Evaluation of Anonymity of Practical Anonymous Communication Networks

Shigeki Kitazawa[1], Masakazu Soshi[2], and Atsuko Miyaji[2]

[1] Mitsubishi Electric Corporation Information Technology R&D Center
5-1-1 Ofuna, Kamakura, Kanagawa, 247-8501 JAPAN
shigeki@iss.isl.melco.co.jp

[2] School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Tatsunokuchi, Nomi, Ishikawa 923-1292, JAPAN
{soshi,miyaji}@jaist.ac.jp

Abstract. In the paper we shall evaluate various aspects of anonymity properties afforded by practical anonymous communication networks. For that purpose, first we propose two novel anonymity metrics for practical anonymous communication networks. Next we shall discuss whether or not deterministic protocols can provide anonymity efficiently in terms of computational complexity. Unfortunately, we can show that it is difficult to build efficient anonymous networks only by means of deterministic approaches. We also run simulation experiments and discuss the results.

1 Introduction

Anonymous communication networks are indispensable to protect privacy of users in open networks such as the Internet. Therefore they have wide application, e.g., electronic voting, and enormous research has been conducted on them [1–8]. The simplest way of establishing anonymous networks is given as follows. When Alice sends a message to Bob anonymously, she first dispatches it to a trusted *proxy* (or anonymizer) and then the proxy forwards the message to Bob. Consequently Bob cannot know who originally injected the message into the network and thus anonymous communication is achieved. This is essentially the same as what Anonymizer does [2]. In this paper, an entity which initiates anonymous communication is called an *initiator*, and an entity for which messages of the initiator are destined is called a *responder*. Furthermore, we use the terms 'proxy' and 'node' interchangeably in this paper.

Anonymous networks, however, could not be useful unless we can evaluate anonymity properties provided by them. Unfortunately, although we can analyze anonymity of some anonymous protocols in a rigorous manner [1, 4, 5], from a practical point of view, such protocols often degrade efficiency or incurs some cost. For example, [5] requires a lot of servers and [4] is quite ineffective.

On the other hand, with respect to practical anonymous communication networks [2, 6, 7], it is difficult to evaluate anonymity attained in the networks. This is mainly due to the lack of *anonymity metrics* for practical anonymous networks. However, since it is difficult to devise general anonymity metrics for practical anonymous networks such as Crowds [7], few attempts have been made so far to evaluate anonymity provided by anonymous networks [7, 10, 11]. Consequently it is difficult to discuss advantages and disadvantages of various practical anonymous networks proposed so far. Therefore in this paper we shall evaluate various aspects of anonymity properties afforded by practical anonymous communication networks.

For that purpose, first we propose two novel anonymity metrics for practical anonymous communication networks. Anonymity metrics proposed in this paper are based on the following observations:

1. Generally speaking, anonymous networks have several intermediate proxies en route from an initiator to a responder. In such a case, as the number of the intermediate proxies increases, i.e., the path that messages in anonymous communication follows becomes longer, anonymity provided by the anonymous protocols becomes higher [3, 6–8]. On the other hand, communication costs, which can be represented by communication paths, cannot be infinitely high. Due to performance reasons, constraints by network architectures, and so on, the costs are under restriction to some degree.
2. As discussed above, the most primitive form of anonymous communication is via one or more proxies. Hence if an initiator can choose a trusted proxy to which she first dispatches her messages whenever she initiates anonymous communication with various responders, then such anonymous networks can afford desirable anonymity.

Later in this paper, from the viewpoints as discussed above, we formalize the anonymity metrics for practical anonymous networks.

Next we shall discuss whether or not deterministic anonymous networks can provide anonymity efficiently in terms of computational complexity. Unfortunately, we can show that we have little hope of efficient anonymous networks only by means of deterministic approaches.

As a result we need to invent practical anonymous networks by probabilistic or heuristic means. Hence we consider several possible (practical) anonymous protocols, run simulation experiments for them, and discuss the results. Simulation results show that we can enhance anonymity only by taking into consideration the neighborhood nodes. Especially, anonymous protocols considered in the experiments suggest some possible extensions to the famous Crowds anonymous system.

The rest of the paper is organized as follows. In Sect. 2 we propose novel anonymous metrics and discuss various aspects of them. In Sect. 3 we run simulation experiments to investigate heuristic approaches. Finally, we conclude this paper in Sect. 4.

2 Proposed Anonymity Metrics and Evaluation

In this section, we propose and discuss two anonymity metrics for practical anonymous networks.

