Journal of Information Assurance and Security
ISSN 1554-1010 Volume 6 (2011) pp. 124-131
© MIR Labs, www.mirlabs.net/jias/index.html

Anonymous Communication System Based on Multiple Loopbacks

Kazuhiro Kono[†], **Shinnosuke Nakano**[††1], **Yoshimichi Ito**[††2], **and Noboru Babaguchi**[††3]

[†]Faculty of Safety Science, Kansai University
7-1 Hakubai-cho, Takatsuki, Osaka, Japan
k-kono@kansai-u.ac.jp

[††]Graduate School of Engineering, Osaka University
2-1 Yamadaoka, Suita, Osaka, Japan
[1]*nakano@nanase.comm.eng.osaka-u.ac.jp*, {[2]*ito*, [3]*babaguchi*}*@comm.eng.osaka-u.ac.jp*

Abstract: **This paper proposes a new anonymous communication system based on multiple loopbacks, which uses probabilistic choice of actions. Our system provides both sender anonymity and receiver anonymity. Our system also decreases the computation load of each relay node, because there exist no multiple-encryption process in our system. Applying an analysis method in an anonymous communication system called 3-Mode Net, we evaluate the number of relay nodes required for communication and sender anonymity. In addition, we evaluate receiver anonymity by using a probability generating function and its properties. From these results, we investigate the relationship between the number of relay nodes and anonymity.**

Keywords: anonymous communication, 3-Mode Net, multiple loopbacks, performance analysis.

I. Introduction

As Information Technology (IT) has developed rapidly, we use various IT systems and services. In the deep penetration of IT into our lives, one of the important issues is to provide services where high anonymity is required such as medical consultation and whistle-blowing on the Internet. This is because anonymity is not guaranteed on the Internet, although we can protect data in a communication by using encryption protocols.

In order to provide anonymity to a sender and a receiver in a communication, several anonymous communication systems, which hide the identities of the sender and the receiver, have been proposed in the past. In general, these systems provide anonymity by forwarding a message from its sender to its receiver through several relay nodes. As forwarding methods, there are three ways as follows:

1. multiple encryption of a message [1, 2],
2. probabilistic choice of actions in relay nodes [3, 4],
3. cyclic routes [5].

For example, two well-known anonymous communication systems Onion Routing [1] and Crowds [3] provide anonymity by multiple encryption and probabilistic choice of actions, respectively. An anonymous communication system using cyclic routes is also proposed in [5], which uses elementary cyclic routes.

These three systems have disadvantages: in Onion Routing, the computation load of relay nodes by encryption/decryption of a message is very large; Crowds does not provide receiver anonymity; in the anonymous communication system using elementary cyclic routes, network topology is restricted. In addition, Onion Routing has an disadvantage that the size of transmitted data is larger than those of the other systems because the size of the data changes by encryption/decryption.

Anonymous communication systems using multiple encryption or cyclic routes never overcome the above shortcomings. For example, an anonymous communication system called 3-Mode Net (3MN) [6], which can be regarded as an extension of Crowds-based anonymous communication systems, enables us to provide receiver anonymity unlike Crowds by introducing multiple encryption of a message. In contrast, 3MN has a shortcoming that the computation load of relay nodes is larger than that of Crowds. Therefore, it is desirable to develop an anonymous communication system using probabilistic choice of actions only.

In this paper, we propose a new anonymous communication system using probabilistic choice of actions and multiple loopbacks. We introduce loopbacks that mean that a message which a node transmits returns to itself through several relay nodes. When a relay node chooses the action of loopbacks, the relay node does the following actions: first, changes the destination of its message to itself; second, forwards the message to another destination. The actions about loopbacks enable us to avoid the situation where the destination of a message indicates the proper receiver of the message. Compared to 3MN, our method decreases the computation load of relay nodes because of no encryption/decryption of a message.

We also analyze the performance of our method. First, we evaluate the number of relay nodes required for communication and sender anonymity because it is shown that a method for analyzing the performance of 3MN in [7, 8] is applicable to that of our method. Next, we evaluate receiver anonymity by using a probability generating function and its properties. From these results, we investigate the relationship between the number of relay nodes and anonymity.

Some of the results described in this paper have been reported in [9]. The main contribution of this paper is to introduce receiver anonymity against collaborating nodes as a measure for evaluating our method and to analyze our method from the viewpoints of the number of relay nodes, sender anonymity, and receiver anonymity.

This paper is organized as follows. Section II briefly presents overviews of Crowds and 3-Mode Net, which are anonymous communication systems where each relay node decides its action with predefined probabilities. In Section III, we propose an anonymous communication system using probabilistic choice of actions and multiple loopbacks. In Section IV, we analyze the performance of our proposed system. In Section V, we consider the influence of the probabilities of mode choice on the performance of our system through numerical examples. We conclude this paper in Section VI.

II. Existing Anonymous Communication Systems Based on Probabilistic Choice of Actions

Existing anonymous communication systems are almost regarded as communication systems, which forward a data set from its sender to its 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 an encrypted message. Since sender anonymity and receiver anonymity depend on the ways of forwarding and creating a data set, several anonymous communication systems have been proposed. In this section, we describe two anonymous communication systems Crowds [3] and 3-Mode Net [6], which provide anonymity by the probabilistic choice of actions.

