrate the inheritance relation $s_1 \to s_2$ in subject category and the inheritance relations given by Eq. (3.3). Since Eq. (3.3) holds for $i = 1, 2$, we obtain the following eight inheritance relations:

$$\begin{aligned} &s_1r_1a_1 \to s_1r_1a_2,\ s_1r_2a_1 \to s_1r_2a_2,\ s_1r_1a_1 \to s_1r_2a_1,\ s_1r_1a_2 \to s_1r_2a_2, \\ &s_2r_1a_1 \to s_2r_1a_2,\ s_2r_2a_1 \to s_2r_2a_2,\ s_2r_1a_1 \to s_2r_2a_1,\ s_2r_1a_2 \to s_2r_2a_2, \end{aligned} \tag{3.4}$$

where $s_ir_ha_m$ is an abbreviation of (s_i, r_h, a_m). In addition, the access rights for s_1 are inherited to s_2 whatever resources and actions may be, because inheritance relations in subject category are independent from action and resource categories. As a result, we obtain four inheritance relations as

$$s_1r_1a_1 \to s_2r_1a_1,\ s_1r_1a_2 \to s_2r_1a_2,\ s_1r_2a_1 \to s_2r_2a_1,\ s_1r_2a_2 \to s_2r_2a_2. \tag{3.5}$$

It then follows that the graph representing the integrated inheritance relations of action and resource categories is composed of the set of edges given by Eqs. (3.4) and (3.5), and the set of vertices $\{s_1r_1a_1, s_1r_1a_2, s_1r_2a_1, s_1r_2a_2, s_2r_1a_1, s_2r_1a_2, s_2r_2a_1, s_2r_2a_2\}$.

As shown above, the integration of inheritance relations of different categories is too complicated to implement even for the above simple case. In Section 3.5.3, we present a general procedure for integrating inheritance relations of different categories, and show that the procedure can be implemented easily by using adjacency matrices.

3.5 Integration of Inheritance Relations Using Adjacency Matrices

We present a matrix-based algorithm for integrating inheritance relations in a single category, as well as a method for integrating inheritance relations of all categories. The expression of graphs via *adjacency matrices* plays a crucial role.

3.5.1 Adjacency Matrices and Their Properties

We first provide the definition of adjacency matrices, as well as the definitions of summation and multiplication of matrices. We also introduce a useful property of adjacency matrices.

Definition 3.2 *Let G denote a directed graph with a vertex set $V = \{v_1, v_2, \ldots, v_n\}$. The adjacency matrix A of the graph G is defined as $A = [a_{ij}]$ $(i = 1, \ldots, n;\ j = 1, \ldots n)$ where each element a_{ij} is given by*

$$a_{ij} = \begin{cases} 1 \ (\textit{if there is an edge from } v_i \textit{ to } v_j), \\ 0 \ (\textit{otherwise}). \end{cases} \tag{3.6}$$

Summation and multiplication of adjacency matrices are the same as usual matrix operations except for the elementwise summation.

Definition 3.3 *Sum of elements a and b is defined as the logical sum, that is,*

$$a + b = \begin{cases} 0 \ (a = b = 0), \\ 1 \ (\textit{otherwise}). \end{cases} \tag{3.7}$$

The following lemma concerns the condition for the existence of a path connecting two vertices, and is exploited for detecting circuits and redundant edges.

Lemma 3.1 ([46, 47]) *Let A denote an adjacency matrix of a directed graph G whose vertex set is given by $V = \{v_1, v_2, \ldots, v_n\}$. Then, there exists a path of length p directed from v_i to v_j if and only if the (i, j)th element of A^p is equal to* 1.

Remark 3.4 *As mentioned in Section 3.4, we only deal with a directed graph without loops. Thus diagonal elements of adjacency matrices of directed graphs are equal to* 0.

3.5.2 Integration in a Single Category Using Adjacency Matrices

We give a matrix-based method for integrating inheritance relations in a single category. Although we derive the method for action category, the method can be applied to other categories in a similar way.

Let H denote the number of systems, and inheritance relations in action category of system h are represented by the graph $G_h(\mathrm{ACT}_h, \mathrm{IR}_h^{\mathrm{ACT}})$ $(h = 1, \ldots, H)$. In integrating these inheritance relations, the size of the corresponding adjacency matrices of all systems must be the same because matrix operations include matrix summation. Therefore, we first rewrite graphs by replacing the vertex set ACT_h by ACT given by Eq. (3.1).

In the following, we give a matrix-based algorithm for the integration according to the procedure at the end of Section 3.4.1.

a) The First Operation : Taking a Union

The first operation is to take a union of inheritance relations. This is accomplished by adding each adjacency matrix corresponding to inheritance relations of each system. Let A_h denote an adjacency matrix representing inheritance relations in action category of system h. Then, the adjacency matrix A corresponding to the integrated inheritance relation in action category is given by

$$A = A_1 + A_2 + \cdots + A_H. \tag{3.8}$$

b) The Second Operation : Detection and Elimination of Circuits

We can detect and eliminate all circuits by repeated use of the following procedure until all circuits in the graph are eliminated.

1. Remove all vertices composing a circuit and introduce a new vertex instead.
2. Connect the edges associated with removed vertices to the new vertex.

For this procedure, the following theorems are useful.

Theorem 3.1 *Let G denote a directed graph with n vertices and A denote an adjacency matrix of G. Then, G does not include circuits if and only if $A^n = 0$.*

Theorem 3.2 *Let G denote a directed graph and A denote its adjacency matrix. Then, there exists a circuit of length l including v_i if and only if the (i, i)th element of A^l is equal to* 1.

These theorems are easily proved with Lemma 3.1, so that they are omitted.

Based on the above two theorems, a procedure is carried out by the following matrix operations:

Step. 1 Let $l = 2$.

Step. 2 Compute A^l.

Step. 3 If any diagonal elements of A^l are equal to 1, at least one circuit exists. In this case, go to the following steps. Otherwise, go to Step. 4.

Step. i Choose a vertex v_i consisting a circuit. After that, identify the circuit $v_i v_{j_1} v_{j_2} \dots v_{j_{l-1}} v_i$ which includes v_i by tracing $A, A^2, \dots, A^l$. Let c_i denote the set of vertices $\{v_{j_1}, v_{j_2}, \dots, v_{j_{l-1}}\}$.

Step. ii Replace the i-th column by the logical sum of the columns corresponding to vertices in c_i. After that, replace the i-th row by the logical sum of the rows corresponding to vertices in c_i.

Step. iii Replace the (i, i)th element of A by 0.

Step. iv Remove all columns and rows corresponding to vertices in c_i.

Step. v Redefine A as the new generated matrix and return to Step. 1.

Step. 4 If $A^l = 0$, the graph does not include any circuits, and the algorithm is finished. Otherwise, add 1 to l and return to Step. 2.

The implication of Step. 3 in this algorithm is as follows: In Step. i, a selected vertex is regarded as a new vertex. Step. ii adds every edge of each vertex in a circuit to the selected vertex in the circuit. Step. iii removes all loops because we have assumed that every graph has no loop. Finally, Step. iv removes all the vertices in the circuit except for the selected vertex.

Remark 3.5 *One may think that the graph obtained by the above algorithm depends on the choice of a vertex in Step. i, and the graph is not determined uniquely. As seen from Step. ii to Step. iv, however, the graph is uniquely determined whatever a chosen vertex of a circuit may be, since the newly generated vertex has every edge of each eliminated vertex.*

c) The Third Operation : Detection and Removal of Redundant Edges

The following theorem is used for detecting redundant edges, which is derived easily from Lemma 3.1.

Theorem 3.3 *Let G denote a directed graph without circuits and A denote an adjacency matrix of G. Then, the edge directed from v_i to v_j is redundant if and only if the (i, j)th element of A^l is equal to 1 for some $l \geq 2$.*

Based on Theorem 3.3, a procedure for detecting and removing redundant edges via matrix operations is obtained as follows:

Step. 1 Let $l = 2$.

Step. 2 Compute A^l.

Step. 3 If the (i, j)th elements of A and A^l are equal to 1, replace the (i, j)th element of A by 0, add 1 to l, and return Step. 2.

Step. 4 If $A^l = 0$, the graph does not include any redundant edges, and the algorithm is finished. Otherwise, add 1 to l and return to Step. 2.

We note that Theorem 3.1 ensures that the algorithms for detection and removal of circuits and redundant edges finish within finite steps. Both algorithms stop when l is equal to n where n represents the number of vertices in G.

Remark 3.6 *The graph obtained by the above algorithm is uniquely determined because the output of the above algorithm is independent of the choice of vertices in Step 3.*

3.5.3 Integration of Inheritance Relations for Different Categories Using Adjacency Matrices

Before presenting a method for integrating inheritance relations of three categories, we present a method for integrating inheritance relations of two different categories using adjacency matrices. Let $G(\mathrm{RES}, \mathrm{IR}^{\mathrm{RES}})$ and $G(\mathrm{ACT}, \mathrm{IR}^{\mathrm{ACT}})$ denote the graphs associated with the inheritance relations in subject category and action category, respectively, where $\mathrm{RES} = \{r_1, \ldots, r_{n_R}\}$ and $\mathrm{ACT} = \{a_1, \ldots, a_{n_A}\}$, and the corresponding adjacency matrices denote R and A, respectively.

We give a method for generating the adjacency matrix X_{RA} representing the integrated inheritance relations of resource and action categories. As seen from the observation in Section 4.2, the generated graph representing the integrated inheritance relations of resource and action categories is defined on the vertex set given by

$$\mathrm{RES} \times \mathrm{ACT} = \{r_1 a_1, \ldots, r_1 a_{n_\mathrm{A}}, \ldots, r_{n_\mathrm{R}} a_1, \ldots, r_{n_\mathrm{R}} a_{n_\mathrm{A}}\}.$$

Accordingly, the order of the adjacency matrix associated with the integrated inheritance relations of resource and action categories is given by $n_\mathrm{R} \times n_\mathrm{A}$.

Note first that the inheritance relations in resource category hold whatever action may be. Consequently, we obtain the following adjacency matrix X_{RA_1}, which corresponds to the dashed line of Fig. 3.4:

$$X_{\mathrm{RA}_1} = R \otimes I_{n_\mathrm{A}} = \begin{bmatrix} r_{11} I_{n_\mathrm{A}} & \cdots & r_{1 n_\mathrm{R}} I_{n_A} \\ \vdots & \ddots & \vdots \\ r_{n_\mathrm{R} 1} I_{n_\mathrm{A}} & \cdots & r_{n_\mathrm{R} n_\mathrm{R}} I_{n_\mathrm{A}} \end{bmatrix}, \tag{3.9}$$

where r_{ij} is the (i, j)th element of matrix R, I_{n_A} stands for the identity matrix of order n_A, and $\otimes$ denotes Kronecker product defined as follows: For a matrix $X = [x_{ij}]$ of the size p-by-q and a matrix Y, the Kronecker product of X and Y is defined as

$$X \otimes Y := \begin{bmatrix} x_{11} Y & \cdots & x_{1q} Y \\ \vdots & & \vdots \\ x_{p1} Y & \cdots & x_{pq} Y \end{bmatrix}. \tag{3.10}$$

Note also that inheritance relations in action category hold whatever resource may be. Thus we obtain the following adjacency matrix X_{RA_2}, which corresponds to the solid line of Fig. 3.4:

$$X_{\mathrm{RA}_2} = I_{n_\mathrm{R}} \otimes A = \begin{bmatrix} A & & 0 \\ & \ddots & \\ 0 & & A \end{bmatrix}. \tag{3.11}$$

With Eqs. (3.9) and (3.11), we obtain the adjacency matrix X_{RA} of the integrated inheritance relations of resource and action categories as follows:

$$X_{\mathrm{RA}} = X_{\mathrm{RA}_1} + X_{\mathrm{RA}_2} = R \otimes I_{n_\mathrm{A}} + I_{n_\mathrm{R}} \otimes A. \tag{3.12}$$

Notice that the right-hand side of Eq. (3.12) is a well-known matrix operation called Kronecker sum. The Kronecker sum of the matrices X ($n \times n$) and Y ($m \times m$) is defined as

$$X \oplus Y := X \otimes I_m + I_n \otimes Y. \tag{3.13}$$

Summarizing the above, we obtain the following theorem.

Theorem 3.4 *Let R and A denote adjacency matrices representing inheritance relations of resource category and action category, respectively. Then, the adjacency matrix representing integrated inheritance relations of these two categories is given by $R \oplus A$, i.e., Kronecker sum of matrices R and A.*

Example 3.1 *Consider the integration of inheritance relations of resource and action categories shown in Fig. 3.4. The adjacency matrix R associated with the inheritance relations of resource category defined on* {a, b, c}*, and the adjacency matrix A for the action category defined on* {E, P, S} *are given, respectively by*

$$R = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}. \tag{3.14}$$

In this case, the integrated inheritance relations are defined on the vertex set

$$\mathrm{RES} \times \mathrm{ACT} = \{\mathrm{aE}, \mathrm{aP}, \mathrm{aS}, \mathrm{bE}, \mathrm{bP}, \mathrm{bS}, \mathrm{cE}, \mathrm{cP}, \mathrm{cS}\}, \tag{3.15}$$

and the corresponding X_{RA_1}, X_{RA_2}, and X_{RA} are obtained as follows:

$$X_{\mathrm{RA}_1} = \begin{bmatrix} 0 & I_3 & I_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad X_{\mathrm{RA}_2} = \begin{bmatrix} A & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & A \end{bmatrix}, \quad X_{\mathrm{RA}} = \begin{bmatrix} A & I_3 & I_3 \\ 0 & A & 0 \\ 0 & 0 & A \end{bmatrix}. \tag{3.16}$$

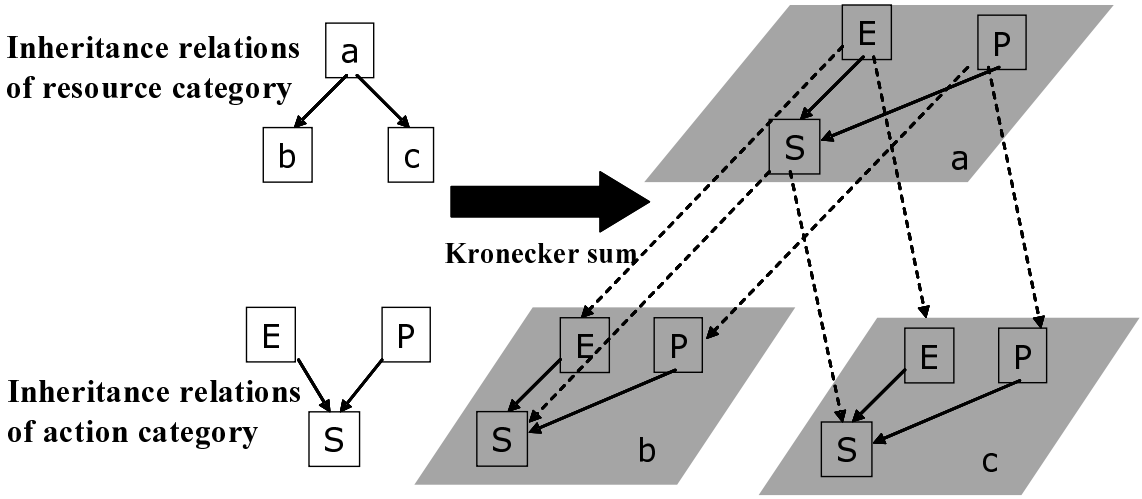


Figure 3.4: An example of Kronecker sum between resource and action categories.

