ON THE ANONYMITY OF TIMED POOL MIXES

Andrei Serjantov
University of Cambridge
Computer Laboratory
15 JJ Thomson Avenue
Cambridge CB3 0FD, United Kingdom
Andrei.Serjantov@cl.cam.ac.uk

Richard E. Newman
Department of Computer & Information Science & Engineering
PO Box 116120 University of Florida
Gainesville, FL 32611-6120, USA
nemo@cise.ufl.edu

Abstract This paper presents a method for calculating the anonymity of a timed pool mix. Thus we are able to compare it to a threshold pool mix, and any future mixes that might be developed. Although we are only able to compute the anonymity of a timed pool mix after some specific number of rounds, this is a practical approximation to the real anonymity.

1. Introduction

Many anonymity systems use the notion of a mix as introduced by Chaum (Chaum, 1981). The purpose of a mix is to hide the correspondence between incoming and outgoing messages, so that the attacker (who is not able to see the inner workings of a mix) cannot tell who sends messages to whom.

Thus, a mix is a proxy that collects some number of messages inside it (thus introducing a delay), "mixes them up" and forwards them on. There are two fundamental characteristics of a mix: the anonymity it provides, or roughly speaking, the number of messages it collects, and the time for which it delays messages. The former should be maximized while minimizing the latter.

Mixing can be done in a variety of ways. For example, a mix may wait for a particular number of messages to arrive (threshold mix) before forwarding the messages on, or a particular time interval (timed mix). Chaum's original system used a simple threshold mix, but over the last few years many mixes have been proposed in the literature (Kesdogan et al., 1998; Jerichow, 2000; Cottrell, 1994). A survey of some mixing strategies with an emphasis on their properties under active $(n-1)$ attacks has been presented by Serjantov et al. (Serjantov et al., 2002). Although the minimum and maximum anonymity of several mixes were presented there, the average anonymity of timed mixes was not. In fact, the authors stated that the anonymity of a timed mix depends on the entire history of message arrivals at this mix, but do not go further in exploring this idea. In this paper, we look at this issue in detail and show how working out the anonymity of timed pool mixes can be achieved.

First, we describe the timed pool mix itself. We then describe the threshold pool mix and give an outline of a method which will enable a comparison of the two. We then proceed to give a general outline of how to analyse the anonymity of the timed mix. Finally, we describe our implementation of the analysis and give some suggestions for a fair comparison of the two mixes.

2. Description of the Timed Pool Mix

Timed Pool Mix

Parameters: t, period; n, pool.

Flushing Algorithm: The mix fires every t seconds. If N_i messages have arrived since the last time the mix fired, then a pool of n messages[1] chosen uniformly at random from the $N_i + n$ is retained in the mix. The others are forwarded. If $N_i = 0$, the mix does not send out any messages.

It is interesting to compare the mixing strategy of this mix to that of a threshold pool mix.

Threshold Pool Mix

Parameters: N, threshold; n, pool.

Flushing Algorithm: The mix fires when $N + n$ messages accumulate in the mix (or when N messages have arrived since the last time the mix fired). A pool of n messages, chosen uniformly at random from all the messages, is retained in the mix. The other N are forwarded on.

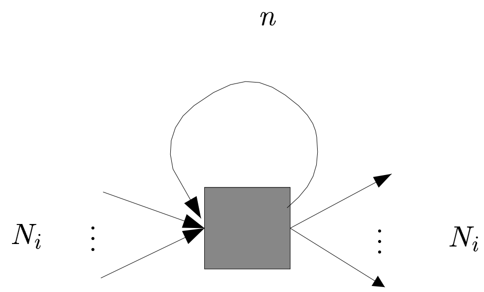

Figure 1. A Timed Pool Mix

It is clear that the anonymity set of the timed pool mix at round i (the set of senders that have a non-zero probability of having sent a message present in the mix at round i) includes the senders of all the messages that have passed through it since round 0. This is also the case for the threshold pool mix. Thus, using anonymity sets (or, to be more precise, the size of the anonymity sets) for measuring anonymity does not distinguish the two mix types.

Thus, we follow the approach taken in (Serjantov and Danezis, 2002)[2] and calculate anonymity of the timed pool mix using the information theoretic metric proposed in that paper. That work has already analysed the threshold pool mix, so that will not be repeated here.

3. Method

First of all, assume that all messages arrive at the pool mix directly from senders. Furthermore, for the purposes of comparison, assume that the senders of all the messages are distinct.