As stated in Sect. 1, our proposed metrics are briefly summarized as follows:

1. anonymity properties with respect to communication paths, and
2. anonymity properties with respect to the possibility of selecting trusted proxies.

In this section, we first present the background of each anonymity metrics and then formalize the two metrics. Next we discuss whether or not deterministic protocols can provide anonymity efficiently in terms of computational complexity. Unfortunately, we can show that we have little hope of efficient anonymous networks only by means of deterministic approaches.

2.1 Anonymity Metric (1)

Background.

1. Generally speaking, as the number of the intermediate proxies on anonymous communication path increases, anonymity provided by the anonymous network becomes higher.
2. On the other hand, communication costs cannot be infinitely high. Due to performance reasons, constraints by network architectures, and so on, the costs are under restriction to some length.

Now we model anonymity properties just discussed. First, in order to develop an abstract model for evaluating anonymity from the viewpoint of the item 1 above, assume that some value is assigned to each node in the network. Moreover, assume that anonymity afforded by anonymous communication can be estimated by adding the value of each node on the path. In other words, a larger value of the sum indicates a higher level of anonymity. Henceforth we suppose that we are given a function which assigns the value to each node and we call the function *privacy function*. The assigned values are called *privacy values*.

On the other hand, communication costs are usually constrained from the viewpoint of the item 2. To express such a situation, let us assume that some value, which is apart from the privacy value of the node, is assigned to each node. Then as the sum of the values on a path becomes larger, the cost of the path becomes larger. Henceforth we suppose that we are given a function which assigns the value to each node and we call the function *cost function*[3]. The assigned values are called *cost values*.

Given a network system, it would be straightforward to define cost functions. With respect to privacy functions, we can consider various ways of deriving them. For instance, we can take advantage of various available rating methods such as those used in reputation systems [12, 13].

For another example, ISO defines the international standard for security system evaluation, i.e., ISO 15408 [14]. In the standard, 'Security Functional Requirements' prescribes several privacy conditions, that is,

[3] Alternatively, we can assign a cost value to a communication link. However, for simplicity, in this paper we model it as is given above.

FPR_AANO.1 Anonymity, FPR_PSE Pseudonymity, FPR_UNO.1 Unobservability, and FPR_UNL.1 Unlinkability. They are good candidates for the basis on which privacy functions are developed.

In summary, privacy and cost functions provide abstraction for anonymity properties as mentioned above. In the rest of this paper, we suppose that these two functions are given in advance. However, as shown later, it is shown that even if such functions are available, effective anonymous communication networks can hardly be built only by means of deterministic methods.

Formalization and Evaluation. In this section, we formalize the anonymity metric discussed in Sect. 2.1.

First, a network is supposed to be represented by a directed graph $G = (V, E)$. Here V is a set of nodes (proxies) and E is a set of directed edges. In general $E \subseteq V \times V$ holds.

Hereinafter in this paper we denote an initiator and a responder by s and d ($\in V$), respectively, unless explicitly stated. Moreover, as discussed in Sect. 2.1, we assume that privacy function $\mathcal{P} : V \to N$ and cost function $\mathcal{C} : V \to N$ are given. Here N is a set of non-negative integers.

Now we are ready to define the anonymity metric as follows:

Definition 1. *In a directed graph G, with respect to a path v_1 $(= s)$, v_2, …, v_n $(= d)$, if $\sum_{i=1}^{n} \mathcal{P}(v_i) \geq p$ and $\sum_{i=1}^{n} \mathcal{C}(v_i) \leq c$, then we call the path* (p, c)-anonymous.

Next we shall discuss whether or not deterministic protocols can efficiently provide anonymity in terms of the metric. For that purpose, we define a decision problem corresponding to the anonymity metric as follows:

Problem 1. ((p, c)-anonymity)

[INSTANCE] A directed graph $G = (V, E)$, s, $d \in V$, and privacy function $\mathcal{P} : V \to N$ and cost function $\mathcal{C} : V \to N$.

[QUESTION] Is there a (p, c)-anonymous path from s to d?

Now we can prove Theorem 1.

Theorem 1. *(p, c)-anonymity is NP complete.*

Proof. Given G, privacy function $\mathcal{P}$, cost function $\mathcal{C}$, and a path v_1 $(= s)$, $v_2, \ldots,$ v_n $(= d)$, it should be obvious that in polynomial time we can determine whether or not the path is (p, c)-anonymous. Thus we can construct a non-deterministic algorithm to solve (p, c)-anonymity in polynomial time and consequently (p, c)-anonymity is in NP.