Kronecker sum is also applied to integrating inheritance relations of three categories. Let S denote the adjacency matrix of the corresponding inheritance relations of subject category. Then, in a similar way to the above discussion, it is shown that the adjacency matrix X_{SRA} of the integrated inheritance relations of three categories is given by

$$X_{\mathrm{SRA}} = S \oplus X_{\mathrm{RA}} = S \oplus R \oplus A. \tag{3.17}$$

Using Kronecker sum, we can easily integrate all inheritance relations of three categories.

Remark 3.7 ([48]) *In Eq. (3.17), one may think that the right-hand side of the equation should be $S \oplus (R \oplus A)$. However, Eq. (3.17) is correct because Kronecker product is associative, and thus, Kronecker sum is also associative, that is, $(S \oplus R) \oplus A = S \oplus (R \oplus A)$.*

Remark 3.8 *The graph obtained by Kronecker sum is uniquely determined. Thus, together with Remark 3.5 and Remark 3.6, it is concluded that the graph obtained by our algorithm is uniquely determined.*

Here, we have to check whether integrated inheritance relations include circuits or redundant edges. Concerning these points, the following theorems hold:

Theorem 3.5 *Let G_X and G_Y denote graphs with adjacency matrices X and Y, respectively. If G_X and G_Y do not include circuits, the graph associated with the adjacency matrix $X \oplus Y$ does not include circuits.*

Theorem 3.6 *Let G_X and G_Y denote graphs with adjacency matrices X and Y, respectively. If G_X and G_Y do not include circuits and redundant edges, the graph associated with the adjacency matrix $X \oplus Y$ does not include redundant edges.*

The proofs of Theorem 3.5 and Theorem 3.6 are given in Appendix A.1 and Appendix A.2, respectively. Combining the above theorems, the following corollary is obtained:

Corollary 3.1 *Let G_X and G_Y denote graphs with adjacency matrices X and Y, respectively. If both G_X and G_Y include neither circuits nor redundant edges, the graph associated with the adjacency matrix $X \oplus Y$ includes neither circuits nor redundant edges.*

By this corollary, it is assured that the inheritance relations obtained by Kronecker sum have no circuits and redundant edges provided that inheritance relations of all categories include no circuits and redundant edges.

3.6 Conclusion

In this chapter, we have discussed our policy integration framework. In particular, in order to generate integrated access control policies, we have presented a matrix-based algorithm for integrating inheritance relations of access rights under the assumption that inheritance relations in a single category are independent from other categories. By applying our algorithm, we can generate integrated inheritance relations automatically, and the results can be applied to centralized access rights management systems, such as Microsoft Windows RMS and Adobe LiveCycle Rights Management ES, where administrators have to prepare integrated inheritance relations manually in the conventional framework.

One of the most significant problems is that our algorithm will stop if an administrator thinks that the conflicts caused by the integration yield serious problems. We have to improve the algorithm so as to preclude inheritance relations which cause conflicts, and to continue the integration process without such inheritance relations.

Our proposed method will be especially useful for unifying management of access rights management systems which possess various inheritance relations satisfying the independence assumption. We cannot, however, apply the algorithm when we want to set up complex access control policies which do not satisfy the independence assumption.

In such a situation, the Role-Based Access Control (RBAC) model [49, 50] will be useful. Using the RBAC model, administrators can establish policies in detail. However, the RBAC model can deal only with inheritance relations in subject category, and they have to address resources and actions individually. Therefore, it is desirable to use our method together with the RBAC model in a mutually complementary manner. In this case, we have to check whether the policies established by the RBAC model and the policies generated by our proposed method do not conflict by using policy conflict detection tools [9, 10, 11, 13].

Chapter 4

Analysis of the Number of Relay Nodes and the Number of Encryption for Anonymous Communication System 3-Mode Net

4.1 Introduction

In this and next chapters, we analyze the performance of an anonymous communication system 3-Mode Net (3MN). In Chapter 4, we analyze the number of relay nodes and the number of encryption required for communication in 3MN. In particular, we show the probability distributions, the expectations, and the variances of the above two performance measures. The formulas are derived from random walk theory, probability generating functions, and their properties. From these results, we investigate the impacts of the probabilities of mode selections and the initial multiplicity of encryption on the behavior of 3MN. Using these results, we give some conditions to avoid such a situation that the number of relay nodes and the number of encryption become extremely large. This chapter is related to the work published in [51, 52].

4.2 Modeling of 3-Mode Net by Random Walk

When D-Mode, E-Mode, or T-Mode is chosen in 3MN, the multiplicity of encryption decreases by one, increases by one, and remains unchanged, respectively. Accordingly, the behavior of 3MN can be modeled by a random walk, because the multiplicity of the encryption of a data set changes in a probabilistic manner. The analysis method by random walk has been used commonly, and many useful results are known at present [53]. By applying these useful results, we analyze the number of relay nodes and the number of encryption.

A random walk is defined as a stochastic process on a set of integers, which starts at the origin and walks one step to the positive or negative direction with predefined probabilities independent of its location. As seen from such a viewpoint, the behavior of 3MN is regarded as the following stochastic process.
Modeling of 3MN with a random walk: Let k denote the initial multiplicity of encryption. Then, 3MN is regarded as a random walk on the integers which starts at a position k and at each point, moves one step to the negative direction with probability p_{D}, moves one step to the positive direction with probability p_{E}, or stays on its position with probability p_{T}. Once the walk arrives at the origin, i.e., when the multiplicity of encryption is equal to 0, the walk finishes.

In order to derive the probability distributions of the number of relay nodes and the number of encryption in the next section, we calculate the total number of all communication paths until the walk first arrives at 0 from k. The number can be obtained by exploiting the "first passage time distribution", which is well known in random walk theory [53].

Lemma 4.1 *Let k denote the initial multiplicity of encryption, and d, e, t denote the numbers of which D-Mode, E-Mode, and T-mode are chosen, respectively (therefore, $k = d - e$). Then, the total number n of paths which first reach 0 from k is given as follows:*

$$n = \frac{k}{d+e+t}\frac{(d+e+t)!}{d!e!t!}. \tag{4.1}$$

Proof of Lemma 4.1: From the result of the first passage time distribution [53], the total number n_{DE} of paths which first reach 0 from k after D-Mode and E-Mode are chosen d and e times, respectively, is as follows:

$$n_{\mathrm{DE}} = \frac{d-e}{d+e}\frac{(d+e)!}{d!e!}.$$

Next, we consider the combination of the above case when D-Mode and E-Mode are chosen $d+e$ times and the case when T-Mode is chosen t times. Since T-Mode is not chosen at the last position, the total number of the combination is equal to ${}_{d+e+t-1}C_t$. As a result, the total number n of paths which first reach 0 from k after D-Mode, E-Mode, and T-Mode are chosen d, e, and t times, respectively, is as follows:

$$n = {}_{d+e+t-1}C_t \times n_{\mathrm{DE}} = \frac{d-e}{d+e+t}\frac{(d+e+t)!}{d!e!t!}. \qquad \blacksquare$$

4.3 Probability Distributions of the Number of Relay Nodes and the Number of Encryption

In this section, we derive the probability distributions of the number of relay nodes as well as the number of encryption. From Lemma 4.1, we can obtain the following theorems which give the probability distributions of the number of relay nodes and the number of encryption. We define K as a random variable representing the initial multiplicity of encryption.

Theorem 4.1 *Let N denote a random variable representing the number of relay nodes required for communication. Then, the probability distribution $P(N = r \mid K = k)$ is given by the following equation:*

$$P(N = r \mid K = k) = \sum_{t \in F(k,r)} \frac{k}{r} \frac{r!}{d!e!t!} {p_{\mathrm{D}}}^d {p_{\mathrm{E}}}^e {p_{\mathrm{T}}}^t, \tag{4.2}$$

where $d = (r - t + k)/2$, $e = (r - t - k)/2$, and $F(k, r)$ is a set of integers defined as

$$F(k, r) = \{t \mid 0 \leqq t \leqq r - k\ ,\ t \equiv r - k \pmod 2\}.$$

Theorem 4.2 *Let N_{e} denote a random variable representing the number of encryption required for communication, that is, the sum of the number of the selection of E-Mode and the initial multiplicity of encryption. Then, the probability distribution $P(N_{\mathrm{e}} = r \mid K = k)$ is given by the following equation:*

$$P(N_{\mathrm{e}} = r \mid K = k) = \frac{k}{2r - k} \frac{(2r - k)!}{r!(r - k)!} \frac{{p_D}^r {p_E}^{(r-k)}}{(p_{\mathrm{D}} + p_{\mathrm{E}})^{(2r-k)}}. \tag{4.3}$$

Proof of Theorem 4.1: Suppose that D-Mode, E-Mode, and T-Mode are chosen d, e, and t times, respectively. The probability of this event until a message reaches the proper receiver is the product of ${p_{\mathrm{D}}}^d {p_{\mathrm{E}}}^e p_{\mathrm{T}}^t$ and the total number of the above event. The total number of this event is given by Lemma 4.1. Therefore, in order to prove Theorem 4.1, we need to derive the constraints on d, e, and t.

Since r represents the number of relay nodes required for communication, the following equations hold: $r = d+e+t$ and $k = d-e$, or equivalently, $d = (r-t+k)/2$ and $e = (r-t-k)/2$. Since d and e are integers, the following condition on t has to be satisfied: $t \equiv r - k \pmod 2$. The maximum value of t is also equal to $r - k$ because $e \geqq 0$. Consequently we obtain the above set $F(k, r)$ on t and Eq. (4.2). ■

Proof of Theorem 4.2: Suppose that D-Mode, E-Mode, and T-Mode are chosen d, e, and t times, respectively. The probability of this event until a message reaches the proper receiver is the product of ${p_{\mathrm{D}}}^d {p_{\mathrm{E}}}^e p_{\mathrm{T}}^t$ and the total number of the above event. The total number of this event is given by Lemma 4.1. Therefore, in order to prove Theorem 4.2, we need to derive the constraints on d, e, and t.

Since r represents the number of encryption required for communication, this is the sum of the number of selection of E-Mode and the initial multiplicity of encryption. This is equal to the number of selection of D-Mode. Therefore, we obtain two equations $e = r - k$ and $d = r$. In addition, t takes all nonnegative integers because the number of selection of

T-Mode is independent of the number of encryption. As a result, we obtain

$$\begin{aligned} P(N_{\mathrm{e}} = r \mid K = k) &= \sum_{t=0}^{\infty} \frac{k}{d+e+t} \frac{(d+e+t)!}{d!e!t!} {p_{\mathrm{D}}}^{d} {p_{\mathrm{E}}}^{e} {p_{\mathrm{T}}}^{t} \\ &= \frac{k}{2r-k} \frac{(2r-k)!}{r!(r-k)!} {p_{\mathrm{D}}}^{r} {p_{\mathrm{E}}}^{r-k} \times \sum_{t=0}^{\infty} \frac{(2r-k+t-1)!}{(2r-k-1)!t!} {p_{\mathrm{T}}}^{t} \\ &= \frac{k}{2r-k} \frac{(2r-k)!}{r!(r-k)!} \frac{{p_D}^{r} {p_E}^{(r-k)}}{(p_{\mathrm{D}}+p_{\mathrm{E}})^{(2r-k)}}, \end{aligned}$$

where we use the following identity:

$$\sum_{t=0}^{\infty} \frac{(a+t)!}{a!t!} x^t = \frac{1}{(1-x)^{a+1}}. \qquad \blacksquare$$

4.4 Expectations and Variances of the Number of Relay Nodes and the Number of Encryption

We show the expectations and the variances of the number N of relay nodes and the number N_{e} of encryption in order to consider the impacts of the probabilities of mode selections and the initial multiplicity of encryption. We derive them by using probability generating functions and their properties [53].

4.4.1 Probability Generating Functions

Let X denote a nonnegative integer random variable. Then, a probability generating function $g_X(\lambda)$ is defined as

$$g_X(\lambda) = E(\lambda^X) = \sum_{r=0}^{\infty} P(X = r)\lambda^r, \tag{4.4}$$

where $E(\cdot)$ is the expectation operator. Using the probability generating function, the expectation $E(X)$ and the variance $V(X)$ of X are obtained as follows [53]:

$$E(X) = g'_X(1), \tag{4.5}$$
$$V(X) = g''_X(1) + g'_X(1) - (g'_X(1))^2. \tag{4.6}$$

Therefore, the calculation of the expectations and the variances of the number of relay nodes and the number of encryption is reduced to the derivation of those probability generating functions. As for the probability generating functions, the following lemmas are derived:

Lemma 4.2 *Let τ_k denote a random variable representing the number of relay nodes required for communication under the condition that the initial multiplicity of encryption is k. Then, the probability generating function for τ_k is given by*

$$g_{\tau_k}(\lambda) = \begin{cases} \left(\dfrac{1 - p_{\mathrm{T}}\lambda - \sqrt{(1-p_{\mathrm{T}}\lambda)^2 - 4p_{\mathrm{D}}p_{\mathrm{E}}\lambda^2}}{2p_{\mathrm{E}}\lambda} \right)^k & (p_{\mathrm{E}} \neq 0), \\ \left(\dfrac{p_{\mathrm{D}}\lambda}{1 - p_{\mathrm{T}}\lambda} \right)^k & (p_{\mathrm{E}} = 0). \end{cases} \tag{4.7}$$

Lemma 4.3 *Let ϵ_k denote a random variable representing the number of encryption required for communication under the condition that the initial multiplicity of encryption is k. Then, the probability generating function for ϵ_k is given by*

$$g_{\epsilon_k}(\lambda) = \begin{cases} \left(\dfrac{1 - p_{\mathrm{T}} - \sqrt{(1-p_{\mathrm{T}})^2 - 4p_{\mathrm{D}}p_{\mathrm{E}}\lambda}}{2p_{\mathrm{E}}} \right)^k & (p_{\mathrm{E}} \neq 0), \\ \left(\dfrac{p_{\mathrm{D}}\lambda}{1 - p_{\mathrm{T}}} \right)^k & (p_{\mathrm{E}} = 0). \end{cases} \tag{4.8}$$

The proofs of Lemma 4.2 and Lemma 4.3 are given in Appendix A.3 and Appendix A.4, respectively. In the next subsection, we derive the expectations and the variances of the number of relay nodes and the number of encryption from Lemma 4.2 and Lemma 4.3.

At the end of this subsection, we derive a condition about the probabilities of mode selections by considering the above two probability generating functions. As shown in Eq. (4.4), the value of the probability generating functions at $\lambda = 1$ means the probability that messages reach receivers because $g_X(1) = \sum_{r=0}^{\infty} P(X = r)$. From Eqs. (4.7) and (4.8), we obtain

$$g_{\tau_k}(1) = g_{\epsilon_k}(1) = \begin{cases} \left(\dfrac{p_{\mathrm{E}} + p_{\mathrm{D}} - | p_{\mathrm{E}} - p_{\mathrm{D}} |}{2p_{\mathrm{E}}} \right)^k & (p_{\mathrm{E}} \neq 0), \\ 1 & (p_{\mathrm{E}} = 0). \end{cases} \tag{4.9}$$

In order to ensure that messages certainly reach receivers, the value of Eq. (4.9) should be equal to 1. Thus, we require the following assumption [53].

Assumption 4.1 $p_{\mathrm{D}} \geq p_{\mathrm{E}}$.

4.4.2 Expectations and Variances

From Eqs. (4.5), (4.6), (4.7), and (4.8), we obtain the next theorems about the expectations and the variances of the number of relay nodes and the number of encryption. The detailed derivations for these values are given in Appendix A.5 and Appendix A.6.

