Local View Attack on Anonymous Communication

Marcin Gogolewski, Marek Klonowski, and Mirosław Kutyłowski

Institute of Mathematics and Computer Science, Wrocław University of Technology,
ul. Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
klonowski@im.pwr.wroc.pl
Miroslaw.Kutylowski@pwr.wroc.pl
marcing@amu.edu.pl**

Abstract. We consider anonymous communication protocols based on onions: each message is sent in an encrypted form through a path chosen at random by its sender, and the message is re-coded by each server on the path. Recently, it has been shown that if the anonymous paths are long enough, then the protocols provide provable security for some adversary models. However, it was assumed that all users choose intermediate servers uniformly at random from the same set of servers.

We show that if a single user chooses only from a constrained subset of possible intermediate servers, anonymity level may dramatically decrease. A thumb rule is that if Alice is aware of much less than 50% of possible intermediate servers, then the anonymity set for her message becomes surprisingly small with high probability. Moreover, for each location in the anonymity set an adversary may compute probability that it gets a message of Alice. Since there are big differences in these probabilities, in most cases the true destination of the message from Alice is in a small group of locations with the highest probabilities.

Our results contradict a quite common belief that the protocols mentioned guarantee anonymity provided that the set of possible intermediate servers for each user is large.

1 Introduction

1.1 Background

There is a growing need for anonymity in electronic communication. Many anonymity protocols have been proposed – their aim is not only to hide the contents of messages sent, but also who is communicating with whom.

Application area of such protocols is much broader than implementing point-to-point anonymous communication. For instance, they are essential components of various voting schemes [3],some auction protocols [15], anonymous browsing or even they serve as a building block in some secure function evaluation protocols [11].

Existing solutions are generally based on two fundamental ideas - *MIXes* introduced by David Chaum in [2] and *Onions* that appeared in a number of papers [16, 17, 8]. In these protocols the messages are routed through servers called *MIXes* or *MIX-servers*.

** contact author

Generally, there is a common idea behind both solutions – at the beginning we have a batch of ciphertexts submitted by all users. Then a number of steps is executed. During a step each message is appointed to some server, each server recodes the messages obtained and returns them in a random order. In this way, the encoded messages become more and more “mixed”.

The problem is that users of many anonymity systems are requested to construct a path of randomly and independently chosen servers for each message, called in this paper *anonymity path*. It concerns all onion-based protocols [16, 17, 8] and some MIX network architectures. For such protocols, it is assumed that the users have the same knowledge about the servers that can be used as intermediate servers on the path.

In a large and dynamic system it is hard to achieve that all users have the same view of the network. The problem we address in this paper is how secure are anonymous communication protocols, if the users choose the servers for anonymity paths from different sets of servers.

Related Work There are many papers concerning anonymous communication protocols based on MIXes and Onions. On the other hand, there are only few papers providing rigorous proofs about immunity of such systems against an adversary. One of the very first work in this direction is due to Kesdogan et al. [12]. In this paper cardinality of so-called anonymity set was used as an anonymity measure.

In other papers like [16, 4, 1, 10] sophisticated and very restrictive anonymity measures were used that take into account also correlation between messages. These papers provide rigorous proofs of anonymity in different adversary models: their goal was to show that an anonymity measure reaches appropriate values with high probability for a certain length of an anonymity path.

Still, these proofs use the assumption that all users have exactly the same knowledge of the servers that may be used as intermediate servers on anonymity paths. No attempt has been made to analyze what happens if this assumption is not satisfied. In [7], it has been mentioned that violating this assumption might influence security of the system. However, a common belief is that anonymity of a user A is in danger only if the set of potential intermediate servers is small.

Paper Organization In Section 2 we recall basic facts about anonymous communication protocols. In Section 3 we consider dangers that arise when a user is aware of some extra servers and may use them for creating anonymous paths. Section 4 is devoted to the case when a user is aware of only a constant fraction of servers.

2 Anonymity Protocols and Problem Statement

MIXes MIX is a cryptographic primitive introduced in [2]: assume that users $1, 2, \ldots, n$ wish to publish anonymously messages $m_1, m_2, \ldots, m_n$. For this purpose they submit their messages to a *MIX-server* after encrypting them with MIX-server’s public key k, that is, they submit $E_k(m_1), E_k(m_2), \ldots, E_k(m_n)$ .The MIX-server decrypts the ciphertexts obtained with its private key, chooses a permutation π uniformly at random, and outputs $m_{\pi(1)}, m_{\pi(2)}, \ldots, m_{\pi(n)}$.

If the MIX-server is honest (i.e. does not reveal permutation π), then for an external observer the relation between the input and the output of the MIX-server remains hidden. Of course, some additional requirements must be fulfilled. For example, E has to be a probabilistic encryption scheme - otherwise one can establish this relationship by encrypting the output. For further details see [9].