Now we show that "PARTITION", which is known to be NP-complete, is polynomially reducible to (p, c)-anonymity. PARTITION is defined as follows [15]:

Problem 2. (PARTITION)

[INSTANCE] A finite set $A = \{a_1, a_2, \ldots, a_n\}$ and a size function $w : A \to N$.

[QUESTION] Is there a subset $A' \subseteq A$ such that $\sum_{a \in A'} w(a) = \sum_{a \in A - A'} w(a)$?

When we are given an arbitrary instance of PARTITION, we construct the following graph $G = (V, E)$, and define privacy function $\mathcal{P}$, cost function $\mathcal{C}$, and the privacy and cost values p and c respectively.

$$\begin{aligned} V &= \{a_{1,0}, a_{2,0}, \ldots, a_{n,0}, a_{(n+1),0}\} \cup \left(\bigcup_{i=1}^{n} \{a_{i,1}, a_{i,2}\} \right) \\ E &= \left(\bigcup_{i=1}^{n} \{(a_{i,0}, a_{i,1})\} \right) \cup \left(\bigcup_{i=1}^{n} \{(a_{i,0}, a_{i,2})\} \right) \cup \left(\bigcup_{i=1}^{n} \{(a_{i,1}, a_{(i+1),0})\} \right) \\ &\qquad \cup \left(\bigcup_{i=1}^{n} \{(a_{i,2}, a_{(i+1),0})\} \right) \end{aligned}$$

$$\begin{array}{ll} \mathcal{P}(a_{i,1}) = \mathcal{C}(a_{i,1}) = w(a_i) & (i = 1, \ldots, n) \\ \mathcal{P}(a_{i,2}) = \mathcal{C}(a_{i,2}) = 0 & (i = 1, \ldots, n) \\ \mathcal{P}(a_{i,0}) = \mathcal{C}(a_{i,0}) = 0 & (i = 1, \ldots, n+1) \\ p = c = \dfrac{1}{2} \displaystyle\sum_{i=1}^{n} w(a_i) & \end{array}$$

where $a_{i,0} = s$ and $a_{(n+1),0} = d$.

To illustrate the above reduction, for example we depict in Fig. 1 the reduction where the instance of PARTITION is $A = \{a_1, a_2, a_3\}$, $w(a_1) = 3$, $w(a_2) = 5$, $w(a_3) = 2$.

Let us now suppose that PARTITION has a solution $A' = \{a_{i_1}, a_{i_2}, ..., a_{i_j}\}$. Without loss of generality, assume that $i_1 < i_2 ... < i_j$. In such a case we define a function δ as follows:

$$\delta(k) = \begin{cases} 1 \text{ if } k = i_l \text{ for some } l \\ 2 \text{ otherwise} \end{cases}$$

By using function δ, consider a path $P = a_{1,0}, a_{1,\delta(1)}, a_{2,0}, a_{2,\delta(2)}, ..., a_{n,\delta(n)}, a_{(n+1),0}$. We can readily see that P is (p, c)-anonymous in G. This is because $\sum_{a \in P} \mathcal{P}(a) = \sum_{a \in P} \mathcal{C}(a) = \frac{1}{2} \sum_{i=1}^{n} w(a_i)$ $(= p = c)$. For example, in Fig. 1, the path corresponds to $a_{1,0}$, $a_{1,1}$, $a_{2,0}$, $a_{2,2}$, $a_{3,0}$, $a_{3,1}$, $a_{4,0}$.

Conversely, assume that graph G has a (p, c)-anonymous path. Let us further assume that $a_{i_1,1}$, $a_{i_2,1}$, ..., $a_{i_j,1}$ are all nodes such that the second subscript is one. It is now clear that $\{a_{i_1}, a_{i_2}, ..., a_{i_j}\}$ $(= A')$ is a solution of PARTITION. □

It is well-known that in some NP-complete problems there exist algorithms to find solutions which are never far from optimal ones by more than some