Theorem 4.3 *The expectation M_{N} and the variance V_{N} of the number of relay nodes required for communication are given by*

$$M_{\mathrm{N}} = \frac{k}{p_{\mathrm{D}} - p_{\mathrm{E}}} \quad , \quad V_{\mathrm{N}} = \frac{k\{(1 - p_{\mathrm{T}}) - (p_{\mathrm{D}} - p_{\mathrm{E}})^2\}}{(p_{\mathrm{D}} - p_{\mathrm{E}})^3},$$

respectively.

Theorem 4.4 *The expectation M_{E} and the variance V_{E} of the number of encryption required for communication are given by*

$$M_{\mathrm{E}} = \frac{k}{1 - \frac{p_{\mathrm{E}}}{p_{\mathrm{D}}}} \quad , \quad V_{\mathrm{E}} = \frac{k\frac{p_{\mathrm{E}}}{p_{\mathrm{D}}}\left(1 + \frac{p_{\mathrm{E}}}{p_{\mathrm{D}}}\right)}{\left(1 - \frac{p_{\mathrm{E}}}{p_{\mathrm{D}}}\right)^3},$$

respectively.

Notice that we use Assumption 4.1 in order to derive Theorem 4.3 and Theorem 4.4. Their expectations and variances also include the difference of p_{D} and p_{E}. Thus we need a condition that p_{D} is not equal to p_{E} so that their values become finite values. In what follows, we assume the following condition.

Assumption 4.2 $p_{\mathrm{D}} > p_{\mathrm{E}}$.

Theorem 4.3 states that the expectation of the number of relay nodes depends on the initial multiplicity of encryption and the difference of p_{D} and p_{E}. We also observe that its variance can be controlled without changing its expectation by adjusting p_{T}. From these observations, we set p_{T} to be large in order to keep the expectation unchanged and to reduce the possibility that the number of relay nodes becomes extremely large.

From Theorem 4.4, we observe that both the expectation and the variance of the number of encryption depend on the initial multiplicity of encryption and the proportion of p_{D} and p_{E}. From the two equations, we obtain the following equation:

$$V_{\mathrm{E}} = \frac{M_{\mathrm{E}}(M_{\mathrm{E}} - k)(2M_{\mathrm{E}} - k)}{k^2}.$$

This implies that the expectation and the variance of the number of encryption are not independent. Thus it is shown that we cannot reduce the possibility that the number of encryption becomes large under the condition that the expectation is constant. This equation also indicates that V_{E} increases monotonically with M_{E} since $M_{\mathrm{E}} \geqq k$.

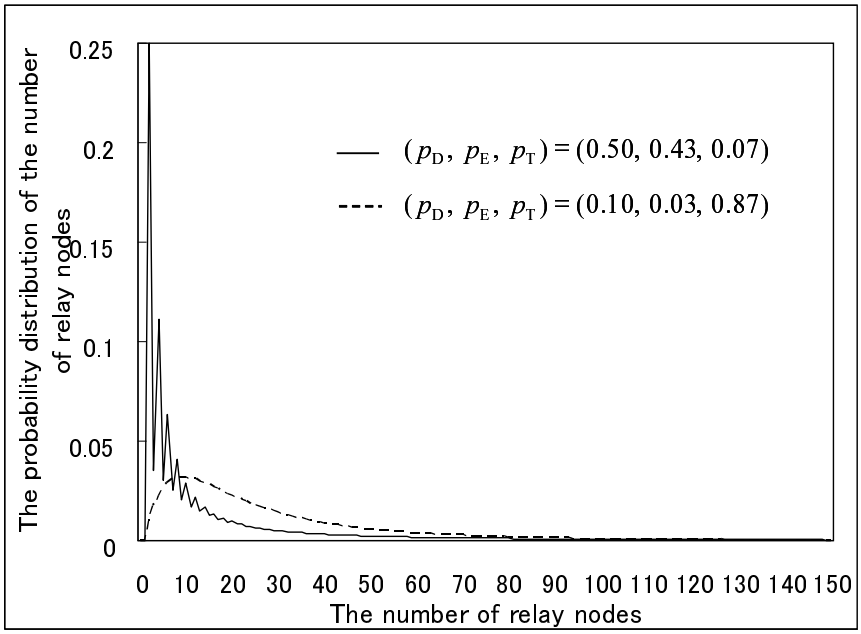


Figure 4.1: Probability distributions of the number of relay nodes.

4.5 Numerical Examples

In this section, we consider the impacts of the probabilities of mode selections through numerical examples. We consider the following two cases under the condition that $k = 2$: Case A: $(p_D, p_E, p_T) = (0.50, 0.43, 0.07)$ and Case B: $(p_D, p_E, p_T) = (0.10, 0.03, 0.87)$.

Note that the differences between p_D and p_E in Case A and Case B are the same. The expectations of relay nodes are the same, whereas p_T and proportions of p_D and p_E are quite different.

First, we compare the probability distributions of the number of relay nodes. We calculate cumulative probabilities in order to compare those distributions in the case where the number of relay nodes is large. The results of the probability distributions and the cumulative probabilities of the number of relay nodes are shown in Fig. 4.1 and Table 4.1, respectively. From Fig. 4.1, we observe that the probability distributions in Case A and Case B are quite different, although their expectations are identical. Also, from Table 4.1, the probabilities that the numbers of relay nodes are more than 50, 100, and 150 in Case A are about 13%, 6%, and 4%, respectively. In contrast, in Case B, those probabilities are about 14%, 3%, and less than 1%, respectively. These results indicate that the spread of the distribution of the number of relay nodes in Case B is smaller than that of Case A.

Furthermore, Table 4.2 compares those expectations and variances of the number of relay nodes in these cases. The variance in Case B is quite smaller than that of Case A, whereas

Table 4.1: Cumulative probabilities of the number of relay nodes.

The number of relay nodes	(p_D, p_E, p_T)	
	(0.50, 0.43, 0.07)	(0.10, 0.03, 0.87)
20	0.7414	0.5037
50	0.8723	0.8551
100	0.9354	0.9735
150	0.9603	0.9940

Table 4.2: Expectations and variances of the number of relay nodes.

(p_D, p_E, p_T)	M_N	V_N
(0.50, 0.43, 0.07)	28.57	5394
(0.10, 0.03, 0.87)	28.57	729.4

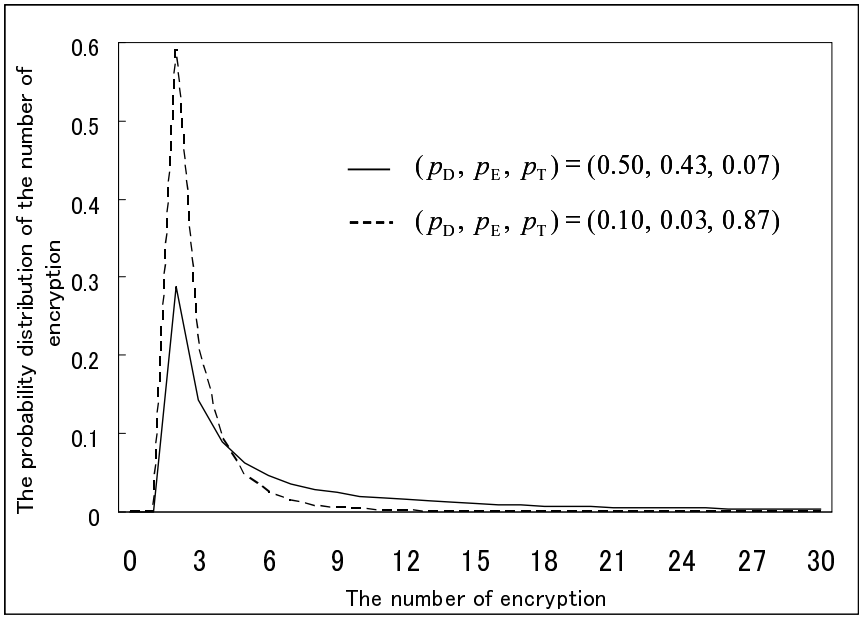


Figure 4.2: Probability distributions of the number of encryption.

the expectations in Case A and Case B are identical. This means that, as indicated by Theorem 4.3, we can reduce the variance of the number of relay nodes without changing the expectation by setting p_T to be larger.

Next, we consider the probability distributions of the number of encryption. The results of the probability distributions and the cumulative probabilities of the number of encryption

Table 4.3: Cumulative probabilities of the number of encryption.

The number of encryption	(p_D, p_E, p_T)	
	(0.50, 0.43, 0.07)	(0.10, 0.03, 0.87)
10	0.7408	0.9953
20	0.8502	0.9999
30	0.8964	0.9999
50	0.9395	0.9999

Table 4.4: Expectations and variances of the number of encryption.

(p_D, p_E, p_T)	M_E	V_E
(0.50, 0.43, 0.07)	14.29	1166
(0.10, 0.03, 0.87)	2.857	2.274

are shown in Fig. 4.2 and Table 4.3, respectively. Fig. 4.2 shows that the distribution in Case A spreads widely than the distribution in Case B. From Table 4.3, the probabilities that the numbers of encryption are more than 20, 30, and 50 in Case A are about 15%, 10%, and 6%, respectively. In contrast, in Case B, the probability that the number of encryption is more than 10 is about 1%, and the probability that it exceeds 20 is small. In addition, Table 4.4 compares the expectations and the variances of the number of encryption in these cases. From Table 4.4, the expectation and the variance in Case B are quite smaller than those of Case A. This means that, as indicated by Theorem 4.4, it is shown that the expectation and the variance of the number of encryption are quite different by the proportion of p_D and p_E.

4.6 Conclusion

In this chapter, we have derived the probability distributions, the expectations, and the variances of the number of relay nodes and the number of encryption. These formulas have been derived based on random walk theory, probability generating functions, and their properties. Using these formulas, it is possible to analyze the effect of the probabilities of mode selections on the performance of 3MN. Theorem 4.3 indicates that the expectation of the number of relay nodes depends on the difference of the probabilities to choose D-Mode and E-Mode, and the expectation becomes large when its difference is small. Further, it is shown that the probability to choose T-Mode is set to be large in order to avoid the situation where the number of relay nodes is extremely large. Theorem 4.4 implies that the expectation and the variance of the number of encryption depend on the proportion of the probabilities to choose D-Mode and E-Mode, and the expectation and the variance become large when its proportion is large.

Chapter 5

Analysis of Anonymity for Anonymous Communication System 3-Mode Net Against Collaborating Nodes

5.1 Introduction

In this chapter, we evaluate sender anonymity against collaborating nodes who collude with each other in order to identify a message sender. We refer to a node who forwards a message to the next node as an immediate predecessor, and consider the probability that the first immediate predecessor among all the collaborating nodes on the communication path coincides with the message sender. In what follows, we call this probability the probability of the message sender.

The reason why we use the probability of the message sender is that the probability of the message sender has been introduced in the literature as a standard measure for sender anonymity in anonymous communication systems [30, 54, 55, 56]. This probability was first employed in [4] for the analysis of sender anonymity in Crowds. The evaluation method is very simple because it only uses the probabilities of mode selections, the number of collaborating nodes, and the number of 3MN members, and it does not consider other attacks such as eavesdropping and timing attacks [57]. Thus we employ this measure for analyzing sender anonymity in 3MN.

The probability of the message sender is derived from a probability generating function. Using the result, we consider the influences of the probabilities of mode selections and the initial multiplicity of encryption on sender anonymity against collaborating nodes. From the results described in Chapter 4, we also describe useful results for a better understanding of the influences of the probabilities of mode selections on the performance of 3MN by showing the relationship between the number of relay nodes and sender anonymity as well as their properties. This chapter is related to the work published in [58].

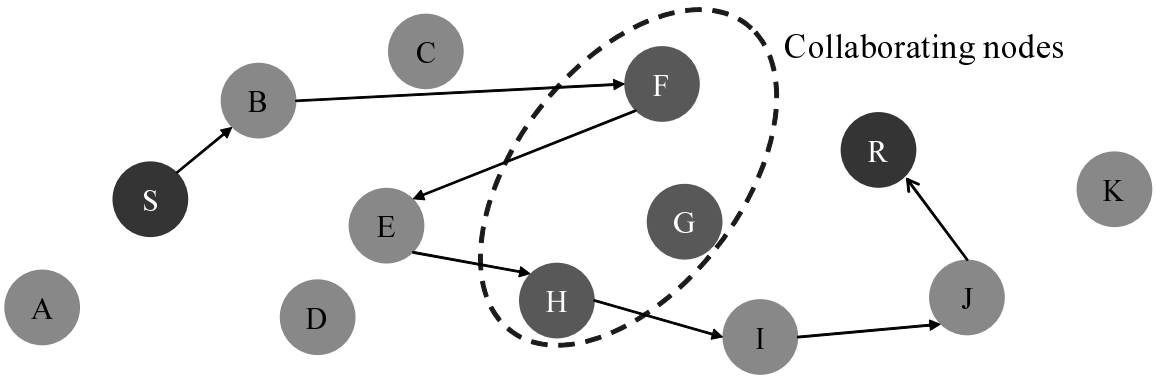


Figure 5.1: An example of collaborating nodes.

5.2 Analysis of Anonymity of 3-Mode Net

In this section, we evaluate sender anonymity against collaborating nodes. To simplify the discussion, we assume that the number of all nodes in 3MN network is constant. In addition, suppose that collaborating nodes do not carry out any other attacks, e.g., eavesdropping, timing attacks [57], passive attacks [59], and so on. We also assume that the sender and the receiver are not the members of collaborating nodes.

5.2.1 Collaborating Nodes

Collaborating nodes are members in 3MN network that collude with each other in order to illegally acquire the identity of a message sender. An example of collaborating nodes on a communication path is shown in Fig. 5.1, where node F, node G, and node H are collaborating nodes and the other nodes are non-collaborating nodes. In this example, sender S creates a communication path to receiver R, which includes collaborating nodes F and H. The collaborating nodes, including collaborating node G who does not appear on the communication path, want to know which of the non-collaborating node is a message sender.

Collaborating nodes know from whom they receive a data set and to whom they send its data set besides the common information of all members: the probabilities of mode selections, the initial multiplicity of encryption, the number of 3MN members, and the number of collaborating nodes in 3MN. They gain no other useful information because they perform no other attacks such as traffic analysis. Therefore, they only investigate the traffic flow of a data set on 3MN network.

No receivers of messages are collaborating nodes. The reason is that, in 3MN, receivers obtain special information relevant to senders in order to reply to received messages. This case when receivers are collaborating nodes is beyond the scope of this dissertation.

5.2.2 Sender Anonymity Against Collaborating Nodes

In order to measure sender anonymity, we derive the probability of the message sender that means that the first immediate predecessor among all the collaborating nodes on the communication path is indeed a message sender, where we refer to a node who forwards a message to the next node as an immediate predecessor. In the example of Fig. 5.1, the probability of the message sender indicates the probability that node B is a message sender because the first immediate predecessor among collaborating nodes F, G, and H is node B. This measure is very simple because it only uses the probabilities of mode selections, the number of collaborating nodes, and the number of 3MN members. This is also a standard measure for evaluating sender anonymity in anonymous communication systems [4, 30, 54, 55, 56]. Our approach is the same as that in [4] for Crowds. Thus we derive the probability of the message sender in a way similar to [4].

Let L_i ($i \geqq 1$) denote the event where the first collaborating node on the communication path appears at the i-th node on the path, and define $L_{i+} = L_i \vee L_{i+1} \vee L_{i+2} \vee \cdots$. Also, let J denote the event where the first immediate predecessor among the immediate predecessors on the communication path is a message sender.

Note here that the 0-th node indicates a message sender and the sender might appear on the communication path several times. Further, the events J and L_1 are different. The reason is that J includes a situation that the first immediate predecessor on the communication path is the message sender by chance because the message sender appears several times.