In order to avoid full dependence on a single MIX, systems consisting of many MIX-servers were proposed. A so-called *MIX cascade* was introduced together with MIXes in [2]. In that case a message is encrypted multiple times with public keys of consecutive MIXes. The encrypted messages are processed by the cascade of MIXes – each MIX removes one encryption layer and permutes the ciphertexts at random.

For better scalability a parallel MIX-cascades can be used: in this case each step is executed by a number of MIXes working in parallel.

Onions It is a core idea of various theoretical systems as well as working implementations (see e.g. [16, 17, 8, 7]). There are many variants of this protocol. A basic one works as follows:

We assume that a message m has to be sent from a node A to a node B. For this purpose node A chooses at random λ intermediate nodes, say, $J_1, J_2, \dots, J_\lambda$ and random strings $r_1, r_2, \dots, r_{\lambda+1}$. Then an *onion* O is built according to the following recursive formula (Enc_X means encryption with the public key of X):

$$\begin{aligned} O_\lambda &= \mathrm{Enc}_B(m, r_{\lambda+1}) \\ O_i &= \mathrm{Enc}_{J_i}(J_{i+1}, O_{i+1}, r_{i+1}) \quad \text{for } i < \lambda \\ O &= O_1 \end{aligned}$$

Then O is sent by A to J_1. Node J_1 "peels off" the first layer by decryption and receives onion O_2, the name of the next server on the path, J_2, and a random string. Then J_2 becomes O_2 and processes it in a similar way. This procedure is repeated until the plaintext m appears after decryption.

Anonymity mechanism of onions is very similar to MIXes. Messages entering the same server at in the same time are recoded and permuted at random - just as for a MIX.

Adversary Model There are many adversary models for anonymity protocols. We consider a passive adversary monitoring traffic of messages in a network. The adversary cannot influence the traffic (for example: insert, duplicate, remove, or modify messages).

The adversary considered in this this paper is global, in the sense that he can eavesdrop at the same time all connections, but can neither corrupt a server nor trace its internal work.

Let us remark that the strongest anonymity results were obtained for the model introduced by Berman et al. in [1], where only a fraction of connection is under adversary's control.

Definitions of Anonymity There are many definitions and measures of anonymity (see e.g. [5] or [13]). The very first definition and the weakest one is based on already mentioned *anonymity set* described in [12]: We consider a single message A from the input of a system. Then we consider the set of all output positions that, from the point of view of the adversary, may contain recoded A with a positive probability. Cardinality of this set divided by number of all messages processed by the system is a measure of anonymity.

The definition based on anonymity set does not take into account that different output positions can be more or less likely to be linked with a particular input. This shortcoming is solved by definition based on entropy [5]. Unfortunately even this definition is not perfect - it does not take into account dependencies between probability distribution of different messages. These dependencies can be crucial for some applications.

The strongest definitions are based on *variation distance* between distribution of all possible permutations of input messages on output positions and the uniform distribution or a priori distribution [1]. Then all dependencies between different messages are taken into account.

In this paper we use the weak measure based on anonymity set. The reason is simple – we show that in some situations even for this weak measure only a low anonymity level is achieved.

Local versus Global View A common assumption in papers dealing with onions as well as MIXes is that the servers on anonymity path are chosen independently uniformly at random over the same set by all users.

It is often believed that even if the users choose from different sets of servers, it does not impact anonymity very much provided that the number of potential servers for each user is sufficiently large. We show that this intuition is wrong – different local views of the network can cause degradation of anonymity in some cases, despite strong results for the case when all local views are the same [1, 10]. This has important practical implications, since it is extremely difficult to provide the same view of the network in a large dynamic system with servers joining and leaving the network.

3 Dangers of Using Extra Servers

In this section we consider a simple scenario, in which anonymity breaks completely down or at least is strongly limited. Our considerations here serve as a kind of warm-up before the next section with a more involved analysis for a more practical setting.

We consider the case with n users, each of them sending a single message. The number of servers in the system that can be used as intermediate nodes is also n. However, all users, except Alice, know only $n-k$ of these servers, while Alice is aware of all of them. Let k servers known only to Alice be called *additional* servers. (Our choice of parameters might be different, for instance each server may send more than one message, but we fix the setting for the sake of simplicity.)

The messages are processed as onions. Each sender fixes a random path of length λ choosing each server independently and uniformly at random. So Alice may choose additional servers while the other servers cannot use them. We consider here a global passive adversary who wants to detect the destination of the message sent by Alice. We assume that for each single user the adversary knows the set of servers in the system known by the user. So in particular, the adversary knows that if an onion is processed through an additional server, then it must be an encoded message of Alice. If the message of Alice does not go through an additional server, then it remains hidden inside the crowd of other messages. However, even then its location might be limited to a small anonymity set, when the message went recently through an additional server.

By evaluating level of anonymity provided by various systems based on onions, the crucial question is how long must be the random path of each message (see [1], [10]). The main idea is that anonymity improves when the length of the random path increases. However, we shall see that it is false for the scenario considered here.

