The Traffic Analysis of Continuous-Time Mixes

George Danezis

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

Abstract. We apply the information-theoretic anonymity metrics to continuous-time mixes, that individually delay messages instead of batching them. The anonymity of such mixes is measured based on their delay characteristics, and as an example the exponential mix (sg-mix) is analysed, simulated and shown to use the optimal strategy. We also describe a practical and powerful traffic analysis attack against connection based continuous-time mix networks, despite the presence of some cover traffic. Assuming a passive observer, the conditions are calculated that make tracing messages through the network possible.

1 Introduction

Building blocks for anonymous communication operating by batching input messages in rounds, such as threshold or pool mixes, have recently been the subject of extensive study [15, 16, 6, 17]. The same is not true for mixes that operate in continuous-time, by individually delaying messages. An example of these is the *sg-mix* construction presented by Kesdogan *et al* [10]. Its inventors present an analysis of its anonymity, but this cannot easily be generalised to other mix strategies.

We will present a new framework for analysing the anonymity provided by mix strategies that individually delay messages. In order to make the analysis easier, we assume that the rate of arrival of messages to the mixes is Poisson distributed. Using the work presented here, different mix strategies can be analysed but we choose to illustrate our method with an analysis of the exponential mix (sg-mix), both because it is relatively simple and because it has been extensively mentioned in the literature. Furthermore, a section is devoted to showing that given some latency constraints the exponential mix is the mixing strategy providing maximal anonymity.

We then present a powerful attack that given enough packets, can break the anonymity provided by connection-based mix networks functioning in continuous-time. The attack relies on detecting an input traffic pattern, at the outputs of the mixes or network, using signal detection techniques. A detailed description is given on how to perform this attack, and confidence intervals are provided to assess the reliability of the results. The attack can be used effectively against many proposed anonymous communications systems such as Onion Routing [13], Freedom [4], TARZAN [7] or MorphMix [14].

2 Delay characteristic and anonymity

The main aim of a mix, as introduced by Chaum [5], is to hide the correspondence between the input and output messages it relays. First it makes its inputs and outputs bitwise unlinkable, which means that a third party cannot link them by observing their bit patterns without knowledge of the cryptographic keys used to perform the transform. Secondly it blurs the timing correlations between inputs and outputs by batching, introducing appropriate random delays and reordering messages. Continuous-time mixes achieve this by delaying each message individually and independently of the others.

We can say that a particular mix strategy is described by its *delay characteristic*. This is a function $f(\beta|\alpha)$ that represents the probability a message injected in the mix at time α leaves the mix at time β, where $\alpha \leq \beta$. Since $f(\beta|\alpha)$ is a conditional probability distribution, it is normalised.

$$\forall \alpha. \int_{\alpha}^{+\infty} f(\beta|\alpha)\, d\beta = 1\,. \tag{1}$$

The *inverse delay characteristic*, $f'(\alpha|\beta)$, of the same mix strategy is a probability distribution that describes the likelihood a message being ejected at time β was injected at time α. Again because it is a conditional probability distribution it is normalised.

$$\forall \beta. \int_{-\infty}^{\beta} f'(\alpha|\beta)\, d\alpha = 1\,. \tag{2}$$

The two characteristics are related, since the second f' can be calculated using Bayes theorem from f. Some knowledge of the probability of arrivals at particular times is necessary to perform this conversion. To simplify things, we will consider that arrivals are Poisson distributed with a rate λ_α. In a Poisson process, the probability of an arrival is independent from other arrivals or the time α.

$$f'(\alpha|\beta) = \frac{f(\beta|\alpha)\mathrm{Pr}[\text{Arrival at } \alpha]}{\int_{-\infty}^{\beta} f(\beta|\alpha)\mathrm{Pr}[\text{Arrival at } \alpha]\, d\alpha} \tag{3}$$

$$= \frac{f(\beta|\alpha)}{\int_{-\infty}^{\beta} f(\beta|\alpha)\, d\alpha} \tag{4}$$

Therefore, given the delay characteristics and some assumptions about the traffic in the network we can calculate the inverse delay characteristic. These will allow us to measure the effective sender and receiver anonymity for this mix strategy.