A. Overview of Crowds

In Crowds, sender S first prepares a data set that consists of the address of a proper receiver R and an encrypted message. Next, S transmits the created data set to another node except for S. When a relay node receives a data set, the node chooses its action with predefined probabilities whether the node sends the received data set to R or another node. Finally, the proper receiver R receives a data set by choosing an action that a relay node sends the data set to R, and the transmission of the message finishes.

Crowds provides sender anonymity because each relay node cannot distinguish whether the immediate predecessor of the node is a message sender or one of relay nodes. Crowds also has an advantage that the computation load of relay nodes is very small because of no encryption process except for encryption of a message that a sender performs. Crowds does not, however, hide the identity of a proper receiver because the destination of a data set always indicates the receiver.

B. Overview of 3-Mode Net

3-Mode Net (3MN) is one of the anonymous communication systems where a sender forwards a data set to a proper receiver with encryption and decryption. In 3MN, each relay node chooses one of three modes as shown in Fig. 1 randomly with predefined probabilities. We, here, refer to the data set as the set of data composed of a multiple-encrypted message and the address of the next destination.

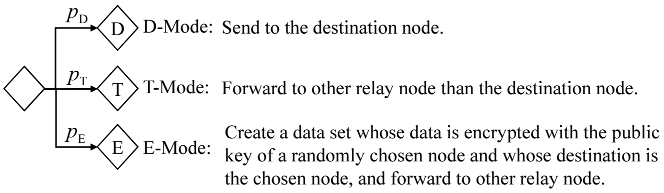

p_D
D
D-Mode: Send to the destination node.
p_T
T
T-Mode: Forward to other relay node than the destination node.
p_E
E
E-Mode: Create a data set whose data is encrypted with the public key of a randomly chosen node and whose destination is the chosen node, and forward to other relay node.

Figure. 1: Actions of a node in 3-Mode Net.

1) Three Modes in 3-Mode Net

In Fig. 1, Decryption Mode (D-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 its decryption key, and produces a new data set.

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

Finally, Encryption Mode (E-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.

The destination of a data set does not always indicate the proper receiver of a message because of the existence of E-Mode, and thus, each relay node cannot judge whether the destination of a received data set indicates a proper receiver or one of relay nodes. 3MN, therefore, guarantees receiver anonymity. This makes sharp contrast with the case of Crowds. Sender anonymity is also provided because no relay node understands whether the immediate predecessor of the node is the sender of a message or not, like Crowds.

2) Issues on 3-Mode Net

3MN has a disadvantage that the computation load of relay nodes is larger than that of Crowds owing to the existence of multiple-encryption, although 3MN has a big advantage that receiver anonymity is provided. 3MN also takes more time required for communication than Crowds by the processes of encryption and decryption when the number of relay nodes required for communication in 3MN is equal to that of Crowds.

When a relay node chooses E-Mode, the size of the created data set is larger than that of a received data set because a new destination is added, and thus, the size of a data set is not constant. Compared to Crowds where a data set consists of the destination of a proper receiver R and an encrypted message, there exists a shortcoming that the size of a data set on 3MN network becomes large.

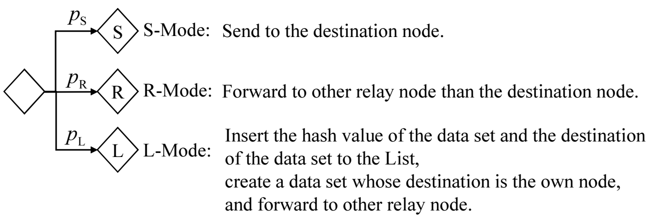

p_S
S
S-Mode: Send to the destination node.
p_R
R
R-Mode: Forward to other relay node than the destination node.
p_L
L
L-Mode: Insert the hash value of the data set and the destination of the data set to the List, create a data set whose destination is the own node, and forward to other relay node.

Figure. 2: Actions of a node in our proposed system.

III. Anonymous Communication System Using Probabilistic Choice of Actions and Multiple Loopbacks

In Section III, we propose a new anonymous communication system which enables us to provide sender anonymity and receiver anonymity using only the probabilistic choice of actions in relay nodes without multiple-encryption in order to overcome disadvantages of Crowds and 3-Mode Net. In our system, it is important how the identity of a proper receiver is protected without encryption because our system is based on the framework of Crowds like 3MN. In the next subsection, we proceed to introduce a "list".

A. List

Routers and relay servers can record transmission and reception histories. We here define a record called "list".

A list is a database composed of the hash value of an encrypted message and next destination in a data set. A data set consists of an encrypted message and next destination, like Crowds. For example, assume that a node receives a data set whose destination is node A. In this case, the node can record the hash value of an encrypted message and next destination A in the data set on its own list.

It is general to use the record of data like a list in other anonymous communication systems. For example, relay nodes in Onion Routing and Crowds record previous and next nodes in order to reply from a receiver to a sender [1, 3]. Throughout this paper, we assume that each node has its own list.