Let D be a random variable denoting the number of steps between the last moment when the message from Alice hits an additional server and delivery of this message. We call D *effective length* of anonymity path. Since each time a message hits an additional server the adversary knows that it is a message from Alice, for providing anonymity against a global adversary only the *effective length* counts and not λ. For this reason we analyze behavior of the random variable D.

Claim 1 *For each t, $0 \leq t \leq \lambda$,*

$$\Pr[D > t] = \left(1 - \tfrac{k}{n}\right)^{t+1} .$$

Indeed, treating probability of hitting an additional server as a failure in Bernoulli trails we obtain:

$$\begin{aligned}
\Pr[D > t] &= \Pr[D = t+1] + \Pr[D = t+2] + \ldots + \Pr[D = \lambda] \\
&= \left(1 - \tfrac{k}{n}\right)^{t+1} \cdot \tfrac{k}{n} + \left(1 - \tfrac{k}{n}\right)^{t+2} \cdot \tfrac{k}{n} + \ldots + \left(1 - \tfrac{k}{n}\right)^{\lambda-1} \cdot \tfrac{k}{n} + \left(1 - \tfrac{k}{n}\right)^{\lambda} \\
&= \left(1 - \tfrac{k}{n}\right)^{\lambda} + \left(1 - \tfrac{k}{n}\right)^{t+1} \cdot \left(1 - \left(1 - \tfrac{k}{n}\right)^{\lambda-t-1}\right) \\
&= \left(1 - \tfrac{k}{n}\right)^{t+1} .
\end{aligned}$$

Let us note that $\Pr[D > t]$ does not depend on parameter λ, except for the maximal range of the random variable D.

Now we can compute the expected value of D:

$$\begin{aligned}
\mathrm{E}[D] &= \textstyle\sum_{t=0}^{\lambda-1} \Pr[D > t] \\
&= \textstyle\sum_{t=1}^{\lambda} \left(1 - \tfrac{k}{n}\right)^{t} = \left(\tfrac{n}{k} - 1\right) \cdot \left(1 - \left(1 - \tfrac{k}{n}\right)^{\lambda}\right) < \tfrac{n}{k} .
\end{aligned}$$

Let us discuss these estimations of D. First assume that $k/n = \frac{1}{4}$. Then $D > 4$ with probability lower than 0.24. Hence, also anonymity set of the Alice's message is very small with high probability. If $k = \frac{n}{\log n}$ (which is a more realistic scenario), then $\mathrm{E}[D] < \log n$. Moreover,

$$\Pr[D > \log n - 1] < \left(1 - \frac{1}{\log n}\right)^{\log n} \approx \frac{1}{e}$$

So, in majority of cases effective length of anonymity path is below $\log n$. On the other hand, for a global adversary model a guaranteed level of anonymity is reached for $\lambda = \Theta(\log^2 n)$ [4] (in fact, after slight changes in the protocol are done). So it may happen that for a given value of k it is impossible to reach a high anonymity level – increasing λ in this case does not help at all since the effective length of anonymity path essentially will not increase. Sad but true!

4 Dangers of a Limited Local View

In this section we consider the case that all users except Alice are choosing intermediate servers from a set N, while Alice is aware of only a subset of N of cardinality $c \cdot |N|$, for some $c < 1$. Later in this section we discuss shortly the case that each user has some limited knowledge of the servers from N.

We consider a global passive adversary who knows N and the sets of servers known by Alice. The goal of the adversary is to determine the destination of a message sent by Alice based on information gained from observing the traffic.

We shall show that anonymity set of the message of Alice might be surprisingly small and therefore the protocol offers a low level anonymity against a global adversary. A very important point is that increasing the length of anonymity path <u>does not help much</u>: after an initial phase the size of the anonymity set fluctuates around a relatively small value.

These results are quite surprising in view of the results concerning the case when all users choose intermediate servers uniformly at random from the same set N. Namely, then increasing the path length <u>improves</u> anonymity level so that finally we get very strong anonymity expressed by <u>a small</u> variation distance between distribution of all permutations of messages and the stationary distribution. There are also results suggesting that the necessary path length is relatively small [16, 4] even in the case of presence of global passive adversary.

4.1 Process definition

N denotes the set of all servers that can be used as intermediate servers on anonymity paths. Let $W \subset N$ be the set of servers known to Alice. Let M be the set of messages sent by all users. We assume that $|M| = |N|$ and that each user sends one message.

At each step of the protocol the adversary may observe positions of M encoded messages, but the problem is to indicate the position of the message sent by Alice – the messages are recoded at each step in such a way that if two or more of them enter the same server, they cannot be linked to the messages leaving this server after recoding. It is exactly the same mechanism as in the case of a MIX-sever [2].

Let a set N_i be the set consisting of all messages u after step i such that it is possible that the message u is the recoded message sent by Alice. More precisely, there exist a (hypothetical) linking between messages entering and leaving each server so that the message sent by Alice at the first step leads to message u after step i. In other words, from the adversary point of view, it cannot be excluded that u is a recoded Alice's message and N_i is the anonymity set of the Alice's message after step i.