We will use the metric introduced in [15] to calculate the sender anonymity provided by a mixing strategy. This metric is based on defining a random variable that describes the possible senders of a message and calculating the entropy of its underlying probability distribution. The entropy is then a measure of the anonymity provided, and can be interpreted as the amount of information an attacker is missing to deterministically link the messages to a sender.

We assume that in a time interval $(\beta - T, \beta)$, K messages arrive at the mix, where K is distributed according to a Poisson distribution with parameter λ_α. These messages arrive at times $X_{1\ldots K}$ each distributed according to a uniform distribution $U(t)$ over the time interval of length T (as required by the Poisson distribution).

Given the inverse delay characteristic of the mix $f'(\alpha|\beta)$, the sender anonymity $\mathcal{A}$ provided by the mix can be calculated. It represents the entropy of the probability distribution describing how likely each of the inputs X_i is to be output at a particular time β.

$$\mathcal{A} = \sum_{i=1}^{K} \frac{f'(X_i|\beta)}{\sum_{j=1}^{K} f'(X_j|\beta)} \log \frac{f'(X_i|\beta)}{\sum_{j=1}^{K} f'(X_j|\beta)} = \tag{5}$$

$$= \frac{1}{\sum_{j=1}^{K} f'(X_j|\beta)} \left(\sum_{i=1}^{K} f'(X_i|\beta) \log f'(X_i|\beta) \right) - \log \sum_{j=1}^{K} f'(X_j|\beta) \tag{6}$$

From the Law of Large Numbers[1] we know that the sums converge to:

$$\sum_{j=1}^{K} f'(X_j|\beta) \to \frac{K}{T} \to \lambda_\alpha \tag{7}$$

$$\sum_{i=1}^{K} f'(X_i|\beta) \log f'(X_i|\beta) \to \frac{K}{T} \int_{\beta-T}^{\beta} f'(t|\beta) \log f'(t|\beta) dt \to \lambda_\alpha \mathcal{E}[f'(\alpha|\beta)] \tag{8}$$

Thus the fraction K/T converges to λ_α, which is the rate of arrival of messages to the mix and the integral (8) reduces to the entropy of the inverse delay characteristic function $\mathcal{E}[f'(\alpha|\beta)]$. Therefore the sender anonymity of a continuous mix with delay characteristic f' and a rate of arrival λ_α can be expressed.

$$\mathcal{A} \to \mathcal{E}[f'(\alpha|\beta)] - \log \lambda_\alpha \tag{9}$$

Putting this into words, the effective sender anonymity set size of the mixing strategy will converge towards the relative entropy of the inverse delay characteristic, as defined by Shannon [19], minus the logarithm of the rate at which messages are received. Similarly the recipient anonymity set size can be calculated using the same techniques and the delay characteristic of the mix strategy.

2.1 The exponential mix

In order to illustrate the calculations above we analyse the exponential mix. The exponential mix has been presented as a mixing strategy by Kesdogan *et al* [10]. In their design additional features are implemented to avoid $(n-1)$ attacks [8, 16], that we are not concerned with in this work.

[1] For large K and $T, \lim_{K\to\infty} \sum_{j=1}^{K} f'(X_j|\beta) = K \int_{\beta-T}^{\beta} U(t) f'(t|\beta)\, dt \to \frac{K}{T}$. Note that for the approximation we do not assume the rate to be large, but simply the observation period T to be large enough to observe some traffic.

The exponential mix can be abstracted as an $M/M/\infty$ queue. We assume, as required from the calculations above, the arrival rates of messages to be Poisson distributed with rate λ_α. Each of the messages that arrives at the mix is delayed according to a random variable that follows the exponential distribution with parameter μ. Therefore the delay characteristic of the exponential mix is:

$$f(\beta|\alpha) = \mu e^{-\mu(\beta-\alpha)}\,. \tag{10}$$

From equation (4) we can calculate the inverse delay characteristic f'. Due to the nature of the exponential distribution, it is equal to the delay characteristic f.