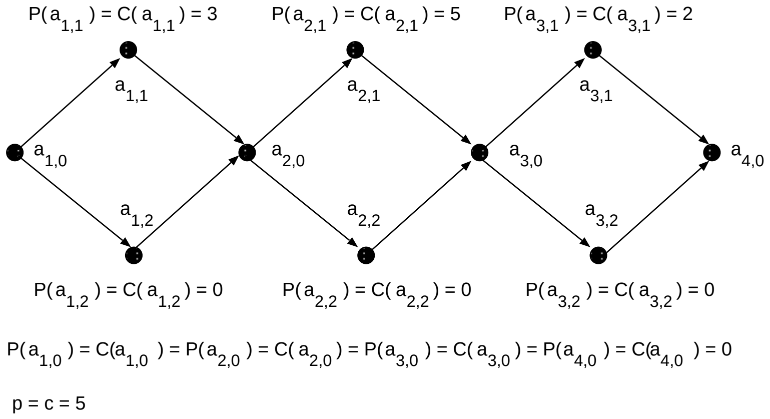

Fig. 1. Reduction from PARTITION

specific bounds. They are called *approximation algorithms* [15]. Unfortunately, (p, c)-anonymity is so difficult that there does not exist such an approximation algorithm. This will be shown below.

We denote by I an instance of (p, c)-anonymity with some fixed c. Furthermore, a solution by an optimization algorithm that maximizes $\sum_{i=1}^{n} \mathcal{P}(v_i)$ of I is denoted by $OPT(I)$ (such problems are called *optimization problems*). Obviously (p, c)-anonymity is no more difficult than a problem to find $OPT(I)$ and the latter problem is thus NP-hard.

Now we can prove Theorem 2.

Theorem 2. *If $P \neq NP$, then there does not exist a deterministic polynomial algorithm (approximation algorithm) A which can guarantee $|OPT(I) - A(I)| \leq p'$ for a fixed p' and all instances I of the optimization problems for (p, c)-anonymity.*

Proof. We prove this theorem by contradiction. Without loss of generality, assume that p' is a positive integer.

Suppose that there is an A in Theorem 2. In such a case, by using A, we can construct a deterministic polynomial algorithm B which can solve (p, c)-anonymity, which contradicts the assumption P $\neq$ NP.

B is actually constructed as follows. First we denote by I' a new instance where privacy function $\mathcal{P}$ is replaced by $\mathcal{P}'$, which is defined as $\mathcal{P}'(v) = (p' + 1)\mathcal{P}(v)$.

Then candidate solutions for I' are clearly the same as those for I and the privacy value of a solution for I' is $(p' + 1)$ times the corresponding value for I. Now note that every solution for I' is a multiple of $p' + 1$ and that $|OPT(I')$ $-A(I')| \leq p'$ holds. So it must hold that $|OPT(I')\ -A(I')| = 0$ and finally we can conclude that $|OPT(I)\ -B(I)| = |OPT(I')\ -A(I')|\ /(p'+1) = 0$. However, the fact also means that we can find $OPT(I)$ in polynomial time. This is a contradiction. □

2.2 Anonymity Metric (2)

In this section we propose and discuss another anonymity metric.

Background. As discussed in Sect. 1, the most primitive form of anonymous communication is via one or more proxies. However, in practical anonymous communication networks, it is not always possible to select a trusted proxy when an initiator anonymously sends messages to a responder because of the network topology or the locations of the initiator or responder. Moreover, we cannot trust all proxies in the network.

Hence if an initiator can choose a trusted proxy to which she first dispatches her messages whenever she initiates anonymous communication with various responders, then such anonymous networks can afford a desired level of anonymity. Therefore we can consider anonymity metric whether or not we can arrange trusted proxies in the anonymous network in such a way as stated above.

Formalization and Evaluation. Here we formalize anonymity metric discussed in Sect. 2.2.

First, as in Sect. 2.1, we regard a network as a directed graph $G = (V, E)$. Moreover, let $s \in V$ and a set of nodes $\{d_1, d_2, ..., d_j\} \subseteq V$ be an initiator and a set of responders, respectively.

Now we can formalize anonymity metric discussed in Sect. 2.2 as follows:

Definition 2. *Suppose that we are given a directed graph $G = (V, E)$, $s \in V$, and a set of nodes $\{d_1, d_2, ..., d_j\} \subseteq V$. In such a case, if for a fixed positive value t ($\leq |V|$) there exist a subset $T \subseteq V$, $|T| \leq t$ such that we can always find some $n \in T$ on paths from s to every $d \in \{d_1, ..., d_j\}$, then we call G is* t-locatable *with respect to s and $\{d_1, ..., d_j\}$.*

Next we shall discuss whether or not deterministic protocols can efficiently provide anonymity in terms of the metric. For that purpose, we define a decision problem corresponding to the anonymity metric as follows:

Problem 3. (t-locatability)

[**INSTANCE**] A directed graph $G = (V, E)$, $s \in V$, a responder set $\{d_1, d_2, ..., d_j\} \subseteq V$, and a positive value t ($\leq |V|$).