Let S_i be the set of all servers where the messages of N_i occur.

At the very beginning every user sends exactly one message. So $|N_0| = 1$, since the adversary exactly knows which message is sent by Alice. Let us consider step i of the protocol. Our goal is to estimate the size of N_i based on the size of N_{i-1}. The set N_i consists of two kinds of messages:

the first kind: the messages that were in N_{i-1} and are sent at step i to servers within W,

the second kind: the messages that were outside N_{i-1}, but went to some servers, where the message of the first kind occur after step i.

Let us note that there is at least one message from N_{i-1} that remains within W, namely, the encrypted message of Alice. The remaining messages from N_{i-1} that go to servers in the set $N \setminus W$ cannot hold the message from Alice.

In order to estimate the number of messages of the second kind we have to find cardinality of the set S_i. The random variable denoting the size of S_i is given by a combination of binomial distribution and so-called bins and balls process. Let $\mathrm{Bin}(v,p)$ be a random variable denoting the number of successes in a Bernoulli process with v trials and success probability p for a single trial. Let $\mathrm{BiBa}(v,u)$ be a random variable denoting the number of non-empty bins (i.e. with at least one ball) after throwing v balls uniformly and independently at random into u bins.

It is easy to see that the number of the messages from N_{i-1} which remain within W at step i is given by the random variable:

$$\mathrm{Bin}(|N_{i-1}|-1, |W|/|N|)+1$$

($|N_{i-1}|-1$ messages are not from Alice, so each of them chooses to stay within W with probability $|W|/|N|$, the term "$+1$" corresponds to the message of Alice). So finally the size of S_i is a random variable with the same distribution as

$$\mathrm{BiBa}(\mathrm{Bin}(|N_{i-1}|-1, |W|/|N|)+1, |W|) \ .$$

A message from the set $M \setminus N_{i-1}$ becomes a member of N_i, if at step i it hits one of the servers of S_i. So the number of messages of the second kind joining N_i is described a random variable with binomial distribution

$$\mathrm{Bin}(|M|-|N_{i-1}|, |S_i|/|N|) \ .$$

Finally, we have got the following recursive formulas on random variables:

$$\begin{aligned}
|S_i| &= \mathrm{BiBa}(\mathrm{Bin}(|N_{i-1}|-1, |W|/|N|)+1, |W|) \\
|N_i| &= \mathrm{Bin}(|M|-|N_{i-1}|, |S_i|/|N|) + \mathrm{Bin}(|N_{i-1}|-1, |W|/|N|)+1. \\
|N_0| &= |S_1| = 1 \ .
\end{aligned}$$

In fact, in the above formulas the sign "$=$" means that the random variables on the left and right side have the same probability distribution.

Our goal is to estimate the size of N_λ, which is anonymity set of the message of Alice after λ steps of processing the messages.

4.2 Analysis

Let us recall that the expected value of a random variable $\mathrm{BiBa}(v,u)$ equals

$$u(1-(1-\tfrac{1}{u})^v) \approx u(1-e^{-\frac{u}{v}}) \ .$$

The expected value of random variable $\mathrm{Bin}(u,p)$ is $u \cdot p$.

Since we assume that $|N| = |M|$, we simplify the formulas by substituting $|N|$ and $|M|$ by a single symbol n. Hence the expected sizes of the sets S_i and N_i are expressed by the following formulas:

$$\mathrm{E}[|S_i|] \approx |W|\Big(1 - \Big(1 - \frac{1}{|W|}\Big)^{\frac{(|N_{i-1}|-1)\cdot|W|}{n}+1}\Big)$$

and

$$\mathrm{E}[|N_i|] \approx \frac{(|M| - |N_{i-1}|)\mathrm{E}[|S_i|] + (|N_{i-1}| - 1)|W|}{n} + 1 \; .$$

(In the first formula we have written $\approx$ instead of $=$, since we have assumed that the number of messages from N_{i-1} that remain in W equals to the expected number of such messages. Similarly, in the second formula we have replaced $|S_i|$ by $\mathrm{E}[|S_i|]$.) After applying the approximation $(1 - \frac{1}{a})^b \approx e^{-b/a}$, we get

$$E[|N_i|] \approx \tfrac{n-|N_{i-1}|}{n} \cdot |W| \cdot \left(1 - e^{-\frac{|N_{i-1}|-1}{n} - \frac{1}{|W|}}\right) + (|N_{i-1}| - 1) \cdot \tfrac{|W|}{n} + 1 \; .$$

Let $\Delta(|N_{i-1}|) = E[|N_i| - |N_{i-1}|]$. Hence

$$\begin{aligned} \Delta(|N_{i-1}|) \approx \;& (n - |N_{i-1}|) \cdot \frac{|W|}{n} \cdot (1 - e^{-\frac{|N_{i-1}|-1}{n} - \frac{1}{|W|}}) \\ & -|N_{i-1}| \cdot (1 - \frac{|W|}{n}) + (1 - \frac{|W|}{n}) \end{aligned} \tag{1}$$