$$f'(\alpha|\beta) = \frac{f(\beta|\alpha)}{\int_{-\infty}^{\beta} f(\beta|\alpha)\,d\alpha} = f(\beta|\alpha) = \mu e^{-\mu(\beta-\alpha)} \tag{11}$$

Using the inverse delay characteristic, and (9) we can now calculate the expected sender anonymity ($\mathcal{E}[\cdot]$ is the entropy function).

$$\mathcal{A} = \mathcal{E}[\Pr[\alpha]] \to \mathcal{E}[f'(\alpha|\beta)] - \log\lambda_\alpha = \tag{12}$$

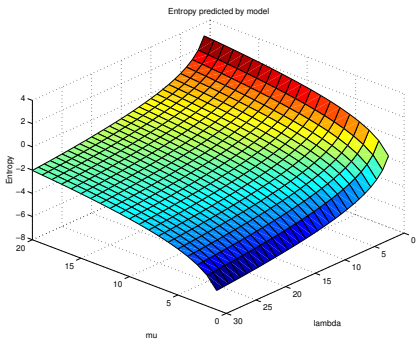
$$= \int_{-\infty}^{\beta} \mu e^{-\mu(\beta-\alpha)} \log \mu e^{-\mu(\beta-\alpha)}\,d\alpha - \log\lambda_\alpha = -\log\frac{\lambda_\alpha e}{\mu} \tag{13}$$

To check the above result (since it relies on the approximations (7) and (8)) a simulation was run for some values of λ_α and μ, and the results were compared with the metric predictions in equation (13). The inverse delay characteristic was used to calculate the probability assigned to a number of messages arriving at a mix. The number of messages was Poisson distributed according to λ_α, and their time of arrival was chosen uniformly. Their delay was a random variable distributed according to the exponential distribution with rate μ. The absolute difference between the predictions (figure 1(a)) and the simulation (figure 1(b)) is shown in figure 1(c).

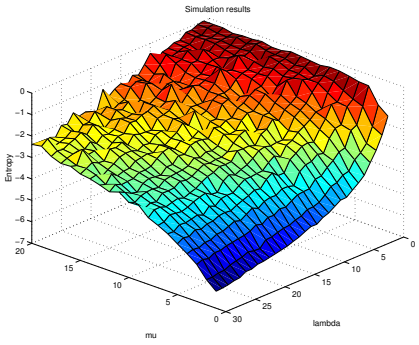
The main divergence of the simulated results from the predicted results, is in the region where the metric predicts positive values for the entropy. This is intuitively impossible and indeed is the largest error from the actual simulation results. The conditions for which the model, that the equation (13) describes, should not be considered accurate is described by:

$$-\log\frac{\lambda_\alpha e}{\mu} > 0 \Rightarrow \mu > \lambda_\alpha e \tag{14}$$

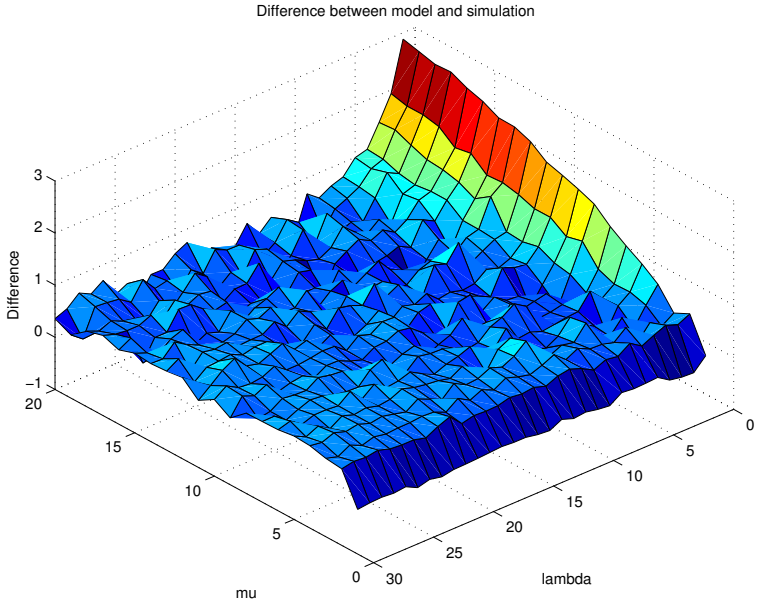
It is clear that an $M/M/\infty$ queue with a departure rate μ larger than the arrival rate λ_α would not provide much anonymity most of the time. The average time a message would spend in the mix is $\frac{1}{\mu}$ while the average time between message arrivals is $\frac{1}{\lambda_\alpha}$, which is larger. Therefore the mix would behave on average as a first-in first-out queue.