B. Three Modes in Our Proposed System

Like 3MN, our proposed anonymous communication system has three modes as shown in Fig. 2, Straight Mode (S-Mode), Relay Mode (R-Mode), and Loopback Mode (L-Mode). Each relay node chooses one of the three modes randomly with predefined probabilities.

In Fig. 2, the first mode is the mode where a node sends a received data set to its destination directly. Since a data set is transmitted to its indicated destination, this mode is called Straight Mode (S-Mode).

The second mode is the mode where a node forwards a received data set to a node other than the destination node. Since a data set is forwarded to another node, this mode is called Relay Mode (R-Mode).

In the third mode, a relay node performs the following three processes: 1) first, insert the hash value of the data set and the destination of the data set to its own list; 2)second, create a data set whose destination is the own node; 3)finally, forward the data set to another node. When a relay node chooses the third mode, the relay node forwards the data set to itself because the destination of the data set is changed to the address of its own node. That is, the relay node performs the action of "loopback". Therefore, the third mode is called Loopback Mode (L-Mode).

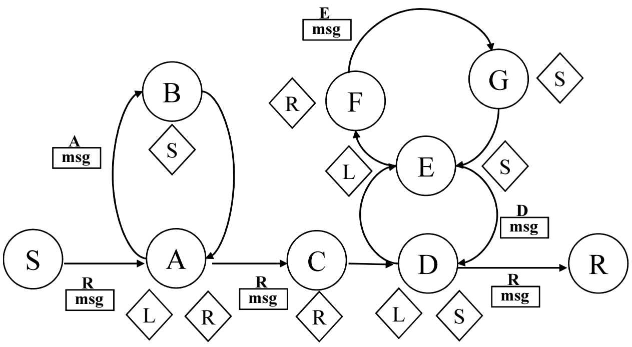


Figure. 3: Actions of a node in our proposed system.

Because of the existence of L-Mode, the destination of a data set is changed, and thus, a relay node cannot understand whether the destination of a data set indicates a proper receiver or not. That is, our proposed system guarantees receiver anonymity. This is similar to the case of 3MN. In addition, a relay node cannot judge whether the immediate predecessor of the node indicates a proper sender or one of relay nodes. Our method also provides sender anonymity like Crowds and 3MN.

As shown in Fig. 2, there exist no encryption and decryption processes in our proposed system although anonymity in our system is similar to the case of 3MN. This is sharp contrast to 3MN. Therefore, we show that, in our system, the computation load of relay nodes is very small like Crowds. Furthermore, the size of a data set in our system is equal to that of Crowds because each data set always consists of an encrypted message and the destination of a node. This implies that the size of a data set is smaller than that of 3MN.

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

C. Behavior of Our Proposed System

We describe the behavior of our proposed system with Fig. 3, where data in square frames means an encrypted message and the letters on square frames indicate the next destinations. We refer to the set of an encrypted message and the next destination as a data set. The letters in diamond frames also indicate chosen modes in relay nodes.

Sender S first creates a data set $\mathrm{R}||K_{\mathrm{R}}(\mathrm{msg})$ that 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). After that, S forwards the data set to another node A.

When a relay node has received a data set, the node first checks its destination. If the destination corresponds to the relay node, the node searches a hash value from its list, which is the same as the hash value of an encrypted message in the received data set. Since the node inserts the hash value of the encrypted message to its own list in the case when L-Mode is chosen, the hash value of the encrypted message in the received data set corresponds to one of the hash values in its list (if the hash value of the encrypted message corresponds to no hash value of its list, the relay node indicates the final receiver as discussed below). After searching the hash value, the relay node produces a new data set whose destination indicates the address of a node paired with the same hash value, and chooses one mode randomly with predefined probabilities. Otherwise, the node only chooses one mode randomly with predefined probabilities.

When node A receives the data set, A chooses one mode randomly with predefined probabilities because the destination of the data set does not indicate A. In this example, A chooses L-Mode. Thus, A enters the hash value of the encrypted message and the destination R in A's list, creates a new data set $\mathrm{A}||K_{\mathrm{R}}(\mathrm{msg})$ whose destination is its own node, and forwards the data set to another node B. After B receives the data set, B also chooses one mode randomly with predefined probabilities because the destination of the data set is not B but A. In Fig. 3, B chooses S-Mode, and thus, B transmits the data set to A, which indicates its destination.

When node A receives the data set again, A first checks its destination, and A understands that the destination of the data set indicates itself. Then, A checks the hash value of an encrypted message $K_{\mathrm{R}}(\mathrm{msg})$ in the received data set against A's list, and changes the destination of the data set to R. After that, A chooses and performs one of three modes randomly with predefined probabilities. In this case, A chooses R-Mode, and A forwards the data set to another node C.

