On the Anonymity of Message-based Anonymity Systems

Andrei Serjantov

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

In this paper we present a brief account of our work on analysing the anonymity of message-based anonymity systems based on free-route mix networks. A mix system consists of *senders* $a, a_1, \ldots$ who want to send messages to *receivers* $r, r_1, \ldots$. They try to preserve *unlinkability* (i.e. hide who is sending messages to whom) against a *global passive attacker* capable of observing the entire network by sending each of their messages via a sequence of *mixes*. Each mix collects n messages, performs a cryptographic transform on each, and sends the messages on, ensuring that the adversary is unable to identify which of the incoming messages to a mix corresponds to any of the outgoing ones. A message goes through a series of mixes (we do not discuss the details of routing, cryptographic operations, etc., see [Cha81] for details), and we wish to determine what the anonymity of the message is when it comes out.

What is anonymity and how should it be measured? We borrow the metric of [SD02], and say that the *sender anonymity* of a message arriving at a particular receiver is the entropy of the probability distribution of the senders who, from the point of view of the attacker, could have sent this message. The central contribution of our work, then, is providing a sound way of working out this probability distribution.

First of all, we construct a model of the mix network. We use a simple operational semantics. Given a set of messages which are sent into the network, our semantics determines the final state of the network and a "real" trace. The real trace represents the observation which would have been made by an attacker who could see the entire network and inside all the mixes. The attacker, in reality, cannot see inside the mixes. Hence, we erase the trace to retain only the observation which the global passive attacker can make.

The attacker can see the erased trace, and he does not know which of the possible real traces corresponds to the erased trace. We can then construct the set of all possible real traces which erase to the erased trace observed by the attacker. We take a message M which the attacker observes to have arrived at receiver R. In each of the real traces, we see who sent this message, and with what probability (for simplicity we leave out this complication). Having done this for the entire set of real traces, we can compute a probability distribution of the senders who could have sent the message M. The entropy of that probability distribution gives the anonymity.

We illustrate this point with an example. Suppose the attacker has observed the following erased trace *Obs*:

$$Obs = [(a,m); (a_1,m); (a,m); (a_1,m); \overline{(m,r,c)}; \overline{(m,r,c)}; \overline{(m,r_1,c')}; \overline{(m,r_1,c')}]$$

This represents a message entering the network from a, another from a_1, then the message from a arriving at mix m, then the message from a_1 arriving at mix m, then a message from m with contents c addressed to r leaving mix m, that message leaving the network, etc[1].

The set of (two) real traces which erase to Obs are:

$$
\begin{aligned}
All = [\ & [(a,([m],r,c));(a_1,([m],r_1,c'));(\tau,(a,([m],r,c)));(\tau,(a_1,([m],r_1,c')));\\
& (\tau,\overline{(m,(\varepsilon,r,c))});\overline{(m,(\varepsilon,r,c))};(\tau,\overline{(m,(\varepsilon,r_1,c'))});\overline{(m,(\varepsilon,r_1,c'))}],\\
& [(a,([m],r_1,c'));(a_1,([m],r,c));(\tau,(a,([m],r_1,c')));(\tau,(a_1,([m],r,c)));\\
& (\tau,\overline{(m,(\varepsilon,r,c))});\overline{(m,(\varepsilon,r,c))};(\tau,\overline{(m,(\varepsilon,r_1,c'))});\overline{(m,(\varepsilon,r_1,c'))}]]
\end{aligned}
$$

The first trace represents the run of the mix system (scenario) with the message from a going to r and the message from a_1 going to r_1 while the second trace represents the scenario with message from a going to r_1 and the message from a_1 to r. If the two scenarios are equiprobable, then the message arriving at r is equally likely to have come from a as it is from a_1, hence its sender anonymity is 1.

This example is, naturally, much simplified from the kinds of traces which would be observed in real anonymity systems by real adversaries. However, the method of computing anonymity is valid for traces of any size. To help us manage big traces, we have implemented a tool which is capable of generating a real trace given a particular network and some incoming messages, erasing it to give the attacker's observation Obs, generating the set of real traces All which would erase to Obs and finally compute the anonymity of any of the messages. This enables us to analyse non-trivial runs in non-trivial networks, although, in general, of course, the analysis is exponential in the number of messages travelling through the network. Nevertheless, our results have some practical relevance to real-world email anonymity systems such as Mixminion and Mixmaster. As part of our analysis we have also gained a much better understanding of how the anonymity of messages in mix networks should be computed and the kinds of data that would be required.

This account of our work leaves out a lot of detail, does not explain the abstractions which are made or explain why they are made. Neither do we discuss related work. For full details the reader is invited to refer to Chapter 6 of [Ser04].

References

[Cha81] David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 4(2), February 1981.

[SD02] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In Paul Syverson and Roger Dingledine, editors, *Privacy Enhancing Technologies*, volume 2482 of *LNCS*, San Francisco, CA, April 2002.

[Ser04] Andrei Serjantov. *On the Anonymity of Anonymity Systems.* PhD thesis, University of Cambridge, 2004.

[1] The seemingly unnecessary repetition of events is required in more complex cases.