We consider $\Delta(|N_{i-1}|)$ as a function of $|N_{i-1}|$ and we fix the value of $\frac{|W|}{|N|}$. We see that the first term in (1) forces Δ to be positive and its impact is bigger for small values of $|N_{i-1}|$. The second term in (1) forces Δ to be negative and its impact grows with the size of $|N_{i-1}|$. So, there is a point where both tendencies have the same strength and so the values of $|N_t|$ oscillate around it. Of course, this equilibrium point depends on $\frac{|W|}{|N|}$.

If we put $\alpha = \frac{|W|}{|N|}$, then we would like to find a value $|N_i|$ such that according to (1), $\Delta(|N_{i-1}|) = 0$. This concerns solving the equation:

$$0 = (n - x) \cdot \alpha \cdot (1 - e^{-\frac{x-1}{n} - \frac{1}{\alpha n}}) - x \cdot (1 - \alpha) + (1 - \alpha) \; . \tag{2}$$

We have not found any closed formula for solutions of Equality 2 (as well as some symbolic computation systems). However, one can easily find the solutions numerically. The results for $n = 1000$ and different values of $|W|$ are plotted on Figure 1.

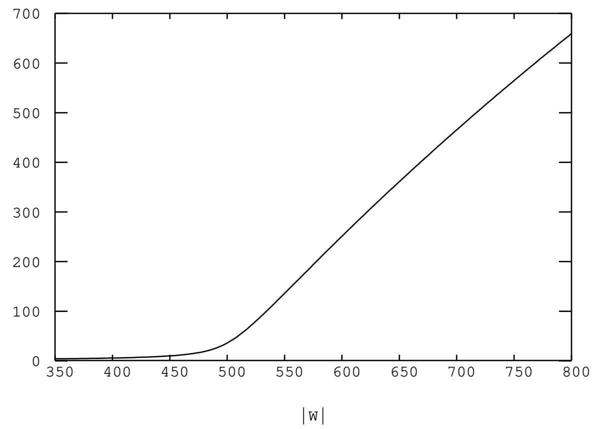

Fig. 1. Zero points of Equality 2 as a function of $|W|$ for $n = 1000$

From the numerical results we can learn a somewhat unexpected phenomena. For values of $|W|$ that are significantly lower than $500 = 0.5n$, the equilibrium point has quite small values and the growth rate is quite slow. Around $0.5n$ there is a radical change of the situation: the growth rate increases until the derivative reaches the value close to 0.5. Then, the function is quite well interpolated by a single line. These results suggest that the sizes of anonymity sets N_i remain small for values of $|W|$ that are not too close to $0.5n$. Then the situation changes abruptly and we should observe that the average size of N_i grows fast with $|W|$.

In the next subsection we compare these results with extensive simulations of the protocol.

4.3 Simulation Results

From the previous considerations we can expect that after some number of steps at the beginning of the protocol anonymity set $|N_t|$ should oscillate around a certain value. Since exact formulas describing the stochastic changes of $|N_t|$ or even their fair approximations seem to be very complex, we performed a number of direct simulations to check these tendencies.

We have performed the experiments for different values of $|N|$, but the results hardly depend on $|N|$, except for small values that are not interesting from our point of view. We have performed simulations for different values of $\frac{|W|}{|N|}$.

The next two figures show dependency between $|W|$ and $|N_i|$ in subsequent steps of the protocol. The figures correspond to the cases $|N| = |M| = 1000$ for $0.3 \cdot |N| \leq |W| \leq |N|$.

Figure 2 is a three dimensional plot. The value plotted is the size of N_i in subsequent steps of the protocol. Different curves correspond to different values of $|W|$ between 30% and 100% of $|N|$ (step 2%).

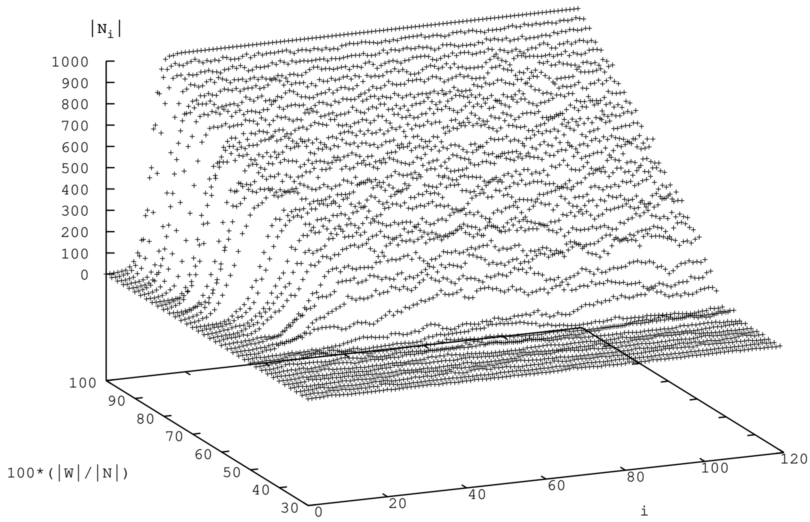

Fig. 2. Simulation results - the size of N_i at different moments for different values of $\frac{|W|}{|N|}$