We proceed as follows:

Consider a message inside the mix at round r (we do not care whether this message leaves the mix or not). Now, calculate the probabilities that it had been sent by each of the senders who sent a message at round j, $j < i$. We now have a probability distribution of senders who could have sent the message. Taking the entropy, $\sum p \log p$, of this probability distribution, will give the anonymity. For a detailed (and a more general) definition of this information theoretic metric see Serjantov and Danezis, 2002.

Given the mix at round r and a history of message arrivals to the mix $[N_1, \ldots, N_r]$, let us calculate the probability of a message from rounds $1 \ldots r$ still being in the mix.

If the message was in the mix before the first flush (round 0), the probability of it staying until round r is:

$$p_0 = \prod_{i=1}^{r} \left(\frac{n}{N_i + n} \right)$$

The probability that a particular message that is in the mix at round r has entered the mix at round r is:

$$p_r = \frac{N_r}{N_r + n}$$

Thus, each of the senders (there were N_r of them) at round r sent this message with probability:

$$p_r = \frac{1}{N_r + n}$$

Similarly, the probability that a message that is in the mix at round r has entered the mix at round $r - 1$ is:

$$p_{r-1} = \frac{n}{N_r + n} \left(\frac{N_{r-1}}{N_{r-1} + n} \right)$$

Now we can calculate the anonymity of the entire probability distribution.

$$E_r = -\frac{N_r}{N_r + n} \log \frac{1}{N_r + n} - \left(\prod_{i=1}^{r} \frac{n}{N_i + n} \right) \log \left(\prod_{i=1}^{r} \frac{n}{N_i + n} \right) - \\ - \sum_{x=1}^{r-1} \left(\frac{N_x}{N_x + n} \left(\prod_{i=x+1}^{r} \frac{n}{N_i + n} \right) \log \left(\frac{1}{N_x + n} \prod_{i=x+1}^{r} \frac{n}{N_i + n} \right) \right)$$

At this point it may be helpful to refer back to Section 5 of Serjantov and Danezis, 2002 and observe how the above expression follows on from the one for anonymity of the threshold pool mix.

We can also derive the anonymity of a timed pool mix in a different way. Recall the formula for composition of mixes from Section 3.2 of Serjantov and Danezis, 2002:

$$E_{total} = E_{mix} + \sum_{0 < x \leq n} p_x E_x$$

Intuitively, this says that the anonymity of a message in the mix is the inherent entropy of the mix (E_{mix}) plus the mean anonymity of all the messages inside the mix.

A message in a pool mix at round i could have come from two places: from the pool remaining after the previous round (call the probability of this p_i) or from a round i sender. So,

$$p_i = \tfrac{n}{N_i+n} \quad \text{and} \quad 1 - p_i = \tfrac{N_i}{N_i+n}$$

Then, the inherent entropy of the mix at round i is

$$E_{i_{mix}} = -p_i \log p_i - (1 - p_i) \log(1 - p_i)$$

Using the above formula we can now obtain the anonymity of a timed pool mix after r rounds:

$$E_r = E_{r_{mix}} + p_r E_{r-1} + (1 - p_r) \log N_r$$

This can be rewritten as:

$$E_r = E_{r_{mix}} + (1-p_r) \log N_r + \sum_{x=1}^{r-1} E_{x_{mix}} \prod_{i=x+1}^{r} p_i + \sum_{x=1}^{r} (1-p_x) \log N_x \prod_{i=x+1}^{r} p_i$$

It is important to notice that to calculate the anonymity of a timed pool mix (using either method), we need to know the number of messages that had arrived at the mix during each of the previous rounds.

Now suppose we wish to analyse the anonymity of a timed pool mix at round R. It is clear that messages which have passed through the mix a long time ago will not contribute much to the current anonymity.[3] Thus, we can approximate the total anonymity by pretending that only r of the R rounds contribute to the anonymity.

Suppose that message interarrival times during time period T follow some probability distribution (an exponential distribution, for example) and that N messages arrive in total. Furthermore, take T to be some multiple of t, the time parameter of the mix, so that the N messages arrive over $r = \frac{T}{t}$ rounds. Now we can simply enumerate all the possible ways (we call these histories) that N messages can arrive in r rounds, and calculate the probability of each one. Now all that is left to do is to calculate the anonymity of each one of the histories, the probability of each history occurring given the exponential (or in fact any other) distribution and compute the mean anonymity.

4. Implementation and Results

Naturally, this is far too tedious to do by hand, so a short program found in Haskell (Peyton Jones et al., 1999) was written to enumerate all the possible histories, calculate their probabilities using an exponential distribution and determine the anonymity of the pool mix. Our program calculates anonymity using both methods presented above, and gives the same results in all our test cases. We hope to make the source code available for download in the near future.