[**QUESTION**] Is a graph G is t-locatable with respect to initiator s and a responder set $\{d_1, d_2, ..., d_j\}$?

Based on the above definitions, we can now prove Theorem 3.

Theorem 3. *t-locatability is NP-complete.*

Proof. t-locatability is in NP. This is because we can construct a non-deterministic polynomial algorithm which arbitrarily chooses a subset of V and determines whether or not the subset satisfies the condition of t-locatability.

Next we show that an NP-complete problem, "VERTEX COVER", is in polynomial time reduced to t-locatability. VERTEX COVER is defined below [15]:

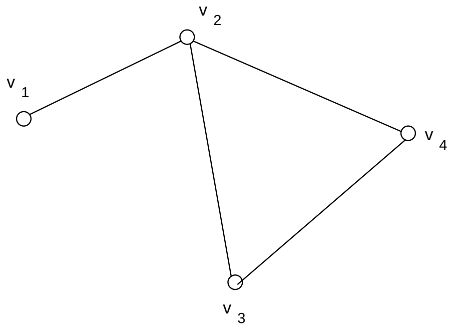

Fig. 2. An Instance of VERTEX COVER

Problem 4. (VERTEX COVER)
[**INSTANCE**] Graph $G = (V, E)$, positive integer $K \leq |V|$.
[**QUESTION**] Is there a vertex cover of size K or less for G, i.e., a subset $V' \subseteq V$ with $|V'| \leq K$ such that for each edge $\{u, v\} \in E$ at least one of u and v belongs to V'?

where G is an undirected graph.

Given an instance of VERTEX COVER, we transform it into an instance of t-locatability, which is defined as follows:

- Graph $G'' = (V'', E'')$, where $V'' = \{s\} \cup V \cup \{v_1v_2 \mid (v_1, v_2) \in E\}$ and $E'' = \{(s, v) \mid v \in V\} \cup \{(v_1, v_1v_2), (v_2, v_1v_2) \mid (v_1, v_2) \in E\}$.
- initiator $= s$,
- responder set $= \{v_1v_2 \mid (v_1, v_2) \in E\}$
- $t = K$

Note that in the above reduction, v_1v_2 ((v_1,v_2) $\in E$) represents a single node in G''.
To demonstrate an example of the reduction, we pay attention to an instance of VERTEX COVER as shown in Fig. 2. The instance is reduced to the instance of t-locatability depicted in Fig. 3.

Below we show that the reduction given above is actually polynomial time reduction from VERTEX COVER to t-locatability.

First let us suppose that VERTEX COVER has a solution $V' = \{v_{i_1}, v_{i_2}, ..., v_{i_j}\}$ $(j \leq K)$. In such a case $T = V'$ $(\subseteq V'')$ is a solution for the corresponding instance of t-locatability. The reason is as follows. Let $R = \{v_1v_2 \mid (v_1, v_2) \in E\}$ be a responder set of G''. Then a path from s to a responder $v_1v_2 \in R$ is either $s \to v_1 \to v_1v_2$ or $s \to v_2 \to v_1v_2$. Keeping in mind that $(v_1, v_2) \in E$ and V' is a solution of VERTEX COVER, we can see that either $v_1 \in V'$ $(= T)$ or $v_2 \in V'$ $(= T)$ holds. Consequently all we have to do is to place a proxy on v_1 $(\in V'')$ or v_2 $(\in V'')$, respectively in the former case or the latter.

Conversely, if an instance of t-locatability given above has a solution $T = \{v_{i_1}, v_{i_2}, ..., v_{i_j}\}$, then we can conclude that $V' = T$ in a similar manner.

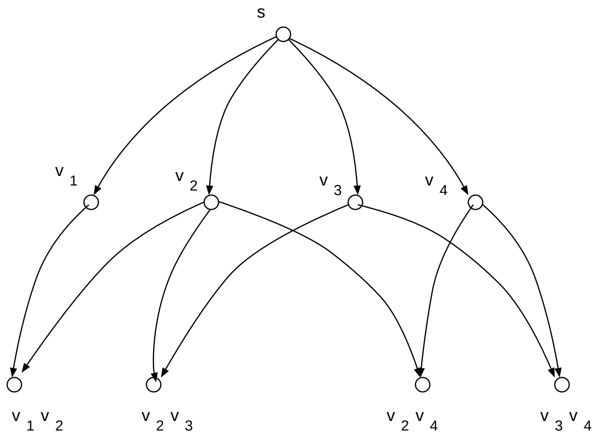

Fig. 3. Reduction from VERTEX COVER

At this stage it should be clear that the above reduction can be done in polynomial time. □

Now we consider an approximation algorithm to the optimization problem for t-locatability. In a similar way as in the case of (p, c)-anonymity, given an instance I of the optimization problem for t-locatability, we denote by $OPT(I)$ a solution found by an optimization algorithm of the problem that minimizes t. Then we can show that it is difficult to even find a solution approximate to the optimal one.