From Figure 2 one can see that for some values of $\frac{|W|}{|N|}$ anonymity set remains small for all i. Then there is a cut-off point where $|N_i|$ starts to grow until it reaches a stable level. This level depends on $\frac{|W|}{|N|}$, just as predicted before (see Figure 1).

Figure 3 presents planar visualization of the same simulation for $|W| = 0.4 \cdot |N|$, $0.5 \cdot |N|$, $0.6 \cdot |N|$, $0.7 \cdot |N|$, $0.8 \cdot |N|$, $0.9 \cdot |N|$. At this point our aim is to convince the reader that there is qualitative gap between anonymity for these values of parameter $|W|$ and relation between sizes of $|W|$ and the average size of N_i is <u>not</u> linear. For $|W| = 0.4 \cdot |N|$, $0.5 \cdot |N|$ the anonymity set has a very small size even if some small deviations occur.

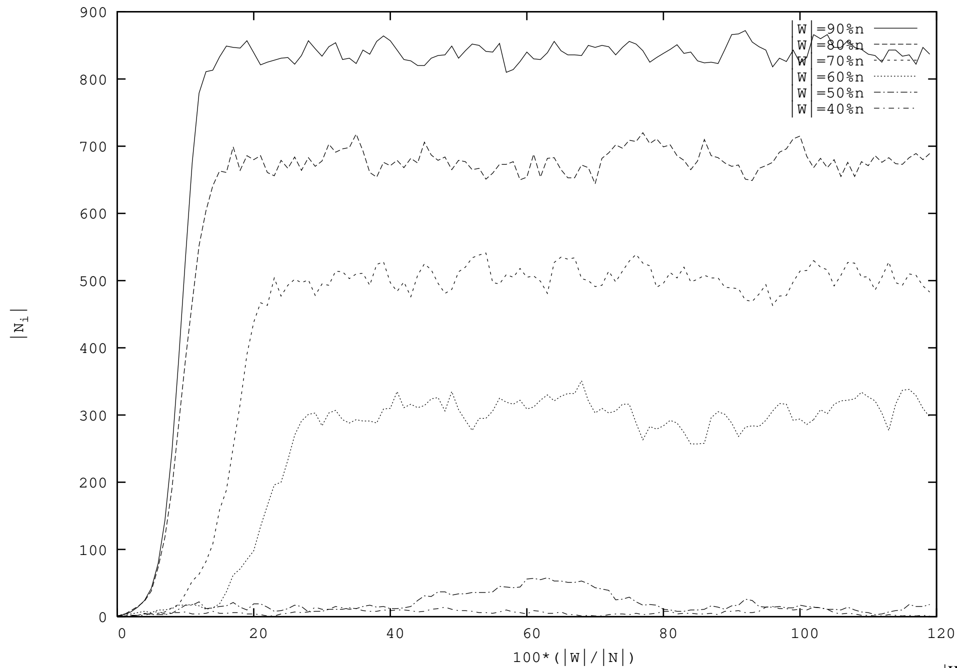

Fig. 3. The size of N_i for the same simulations - the results for different values of $\frac{|W|}{|N|}$

4.4 Statistical analysis

So far we were concerned with the size of N_i only, ignoring probability distribution that for each message from N_i describes the chance that it is the message sent by Alice. It turns out that this probability distribution is highly nonuniform. This property reduces anonymity level of the protocol even more.

We consider the case: $n = 1000$ and the situation after 20 steps. Figure 4 shows the probabilities mentioned for the case when $W = N$, after sorting them. The dashed line shows probability of the message from Alice. As one can expect, the probability distribution is not completely uniform – each probability depends on the number and location of paths leading from Alice to each position after step 20. It is not a big surprise that the message of Alice is in a position that is hard to guess from the probability distribution.