In a similar fashion, node C and the following nodes forward a data set with replace of destinations of data sets. Finally, the proper receiver R receives a data set $\mathrm{R}||K_{\mathrm{R}}(\mathrm{msg})$. R first checks its destination and confirms that the destination indicates R and the hash value of an encrypted message corresponds to no hash value in R's list. Therefore, R recognizes that the proper destination of the data set is itself. R then acquires the message msg by decrypting the data set $\mathrm{R}||K_{\mathrm{R}}(\mathrm{msg})$. The transmission of the message finishes.

IV. Performance Analysis

In this section, we analyze the performance of our system. We first model our system, and investigate the relationship between our system and 3-Mode Net. We then analyze the number of relay nodes required for communication and sender anonymity by applying an analysis method in 3MN. We also define and derive a measure for receiver anonymity.

A. Modeling

In our system, loopbacks occur repeatedly as shown in Fig. 3. We here define a value called the multiplicity of loopbacks, which indicates the difference of the number of times when L-Mode is chosen and the number of times when S-Mode is chosen. We set the initial number of the multiplicity of loopbacks to 1. We also refer to this initial number as the initial multiplicity of loopbacks.

For example, in Fig. 3, the multiplicity of loopbacks in node A is equal to 2 when A receives a data set from node S and chooses L-Mode, and the multiplicity of loopbacks in A is equal to 1 when A receives a data set from node B and chooses R-Mode. In node F, the multiplicity of loopbacks is equal to 3. The multiplicity of loopbacks indicates 0 in the proper receiver R.

When S-Mode, L-Mode, or R-Mode is chosen, the multiplicity of loopbacks decreases by one, increases by one, and remains unchanged, respectively. Thus the behavior of our proposed method is modeled by a random walk, because the multiplicity of loopbacks for a data set changes in a probabilistic manner. A random walk is defined as a stochastic process on a set of integers, which starts at the origin and moves one step on the positive or negative direction with predefined probabilities independent of its location. As seen from such a viewpoint, the behavior of our system is regarded as the following stochastic process.

Modeling of our proposed system: our proposed system is regarded as a random walk on the integers which starts at a position 1 and at each point, moves one step to the negative direction with probability p_{S}, moves one step to the positive direction with probability p_{L}, or stays on its position with probability p_{R}. Once the walk arrives at the origin, i.e., when the multiplicity of loopbacks is equal to 0, the walk finishes.

B. Relationship and Comparison with 3-Mode Net

In [7, 8], the performance analysis of 3-Mode Net is discussed based on the multiplicity of encryption. From Fig. 1, it is shown that when D-Mode, E-Mode, or T-Mode is chosen, the multiplicity of encryption decreases by one, increases by one, and remains unchanged, respectively. That is, the behavior of 3MN can also be modeled by a random walk as presented above.

We here investigate relationships between the change of the multiplicity of loopbacks in our system and the change of the multiplicity of encryption in 3MN. This is because the behaviors of both 3MN and our system are discussed from the viewpoint of multiplicity. The corresponding table is shown in Table 1 (k represents the initial number of multiplicity of encryption). Table 1 means that our system has the same structure as 3MN in the case where p_{D}, p_{E}, p_{T}, and k in 3MN are equal to p_{S}, p_{L}, p_{R}, and 1 in our system, respectively. Therefore, we analyze our system in an exactly similar way as an analysis method of 3MN.

We also compare our system with Crowds and 3MN. As shown in Table 2, the computation load of each relay node is smaller than that of 3MN because our system needs no encryption/decryption processes although our system and 3MN have the same structure. Further, the size of a data set in our system is smaller than that of 3MN and is equal to that of Crowds because the size of a data set always consists of the address of next destination and an encrypted message. Thus, our system can decrease network traffic. Consequently, our system only inherits above merits of Crowds and a merit of 3MN that sender/receiver anonymity is provided, and thus, it is said that our system is superior to 3MN and Crowds.

Table 1: Relationship between 3-Mode Net and our system.

	Probabilities of mode choice			Initial multiplicity
3-Mode Net	p_{D}	p_{E}	p_{T}	k
Our system	p_{S}	p_{L}	p_{R}	1

Table 2: Comparison among Crowds, 3-Mode Net, and our system.

	Crowds	3-Mode Net	Our system
Sender anonymity	Yes	Yes	Yes
Receiver anonymity	No	Yes	Yes
The number of encryption	None	High	None
The computation load	Very small	Large	Small
The size of a data set	Unchanged	Changed	Unchanged
Network traffic	Small	Large	Small
Storage	Need	None	Need

C. The Number of Relay Nodes

Applying the method in [7], we can obtain the following theorems which give the probability distribution, the expectation, and the variance of the number of relay nodes in our proposed system. We define s, l, and r as the numbers of which S-Mode, L-Mode, and R-Mode are chosen, respectively.