Theorem 4. *If $P \neq NP$, then no polynomial time approximation algorithm A for t-locatability can guarantee $|OPT(I) - A(I)| \leq t'$ for a fixed constant t'.*

Proof. We can show this theorem again by contradiction as Theorem 2. Without loss of generality, we assume that t' is a positive integer.

Suppose that A is actually such an approximate algorithm in Theorem 4. Then by using A we can construct a deterministic polynomial time algorithm B which can solve t-locatability.

B is constructed as follows. First we consider a new instance I' of t-locatability, where graph G is replaced by a new graph $G' = (V', E')$. G' has the same s as I and consists of $(t'+1)$ graphs, each of which is isomorphic to G.

More precisely, G' is defined as follows:

$$\begin{aligned}V' &= \{s\} \cup \{v[i] \mid v \in V, 1 \le i \le t'+1\}\\ &\quad \cup \{v_1[i]v_2[i] \mid v_1v_2 \in V, 1 \le i \le t'+1\}\\ E' &= \{(s, v[i]) | v \in V, 1 \le i \le t'+1\}\\ &\quad \cup \{(v_1[i], v_1[i]v_2[i]) \mid v_1v_2 \in V, 1 \le i \le t'+1\}\\ &\quad \cup \{(v_2[i], v_1[i]v_2[i]) \mid v_1v_2 \in V, 1 \le i \le t'+1\}\end{aligned}$$

Then the corresponding nodes in G to candidate solutions for I' are clearly the candidate solutions for I and the number of proxies of a solution for I' is $(t'+1)$ times the corresponding value for I. Now note that every solution for I' is a multiple of $t'+1$ and that $|OPT(I') - A(I')| \le t'$ holds. So it must hold that $|OPT(I') - A(I')| = 0$ and finally we can conclude that $|OPT(I) - B(I)| = |OPT(I') - A(I')| / (t'+1) = 0$. However, the fact also means that we can find $OPT(I)$ in polynomial time. Contradiction. □

2.3 Discussion

So far we have proposed and thoroughly discussed two new anonymity metrics for practical anonymous networks ((p, c)-anonymity, t-locatability). It is also possible to develop other anonymity metrics. In this section we take into consideration some anonymity metrics other than (p, c)-anonymity and t-locatability.

For example, as mentioned before, it is the most intuitive to adopt anonymity metric which considers anonymity properties where the level of anonymity becomes higher as the path which anonymous communication follows becomes longer. Generally speaking, this can be formalized by a decision problem "LONGEST PATH" [15].

Problem 5. (LONGEST PATH)

[**INSTANCE**] Graph $G = (V, E)$, length $l(e) \in Z^+$ for each $e \in E$, positive integer K, specified vertices s, $t \in V$.

[**QUESTION**] Is there a simple path in G from s to t of length K or more, i.e., whose edge lengths sum to at least K?

LONGEST PATH is also known to be NP-complete [15].

Let us consider another anonymity metric. When an initiator sends many messages to a responder anonymously, as the number of paths the messages follow becomes larger, it becomes more difficult for attackers to gather information about the initiator and the responder and thus more anonymity would be provided.

Anonymity metric for the case just mentioned can be formalized by (bounded) disjoint paths [15]. Bounded disjoint paths are a set of paths whose lengths are limited by some bound and no pair of which have a node in common. More formally, the problem is defined as follows.

Problem 6. (MAXIMUM LENGTH-BOUNDED DISJOINT PATHS)

[INSTANCE] Graph $G = (V, E)$, specified vertices s and t, positive integers J, $K \leq |V|$.

[QUESTION] Does G contain J or more mutually vertex disjoint paths from s to t, none involving more than K edges?

MAXIMUM LENGTH-BOUNDED DISJOINT PATHS is NP-complete [15].

We also considered other anonymity metrics, which are omitted from this paper due to the lack of space, most of which are in NP-complete. Needless to say, it is strongly believed that it is almost impossible to solve NP-complete problems efficiently (i.e., in polynomial time). This implies that NP $\neq$ P.