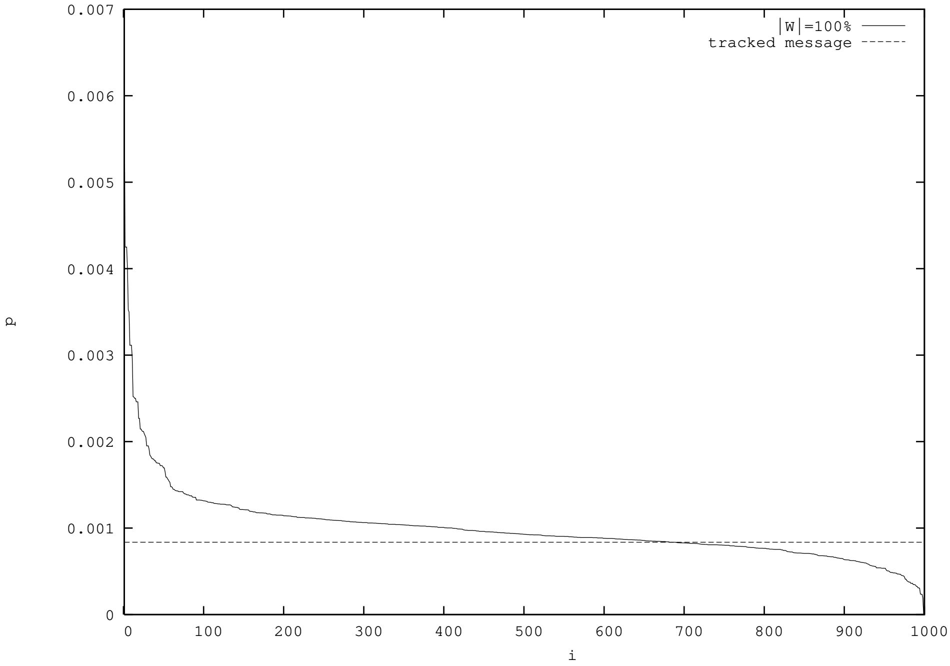

Fig. 4. Probabilities after simulation of 20 steps for the case $W = N$

Figure 5 describes the simulation results for the same $|N| = 1000$, but this time the size of W is $0.7n$. In this case the anonymity set N_{20} is quite big (near 60% of N), so finding N_{20} does not help much. However, the actual Alice's message is very likely to be within 3% positions with the highest probabilities.

It is worth to say that such a reduction of the set of suspects for being the message from Alice based on probabilities works when we speed up the computations and determine the probabilities in an incorrect way. Namely, if some message from N_{i-1} leaves W, then we do not recompute the probabilities of all members of the anonymity set based on the routes that have occurred in the past, but simply re-scale all probabilities.

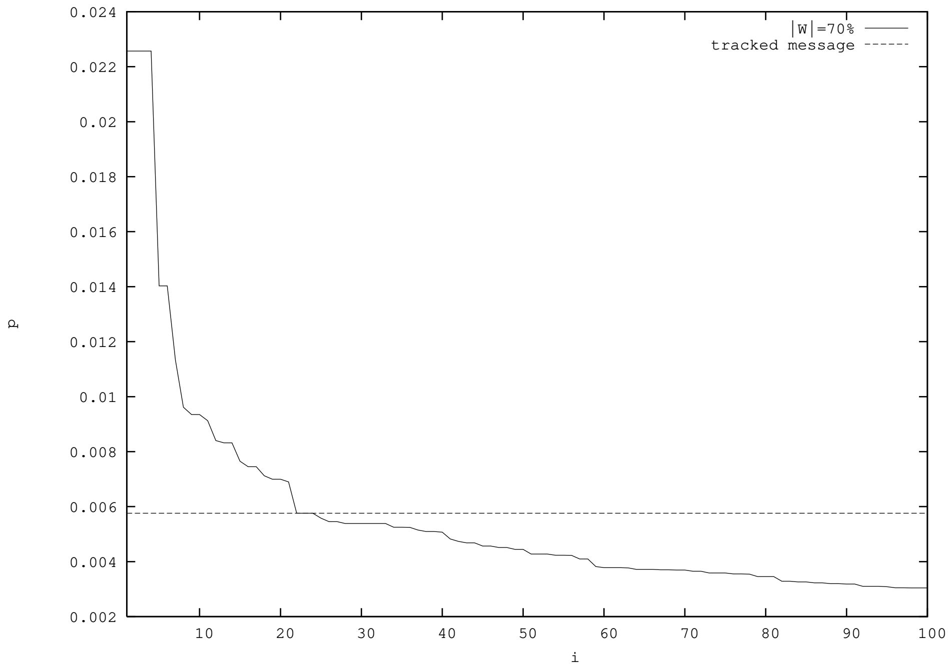

Fig. 5. Probabilities after simulation of 20 steps for the case $|W| = 0.7|N|$

4.5 Attack extensions

Intersection attack Let us assume that Alice sends many messages to the same destination in a row. Then we may apply intersection attack [6] and reduce rapidly the anonymity set. The attack has good chances to succeed, since in different rounds the anonymity sets obtained are almost independent, except that the true destination is always the member of the anonymity set. For instance, assume that $|N_\lambda| = |S_\lambda| = 30$, $N = 1000$, and that during the next execution $|N_\lambda| = 30$. Then the size of intersection of both anonymity sets is given by the random variable $\mathrm{Bin}(29, 30/1000) + 1$ and has expected value about 2. Of course, it does not help Alice, if she sends the messages through different routes.