We consider the conditional probability $P(J \mid L_{1+})$ that the first immediate predecessor among all the collaborating nodes is the message sender, under the condition that one of the collaborating nodes receives a data set. Unlike the simple situation such as Crowds, it is rather hard to derive the probability because we have to compute infinite series concerning L_{1+} which is very complicated for 3MN case. In order to avoid the computation of the infinite series, we use a probability generating function and its properties. Using the function, we obtain the following theorem that concerns the probability of the message sender.

Theorem 5.1 *Let n_t and n_c denote the number of 3MN members and that of collaborating nodes in 3MN, respectively. Then, the conditional probability $P(J|L_{1+})$ is given by*

$$P(J \mid L_{1+}) = \frac{(n_t - n_c)(n_c + 1) - n_t \times g_{\tau_k}\left(\frac{n_t - n_c}{n_t}\right)}{n_t(n_t - n_c)\left\{1 - g_{\tau_k}\left(\frac{n_t - n_c}{n_t}\right)\right\}}, \tag{5.1}$$

where $g_{\tau_k}(\lambda)$ is a probability generating function for a random variable τ_k representing the number of relay nodes, defined by Eq. (4.7) in Chapter 4.

Proof of Theorem 5.1: The conditional probability $P(J \mid L_{1+})$ is obtained by the following equation:

$$P(J \mid L_{1+}) = \frac{P(J \wedge L_{1+})}{P(L_{1+})} = \frac{P(J)}{P(L_{1+})} = \frac{P(J \mid L_1)P(L_1) + P(J \mid L_{2+})P(L_{2+})}{P(L_{1+})}. \tag{5.2}$$

In the second equality, we use $P(J \wedge L_{1+}) = P(J)$ because J implies L_{1+}. Note also that $P(L_1) = n_c/n_t$, $P(J \mid L_1) = 1$, and $P(J \mid L_{2+}) = 1/(n_t - n_c)$. The third equality indicates that if the first collaborating node receives a data set through several nodes, the first immediate predecessor of the collaborating node is randomly chosen among non-collaborating nodes equally likely [4].

In order to calculate Eq. (5.2), we have to compute $P(L_{1+})$. This value is calculated as follows:

$$P(L_{1+}) \;=\; 1 - g_{\tau_k}\left(\frac{n_t - n_c}{n_t}\right). \tag{5.3}$$

The derivation of this equation is given in Appendix A.7.

From Eqs. (5.2) and (5.3), $P(J \mid L_{1+})$ is computed as follows:

$$P(J \mid L_{1+}) \;=\; \frac{(n_t - n_c)(n_c + 1) - n_t \times g_{\tau_k}\left(\frac{n_t - n_c}{n_t}\right)}{n_t(n_t - n_c)\left\{1 - g_{\tau_k}\left(\frac{n_t - n_c}{n_t}\right)\right\}}. \tag{5.4}$$

■

As seen from Eq. (5.1), $P(J \mid L_{1+})$ depends on the probability generating function $g_{\tau_k}(\lambda)$. In order to evaluate $P(J \mid L_{1+})$, we need to analyze $g_{\tau_k}(\lambda)$. In the following subsections, we derive properties of $g_{\tau_k}(\lambda)$, and consider the influences of the probabilities of mode selections and the initial multiplicity of encryption on the sender anonymity in 3MN.

5.2.3 Properties of a Probability Generating Function

When we need to indicate p_{D} and p_{E} explicitly for expressing $g_{\tau_k}(\lambda)$, we use the notation $g_{\{\tau_k, p_{\mathrm{D}}, p_{\mathrm{E}}\}}(\lambda)$. Concerning the probability generating function $g_{\tau_k}(\lambda)$, the following lemma is obtained.

Lemma 5.1 *Under Assumption 4.2, the probability generating function $g_{\tau_k}(\lambda)$ has the following properties:*

(i) *$g_{\tau_k}(\lambda)$ is a monotonically increasing function of λ on $0 \le \lambda \le 1$ where $g_{\tau_k}(0) = 0$ and $g_{\tau_k}(1) = 1$.*

(ii) *$0 < g_{\tau_k}(\lambda) < 1 \quad (\lambda \in (0,1))$.*

(iii) *There exists $\lambda \in (0,1)$ such that $g_{\{\tau_k, p_{\mathrm{D}_1}, p_{\mathrm{E}_1}\}}(\lambda) = g_{\{\tau_k, p_{\mathrm{D}_2}, p_{\mathrm{E}_2}\}}(\lambda) = \alpha$ if and only if α is a solution of $(p_{\mathrm{E}_1} - p_{\mathrm{E}_2})x - (p_{\mathrm{D}_1} - p_{\mathrm{D}_2}) = 0$ satisfying $\alpha \in (0,1)$.*

(iv) *One of the following inequalities holds for every $\lambda \in (0,1)$ if and only if there is no $\alpha \in (0,1)$ satisfying $(p_{\mathrm{E}_1} - p_{\mathrm{E}_2})\alpha - (p_{\mathrm{D}_1} - p_{\mathrm{D}_2}) = 0$.*

$$\begin{aligned} g_{\{\tau_k, p_{\mathrm{D}_1}, p_{\mathrm{E}_1}\}}(\lambda) &> g_{\{\tau_k, p_{\mathrm{D}_2}, p_{\mathrm{E}_2}\}}(\lambda), \\ g_{\{\tau_k, p_{\mathrm{D}_1}, p_{\mathrm{E}_1}\}}(\lambda) &< g_{\{\tau_k, p_{\mathrm{D}_2}, p_{\mathrm{E}_2}\}}(\lambda). \end{aligned}$$

(v) *When p_{E} is constant, $g_{\tau_k}(\lambda)$ increases with p_{D} for every $\lambda \in (0,1)$.*

(vi) *When p_{D} is constant, $g_{\tau_k}(\lambda)$ decreases as p_{E} increases for every $\lambda \in (0,1)$.*

(vii) *When $p_{\mathrm{D}} - p_{\mathrm{E}}$ is constant, $g_{\tau_k}(\lambda)$ increases with p_{D} (and therefore p_{E}) for every $\lambda \in (0,1)$.*

(viii) *$g_{\tau_k}(\lambda)$ decreases as k increases for every $\lambda \in (0,1)$.*

Proof of Lemma 5.1: In order to prove (i) - (vii) of Lemma 5.1, it suffices to show for the case of $k = 1$ because $g_{\tau_k}(\lambda) = (g_{\tau_1}(\lambda))^k$. Further (viii) is a direct consequence of (ii) for $k = 1$. Consequently we prove statements (i) - (vii) for $k = 1$.
(i) It is straightforward from the definition (4.4) in Chapter 4, because $P(\tau_1 = r) \geq 0$ ($r = 0, 1, \cdots$), $P(\tau_1 = 0) = 0$, and $\sum_{r=0}^{\infty} P(\tau_1 = r) = 1$.
(ii) It is a direct consequence of (i).
(iii) First, we show "only if" part. Note that, as seen from Eq. (A.14), $g_{\tau_1}(\lambda)$ is a solution of the following equation:

$$p_{\mathrm{E}} x^2 + (1 - p_{\mathrm{D}} - p_{\mathrm{E}} - 1/\lambda)x + p_{\mathrm{D}} \quad = \quad 0. \tag{5.5}$$

From the condition of the left-hand side of (iii), we obtain the following equation:

$$p_{\mathrm{E}_1}\alpha^2 + (1 - p_{\mathrm{D}_1} - p_{\mathrm{E}_1} - 1/\lambda)\alpha + p_{\mathrm{D}_1} = p_{\mathrm{E}_2}\alpha^2 + (1 - p_{\mathrm{D}_2} - p_{\mathrm{E}_2} - 1/\lambda)\alpha + p_{\mathrm{D}_2}. \tag{5.6}$$

From this equation, we obtain

$$\{(p_{\mathrm{E}_1} - p_{\mathrm{E}_2})\alpha - (p_{\mathrm{D}_1} - p_{\mathrm{D}_2})\}(\alpha - 1) \quad = \quad 0. \tag{5.7}$$

Since $\alpha \neq 1$ for $\lambda \in (0,1)$, we obtain the condition of the right-hand side of (iii).

Next, we prove "if" part. We show that α is the smaller solution of the following two equations for some $\lambda \in (0,1)$ when there exists $\alpha \in (0,1)$ satisfying $(p_{\mathrm{E}_1} - p_{\mathrm{E}_2})\alpha - (p_{\mathrm{D}_1} - p_{\mathrm{D}_2}) = 0$.

$$p_{\mathrm{E}_1} x^2 + (1 - p_{\mathrm{D}_1} - p_{\mathrm{E}_1} - 1/\lambda)x + p_{\mathrm{D}_1} \quad = \quad 0, \tag{5.8}$$

$$p_{\mathrm{E}_2} x^2 + (1 - p_{\mathrm{D}_2} - p_{\mathrm{E}_2} - 1/\lambda)x + p_{\mathrm{D}_2} \quad = \quad 0. \tag{5.9}$$

To this end, we show that there exists $\lambda \in (0,1)$ such that the value of Eq. (5.6) is equal to 0, and that the smaller solution of Eq. (5.5) is less than 1 when $\lambda \in (0,1)$. From $(p_{\mathrm{E}_1} - p_{\mathrm{E}_2})\alpha - (p_{\mathrm{D}_1} - p_{\mathrm{D}_2}) = 0$, it follows that $\{(p_{\mathrm{E}_1} - p_{\mathrm{E}_2})\alpha - (p_{\mathrm{D}_1} - p_{\mathrm{D}_2})\}(\alpha - 1) = 0$, and thus, we obtain

$$p_{\mathrm{E}_1}\alpha^2 + (1 - p_{\mathrm{D}_1} - p_{\mathrm{E}_1})\alpha + p_{\mathrm{D}_1} \quad = \quad p_{\mathrm{E}_2}\alpha^2 + (1 - p_{\mathrm{D}_2} - p_{\mathrm{E}_2})\alpha + p_{\mathrm{D}_2}. \tag{5.10}$$

Here let $\lambda = \alpha/l$ where l represents the value of Eq. (5.10). In this case, from Eq. (5.10), it is easy to show that the value of Eq. (5.6) is equal to 0. Next, we show $\lambda = \alpha/l \in (0,1)$, that is, $l - \alpha > 0$. Since $0 < \alpha < 1$ and $p_{\mathrm{D}_1} > p_{\mathrm{E}_1}$, we obtain the following inequality:

$$l - \alpha \quad = \quad p_{\mathrm{E}_1}\alpha^2 - (p_{\mathrm{D}_1} + p_{\mathrm{E}_1})\alpha + p_{\mathrm{D}_1} \quad = \quad (p_{\mathrm{D}_1} - p_{\mathrm{E}_1}\alpha)(1 - \alpha) \quad > \quad 0. \tag{5.11}$$

Finally, we show that the smaller solution of Eq. (5.5) is less than 1 if $\lambda \in (0,1)$. When $p_{\mathrm{E}} = 0$, $x = p_{\mathrm{D}}/\{p_{\mathrm{D}} + (1-1/\lambda)\}$ is the only solution of Eq. (5.5) and it is easy to see that the solution is less than 1. When $p_{\mathrm{E}} = 0$, it suffices to show that

$$-\frac{1-p_{\mathrm{D}}-p_{\mathrm{E}}-1/\lambda}{2p_{\mathrm{E}}} > 1, \tag{5.12}$$

when $\lambda \in (0,1)$. Since $0 < \lambda < 1$ and $p_{\mathrm{D}} > p_{\mathrm{E}}$, we obtain the following equation.

$$-\frac{1-p_{\mathrm{D}}-p_{\mathrm{E}}-1/\lambda}{2p_{\mathrm{E}}} - 1 = \frac{1}{2p_{\mathrm{E}}}\left\{\frac{1}{\lambda}(1-\lambda) + p_{\mathrm{D}} - p_{\mathrm{E}}\right\} > 0. \tag{5.13}$$

This completes the proof of (iii) of Lemma 5.1.
(iv) This is a direct consequence of (iii).
(v) Suppose that $p_{\mathrm{E}_1} = p_{\mathrm{E}_2} = p_{\mathrm{E}}$ and $p_{\mathrm{D}_1} < p_{\mathrm{D}_2}$. In this case, there is no $\alpha \in (0,1)$ satisfying the equation $(p_{\mathrm{E}_1} - p_{\mathrm{E}_2})\alpha - (p_{\mathrm{D}_1} - p_{\mathrm{D}_2}) = 0$. Thus, one of the inequalities in (iv) holds. Note here that, from Eq. (4.5) and Theorem 4.3 in Chapter 4, we have $g'_{\tau_1}(1) = 1/(p_{\mathrm{D}} - p_{\mathrm{E}})$. Thus, we obtain $g'_{\{\tau_1,p_{\mathrm{D}_1},p_{\mathrm{E}}\}}(1) > g'_{\{\tau_1,p_{\mathrm{D}_2},p_{\mathrm{E}}\}}(1)$. Using these facts, we obtain $g_{\{\tau_1,p_{\mathrm{D}_1},p_{\mathrm{E}}\}}(\lambda) < g_{\{\tau_1,p_{\mathrm{D}_2},p_{\mathrm{E}}\}}(\lambda)$ for every $\lambda \in (0,1)$.
(vi) In a manner similar to the proof of (v), we can obtain $g_{\{\tau_1,p_{\mathrm{D}},p_{\mathrm{E}_1}\}}(\lambda) > g_{\{\tau_1,p_{\mathrm{D}},p_{\mathrm{E}_2}\}}(\lambda)$ for every $\lambda \in (0,1)$, when $p_{\mathrm{E}_1} < p_{\mathrm{E}_2}$.
(vii) Suppose that $p_{\mathrm{D}_1} - p_{\mathrm{E}_1} = p_{\mathrm{D}_2} - p_{\mathrm{E}_2}$, $p_{\mathrm{D}_1} < p_{\mathrm{D}_2}$, and $p_{\mathrm{E}_1} < p_{\mathrm{E}_2}$. In this case, we have $p_{\mathrm{D}_1} - p_{\mathrm{D}_2} = p_{\mathrm{E}_1} - p_{\mathrm{E}_2}$, and thus, there is no $\alpha \in (0,1)$ satisfying the equation $(p_{\mathrm{E}_1} - p_{\mathrm{E}_2})\alpha - (p_{\mathrm{D}_1} - p_{\mathrm{D}_2}) = 0$. Therefore, the condition (iv) is satisfied also in this case, and one of the inequality of (iv) holds. Note here that, from the definition (4.4), we have $g'_{\tau_1}(0) = P(\tau_1 = 1) = p_{\mathrm{D}}$. Consequently, we obtain $g'_{\{\tau_1,p_{\mathrm{D}_1},p_{\mathrm{E}_1}\}}(0) < g'_{\{\tau_1,p_{\mathrm{D}_2},p_{\mathrm{E}_2}\}}(0)$ in this case. Using these facts, we obtain $g_{\{\tau_1,p_{\mathrm{D}_1},p_{\mathrm{E}_1}\}}(\lambda) < g_{\{\tau_1,p_{\mathrm{D}_2},p_{\mathrm{E}_2}\}}(\lambda)$ for every $\lambda \in (0,1)$. ■

5.2.4 Evaluation of Sender Anonymity

From the results of the preceding subsection, we evaluate sender anonymity in 3MN.

a) Influences of the Probabilities of Mode Selections

Using Theorem 5.1 and Lemma 5.1, we investigate the influences of the probabilities of mode selections. From Theorem 5.1, $P(J \mid L_{1+})$ is rewritten to be

$$P(J \mid L_{1+}) = \frac{1}{1-g_{\tau_k}(\lambda)}\left(\frac{n_c+1}{n_t} - \frac{1}{n_t-n_c}\right) + \frac{1}{n_t-n_c}. \tag{5.14}$$

Since $n_t \geq n_c + 2$, the following equation holds.

$$\frac{n_c+1}{n_t} - \frac{1}{n_t-n_c} = \frac{n_c(n_t-n_c-1)}{n_t(n_t-n_c)} > 0. \tag{5.15}$$

From the above equations and the statement (ii) of Lemma 5.1, we have the following theorem.