Theorem 1 *Let N denote a random variable representing the number of relay nods required for communication. Then the probability distribution $P(N = x)$ is given by*

$$P(N = x) = \sum_{r \in I(x)} \frac{1}{x} \frac{x!}{s!l!r!} {p_{\mathrm{S}}}^{s} {p_{\mathrm{L}}}^{l} {p_{\mathrm{R}}}^{r},$$

where $s = (x - r + 1)/2$, $l = (x - r - 1)/2$, and $I(x)$ is a set of integers defined as

$$I(x) \quad = \quad \{r | 0 \leqq r \leqq x - 1\ ,\ r \equiv x - 1 \pmod 2\}.$$

Theorem 2 *The expectation M_{N} and the variance V_{N} of the number of relay nodes required for communication are given by*

$$M_{\mathrm{N}} = \frac{1}{p_{\mathrm{S}} - p_{\mathrm{L}}}, \quad V_{\mathrm{N}} = \frac{(1 - p_{\mathrm{R}}) - (p_{\mathrm{S}} - p_{\mathrm{L}})^2}{(p_{\mathrm{S}} - p_{\mathrm{L}})^3},$$

respectively.

Theorem 2 indicates that the expectation of the number of relay nodes depends on the differences of p_{S} and p_{L}. We also observe that its variance can be controlled without changing its expectation by adjusting p_{R}. From a practical perspective, it is not desirable to set its variance to be large because the number of relay nodes may become extremely large. Consequently, we set p_{R} to be large in order to keep the expectation unchanged and to reduce the possibility that the number of relay nodes becomes extremely large.

D. Sender Anonymity

We evaluate sender anonymity against collaborating nodes who collude with each other in order to identify a message sender. The measure of sender anonymity is defined as 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 sender under the condition that a collaborating node receives a data set [3, 10, 11].

Let H_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 (note that the 0-th node indicates a message sender), and define $H_{i+} = H_i \vee H_{i+1} \vee H_{i+2} \vee \cdots$. Also, let I denote the event where the first immediate predecessor among the immediate predecessors on the communication path is a message sender.

We consider the conditional probability $P(I \,|\, H_{1+})$. In 3MN, $P(I \,|\, H_{1+})$ is derived from a probability generating function [8]. Using its method, we obtain the following theorem that concerns the probability of the message sender in our proposed system.

Theorem 3 *Let n_{t} and n_{c} denote the number of all nodes and that of collaborating nodes in our proposed system. Then, the conditional probability $P(I \mid H_{1+})$ is given by*

$$P(I \,|\, H_{1+}) = \frac{(n_t - n_c)(n_c + 1) - n_t \times g_{\tau_1}\left(\dfrac{n_t - n_c}{n_t}\right)}{n_t(n_t - n_c)\left\{1 - g_{\tau_1}\left(\dfrac{n_t - n_c}{n_t}\right)\right\}}, \quad (1)$$

where $g_{\tau_1}(\lambda)$ is a probability generating function for a random variable τ_1 representing the number of relay nodes under the condition that the initial multiplicity of loopbacks is 1, defined by

$$g_{\tau_1}(\lambda) = \frac{1 - p_{\mathrm{R}}\lambda - \sqrt{(1 - p_{\mathrm{R}}\lambda)^2 - 4p_{\mathrm{S}}p_{\mathrm{L}}\lambda^2}}{2p_{\mathrm{L}}\lambda}. \quad (2)$$

E. Receiver Anonymity

At the end of this section, we evaluate receiver anonymity against collaborating nodes who collude with each other in order to identify a message receiver. The measure of receiver anonymity is defined as the probability of the message receiver that means that the destination of a data set which the last collaborating node among all the collaborating nodes on the communication path receives is indeed a message receiver under the condition that a collaborating node receives a data set.

Note that it is highly possibility that the destination of a data set is indeed its proper receiver when a relay node receives the data set compared to any other nodes. As a typical example, in Crowds, it is shown that the destination of a data set always indicates its receiver although relay nodes forwards a data set to another node except for its receiver.

Let L_i $(i \geqq 1)$ denote the event where the last 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 destination of a data set which the last collaborating node among all the collaborating nodes on the communication path receives is a message receiver.

We consider the conditional probability $P(J \mid L_{1+})$ that the destination of a data set which the last collaborating node among all the collaborating nodes on the communication path receives is indeed a message receiver under the condition that a collaborating node receives a data set. Similar to the case of the derivation of sender anonymity, we use a probability generating function and its properties [12]. Using the function, we obtain the following theorem that concerns the probability of the message receiver.

Theorem 4 *Let n_{t} and n_{c} denote the number of all nodes and that of collaborating nodes in our proposed system. Then, the conditional probability $P(J \mid L_{1+})$ is given by*

$$P(J \,|\, L_{1+}) = \frac{n_c(n_t - n_c - 1) \times g_{\tau_1}\left(\dfrac{n_t - n_c}{n_t}\right)}{(n_t - n_c)^2 p_{\mathrm{S}}\left\{1 - g_{\tau_1}\left(\dfrac{n_t - n_c}{n_t}\right)\right\}} + \frac{1}{n_t - n_c}, \quad (3)$$

where $g_{\tau_1}(\lambda)$ is a probability generating function for a random variable τ_1 representing the number of relay nodes under the condition that the initial multiplicity of loopbacks is 1, defined by (2).