Model extension So far we have assumed that all servers except Alice choose the intermediate servers uniformly at random from the whole set N. However, one can see that essentially the same attack can be used with high probability if the sets of intermediate servers of the users are chosen in a stochastically independent way. On the other hand, if Alice shares the same set of intermediate servers with other users, then at least it becomes hard to distinguish between them.

5 Countermeasures and Conclusions

A common intuition is that in order to achieve a better level of anonymity each user should use as many servers as possible for choosing intermediate servers on anonymity paths. However, in large and dynamic networks this will lead inevitably to a situation that different users will use different sets of servers. Some of them will stay behind and use relatively few servers and some will be fast in changes and use a larger set of servers. As we have shown, both cases are dangerous. The threats do not disappear even if the sizes of the sets used by different users are the same: if a set of servers used by a user is "independent" from the sets used by the other users, then the same attack applies.

The problems may be avoided if the sets of servers used by different users are the same. However, it is hard to achieve in a dynamic, large scale network without a central control. So it is a strong argument to use small collections of servers that are stable and remain in service as anonymizers for a long time. However, this is a step from a fully distributed system with no control to a more centralized one.

The third solution could be to use onion encoding with so called-navigators [14]. In this case the anonymity paths are chosen dynamically: a skeleton is established by the user, but subpaths are determined on-the-fly by the servers on the route. In this case, if a particular onion leaves the set W discussed before, then there is no reason to exclude this onion from anonymity set of the message sent by Alice.

References

1. Berman, R., Fiat, A., Ta-Shma, A.: Provable Unlinkability Against Traffic Analysis. Financial Cryptography 2004, Lecture Notes in Computer Science 3110, Springer-Verlag: 266-280
2. Chaum, D.: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Communication of the ACM 24(2)(1981): 84-88
3. Chaum, D.: Secret-Ballot Receipts and Transparent Integrity. Better and less-costly electronic voting and polling places. Available at `http://theory.lcs.mit.edu/~rivest/voting/papers/Chaum-SecretBallotReceiptsTrueVoterVerifiableElections.pdf`
4. Czumaj, A., Kanarek, P., Kutyłowski, M., Loryś K.: Distributed Stochastic Processes for Generating Random Permutations. ACM-SIAM Symposium on Discrete Algorithms (SODA) '99, 271-280
5. Danezis, G., Serjantov A.: Towards an Information Theoretic Metric for Anonymity. Privacy Enhancing Technologies 2002, Lecture Notes in Computer Science 2482, Springer-Verlag: 41-53
6. Danezis, G., Serjantov, A.: Statistical Disclosure or Intersection Attacks on Anonymity. Information Hiding '2004, Lecture Notes in Computer Science 3200, Springer-Verlag: 293-308
7. Dingledine, R., Mathewson, N., Syverson P.: Tor: the Second Generation Onion Router. USENIX Security, 2004
8. Gülcü, C., Tsudik, G.: Mixing E-mail with BABEL. ISOC Symposium on Network and Distributed System Security, IEEE 1996: 2-16
9. Goldschlag, D. M., Reed, M. G.,Syverson P.: Private Web Browsing. Journal of Computer Security, Special Issue on Web Security 5(1997): 237-248

10. Gomułkiewicz, M., Klonowski, M., Kutyłowski, M.: Provable Unlinkability Against Traffic Analysis Already After $O(\log(n))$ Steps! Information Security Conference 2005, Lecture Notes in Computer Science 3381, Springer-Verlag: 229-238
11. Jakobsson, M., Juels, A.: Mix and Match: Secure Function Evaluation via Ciphertexts. Advances in Cryptology - Asiacrypt 2000, Lecture Notes in Computer Science 1976, Springer-Verlag: 162–177
12. Kesdogan, D., Egner, J., Büschkes, R.: Stop-and-Go-MIXes Providing Probabilistic Anonymity in an Open System. Information Hiding '1998, Lecture Notes in Computer Science 1525, Springer-Verlag: 83-98
13. Köhntopp, M., Pfitzmann, A.: Anonymity, Unobservability, and Pseudonymity: A Proposal for Terminology. Workshop on Design Issues in Anonymity and Unobservability 2000, Lecture Notes in Computer Science 2009, Springer-Verlag: 1-9
14. Klonowski, M., Kutyłowski, M., Zagórski, F.: Anonymous Communication with On-line and Off-line Onion Encoding. SOFSEM'2005, Lecture Notes in Computer Science 3381, Springer-Verlag: 229-238
15. Kurosawa, K., Ogata, W.: Bit- Slice Auction Circuit. ESORICS 2002, Lecture Notes in Computer Science 2502, Springer-Verlag: 24-38
16. Rackoff, C., Simon, D.R.: Cryptographic Defense Against Traffic Analysis. ACM Symposium on Theory of Computing (STOC) '25 (1993): 672-681
17. Syverson, P., Reed, M. G., Goldschlag, D. M.: Anonymous Connections and Onion Routing. IEEE Journal on Selected Areas in Communication, 1998, 16(4):482-494