Theorem 5.2

$$P(J \mid L_{1+}) \;>\; \frac{1}{n_t - n_c}. \tag{5.16}$$

The right-hand side of Eq. (5.16) is the probability that a randomly chosen node is the sender. Therefore, Theorem 5.2 implies that the first immediate predecessor among the immediate predecessors of the collaborating nodes is the most presumable in all non-collaborating nodes whatever the probabilities of mode selections and the initial multiplicity of encryption may be.

We also observe from Eqs. (5.14) and (5.15) that $P(J \mid L_{1+})$ becomes large with $g_{\tau_k}(\lambda)$. From the statements (v) and (vi) of Lemma 5.1, the following theorem is obtained.

Theorem 5.3 *$P(J \mid L_{1+})$ has the following properties:*

(i) When p_{E} is constant, $P(J \mid L_{1+})$ increases as p_{D} increases.

(ii) When p_{D} is constant, $P(J \mid L_{1+})$ decreases as p_{E} increases.

Consequently, in order to provide high anonymity to a sender, we set p_{D} to be small and p_{E} to be large.

b) Influences of the Initial Multiplicity of Encryption

We consider the influences of the initial multiplicity of encryption on sender anonymity. From the statement (viii) of Lemma 5.1, the following theorem is obtained.

Theorem 5.4 *If k increases, $P(J \mid L_{1+})$ decreases.*

Therefore, by setting k to be large, we can provide high anonymity to a sender.

When k goes to infinity, we can derive the following theorem.

Theorem 5.5 *Let n_t and n_c denote the number of 3MN members and that of collaborating nodes in 3MN, respectively. Then, $P(J \mid L_{1+})$ converges to $(n_c+1)/n_t$ whatever p_{D}, p_{E}, and p_{T} may be as k goes to infinity. In addition, this value is larger than $1/(n_t-n_c)$.*

Proof of Theorem 5.5: From the statement (ii) of Lemma 5.1, together with the fact that $g_{\tau_k}(\lambda) = g_{\tau_1}(\lambda)^k$, $\lim_{k\to\infty} g_{\tau_k}(\lambda) = 0$ $(0<\lambda<1)$. Thus, we obtain

$$\lim_{k\to\infty} P(J \mid L_{1+}) \;=\; \frac{n_c + 1}{n_t}. \tag{5.17}$$

The last statement of Theorem 5.5 is derived from Eq. (5.15). ■

Theorem 5.5 implies that even though the number of the initial multiplicity of encryption goes to infinity, the probability that the first immediate predecessor among the immediate predecessors of the collaborating nodes is the sender does not approach to the probability that a randomly chosen node is the sender.

5.3 Relationship between the Number of Relay Nodes and Sender Anonymity

In this section, we consider the relationship between the number of relay nodes and sender anonymity in 3MN.

First, we consider the relationship between the expectation of the number of relay nodes and sender anonymity. From Theorem 4.3 in Section 4.4.2, in order to reduce the expectation of the number of relay nodes, we have to set p_{D} to be large, p_{E} to be small, and k to be small. In contrast, as discussed in Section 5.2, in order to provide high sender anonymity, we have to set p_{D} to be small, p_{E} to be large, and k to be large. Accordingly, there is a trade-off between the expectation of the number of relay nodes and sender anonymity.

Next, in order to reduce the variance of the number of relay nodes while the expectation remains constant and to provide high sender anonymity, we investigate the variation of $P(J \mid L_{1+})$ when p_{T} increases under the condition that $p_{\mathrm{D}}-p_{\mathrm{E}}$ is constant. Using the statement (vii) of Lemma 5.1, we obtain the following theorem.

Theorem 5.6 *$P(J \mid L_{1+})$ decreases as p_{T} increases, when $p_{\mathrm{D}}-p_{\mathrm{E}}$ is constant.*

Theorem 5.6 indicates that it is possible to reduce the variance of the number of relay nodes with the expectation unchanged and to keep high sender anonymity. Therefore, it is shown that we set p_{T} to be large in order to reduce the situation that the number of relay nodes becomes extremely large and to provide high anonymity to a sender.

5.4 Numerical Examples

5.4.1 Effects of the Probabilities of Mode Selections

In this subsection, we consider the influence of the probabilities of mode selections on sender anonymity against collaborating nodes through numerical examples.

First, we illustrate sender anonymity under the condition that $k = 2$, $n_t = 10$, and $n_c = 1$. Fig. 5.2 shows sender anonymity under the various probabilities of mode selections in the region satisfying $0 < p_{\mathrm{D}} < 1$, $0 < p_{\mathrm{E}} < 1$, $0 < p_{\mathrm{D}} + p_{\mathrm{E}} < 1$ (this corresponds to $0 < p_{\mathrm{T}} < 1$), and $p_{\mathrm{D}} > p_{\mathrm{E}}$. From Fig. 5.2, we observe that $P(J \mid L_{1+})$ becomes large with p_{D} under the condition that p_{E} is constant. Also, as p_{E} becomes small, $P(J \mid L_{1+})$ becomes large under the condition that p_{D} is constant. These results show that sender anonymity degrades when we set p_{D} to be large and p_{E} to be small. In order to provide high anonymity, we must set p_{D} to be small and p_{E} to be large. In such a situation, however, the number of relay nodes becomes large as discussed in Section 4.4. Consequently, as discussed in Section 5.3, there is a performance trade-off between sender anonymity and the number of relay nodes required for communication.

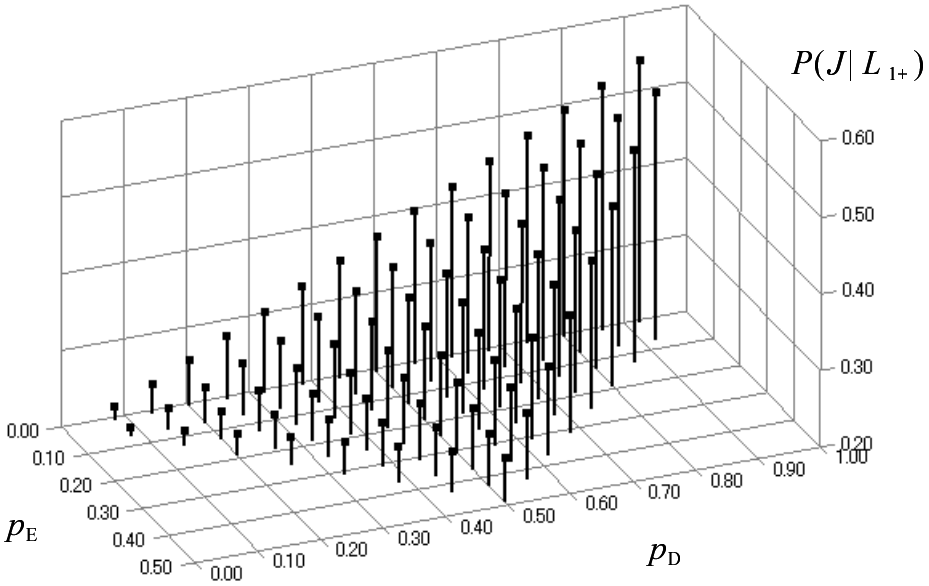


Figure 5.2: Sender anonymity under the various probabilities of mode selections.

Table 5.1: Expectations and variances of the number of relay nodes and the number of encryption, and sender anonymity.

	(p_D, p_E, p_T) $(k = 1, n_t = 10, n_c = 1)$		
	(0.50, 0.43, 0.07)	(0.10, 0.03, 0.87)	(0.75, 0.05, 0.20)
M_N	14.29	14.29	1.429
V_N	2697	364.7	0.9038
M_E	7.143	1.426	1.071
V_E	582.9	1.137	0.08746
$P(J \mid L_{1+})$	0.3728	0.2695	0.7564

Second, in order to investigate the effects of the probabilities of mode selections on sender anonymity together with the number of relay nodes and the number of encryption, we consider the following three cases under the condition that $k = 1$: Case A: (p_D, p_E, p_T)=(0.50, 0.43, 0.07), Case B: (p_D, p_E, p_T)=(0.10, 0.03, 0.87), and Case C: (p_D, p_E, p_T)=(0.75, 0.05, 0.20). Case A and Case B were also considered in Section 4.5. The reason for $k = 1$ is to see the influence of mode probabilities for sender anonymity more clearly. The numerical results of the expectations and the variances of the number of relay nodes and the number of encryption, and the conditional probability $P(J \mid L_{1+})$ are shown in Table 5.1,

Table 5.2: Sender anonymity in the several initial multiplicity of encryption.

	(p_D, p_E, p_T) $(n_t = 10, n_c = 1)$		
	(0.50, 0.43, 0.07)	(0.10, 0.03, 0.87)	(0.75, 0.05, 0.20)
$k = 1$	0.3728	0.2695	0.7654
$k = 2$	0.2687	0.2212	0.4621
$k = 3$	0.2360	0.2082	0.3617
$k = 5$	0.2128	0.2015	0.2827
$k = 10$	0.2014	0.2000	0.2269
$k = 100$	0.2000	0.2000	0.2000

where M_N, V_N, M_E, and V_E stand for the expectation and the variance of the number of relay nodes, and the expectation and the variance of the number of encryption, respectively. From Table 5.1, we observe that $P(J \mid L_{1+}) = 0.3728$ in Case A and $P(J \mid L_{1+}) = 0.2695$ in Case B. If collaborating nodes tried to identify a message sender without any information other than the number of 3MN members and that of collaborating nodes, all non-collaborating nodes would seem to be the message sender equally likely, i.e., $1/(n_t - n_c) = 1/9 = 0.1111$. Compared with this, $P(J \mid L_{1+})$ in Case A and $P(J \mid L_{1+})$ in Case B are larger, yet less than 0.5. Therefore, sender anonymity is maintained in these two examples, because it is shown that sender anonymity is provided in [4] if $P(J \mid L_{1+})$ is less than 0.5.

From Table 5.1, 3MN in Case C hardly maintains sender anonymity because $P(J \mid L_{1+}) = 0.7564$. In Case C, the first relay node sends the message to a message receiver with high probability 0.75, and thus the number of relay nodes and the number of encryption are quite small. When the number of relay nodes and the number of encryption become small, sender anonymity becomes small. Consequently, we conclude that sender anonymity deteriorates when the inappropriate probabilities of mode selections are chosen so that the number of relay nodes and the number of encryption are quite small.

We also observe that sender anonymity for Case A and Case B is different although the expectations of the number of relay nodes are identical. It follows from this observation that there exist the probabilities of mode selections, which yield small numbers of relay nodes and encryption, and high sender anonymity. This is the same as the discussion in Section 5.3. Note here that we have to consider how much anonymity is guaranteed for a message receiver because the behavior of 3MN in Case B is almost similar to that of Crowds and Crowds does not provide receiver anonymity. Therefore, in order to consider the security and the performance of 3MN, we need to analyze 3MN in more detail.

5.4.2 Effects of the Initial Multiplicity of Encryption

In this subsection, using the three cases in the above examples, we examine the influence of the initial multiplicity of encryption on sender anonymity through numerical examples.

Table 5.2 indicates that, in the examples, the probabilities $P(J \mid L_{1+})$ between $k = 1$ and $k = 2$ vary widely. This implies that, considering the performance of 3MN, it is appropriate to select $k = 2$ as the initial multiplicity of encryption. We also observe that $P(J \mid L_{1+})$ converges to 0.2 whatever p_{D}, p_{E}, and p_{T} may be, when k becomes large. This value is larger than the probability $1/(n_t - n_c) = 1/9$, as stated in Theorem 5.5.

5.5 Conclusion

In this chapter, we have analyzed sender anonymity for 3MN against collaborating nodes. In order to evaluate sender anonymity in 3MN, we have considered collaborating nodes who try to identify a message sender, and have derived the conditional probability that the first immediate predecessor among all the collaborating nodes is the message sender when a collaborating node receives the message. This conditional probability has been represented by a probability generating function. Through the analysis, we clarified that

- Our approach needs no infinite series unlike the method employed in Crowds.
- The probability to choose D-Mode is set to be small and the probability to choose E-Mode is set to be large in order to provide high sender anonymity.
- The initial multiplicity of encryption is equal to 2 through simulations.
- The probability to choose T-Mode is set to be large in order to keep high sender anonymity and to reduce the variance of the number of relay nodes under the condition that its expectation is constant.

Although we analyzed sender anonymity in 3MN, we do not investigate receiver anonymity. As shown in Section 5.3, if the probability to choose T-Mode is set very large, receiver anonymity might be lost because the behavior of 3MN is similar to that of Crowds. Thus, we need to consider receiver anonymity as well as sender anonymity.

Chapter 6

Conclusions and Future Directions

In this dissertation, we have discussed policy integration and anonymous communication for protecting personal information. These techniques are essential to protect personal information from the viewpoint of the life-cycle of personal information. The development of policy integration and anonymous communication systems enables us to prevent information leakage as well as to hide the information of a sender and a receiver. The contents and the results of this dissertation are summarized below.

In Chapter 3, we have introduced a matrix-based algorithm for integrating inheritance relations of access rights for policy integration. Since the integration of inheritance relations plays an especially important role in the framework of policy integration, we have focused on the method for integrating inheritance relations of access rights. The main characteristics of our algorithms are as follows: 1) our integration algorithm can deal with inheritance relations in subject, resource, and action categories; 2) our algorithm is easily implementable. Introducing inheritance relations in three categories, we can apply our algorithm for existing access rights management systems, i.e., Microsoft Windows® Rights Management Services and Adobe® Livecycle™ Rights Management ES. We have also proved that operations necessary for integrating inheritance relations can be carried out by basic matrix operations in order to provide the second characteristics.

In order to clarify the influences of the probabilities of mode selections on the behavior of an anonymous communication system named 3-Mode Net (3MN), we have analyzed the performance of 3MN quantitatively. In Chapter 4, we have analyzed the number of relay nodes and the number of encryption required for communication in 3MN. We have given explicit formulas of their probability distributions, expectations, and variances. In Chapter 5, we have evaluated sender anonymity against collaborating nodes who collude with each other in order to identify a message sender. We have derived the conditional probability that the first immediate predecessor of all the collaborating nodes on the communication path coincides with the message sender under the condition that a collaborating node receives a data set. Introducing probability generating functions, our approach needs no infinite series unlike the method employed in Crowds. Through the analysis of the performance of 3MN,

it is shown that

- The probability to choose T-Mode is set to be large in order to keep sender anonymity and to reduce the variance of the number of relay nodes under the condition that its expectation is constant.
- There is no way of selecting the probabilities of mode selections for reducing the variance of the number of encryption under the condition that its expectation is constant.
- The probability to choose D-Mode is set to be small and the probability to choose E-Mode is set to be large in order to provide high sender anonymity.
- The initial multiplicity of encryption is equal to 2 through simulations.