(a) Predictions for exponential mix



(b) Simulation of exponential mix



(c) Absolute difference between prediction and simulation

Fig. 1. Simulation of exponential mix for different μ and λ.

2.2 The latency of a mix strategy

The delay characteristic of a mix can also be used to calculate the latency introduced by a mix strategy and its variance. This can be done trivially since the latency of the mix strategy is the expectation $E[\cdot]$ of the delay characteristic function $f(\beta|\alpha)$.

$$E[f(\beta|\alpha)] = \int_{\alpha}^{+\infty} (\beta - \alpha)\, f(\beta|\alpha)\, d\beta \tag{15}$$

Similarly the variance $V[\cdot]$ of the delay can be calculated using the expectation:

$$V[f(\beta|\alpha)] = \int_{\alpha}^{+\infty} (E[f(\beta|\alpha)] - (\beta - \alpha))^2\, f(\beta|\alpha)\, d\beta \tag{16}$$

For the exponential mix the mean delay is $\frac{1}{\mu}$ and its variance is $\frac{1}{\mu^2}$.

2.3 Optimal mixing strategies

So far, we have described how to measure the anonymity and latency of a continuous-time mix, given its delay strategy. Naturally, the next problem is finding a mix strategy that maximises entropy, and therefore anonymity.

We need to find a distribution f with a particular mean a, which represents the average latency of the mix. Since a packet can only leave the mix after it arrived, the function f can only occupy half the timeline, namely the interval $[0, +\infty)$. We prove that the optimal probability distribution f is the exponential probability distribution. This result was first proved by Shannon [19] using techniques from the calculus of variations [22]. We want to minimise:

$$E[f(x)] = -\int_{0}^{-\infty} f(x) \log f(x) dx \tag{17}$$

Subject to the constraints:

$$a = \int_{0}^{-\infty} x f(x) dx \qquad \text{and} \qquad \int_{0}^{-\infty} f(x) dx = 1 \tag{18}$$

Then by the calculus of variations [22] we must solve:

$$\frac{\partial(-f(x)\log f(x) + \lambda x f(x) + \mu f(x))}{\partial f} = 0 \tag{19}$$

$$\Rightarrow\ -1 - \log f(x) + \lambda x + \mu = 0 \tag{20}$$

$$\Rightarrow\ f(x) = e^{\lambda x + \mu - 1} \tag{21}$$

After incorporating the constraints, the resulting function is:

$$f(x) = \frac{1}{a} e^{-\frac{1}{a}x} \tag{22}$$

This is exactly the exponential mix as analysed in section 2.1, which is therefore optimal.

3 Traffic analysis of continuous mixes

In the previous sections we have considered the anonymity of single packets mixed using a continuous-time mixing strategy. Continuous-time mixes can approximate circuit-based systems that implement minimal mixing, in order to provide real-time communications. In such systems a number of packets, all belonging to the same stream, are quickly routed through the same path in the network.

The Onion Routing project [20] first drew the community's attention to the need for traffic padding to protect against fine-grained traffic analysis. Since then some publications have discussed traffic analysis and possible defences against it [1, 12]. Others refer to the same problem in the context of intersection attacks [3, 2, 9] and present padding as a potential protection.

Some previous work has drawn attention to the vulnerabilities of anonymous systems to "timing" attacks [14], while Kesdogan *et al* [9] present a concrete attack. Serjantov *et al* [18] present a traffic analysis attack based on counting packets on the links, while Levine *et al* [11] uses more fine grained traffic patterns to trace them. We will now present a very general way of performing traffic analysis on streams of packets travelling through the same route in a continuous-time mix network. We show that after a certain number of messages, that can be calculated, the communication can be traced with high confidence.