Proof of Theorem 4: The conditional probability $P(J \mid L_{1+})$ is obtained by the following equation:

$$\begin{aligned} P(J\,|\,L_{1+}) &= \frac{P(J \wedge L_{1+})}{P(L_{1+})} \\ &= \frac{P(J \wedge Z_{1+,1}) + P(J \wedge Z_{1+,2+})}{P(L_{1+})} \\ &= \frac{P(J\,|\,Z_{1+,1})P(Z_{1+,1}) + P(J\,|\,Z_{1+,2+})P(Z_{1+,2+})}{P(L_{1+})}, \end{aligned} \tag{4}$$

where $Z_{i,j}$ is an event where the multiplicity of loopbacks is equal to j when the last collaborating node on the communication path appears at the i-th node on the path and we define $Z_{i,j+} = Z_{i,j} \vee Z_{i,j+1} \vee Z_{i,j+2} \vee \cdots$. In the second quality, we use $L_{1+} = Z_{1+,1+} = Z_{1+,1} \vee Z_{1+,2+}$. Note also that $P(J \mid Z_{1+,1}) = 1$, $P(J \mid Z_{1+,2+}) = 1/(n_t - n_c)$, and $P(L_{1+}) = P(H_{1+}) = 1 - g_{\tau_1}((n_t - n_c)/n_t)$.

In order to calculate Eq. (4), we have to compute $P(Z_{1+,1})$. This value is calculated as follows:

$$P(Z_{1+,1}) \quad = \quad \frac{n_c}{(n_t - n_c)p_{\mathrm{S}}} \times g_{\tau_1}\left(\frac{n_t - n_c}{n_t}\right). \tag{5}$$

The derivation of this equation is given by Appendix A.

From Eqs. (4) and (5), $P(J\,|\,L_{1+})$ is computed as follows:

$$P(J\,|\,L_{1+}) = \frac{n_c(n_t - n_c - 1) \times g_{\tau_1}\left(\frac{n_t - n_c}{n_t}\right)}{(n_t - n_c)^2 p_{\mathrm{S}} \left\{1 - g_{\tau_1}\left(\frac{n_t - n_c}{n_t}\right)\right\}} + \frac{1}{n_t - n_c}.$$

This completes the proof. ■

Remark 1 *It should be noted that $P(L_{1+}) = P(H_{1+})$. L_{1+} indicates an event where the last collaborating node appears on the communication path, and H_{1+} indicates an event where the first collaborating node appears on the communication path. Since both events imply that one or more collaborating nodes appears on the communication path, L_{1+} and H_{1+} mean that a collaborating node appears on the communication path. From [8], $P(H_{1+})$ is given as follows:*

$$P(H_{1+}) = 1 - g_{\tau_1}\left(\frac{n_t - n_c}{n_t}\right).$$

V. Numerical Examples

In this section, we consider the impacts of the probabilities of mode choice through numerical examples. In order to understand the influence of the probabilities of mode choice on sender anonymity and receiver anonymity easily, we illustrate sender anonymity and receiver anonymity under the condition that $n_t = 20$ and $n_c = 3$.

Fig. 4 and Fig. 5 show sender anonymity and receiver anonymity under the various probabilities of mode choice in the region satisfying $0 < p_{\mathrm{S}} < 1$, $0 < p_{\mathrm{L}} < 1$, $0 < p_{\mathrm{S}} + p_{\mathrm{L}} < 1$, and $p_{\mathrm{S}} > p_{\mathrm{L}}$, respectively. Fig. 4 shows that we set p_{S} to be small and p_{L} to be large in order to provide high sender anonymity. Fig. 5 also shows that we set p_{S} to be small and p_{L} to be large in order to provide high receiver anonymity. Therefore, we conclude that we set p_{S} to be small and p_{L} to be large in order to guarantee both sender anonymity and receiver anonymity.

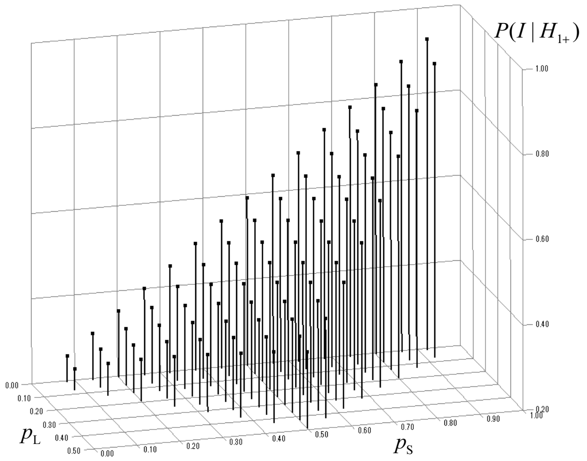


Figure. 4: Sender anonymity under various probabilities of mode choice.

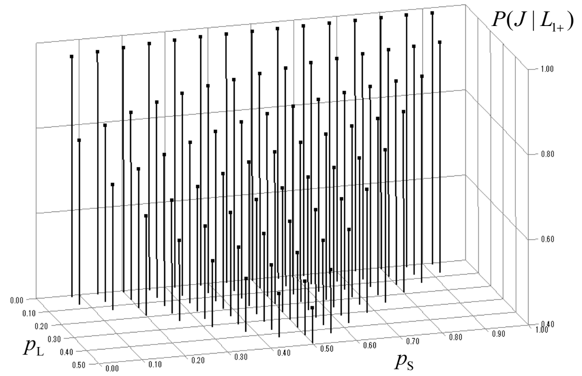


Figure. 5: Receiver anonymity under various probabilities of mode choice.