Let us discuss future work in policy integration and anonymous communication systems. One of the most significant problems about our policy integration framework is that our algorithm will stop if an administrator thinks that the conflicts caused by the integration yield serious problems. Thus we have to improve the algorithm so as to preclude the inheritance relations which may cause conflicts, and to continue the integration process without the inheritance relations. Our framework also does not address inheritance relations of access rights which do not satisfy the independence assumption. Thus we need to extend our framework. Furthermore, access rights management technologies include many technologies, and many systems are used in practice. Therefore, the implementation of our policy integration framework to a centralized access rights management system together with policy refinement and policy conflict detection/resolution technologies [6, 7, 10, 11, 13] is an important future topic of our study.

We next consider future issues on 3MN. Although we evaluate the number of relay nodes, the number of encryption, and sender anonymity, we do not evaluate receiver anonymity against collaborating nodes. To define receiver anonymity against collaborating nodes is necessary for investigating sender-receiver anonymity and the relationship between the number of relay nodes and sender-receiver anonymity. We should also investigate anonymity in 3MN against other attacks such as eavesdropping and timing attacks because sender anonymity and receiver anonymity might not be guaranteed under particular conditions. In order to confirm the utility of 3MN in practice, the implementation of 3MN is now on-going. We need to compare 3MN to anonymous communication systems used on the Internet, such as Tor [36, 60], Mixminion [38], and Freenet [39]. Further, there may exist a lot of challenges on the implementation, i.e., the issue on public key encryption [61] and the issue on the network topology of 3MN [55, 62, 63].

Although we focus on policy integration and anonymous communication, technologies for protecting personal information are closely related to other technologies such as authentication, digital signature, and automated trust negotiation. In particular, we pay attention to automated trust negotiation [64, 65]. In the future, we will consider a new framework for protecting personal information by integrating several technologies.

Finally, through this research, we have developed two technologies for protecting personal information, policy integration and anonymous communication. We hope that this research will make some contributions toward the development of novel technologies for protecting personal information.

Bibliography

[1] A. Komatsu, "Privacy Conscious Architecture in Identity & Access Management," *Journal of Information Processing Society of Japan (<Special Feature> Trends on Information Security Research and Development)*, Vol. 48, No. 7, pp. 737–743, July 2007, (in Japanese).

[2] E. Okamoto, "Recent Research of Basic Technologies for Privacy Protection," *Journal of Information Processing Society of Japan (<Special Feature> Trends on Information Security Research and Development)*, Vol. 48, No. 7, pp. 744–749, July 2007, (in Japanese).

[3] N. Miyake, Y. Ito, and N. Babaguchi, "3-Mode Net: A Bi-directional Anonymous Communication System Based on Multiple Encryption and Probabilistic Selections of Actions," *The Institute of Electronics, Information and Communication Engineers (IEICE) Trans. Fundamentals of Electronics, Communications and Computer Sciences*, Vol. J91-A, No. 10, pp. 949–956, October 2008, (in Japanese).

[4] M. Reiter and A. Rubin, "Crowds: Anonymity for Web Transactions," *ACM Trans. Information and System Security*, Vol. 1, No. 1, pp. 66–92, June 1998.

[5] M. Reed, P. Syverson, and D. Goldschlag, "Anonymous Connections and Onion Routing," *IEEE Journal on Selected Areas in Communications*, Vol. 16, No. 4, pp. 482–494, May 1998.

[6] A. Bandara, E. Lupu, J. Moffettt, and A. Russo, "A Goal-Based Approach to Policy Refinement," in *Proc. 5th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY 2004)*, June 2004, pp. 229–239.

[7] J. Rubio-Loyola, J. Serrat, M. Charalambides, P. Flegkas, G. Pavlou, and A. Lafuente, "Using Linear Temporal Model Checking for Goal-oriented Policy Refinement Frameworks," in *Proc. 6th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY 2005)*, June 2005, pp. 181–190.

[8] J. Rubio-Loyola, J. Serrat, M. Charalambides, P. Flegkas, and G. Pavlou, "A Functional Solution of Goal-Oriented Policy Refinement," in *Proc. 7th IEEE International Work-*

shop on Policies for Distributed Systems and Networks (POLICY 2006), June 2006, pp. 133–144.

[9] B. Shafiq, J. B. D. Joshi, E. Bertino, and A. Ghafoor, "Secure Interoperation in a Multidomain Environment Employing RBAC Policies," *IEEE Trans. Knowledge and Data Engineering*, Vol. 17, No. 11, pp. 1557–1577, November 2005.

[10] D. Agrawal, J. Giles, K. Lee, and J. Lobo, "Policy Ratification," in *Proc. 6th IEEE International Workshop on Policies for Distributed Systems and Networks (Policy 2005)*, June 2005, pp. 223–232.

[11] H. Kamoda, M. Yamaoka, S. Matsuda, K. Broda, and M. Sloman, "Policy Conflict Analysis Using Free Variable Tableaux for Access Control in Web Services Environments," in *Proc. Policy Management for the Web Workshop at the 14th International World Wide Web Conference (WWW 2005)*, March 2005, pp. 5–12.

[12] H. Kamoda, A. Hayakawa, M. Yamaoka, S. Matsuda, K. Broda, and M. Sloman, "Policy Conflict Analysis Using Tableaux for On Demand VPN Framework," in *Proc. 6th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2005)*, June 2005, pp. 565–569.

[13] H. Kamoda, K. Kono, Y. Ito, and N. Babaguchi, "Access Control Policy Inconsistency Check Using Model Checker," in *The Special Interest Group Notes of Information Processing Society of Japan (IPSJ)*, March 2007, pp. 441–446, (in Japanese).

[14] C. Montangero, S. Reiff-Marganiec, and L. Semini, "Logic-Based Conflict Detection for Distributed Policies," *Fundamenta Informaticae*, Vol. 89, No. 4, pp. 511–538, January 2009.

[15] Organization for the Advancement of Structured Information Standards (OASIS), *eXtensible Access Control Markup Language (XACML) Version 2.0*, OASIS Standard, February 2005.

[16] Organization for the Advancement of Structured Information Standards (OASIS), *Core and Hierarchical Role Based Access Control (RBAC) Profile of XACML Version 2.0*, OASIS Standard, February 2005.

[17] N. Damianou, N. Dulay, E. Lupu, and M. Sloman, "The Ponder Policy Specification Language," in *Proc. International Workshop on Policies for Distributed Systems and Networks (POLICY 2001)*, January 2001, pp. 18–38.

[18] Microsoft, "Windows Server 2003 Rights Management Services," Microsoft (online), http://www.microsoft.com/windowsserver2003/technologies/rightsmgmt/default.mspx (accessed 2008-12-24).

[19] Adobe Systems, "Adobe LiveCycle Rights Management ES," Adobe Systems (online), http://www.adobe.com/products/server/policy/ (accessed 2008-12-24).

[20] P. Bonatti, M. Sapino, and V. Subrahmanian, "Merging Heterogeneous Security Orderings," in *Proc. European Symposium on Research in Computer Security (ESORICS 1996)*, September 1996, pp. 183–197.

[21] L. Gong and X. Qian, "Computational Issues in Secure Interoperation," *IEEE Trans. Software Engineering*, Vol. 22, No. 1, pp. 43–52, January 1996.

[22] S. Dawson, S. Qian, and P. Samarati, "Providing Security and Interoperation of Heterogeneous Systems," *Distributed and Parallel Databases*, Vol. 8, No. 1, pp. 119–145, January 2000.

[23] C. Pan, P. Mitra, and P. Liu, "Semantic Access Control for Information Interoperation," in *Proc. 11th ACM Symposium on Access Control Models and Technologies (SACMAT 2006)*, June 2006, pp. 237–246.

[24] P. Mazzoleni, E. Bertino, B. Crispo, and S. Sivasubramanian, "XACML Policy Integration Algorithms," in *Proc. 11th ACM Symposium on Access Control Models and Technologies (SACMAT 2006)*, June 2006, pp. 219–227.

[25] M. Koch, L. Mancini, and F. Parisi-Presicce, "On the Specification and Evolution of Access Control Policies," in *Proc. 6th ACM Symposium on Access Control Models and Technologies (SACMAT 2001)*, May 2001, pp. 121–130.

[26] M. Sugano, S. Tanaka, Y. Sakata, K. Oguma, and N. Shiratori, "Application and Implementation of Policy Control Method "PolicyComputing" in Computer Networks," *Trans. Information Processing Society of Japan*, Vol. 42, No. 2, pp. 126–137, February 2001, (in Japanese).

[27] S. Yau and Z. Chen, "Security Policy Integration and Conflict Reconciliation for Collaborations among Organizations in Ubiquitous Computing Environments," in *Proc. 5th International Conference on Ubiquitous Intelligence and Computing*, June 2008, pp. 3–19.

[28] D. Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms," *Communications of the ACM*, Vol. 24, No. 2, pp. 84–88, February 1981.

[29] D. Chaum, "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability," *Journal of Cryptology*, Vol. 1, pp. 65–75, January 1988.

[30] B. Levine and C. Shields, "Hordes: A Multicast Based Protocol for Anonymity," *ACM Journal of Computer Security*, Vol. 10, No. 3, pp. 213–240, September 2002.

[31] R. Song and L. Korba, "Review of Network-Based Approaches for Privacy," in *Proc. 14th Annual Canadian Information Technology Security Symposium (CITSS 2002)*, May 2002, pp. 1–10.

[32] S. Kitazawa, S. Nagano, M. Soshi, and A. Miyaji, "Anonymous Communication with Elementary Cyclic Routes," *Trans. Information Processing Society of Japan*, Vol. 41, No. 8, pp. 2148–2161, August 2000, (in Japanese).

[33] S. Yamanaka, K. Kobara, and H. Imai, "Valkyrie: An Anonymous Routing Scheme on Unstable Network," *Trans. Information Processing Society of Japan*, Vol. 46, No. 8, pp. 2025–2035, August 2005, (in Japanese).

[34] S. Goel, M. Robson, M. Polte, and E. Sirer, "Herbivore: A Scalable and Efficient Protocol for Anonymous Communication," Cornell University Computing and Information Science, Technical Report 2003-1890, February 2003.

[35] R. Sherwood, B. Bhattacharjee, and A. Srinivasan, "$\mathcal{P}^5$: A Protocol for Scalable Anonymous Communication," in *Proc. IEEE Symposium on Security and Privacy*, May 2002, pp. 58–70.

[36] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The Second-Generation Onion Router," in *Proc. 13th USENIX Security Symposium*, August 2004, pp. 303–320.

[37] J. Boyan, "The Anonymizer: Protecting User Privacy on the Web," *Computer-Mediated Communication Magazine*, Vol. 4, No. 9, September 1997.

[38] G. Danezis, R. Dingledine, and N. Mathewson, "Mixminion: Design of a Type III Anonymous Remailer Protocol," in *Proc. IEEE Symposium on Security and Privacy*, May 2003, pp. 2–15.

[39] I. Clarke, O. Sandberg, B. Wiley, and T. Hong, "Freenet: A Distributed Anonymous Information Storage and Retrieval System," in *Proc. International Workshop on Design Issues in Anonymity and Unobservability*, July 2000, pp. 46–66.

[40] S. Nakano, K. Kono, Y. Ito, and N. Babaguchi, "Reduction of the Number of Relay Nodes in Anonymous Communication System 3-Mode Net," in *Proc. 2009 Institute of Electronics, Information and Communication Engineers (IEICE) General Conference*, A-7-8, March 2009, p. 182, (in Japanese).

[41] M. Reiter and A. Rubin, "Anonymous Web Transactions with Crowds," *Communications of the ACM*, Vol. 42, No. 2, pp. 32–38, February 1999.

[42] M. Gomulkiewicz, M. Klonowski, and M. Kutylowski, "Onions Based on Universal Re-Encryption - Anonymous Communication Immune Against Repetitive Attack," in *Proc. 5th International Workshop on Information Security Applications (WISA 2004)*, August 2004, pp. 400–410.

[43] J. Tamura, K. Kobara, and H. Imai, “New Bi-directional Anonymous Routing Schemes over Dynamic Networks,” *Trans. Information Processing Society of Japan*, Vol. 48, No. 2, pp. 494–504, February 2007, (in Japanese).

[44] K. Kono, Y. Ito, A. Aoyama, H. Kamoda, and N. Babaguchi, “Matrix-Based Algorithm for Integrating Inheritance Relations of Access Rights for Policy Generation,” *Journal of Information Processing*, Vol. 17, pp. 318–327, December 2009.

[45] K. Kono, Y. Ito, A. Aoyama, H. Kamoda, and N. Babaguchi, “An Integration Method of Access Control Policies Using Adjacency Matrix,” in *The Special Interest Group Notes of Information Processing Society of Japan (IPSJ)*, May 2007, pp. 45–50, (in Japanese).

[46] N. Christofides, *Graph Theory : An Algorithmic Approach*. Academic Press, 1975.

[47] C. Berge, *The Theory of Graphs and its Applications*. John Wiley & Sons, Inc., 1962.

[48] J. W. Brewer, “Kronecker Products and Matrix Calculus in System Theory,” *IEEE Trans. Circuits and Systems*, Vol. 25, No. 9, pp. 772–781, September 1978.

[49] D. Ferraiolo and D. Kuhn, “Role Based Access Control,” in *Proc. 15th National Computer Security Conference*, October 1992, pp. 554–563.

[50] D. Ferraiolo, R. Sandhu, S. Gavrila, D. Kuhn, and R. Chandramouli, “Proposed NIST Standard for Role-Based Access Control,” *ACM Trans. Information and System Security*, Vol. 4, No. 3, pp. 224–274, August 2001.

[51] K. Kono, S. Nakano, Y. Ito, and N. Babaguchi, “Performance Analysis of Anonymous Communication System 3-Mode Net,” in *Proc. 5th International Conference on Information Assurance and Security (IAS 2009)*, August 2009, pp. 593–596.

[52] K. Kono, S. Nakano, Y. Ito, and N. Babaguchi, “A Consideration on the Numbers of Relay Nodes and Encryption Required for Anonymous Communication System 3-Mode Net,” *Journal of Information Assurance and Security*, to appear in 2010.

[53] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. John Wiley & Sons, Inc., 1968, Vol. 1.

[54] D. Figueiredo, P. Nain, and D. Towsley, “On the Analysis of the Predecessor Attack on Anonymity Systems,” University of Massachusetts Computer Science Technical Report 04-65, July 2004.

[55] A. Nambiar and M. Wright, “Salsa: A Structured Approach to Large-Scale Anonymity,” in *Proc. ACM Conference on Computer and Communication Security (CCS 2006)*, October 2006, pp. 17–26.

[56] M. Wright, M. Adler, B. Levine, and C. Shields, "The Predecessor Attack: An Analysis of a Threat to Anonymous Communication Systems," *ACM Trans. Information and System Security*, Vol. 7, No. 4, pp. 489–522, November 2004.

[57] P. Kocher, "Timing Attacks on Implementations of Die-Hellman, RSA, DSS, and Other Systems," in *Proc. 16th Annual International Cryptology Conference on Advances in Cryptology*, August 1996, pp. 104–113.

[58] K. Kono, S. Nakano, Y. Ito, and N. Babaguchi, "Security Analysis of Anonymous Communication System 3-Mode Net Against Collaborating Nodes," in *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2009)*, October 2009, pp. 105–110.

[59] P. Mittal and N. Borisov, "Information Leaks in Structured Peer-to-Peer Anonymous Communication Systems," in *Proc. ACM Conference on Computer and Communications Security (CCS 2008)*, October 2008, pp. 267–278.

[60] A. Panchenko, L. Pimenidis, and J. Renner, "Performance Analysis of Anonymous Communication Channels Provided by Tor," in *Proc.3rd International Conference on Availability, Reliability and Security (ARES 2008)*, March 2008, pp. 221–228.