Therefore based on the discussion so far, generally speaking, it is difficult to establish effective practical anonymous networks only by means of deterministic approaches. Hence practical anonymous networks should be built with probabilistic or heuristic approaches.

3 Simulation Experiments

Based on the previous section, we consider several practical anonymous protocols and run the simulation experiments for them in this section. As a result of the experiments, we can show that heuristic approaches for anonymous networks work fine in terms of our anonymity metrics.

3.1 Descriptions of Simulations

In this section we discuss the background for our simulation experiments.

In order to conduct network simulation, first we have to generate network topologies which conform to the existing networks. In our simulation experiments, we have used Inet topology generator developed in Michigan University [16].

Next we focus our attention on privacy and cost functions. Although we can suppose various privacy and cost functions, in the simulation we define the functions according to the following policies:

N1 Privacy and cost values are generated at random.

N2 Larger cost and smaller privacy values are assigned to articulation nodes[4]. Intuitively speaking, articulation nodes are the points where various paths join and so the attacks to such nodes can pose a serious threat to anonymity properties. In other words, N2 is one example scenario in favor of attackers. In our simulation, we assign every articulation node and its adjacent node (i.e., connected through an edge) to more than 0.8 times the maximum cost value with probability 0.8. Furthermore, with probability 0.8, we assign every articulation node and its adjacent node to less than 0.2 times the maximum privacy value.

[4] Articulation nodes are what increase the number of connected components of the graph if they are removed [17].

Table 1. Simulation results (N1)

	P1+N1	P2+N1	P3+N1	P4+N1	P5+N1
length	5.68	5.55	5.86	3.93	3.86
cost	2968.53	2390.54	3112.20	1790.56	2174.50
privacy	3119.67	2973.76	3818.19	2259.76	2232.12

Table 2. Simulation results (N2)

	P1+N2	P2+N2	P3+N2	P4+N2	P5+N2
length	5.51	5.71	5.59	3.77	3.88
cost	5392.66	4893.18	5132.26	3670.01	4019.74
privacy	1125.10	1512.52	1751.69	1039.15	843.36

Furthermore, in our simulation we implement the following anonymous communication protocols.

P1 Crowds protocol [7]
P2 Crowds protocol with a strategy to choose a node with the smallest cost value when forwarding messages.
P3 Crowds protocol with a strategy to choose a node with the largest anonymity value when forwarding messages.
P4 protocol to choose a path with the smallest sum of the cost values of the nodes on the path.
P5 shortest path (just for comparison)

3.2 Evaluation

As stated in Sect. 3.1, there are two ways of network conditions (N1 and N2) and five ways of protocols (P1, P2, P3, P4, and P5). Thus we combinedly have ten ways of simulation experiments. Henceforth we call each of them P1+N1, P1+N2, P2+N1, P2+N2, P3+N1, P3+N2, P4+N1, P4+N2, P5+N1, and P5+N2, respectively.

The network topology used in our simulation was generated by Inet and the number of the nodes is 3037. Privacy and cost values are integers from 0 to 1000. The forwarding probability of Crowds is 2/3. Finally, initiators and responders were chosen randomly and each experiment was run 1000 times. We show the averages of the simulation experiments in Table 1 and 2.

In the following evaluation, we regard P1+N1, i.e., Crowds protocol on a network with random privacy and cost values, as the base for comparison.

First consider the case in Fig. 1. In comparison of P2+N1 with P1+N1, since P2 selects a node with the smallest cost value, the average of the value decreases by 19.5%. At the same time, the average privacy value also decreases by 4.7%. However, the rate of decrease of the cost value is greater than that of the privacy and consequently it shows the effectiveness of P2.

Similarly, in comparison of P3+N1 with P1+N1, since P3 selects a node with the largest privacy value, the average of the value increases by 22.4%. On the other hand, the average cost value also increases, but only by 4.8%. Hence we can see that P3 is also a promising heuristic.

With respect to P4+N1, since P4 selects a path with the smallest sum of the cost values of the nodes on the path, the rate of decrease of cost values is the largest (39.7% decrease in comparison with P1+N1). However, the rate of decrease of the privacy values is also large (27.6% decrease in comparison with P1+N1). This is partly because paths with smaller cost values are often shorter and hence privacy values also become smaller.