In contrast, from Theorem 2, the expectation of the number of relay nodes becomes large when we set p_{S} to be small and p_{L} to be large. There is a performance trade-off between anonymity and the expectation of the number of relay nodes required for communication.

From Fig. 6, which shows a relationship between sender anonymity and p_{R} under the condition that the difference of p_{S} and p_{L} is constant, it is also shown that sender anonymity is guaranteed in the case where p_{R} increases when the difference between p_{S} and p_{L} is constant. However, from Fig. 7, which shows a relationship between receiver anonymity and p_{R} under the condition that the difference of p_{S} and p_{L} is constant, we observe that receiver anonymity is not guaranteed in the case where p_{R} increases when the difference between p_{S} and p_{L} is constant. From Theorem 2, we conclude that, in the case where we reduce the variance of the number of relay nodes under the condition that the expectation of the number of relay nodes is constant, receiver anonymity is lost although we keep high sender anonymity.

As discussed above, we cannot select the probabilities of mode choice so that the number of relay nodes required

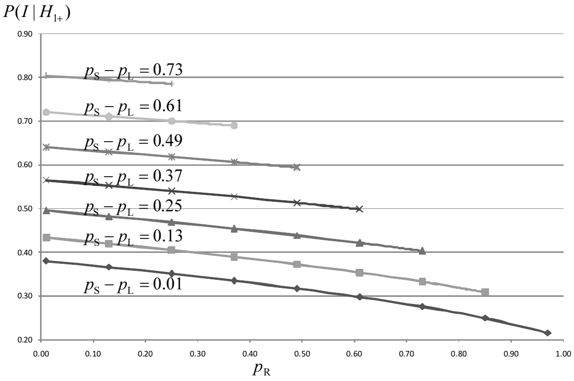


Figure. 6: Relationship between sender anonymity and p_{R} under the condition that $p_{\mathrm{S}} - p_{\mathrm{L}}$ is constant.

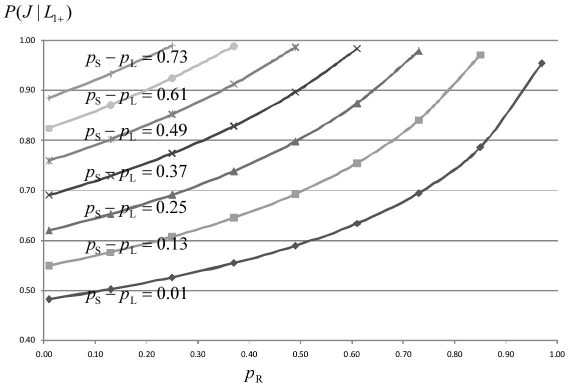


Figure. 7: Relationship between receiver anonymity and p_{R} under the condition that $p_{\mathrm{S}} - p_{\mathrm{L}}$ is constant.

for communication becomes small and high anonymity to a sender and a receiver is guaranteed. Thus, we need to establish the proper probabilities of mode choice according to circumstances.

VI. Conclusion

In this paper, we have proposed an anonymous communication system, which provides sender anonymity and receiver anonymity, using probabilistic choice of actions in relay nodes only. We have introduced loopbacks that indicate that a message, which a relay node sends, returns to itself. The main characteristics of our proposed anonymous communication system is as follows:

- Our system provides both sender anonymity and receiver anonymity.
- Our system does not need multiple-encryption unlike Onion Routing and 3-Mode Net.
- The computational load of each relay node is as small as that of Crowds.
- The size of a data set is always constant.

The remaining works are mainly as follows:

1. realize a bidirectional communication.
2. implement our system, including peer-to-peer (P2P) and DHT approaches to improve scalability such as [13, 14].

In particular, in order to confirm the utility of our system in practice, we also plan to implement our system and to compare our system to 3MN or other anonymous communication systems, such as Tor [2].

Acknowledgment

This work was supported in part by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science.

References

[1] 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.

[2] R. Dingledine, N. Mathewson, and P. Syverson, “Tor: The Second-Generation Onion Router,” in *Proc. 13th USENIX Security Symposium*, August 2004.

[3] 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.

[4] A. Mislove, G. Overoi, A. Post, C. Reis, P. Druschel, and D. Wallach, “AP3: Cooperative, decentralized Anonymous Communication,” in *Proc. ACM SIGOPS European Workshop*, September 2004.

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

[6] 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).

[7] 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*, Vol. 5, No. 3, pp. 276–283, 2010.

[8] K. Kono, S. Nakano, Y. Ito, and N. Babaguchi, “Theoretical Analysis of the Performance of Anonymous Communication System 3-Mode Net,” *The Institute of Electronics, Information and Communication Engineers (IEICE) Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E93-A, No. 7, pp. 1338–1345, July 2010.