[61] K. Kono, Y. Ito, and N. Babaguchi, "A Study on RSA Cryptosystem Using Pseudoprimes," in *Proc. 2005 Society Conference of Institute of Electronics, Information and Communication Engineers (IEICE)*, A-7-2, September 2005, p. 175, (in Japanese).

[62] S. Seys and B. Preneel, "ARM: Anonymous Routing Protocol for Mobile Ad hoc Networks," *International Journal of Wireless and Mobile Computing*, Vol. 3, No. 3, pp. 145–155, October 2009.

[63] A. Durresi, V. Paruchuri, L. Barolli, and R. Kannan, "Anonymous Communication Protocol for Sensor Networks," *International Journal of Wireless and Mobile Computing*, Vol. 3, No. 4, pp. 236–246, November 2009.

[64] R. Tanihira, K. Kono, Y. Ito, and N. Babaguchi, "Fair Automated Trust Negotiation with Credential Disclosure Points," in *Proc. 2007 Institute of Electronics, Information and Communication Engineers (IEICE) General Conference*, A-7-10, March 2007, p. 214, (in Japanese).

[65] T. Hanaoka, K. Kono, Y. Ito, and N. Babaguchi, "Automated Trust Negotiation among Three Agent Based on Service Usage of Client," in *Proc. 2008 Institute of Electronics, Information and Communication Engineers (IEICE) General Conference*, A-7-10, March 2008, p. 182, (in Japanese).

Appendix A

Proofs and Derivations

A.1 Proof of Theorem 3.5

Let X and Y denote square matrices of order n and order m, respectively. Since the graphs associated with X and Y do not include circuits, we obtain the following equations from Theorem 3.1.

$$X^p = 0 \ \ (\forall p \geq n), \quad Y^q = 0 \ \ (\forall q \geq m). \tag{A.1}$$

In order to prove Theorem 3.5, it suffices to show that $(A \oplus B)^{mn} = 0$. To this end, we apply the following basic properties of Kronecker product [48].

$$(X \otimes I_m)^p = X^p \otimes I_m, \tag{A.2}$$

$$(I_n \otimes Y)^q = I_n \otimes Y^q, \tag{A.3}$$

$$(X \otimes I_m) \cdot (I_n \otimes Y) = (I_n \otimes Y) \cdot (X \otimes I_m), \tag{A.4}$$

$$(X \oplus Y)^l = \sum_{p=0}^{l} (X \otimes I_m)^p \cdot (I_n \otimes Y)^{l-p}. \tag{A.5}$$

From Eq. (A.1) together with Eqs. (A.2), (A.3), (A.4), and (A.5), the following equations are obtained.

$$\overline{X}^p = 0 \ \ (\forall p \geq n), \quad \overline{Y}^q = 0 \ \ (\forall q \geq m), \tag{A.6}$$

$$(X \oplus Y)^{mn} = \sum_{p=0}^{mn} \overline{X}^p \overline{Y}^{mn-p}, \tag{A.7}$$

where $\overline{X} = X \otimes I_m$ and $\overline{Y} = I_n \otimes Y$. Note here that $mn - n + 1 \geq m$ because $(m-1)(n-1) \geq 0$. Hence, from Eq. (A.1), we obtain

$$(X \oplus Y)^{mn} = \sum_{p=0}^{n-1} \overline{X}^p \overline{Y}^{mn-p} = \sum_{q=mn-n+1}^{mn} \overline{X}^{mn-q} \overline{Y}^q = 0. \tag{A.8}$$

■

A.2 Proof of Theorem 3.6

Let $X = [x_{ij}]$ and $Y = [y_{ij}]$ denote square matrices of order n and order m, respectively. In this proof, $x_{ij}^{(l)}$ and $y_{ij}^{(l)}$ denote the (i, j)-element of X^l and Y^l, respectively.

Since the graphs associated with X and Y do not include circuits and redundant edges, the following equations are obtained from Theorem 3.1 and Theorem 3.2.

$$x_{ii}^{(l)} = 0 \ \ (1 \le i \le n, \ l \ge 1), \tag{A.9}$$

$$y_{ii}^{(l)} = 0 \ \ (1 \le i \le m, \ l \ge 1). \tag{A.10}$$

From Theorem 3.3, we obtain for $l \neq 2$

$$x_{ij}^{(l)} = 0, \ \text{ if } \ x_{ij} = 1 \ (i \neq j), \tag{A.11}$$

$$y_{ij}^{(l)} = 0, \ \text{ if } \ y_{ij} = 1 \ (i \neq j). \tag{A.12}$$

From Eqs. (A.2), (A.3), (A.4), and (A.5), $(X \oplus Y)^l$ $(l \ge 2)$ is calculated as follows:

$$\begin{aligned}
(X \oplus Y)^l &= \sum_{z=0}^{l} (X \otimes I_m)^l (I_n \otimes Y)^{l-z} = \sum_{z=0}^{l} (X^z \otimes I_m)(I_n \otimes Y^{l-z}) \\
&= \sum_{z=0}^{l} \begin{bmatrix} x_{11}^{(z)} I_m & \cdots & x_{1n}^{(z)} I_m \\ \vdots & \ddots & \vdots \\ x_{n1}^{(z)} I_m & \cdots & x_{nn}^{(z)} I_m \end{bmatrix} \times \begin{bmatrix} Y^{l-z} & & 0 \\ & \ddots & \\ 0 & & Y^{l-z} \end{bmatrix} \\
&= \sum_{z=0}^{l} \begin{bmatrix} x_{11}^{(z)} Y^{l-z} & \cdots & x_{1n}^{(z)} Y^{l-z} \\ \vdots & \ddots & \vdots \\ x_{n1}^{(z)} Y^{l-z} & \cdots & x_{nn}^{(z)} Y^{l-z} \end{bmatrix}.
\end{aligned}$$

On the other hand, $(X \oplus Y)$ is calculated as

$$(X \oplus Y) = (X \otimes I_m) + (I_n \otimes Y) = \begin{bmatrix} Y & x_{12} I_m & \cdots & x_{1n} I_m \\ x_{21} I_m & Y & & \vdots \\ \vdots & & \ddots & x_{n-1n} I_m \\ x_{n1} I_m & \cdots & x_{nn-1} I_m & Y \end{bmatrix}.$$

In order to prove Theorem 3.6, it suffices to show the following propositions:

(i) If $y_{ij} = 1$ $(i \neq j)$, then the (i, j)th element of $\sum_{z=0}^{l} x_{pp}^{(z)} Y^{l-z}$ is equal to 0 for $p = 1, \cdots n$ and $l \ge 2$.

(ii) If $x_{ij} = 1$ $(i \neq j)$, then the diagonal elements of $\sum_{z=0}^{l} x_{ij}^{(z)} Y^{l-z}$ are equal to 0 for $l \ge 2$.

Concerning proposition (i), we obtain $\sum_{z=0}^{l} x_{pp}^{(z)} Y^{l-z} = Y^l$ from Eq. (A.9), because for every p, $x_{pp}^{(z)} = 1$ if and only if $z = 0$. From Eq. (A.12), when $y_{ij} = 1$ ($i \neq j$), the (i, j)th element of Y^l is equal to 0 for $l \geq 2$, from which proposition (i) follows.

Concerning proposition (ii), we obtain $\sum_{z=0}^{l} x_{ij}^{(z)} Y^{l-z} = Y^{l-1}$ from Eq. (A.11), because for $i \neq j$, $x_{ij}^{(z)} = 1$ if and only if $z = 1$. From Eq. (A.10), the diagonal elements of Y^{l-1} are equal to 0 for $l \geq 2$, from which proposition (ii) follows.

■

A.3 Proof of Lemma 4.2

Let τ_k denote a random variable, which represents the number of relay nodes required for communication under the condition that the initial multiplicity of encryption is k. Then, from Eq. (4.4), the probability generating function $g_{\tau_k}(\lambda)$ for τ_k can be written to be

$$g_{\tau_k}(\lambda) = E(\lambda^{\tau_k}).$$

Since τ_k is the sum of k copies of an independent variable τ_1, the probability generating function $g_{\tau_k}(\lambda)$ is calculated as follows:

$$g_{\tau_k}(\lambda) = E(\lambda^{\tau_k}) = E(\lambda^{k\tau_1}) = E(\lambda^{\tau_1})^k = g_{\tau_1}(\lambda)^k. \tag{A.13}$$

This is derived from the relation that $E(\lambda^{X+Y}) = E(\lambda^X)E(\lambda^Y)$ when random variables X and Y are independent. In order to derive $g_{\tau_k}(\lambda)$, we consider $g_{\tau_1}(\lambda)$.

Considering the property of 3MN, we derive several conditions about $g_{\tau_1}(\lambda)$. First, since the events choosing D-Mode, E-Mode, and T-Mode are mutually exclusive, we obtain

$$g_{\tau_1}(\lambda) = E(\lambda^{\tau_1}) = E(\lambda^{\tau_1} \mid \mathrm{D}) + E(\lambda^{\tau_1} \mid \mathrm{E}) + E(\lambda^{\tau_1} \mid \mathrm{T}),$$

where $E(\lambda^{\tau_1} \mid \mathrm{X})$ represents the conditional expectation of λ^{τ_1} under the condition that X-mode is chosen in the first transition.

Next, we consider the above conditional expectations $E(\lambda^{\tau_1} \mid \mathrm{D})$, $E(\lambda^{\tau_1} \mid \mathrm{E})$, and $E(\lambda^{\tau_1} \mid \mathrm{T})$. When D-Mode, E-Mode, and T-Mode are chosen, the multiplicities of encryption become 0, 2, and 1 by one step, respectively. Therefore, the random variables of the conditional expectations in D-Mode, E-Mode, and T-Mode become 1, $1 + \tau_2$, and $1 + \tau_1$, respectively. Since the probabilities of mode selections are p_{D}, p_{E}, and p_{T}, the conditional expectations are given by

$$\begin{aligned}
E(\lambda^{\tau_1} \mid \mathrm{D}) &= p_{\mathrm{D}} E(\lambda^1) = p_{\mathrm{D}}\lambda, \\
E(\lambda^{\tau_1} \mid \mathrm{E}) &= p_{\mathrm{E}} E(\lambda^{1+\tau_2}) = p_{\mathrm{E}}\lambda E(\lambda^{\tau_2}) = p_{\mathrm{E}}\lambda (g_{\tau_1}(\lambda))^2, \\
E(\lambda^{\tau_1} \mid \mathrm{T}) &= p_{\mathrm{T}} E(\lambda^{1+\tau_1}) = p_{\mathrm{T}}\lambda E(\lambda^{\tau_1}) = p_{\mathrm{T}}\lambda (g_{\tau_1}(\lambda)),
\end{aligned}$$

respectively. From these results, the following equation is obtained.

$$g_{\tau_1}(\lambda) = p_{\mathrm{D}}\lambda + p_{\mathrm{T}}\lambda g_{\tau_1}(\lambda) + p_{\mathrm{E}}\lambda(g_{\tau_1}(\lambda))^2. \tag{A.14}$$

When $p_{\mathrm{D}} \neq 0$, we obtain

$$g_{\tau_1}(\lambda) = \frac{1 - p_{\mathrm{T}}\lambda - \sqrt{(1 - p_{\mathrm{T}}\lambda)^2 - 4p_{\mathrm{D}}p_{\mathrm{E}}\lambda^2}}{2p_{\mathrm{E}}\lambda},$$

where we use the condition that $g_{\tau_1}(\lambda)$ has to be finite when $\lambda \to 0$. As a result, we obtain

$$g_{\tau_k}(\lambda) = \left(\frac{1 - p_{\mathrm{T}}\lambda - \sqrt{(1 - p_{\mathrm{T}}\lambda)^2 - 4p_{\mathrm{D}}p_{\mathrm{E}}\lambda^2}}{2p_{\mathrm{E}}\lambda}\right)^k.$$

When $p_{\mathrm{E}} = 0$, we obtain

$$g_{\tau_1}(\lambda) = \frac{p_{\mathrm{D}}\lambda}{1 - p_{\mathrm{T}}\lambda}.$$

In this case,

$$g_{\tau_k}(\lambda) = \left(\frac{p_{\mathrm{D}}\lambda}{1 - p_{\mathrm{T}}\lambda}\right)^k.$$

■

A.4 Proof of Lemma 4.3

Let ϵ_k denote a random variable, which represents the number of encryption required for communication under the condition that the initial multiplicity of encryption is k. Since ϵ_k is the sum of k copies of an independent variable ϵ_1, the probability generating function $g_{\epsilon_k}(\lambda)$ is given as follows:

$$g_{\epsilon_k}(\lambda) = E(\lambda^{\epsilon_k}) = E(\lambda^{k\epsilon_1}) = E(\lambda^{\epsilon_1})^k = g_{\epsilon_1}(\lambda)^k. \tag{A.15}$$

In order to derive $g_{\epsilon_k}(\lambda)$, we consider $g_{\epsilon_1}(\lambda)$.

For $g_{\epsilon_1}(\lambda)$, since the events choosing D-Mode, E-Mode, and T-Mode are mutually exclusive, we obtain

$$g_{\epsilon_1}(\lambda) = E(\lambda^{\epsilon_1}) = E(\lambda^{\epsilon_1} \mid \mathrm{D}) + E(\lambda^{\epsilon_1} \mid \mathrm{E}) + E(\lambda^{\epsilon_1} \mid \mathrm{T}),$$

where $E(\lambda^{\epsilon_1} \mid \mathrm{X})$ represents the conditional expectation of λ^{ϵ_1} under the condition that X-mode is chosen in the first transition.