This consideration is supported by the comparison of P5+N1 with P1+N1. That is, generally speaking, as a path becomes shorter, the sums of privacy and cost values on the path also become smaller. Consequently it is not obvious whether or not P4 and P5 are useful.

In the case in Table 2, we can evaluate the experiments in a similar way as in Table 1. However, in Table 2, we can immediately see that cost and privacy ratios of P2+N2, P3+N2 and P4+N2 with respect to P1+N2 are better than those of P2+N1, P3+N1 and P4+N1 with respect to P1+N1, respectively. This is because N2 is a scenario in favor of attackers and we can obtain greater effects from anonymous protocols which try to enhance anonymity. On the other hand, it is not clear whether or not P5+N2 is more effective than P5+N1.

Note that P1 (Crowds), P2, and P3 do not need information about a whole network, but only about the neighborhood nodes. In other words, from the simulation experiments, we can see that if we monitor only the neighbors and assign appropriate privacy and cost values to the nodes, then we can get significant effects. In particular, since Crowds does not take into consideration such cases, if we slightly modify it as P2 or P3, then we can get more effective version of Crowds anonymous communication protocols.

4 Conclusion

Anonymous communication networks are indispensable to protect users' privacy in open networks such as the Internet. In the paper we have evaluated various aspects of anonymity properties afforded by practical anonymous communication networks.

In this paper we proposed two novel anonymity metrics for practical anonymous communication networks. Furthermore we discussed whether or not deterministic protocols can provide anonymity efficiently in terms of computational complexity. Unfortunately, we can show that we have little hope of efficient anonymous networks only by means of deterministic approaches.

Finally we run simulation experiments and discussed the results. Simulation results show that we can enhance anonymity only by taking into consideration the neighborhood nodes. Especially, anonymous protocols considered in the experiments suggest some possible extensions to the famous Crowds anonymous system.

References

1. Abe, M.: Universally verifiable Mix-net with verification work independent of the number of Mix-servers. IEICE Trans. Fundamentals **E83-A** (2000) 1431–1440
2. Anonymizer: http://www.anonymizer.com/
3. Chaum, D.: Untraceable electronic mail, return addresses, and digital pseudonyms. CACM **24** (1981) 84–88
4. Chaum, D.: The dining cryptographers problem: Unconditional sender and recipient untraceability. Journal of Cryptography **1** (1988) 65–75
5. Desmedt, Y., Kurosawa, K.: How to break a practical MIX and design a new one. EUROCRYPT 2000. Vol. 1807 of LNCS, Springer-Verlag (2000) 557–572
6. Pfitzmann, A.: A switched/broadcast ISDN to decrease user observability. Proc. of 1984 International Zurich Seminar on Digital Communications, Applications of Coding, Channel Coding and Secrecy Coding. (1984) 183–190
7. Reiter, M.K., Rubin, A.D.: Crowds: anonymity for web transactions. ACM Trans. Information and System Security **1** (1998) 66–92
8. Syverson, P.F., Goldschlag, D.M., Reed, M.G.: Anonymous connections and Onion routing. IEEE Symposium on Security and Privacy. (1997) 44–54
9. Mitomo, M., Kurosawa, K.: Attack for flash MIX. ASIACRYPT 2000. Vol. 1976 of LNCS, Springer-Verlag (2000) 192–204
10. Syverson, P.F., Tsudik, G., Reed, M.G., Landwehr, C.E.: Towards an analysis of onion routing security. Workshop on Design Issues in Anonymity and Unobservability. (2000)
11. Wright, M., Adler, M., Levine, B.N., Shields, C.: An analysis of the degradation of anonymous protocols. Network and Distributed System Security Symposium. (2002)
12. Dingledine, R., Freedman, M.J., Hopwood, D., Molnar, D.: A reputation system to increase MIX-net reliability. Information Hiding: 4th International Workshop, IHW 2001. Volume 2137 of LNCS. (2001) 126–140
13. Dingledine, R., Syverson, P.: Reliable MIX cascade networks through reputation. Sixth International Financial Cryptography Conference (FC 2002). (2002)
14. International Organization of Standardization (ISO): International standard ISO/IEC 15408 (1999) Technically identical to Common Criteria version 2.1.
15. Garey, M.R., Johnson, D.S.: Computers and Intractability – A Guide to the Theory of NP-completeness. W. H. Freeman and Co. (1979)
16. Winick, J., Jamin, S.: Inet-3.0: Internet topology generator. Technical Report CSE-TR-456-02, University of Michigan (2002)
17. Harary, F.: Graph Theory. Perseus Publishing (1995)