Note that this approach is not as precise as the one used for the threshold pool mix – we were unable to come up with a closed form for the anonymity of a timed pool mix as the number of rounds grows large. However, we consider the method used above satisfactory and practical. Unfortunately, it relies on the correctness of the implementation.

This program enables a comparison with a threshold pool mix. We take the same scenario and calculate the anonymity of a threshold pool mix after r rounds.

We performed the comparison in the following settings: we took a timed pool mix with a pool of 2 messages over 5 rounds and with 10 messages arriving to it during these 5 rounds. The mean interarrival time for the Poisson process was taken to be 1.

We then compared it to a threshold pool mix with the same volume of traffic flowing through it over 5 rounds. We took the threshold to be 2 messages. Note that this corresponds to the anonymity of the timed mix when the history of the timed mix is $[2, 2, 2, 2, 2]$. The anonymity of the threshold pool mix was 2.91 bits, whilst the anonymity of the timed pool mix was 3.012 bits. Interestingly enough, if we model the arrival of messages using a zipf distribution, the anonymity of the timed pool mix is 3.07 bits.

We do not currently wish to make any general claims about the anonymity of timed mixes versus threshold mixes. Such a statement would be much more a product of the various assumptions we have made than the batching strategies themselves. Indeed, a rigorous comparison of the two mixes would also have to take account of the average delay of the messages through the two mixes, and possibly other factors. In this paper, we merely illustrate that the method which we use to work out the anonymity of a timed mix is powerful enough to enable such a comparison.

5. Related Work

In our opinion, properties of mixes have not been well studied. Many mixes have been proposed (Cottrell, 1994; Jerichow, 2000; Kesdogan et al., 1998), but rigorous descriptions of their properties are lacking.

For instance, Jerichow (Jerichow, 2000) describes "time controlled mixes" and "event controlled mixes" (timed and threshold mixes in our terminology), but only goes as far as providing qualitative statements about their anonymity.

The idea of using a Poisson process to model message arrivals to a mix is due to Kesdogan; he uses it in the analysis of SG mixes (Kesdogan et al., 1998).

6. Conclusions and Future Work

In this paper we have presented an analysis of the anonymity of the timed pool mix. Although we were unable to come up with a simple expression for the anonymity and our methods enumerated all the possible histories of message arrivals, the method still enabled us to do a comparison of the timed and the threshold pool mixes. The sensitivity of timed pool mixes to the distribution of interarrival times of the message arrival process is also of interest.

In the future, we would like to consider more complicated timed pool mixes such as the Cottrell mix (Cottrell, 1994; Serjantov et al., 2002), and a generalised mix where the number of messages sent out onto the network is a function of the number of messages inside it at the time it flushes. A practical method (and an implementation of it) for analysing timed mixes is an important step towards this goal.

We would also like to perform a rigorous comparison of the properties of timed and threshold pool mixes, including details about how they interact with other features of anonymity systems such as dummy traffic.

Notes

1. When the mix starts operating, the pool is filled up with n dummy messages

2. It is worth noting that a very similar metric is proposed in Diaz et al., 2002.

3. Implementors of mixes have also suggested that including a timeout clause such as "message should note be delayed for more than 10 rounds" would make them feel more comfortable(!).

References

Chaum, D. (1981). Untraceable electronic mail, return addresses and digital pseudonyms. *Communications of the A.C.M.*, 24(2):84–88.

Cottrell, L. (1994). Mixmaster and remailer attacks. `http://www.obscura.com/~loki/remailer/remailer-essay.html`.

Diaz, C., Seys, S., Claessens, J., and Preneel, B. (2002). Towards measuring anonymity. In *Privacy Enhancing Technologies*, volume 2482 of *LNCS*.

Jerichow, A. (2000). *Generalisation and Security Improvement of Mix-mediated Anonymous Communication*. PhD thesis, Technischen Universitat Dresden.

Kesdogan, D., Egner, J., and Buschkes, R. (1998). Stop-and-go-MIXes providing probabilistic anonymity in an open system. In *Proceedings of the International Information Hiding Workshop*.

Peyton Jones, S., Hughes, J., et al. (1999). Report on the programming language Haskell 98, a non-strict, purely functional language. `http://www.haskell.org/`.

Serjantov, A. and Danezis, G. (2002). Towards an information theoretic metric for anonymity. In *Privacy Enhacing Technologies*, volume 2482 of *LNCS*, San Francisco, CA.

Serjantov, A., Dingledine, R., and Syverson, P. (2002). From a trickle to a flood: Active attacks on several mix types. In *5th Workshop on Information Hiding*, volume 2578 of *LNCS*.