We consider the above conditional expectations $E(\lambda^{\epsilon_1} \mid \mathrm{D})$, $E(\lambda^{\epsilon_1} \mid \mathrm{E})$, and $E(\lambda^{\epsilon_1} \mid \mathrm{T})$. When D-Mode, E-Mode, and T-Mode are chosen, the multiplicities of encryption become 0, 2, and 1 by one step, respectively. Therefore, the random variables of the conditional

expectations in D-Mode, E-Mode, and T-Mode become 1, ϵ_2, and ϵ_1, respectively. Since the probabilities of mode selections are p_{D}, p_{E}, and p_{T}, the conditional expectations are given by

$$\begin{aligned} E(\lambda^{\epsilon_1} \mid \mathrm{D}) &= p_{\mathrm{D}} E(\lambda^1) = p_{\mathrm{D}}\lambda, \\ E(\lambda^{\epsilon_1} \mid \mathrm{E}) &= p_{\mathrm{E}} E(\lambda^{\epsilon_2}) = p_{\mathrm{E}}(g_{\epsilon_1}(\lambda))^2, \\ E(\lambda^{\epsilon_1} \mid \mathrm{T}) &= p_{\mathrm{T}} E(\lambda^{\epsilon_1}) = p_{\mathrm{T}}(g_{\epsilon_1}(\lambda)), \end{aligned}$$

respectively. From these results, the following equation is obtained.

$$g_{\epsilon_1}(\lambda) = p_{\mathrm{D}}\lambda + p_{\mathrm{T}} g_{\epsilon_1}(\lambda) + p_{\mathrm{E}}(g_{\epsilon_1}(\lambda))^2. \tag{A.16}$$

When $p_{\mathrm{E}} \neq 0$, we obtain

$$g_{\epsilon_k}(\lambda) = \frac{1 - p_{\mathrm{T}} - \sqrt{(1-p_{\mathrm{T}})^2 - 4p_{\mathrm{D}}p_{\mathrm{E}}\lambda}}{2p_{\mathrm{E}}},$$

where we use the condition that $g_{\epsilon_1}(\lambda)$ has to be finite when $\lambda \to 0$. Consequently, we obtain

$$g_{\epsilon_k}(\lambda) = \left(\frac{1 - p_{\mathrm{T}} - \sqrt{(1-p_{\mathrm{T}})^2 - 4p_{\mathrm{D}}p_{\mathrm{E}}\lambda}}{2p_{\mathrm{E}}}\right)^k.$$

When $p_{\mathrm{E}} = 0$, we obtain

$$g_{\epsilon_1}(\lambda) = \frac{p_{\mathrm{D}}\lambda}{1 - p_{\mathrm{T}}}.$$

In this case, we obtain

$$g_{\epsilon_k}(\lambda) = \left(\frac{p_{\mathrm{D}}\lambda}{1 - p_{\mathrm{T}}}\right)^k.$$

■

A.5 Proof of Theorem 4.3

First, we derive the expectation of the number of relay nodes. From Eqs. (4.5) and (A.13), the expectation M_{N} can be written by

$$M_{\mathrm{N}} = E(\tau_k) = g'_{\tau_k}(1) = \left(g^k_{\tau_1}(1)\right)' = k g^{k-1}_{\tau_1}(1) g'_{\tau_1}(1).$$

In order to calculate the above value easily, we define $f_{\mathrm{N}}(\lambda)$ as

$$f_{\mathrm{N}}(\lambda) = (1 - p_{\mathrm{T}}\lambda)^2 - 4p_{\mathrm{D}}p_{\mathrm{E}}\lambda^2. \tag{A.17}$$

Thus, we obtain

$$g_{\tau_1}(\lambda) = \frac{1 - p_T\lambda - f_N^{\frac{1}{2}}(\lambda)}{2p_E\lambda}, \tag{A.18}$$

$$f_N(\lambda) = 1 - p_T\lambda + \frac{1}{2}f_N'(\lambda)\lambda, \tag{A.19}$$

$$f_N(1) = (p_D - p_E)^2, \tag{A.20}$$

$$f_N'(1) = -2(p_D + p_E) + 2(p_D - p_E)^2. \tag{A.21}$$

From Lemma 4.2, we obtain

$$g_{\tau_1}(1) = 1. \tag{A.22}$$

With Eqs. (A.18) and (A.19), $g_{\tau_1}'(\lambda)$ is rewritten to be

$$\begin{aligned} g_{\tau_1}'(\lambda) &= \frac{1}{2\lambda^2 p_E}\Big\{f_N^{-\frac{1}{2}}(\lambda)\Big(-\frac{1}{2}f_N'(\lambda)\lambda + f_N(\lambda)\Big) - 1\Big\} \\ &= \frac{1}{2\lambda^2 p_E}\Big(f_N^{-\frac{1}{2}}(\lambda)(1 - p_T\lambda) - 1\Big) \\ &= \frac{1}{\lambda}f_N^{-\frac{1}{2}}(\lambda)\frac{1 - p_T\lambda - f_N^{\frac{1}{2}}(\lambda)}{2p_E\lambda} \\ &= \frac{1}{\lambda}f_N^{-\frac{1}{2}}(\lambda)g_{\tau_1}(\lambda). \end{aligned} \tag{A.23}$$

From Eqs. (A.20), (A.22), and (A.23), the following equation is obtained.

$$g_{\tau_1}'(1) = f_N^{-\frac{1}{2}}(1)g_{\tau_1}(1) = \frac{1}{p_D - p_E}. \tag{A.24}$$

Using Eq. (A.24), we obtain

$$M_N = kg_{\tau_1}^{k-1}(1)g_{\tau_1}'(1) = \frac{k}{p_D - p_E}.$$

Next, we derive the variance of the number of relay nodes. From Eqs. (4.6) and (A.13), the variance V_N can be written to be

$$\begin{aligned} V_N = V(\tau_k) &= g_{\tau_k}''(1) + g_{\tau_k}'(1) - \Big(g_{\tau_k}'(1)\Big)^2 \\ &= \Big(g_{\tau_1}^k(1)\Big)'' + \Big(g_{\tau_1}^k(1)\Big)' - \Big\{\Big(g_{\tau_1}^k(1)\Big)'\Big\}^2. \end{aligned}$$

The second derivative of $g_{\tau_1}^k(\lambda)$ is expressed by the following equation.

$$\Big(g_{\tau_1}^k(\lambda)\Big)'' = k\Big\{(k-1)g_{\tau_1}^{k-2}(\lambda)\Big(g_{\tau_1}'(\lambda)\Big)^2 + g_{\tau_1}^{k-1}(\lambda)g_{\tau_1}''(\lambda)\Big\}.$$

Using Eq. (A.23), $g''_{\tau_1}(\lambda)$ is calculated as follows:

$$\begin{aligned} g''_{\tau_1}(\lambda) &= \frac{d}{d\lambda}\Big(\frac{1}{\lambda} f_{\mathrm{N}}^{-\frac{1}{2}}(\lambda) g_{\tau_1}(\lambda)\Big) \\ &= -\frac{1}{\lambda^2} f_{\mathrm{N}}^{-\frac{1}{2}}(\lambda) g_{\tau_1}(\lambda) - \frac{1}{2\lambda} f_{\mathrm{N}}^{-\frac{3}{2}}(\lambda) f'_{\mathrm{N}}(\lambda) g_{\tau_1}(\lambda) + \frac{1}{\lambda} f_{\mathrm{N}}^{-\frac{1}{2}}(\lambda) g'_{\tau_1}(\lambda) \\ &= -\frac{1}{\lambda} g'_{\tau_1}(\lambda) - \frac{1}{2} g'_{\tau_1}(\lambda) f'_{\mathrm{N}}(\lambda) f_{\mathrm{N}}^{-1}(\lambda) + \frac{1}{\lambda^2} f_{\mathrm{N}}^{-1}(\lambda) g_{\tau_1}(\lambda). \end{aligned} \tag{A.25}$$

From Eqs. (A.20), (A.21), (A.22), (A.24), and (A.25), we obtain

$$g''_{\tau_1}(1) = -\frac{2}{p_{\mathrm{D}} - p_{\mathrm{E}}} + \frac{1}{(p_{\mathrm{D}} - p_{\mathrm{E}})^2} + \frac{p_{\mathrm{D}} + p_{\mathrm{E}}}{(p_{\mathrm{D}} - p_{\mathrm{E}})^3}. \tag{A.26}$$

Using Eq. (A.26) together with Eqs. (A.22) and (A.24), we obtain

$$V_{\mathrm{N}} = \frac{k\{(1 - p_{\mathrm{T}}) - (p_{\mathrm{D}} - p_{\mathrm{E}})^2\}}{(p_{\mathrm{D}} - p_{\mathrm{E}})^3}.$$

■

A.6 Proof of Theorem 4.4

First, we derive the expectation of the number of encryption. From Eqs. (4.5) and (A.15), the expectation M_{E} can be written to be

$$M_{\mathrm{E}} = E(\epsilon_k) = g'_{\epsilon_k}(1) = \left(g^k_{\epsilon_1}(1)\right)' = k g^{k-1}_{\epsilon_1}(1) g'_{\epsilon_1}(1).$$

In order to calculate the above value easily, we define $f_{\mathrm{E}}(\lambda)$ as

$$f_{\mathrm{E}}(\lambda) = (1 - p_{\mathrm{T}})^2 - 4 p_{\mathrm{D}} p_{\mathrm{E}} \lambda. \tag{A.27}$$

Thus, we obtain

$$g_{\epsilon_1}(\lambda) = \frac{1 - p_{\mathrm{T}} - f_{\mathrm{E}}^{\frac{1}{2}}(\lambda)}{2 p_{\mathrm{E}}}, \tag{A.28}$$

$$f_{\mathrm{E}}(1) = (p_{\mathrm{D}} - p_{\mathrm{E}})^2, \tag{A.29}$$

$$f'_{\mathrm{E}}(\lambda) = f'_{\mathrm{E}}(1) = -4 p_{\mathrm{D}} p_{\mathrm{E}}. \tag{A.30}$$

From Lemma 4.3, we obtain

$$g_{\epsilon_1}(1) = 1. \tag{A.31}$$

Using Eqs. (A.28) and (A.30), $g'_{\epsilon_1}(\lambda)$ is calculated as follows:

$$g'_{\epsilon_1}(\lambda) = -\frac{1}{4 p_{\mathrm{E}}} f_{\mathrm{E}}^{-\frac{1}{2}}(\lambda) f'_{\mathrm{E}}(\lambda) = p_{\mathrm{D}} f_{\mathrm{E}}^{-\frac{1}{2}}(\lambda). \tag{A.32}$$

From Eqs. (A.29) and (A.32), the following equation is obtained.

$$g'_{\epsilon_1}(1) = \frac{p_D}{p_D - p_E}. \tag{A.33}$$

Using Eqs. (A.31) and (A.33), we obtain

$$M_E = k g^{k-1}_{\epsilon_1}(1) g'_{\epsilon_1}(1) = \frac{k}{1 - \frac{p_E}{p_D}}.$$

Next, we consider the variance of the number of encryption. From Eqs. (4.6) and (A.15), the variance V_E can be written to be

$$\begin{aligned} V_E = V(\epsilon_k) &= g''_{\epsilon_k}(1) + g'_{\epsilon_k}(1) - \left(g'_{\epsilon_k}(1)\right)^2 \\ &= \left(g^k_{\epsilon_1}(1)\right)'' + \left(g^k_{\epsilon_1}(1)\right)' - \left\{\left(g^k_{\epsilon_1}(1)\right)'\right\}^2. \end{aligned}$$

The second derivative of $g^k_{\epsilon_1}(\lambda)$ is expressed by the following equation.

$$\left(g^k_{\epsilon_1}(\lambda)\right)'' = k(k-1) g^{k-2}_{\epsilon_1}(\lambda)\left(g'_{\epsilon_1}(\lambda)\right)^2 + k g^{k-1}_{\epsilon_1}(\lambda) g''_{\epsilon_1}(\lambda).$$

From Eqs. (A.28), (A.30), and (A.32), $g''_{\epsilon_1}(\lambda)$ is calculated as follows:

$$g''_{\epsilon_1}(\lambda) = -\frac{1}{2} p_D f_E^{-\frac{3}{2}}(\lambda) f'_E(\lambda) = 2 {p_D}^2 p_E f_E^{-\frac{3}{2}}(\lambda). \tag{A.34}$$

From Eqs. (A.29) and (A.34), we obtain

$$g''_{\epsilon_1}(1) = \frac{2 {p_D}^2 p_E}{(p_D - p_E)^3}. \tag{A.35}$$

Using Eq. (A.35) together with Eqs. (A.31) and (A.33), we obtain

$$V_E = \frac{k \frac{p_E}{p_D}\left(1 + \frac{p_E}{p_D}\right)}{\left(1 - \frac{p_E}{p_D}\right)^3}.$$

■

A.7 Derivation of Equation (5.3)

Let τ_k denote a random variable representing the number of relay nodes required for communication under the condition that the initial multiplicity of encryption is k, and $P_k(i)$ denote the probability that the number of relay nodes required for communication is equal to i, that is, $P_k(i) = P(\tau_k = i)$. If the first collaborating node on the communication path appears at the i-th node on the path, the number of relay nodes required for communication is larger than

or equal to i. Therefore, the probability $P(L_i)$ that the first collaborating node on the path appears at the i-th node on the path is given by

$$\begin{aligned} P(L_i) &= \sum_{j=i}^{\infty} P_k(j) \left(\frac{n_t - n_c}{n_t}\right)^{i-1} \left(\frac{n_c}{n_t}\right) \\ &= \left\{1 - \sum_{j=0}^{i-1} P_k(j)\right\} \lambda^{i-1}(1-\lambda), \end{aligned}$$

where $\lambda = (n_t - n_c)/n_t$. As a result, $P(L_{1+})$ is calculated as follows:

$$\begin{aligned} P(L_{1+}) &= \sum_{i=1}^{\infty} P(L_i) = (1-\lambda) \sum_{i=1}^{\infty} \left\{1 - \sum_{j=0}^{i-1} P_k(j)\right\} \lambda^{i-1} \\ &= 1 - (1-\lambda) \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} P_k(j) \lambda^{i-1} \\ &= 1 - \sum_{j=0}^{\infty} P_k(j) \lambda^j = 1 - g_{\tau_k}(\lambda), \end{aligned} \tag{A.36}$$

where the probability generating function $g_{\tau_k}(\lambda)$ is given by Eq. (4.7) and we use the commutativity of summations, which is guaranteed by the absolute convergence of double series in Eq. (A.36). ■

Publications

A. Journal Papers

1. K. Kono, Y. Ito, A. Aoyama, H. Kamoda, and N. Babaguchi, "Matrix-Based Algorithm for Integrating Inheritance Relations of Access Rights for Policy Generation," *Journal of Information Processing*, Vol. 17, pp. 318–327, December 2009.

2. K. Kono, S. Nakano, Y. Ito, and N. Babaguchi, "A Consideration on the Numbers of Relay Nodes and Encryption Required for Anonymous Communication System 3-Mode Net," *Journal of Information Assurance and Security* (to appear in 2010).

B. International Conference Papers

1. K. Kono, S Nakano, Y. Ito, and N. Babaguchi, "Security Analysis of Anonymous Communication System 3-Mode Net Against Collaborating Nodes," in *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2009)*, pp. 105–110, October 2009.

2. K. Kono, S. Nakano, Y. Ito, and N. Babaguchi, "Performance Analysis of Anonymous Communication System 3-Mode Net," in *Proc. 5th IEEE International Conference on Information Assurance and Security (IAS 2009)*, pp. 593–596, August 2009.

C. Technical Reports

1. S. Nakano, K. Kono, Y. Ito, and N. Babaguchi, "Reduction of the Number of Relay Nodes in Anonymous Communication System 3-Mode Net," in *Proc. 2009 Institute of Electronics, Information and Communication Engineers (IEICE) General Conference*, A-7-8, p. 182, March 2009 (in Japanese).

2. T. Hanaoka, K. Kono, Y. Ito, and N. Babaguchi, "Automated Trust Negotiation among Three Agent Based on Service Usage of Client," in *Proc. 2008 Institute of Electronics, Information and Communication Engineers (IEICE) General Conference*, A-7-10, p. 182, March 2008 (in Japanese).

3. K. Kono, Y. Ito, A. Aoyama, H, Kamoda, and N. Babaguchi, "An Integration Method of Access Control Policies Using Adjacency Matrix," in *The Special Interest Group Notes of Information Processing Society of Japan (IPSJ)*, pp. 45–50, May 2007 (in Japanese).

4. H. Kamoda, K. Kono, Y. Ito, and N. Babaguchi, “Access Control Policy Inconsistency Check Using Model Checker,” in *The Special Interest Group Notes of Information Processing Society of Japan (IPSJ)*, pp. 441–446, March 2007 (in Japanese).

5. R. Tanihira, K. Kono, Y. Ito, and N. Babaguchi, “Fair Automated Trust Negotiation with Credential Disclosure Points,” in *Proc. 2007 Institute of Electronics, Information and Communication Engineers (IEICE) General Conference*, A-7-10, p. 214, March 2007 (in Japanese).

6. K. Kono, Y. Ito, and N. Babaguchi, “A Study on RSA Cryptosystem Using Pseudo-primes,” in *Proc. 2005 Society Conference of Institute of Electronics, Information and Communication Engineers (IEICE)*, A-7-2, p. 175, September 2005 (in Japanese).

D. Award List

1. Best Paper Award
   K. Kono, S. Nakano, Y. Ito, and N. Babaguchi, “Performance Analysis of Anonymous Communication System 3-Mode Net,” in *Proc. 5th IEEE International Conference on Information Assurance and Security (IAS 2009)*, pp. 593–596, August 2009.