[9] K. Kono, Y. Ito, and N. Babaguchi, “Anonymous Communication System Using Probabilistic Choice of Actions and Multiple Loopbacks,” in *Proc. 6th International Conference on Information Assurance and Security*, pp. 210–215, August 2010.

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

[11] 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.

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

[13] M. Prateek and B. Nikita, “ShadowWalker: Peer-to-peer Anonymous Communication Using Redundant Structured Topologies,” in *Proc. ACM Conference on Computer and Communications Security*, pp. 161–172, November 2009.

[14] M. Kondo, S. Saito, K. Ishiguro, H. Tanaka, and H. Matsuo, “Bifrost: A Novel Anonymous Communication System with DHT,” in *Proc. 10th International Conference on Parallel and Distributed Computing, Applications and Technologies*, pp. 324–329, December 2009.

Appendix A: Proof of Equation (5)

Let τ_1 denote a random variable representing the number of relay nodes required for communication under the condition that the initial multiplicity of loopbacks is 1, and $P_1(i)$ denote the probability that the number of relay nodes required for communication is equal to i, that is, $P_1(i) = P(\tau_1 = i)$. Note that $P_1(i)$ means that a receiver appears at the $(i+1)$-th node on the path. Considering that the i-th node sends a data set to a receiver by choosing S-Mode, the probability that the multiplicity of loopbacks of a data set is equal to 1 when a relay node appears at the i-th node on the path is equal to $P_1(i)/p_{\mathrm{S}}$. The probability that the multiplicity of loopbacks of a data set is equal to 1 when a collaborating node appears at the i-th node on the path is equal to $(n_c/n_t) \times P_1(i)/p_{\mathrm{S}}$. We derive the probability $P(Z_{i,1})$ that the multiplicity of loopbacks of a data set is equal to 1 when the last collaborating node on the path appears at the i-th node on the path. Considering that collaborating nodes on the communication path do not appear at larger than or equal to $i+1$ on the path, $P(Z_{i,1})$ is given by

$$\begin{aligned} P(Z_{i,1}) &= \frac{n_c \times P_1(i)}{n_t \times p_{\mathrm{S}}} \times \sum_{l=1}^{\infty} P_1(l) \left(\frac{n_t - n_c}{n_t}\right)^{(l-1)} \\ &= \frac{n_c \times P_1(i)}{n_t \times p_{\mathrm{S}}} \times \frac{n_t}{n_t - n_c} \times g_{\tau_1}(\lambda) \\ &= \frac{n_c P_1(i) g_{\tau_1}(\lambda)}{(n_t - n_c) p_{\mathrm{S}}}, \qquad (6) \end{aligned}$$

where the probability generating function $g_{\tau_1}(\lambda)$ is given by Eq. (2) and $\lambda = (n_t - n_c)/n_t$. $g_{\tau_1}(\lambda)$ is also defined as

$$g_{\tau_1}(\lambda) = \sum_{j=0}^{\infty} P(\tau_1 = j)\lambda^j. \qquad (7)$$

As a result, $P(Z_{1+,1})$ is calculated as follows:

$$\begin{aligned} P(Z_{1+,1}) &= \sum_{i=1}^{\infty} P(Z_{i,1}) \\ &= \frac{n_c g_{\tau_1}(\lambda)}{(n_t - n_c) p_{\mathrm{S}}} \sum_{i=1}^{\infty} P_1(i) \\ &= \frac{n_c g_{\tau_1}(\lambda)}{(n_t - n_c) p_{\mathrm{S}}}. \qquad (8) \end{aligned}$$

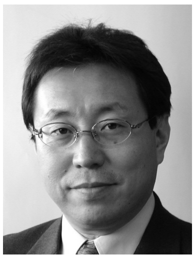
■

Author Biographies

Kazuhiro Kono received the B.E, M.E., and Ph.D. degrees in communication engineering from Osaka University, Japan, in 2005, 2007, and 2010, respectively. He is an Assistant Professor in the Faculty of Safety Science, Kansai University. His research interests include network security, in particular, access control and anonymous communications systems. He is a member of IEICE, IPSJ, IEEE, and ACM.

Shinnosuke Nakano received the B.E. degree in communication engineering from Osaka University, Japan, in 2009. He is currently pursuing the M.E. degree in engineering at Osaka University. He is a student member of IEICE.

Yoshimichi Ito received the B.E. and M.E. degrees in electrical engineering from Kyoto University, Japan, in 1990 and 1992, respectively. He is currently an Assistant Professor in the Graduate School of Engineering, Osaka University. His research interests include system control theory, digital signal processing, and network security. He is a member of IEICE, IEEJ, ISCIE, and SICE.



Noboru Babaguchi received the B.E., M.E., and Ph.D. degrees in communication engineering from Osaka University, Japan, in 1979, 1981, and 1984, respectively. He is currently a Professor in the Graduate School of Engineering, Osaka University. From 1996 to 1997, he was a Visiting Scholar at the University of California, San Diego. His research interests include multimedia computing and artificial intelligence. He is a fellow of IEICE, and a member of IPSJ, ITE, IEEE, ACM, and JSAI.