3.1 Concrete traffic analysis techniques

We denote as $f(t)$ the function that describes the traffic, to be traced, feeding into a continuous mix with delay characteristic $d(x)$. We assume that all messages described by $f(t)$ belong to the same stream, and will therefore be ejected on the same output link. We will assume that there are two output links. The attacker's aim is to determine on which output link the stream is redirected.

On the first link we observe messages coming out at times $X_{1\cdots n}$ and on the second link messages come out at times $Y_{1\cdots m}$ in the time interval $[0, T]$. H_0 represents the hypothesis the input stream $f(t)$ is interleaved in the first channel described by the observations X_i, and H_1 that is in the second corresponding with Y_i.

In order to detect the streams we will make some approximations. We will create two model probability distributions C_X and C_Y and will assume that all messages in the output channels are independent samples out of one of these distributions. The difference between C_X and C_Y is due to our attempt to model the noise in the two output channels. We will also consider that all the other messages are uniformly distributed in the interval $t \in [0, T]$ according to the distribution $U(t) = u$.

When H_0 is true the stream under observation is interleaved in the observations X_i. We will model each of them as following the probability distribution:

$$C_X(t) = \frac{\lambda_f (d * f)(t) + (\lambda_X - \lambda_f) U(t)}{\lambda_X} \tag{23}$$

The probability distribution $(d * f)(t)$ is the convolution of the input signal with the delay characteristic of the mix. The probability a message delayed by $d(x)$ is output at time t given an input stream of messages described by $f(t)$ is described by this convolution.

$$(d * f)(t) = \int d(x) f(t - x) dx \tag{24}$$

Furthermore λ_f is the rate of messages in the input signal, while λ_X is the rate of the output channel. Finally $U(t) = u$ is the uniform distribution in the interval $[0, T]$.

Similarly if hypothesis H_1 is true, the signal is interleaved in the observations Y_i that follow the distribution:

$$C_Y(t) = \frac{\lambda_f (d * f)(t) + (\lambda_Y - \lambda_f) U(t)}{\lambda_Y} \tag{25}$$

In order to decide which of the two hypothesis is valid, H_0 or H_1, we can calculate the likelihood ratio of the two alternative hypothesis.

$$\frac{\mathcal{L}(H_0 | X_i, Y_j)}{\mathcal{L}(H_1 | X_i, Y_j)} = \frac{\prod_{i=1}^{n} C_X(X_i) \prod_{j=1}^{m} u}{\prod_{i=1}^{n} u \prod_{j=1}^{m} C_Y(Y_j)} > 1 \tag{26}$$

We choose to accept hypothesis H_0 if condition (26) is true, and hypothesis H_1 otherwise. Section 3.3 will show how we calculate our degree of confidence when making this choice.

3.2 A simple example

Figure 2 shows six diagrams illustrating the traffic analysis attack. The first column represents, from top to bottom, the signal that we inject in a mix and the two output channels, one of which contains the delayed signal. The right hand side column represents the delay characteristic of the network, an exponential distribution in this case (sg-mix), the "model" that is created by convolving the input signal with the delay characteristic and, at the bottom, the log-likelihood ratio.

The cover traffic or "noise" in the above experiments is assumed to be a Poisson process. Noise is added both to the channel that contains the stream under surveillance (in this case link 1, X_i) and the other link (Y_i). The rate of the signal $f(t)$ in the traffic analysis graphs shown above is 50 messages, while the noise added in X_i has a rate of 150 messages. The second link contains random padding with a rate of 200 messages (Y_i). The delay characteristic $d(x)$ chosen to illustrate the traffic analysis technique is exponential with a departure rate of 30. The graphs therefore illustrate the traffic analysis of a sg-mix node. The decision graph presents the logarithm of the likelihood, ratio $\log \frac{\mathcal{L}(H_0 | X_i, Y_j)}{\mathcal{L}(H_1 | X_i, Y_j)}$, as an attacker would compute it at each point in the simulation time. After 700 simulation ticks the log-likelihood ratio is clearly positive indicating that H_0 should be accepted.

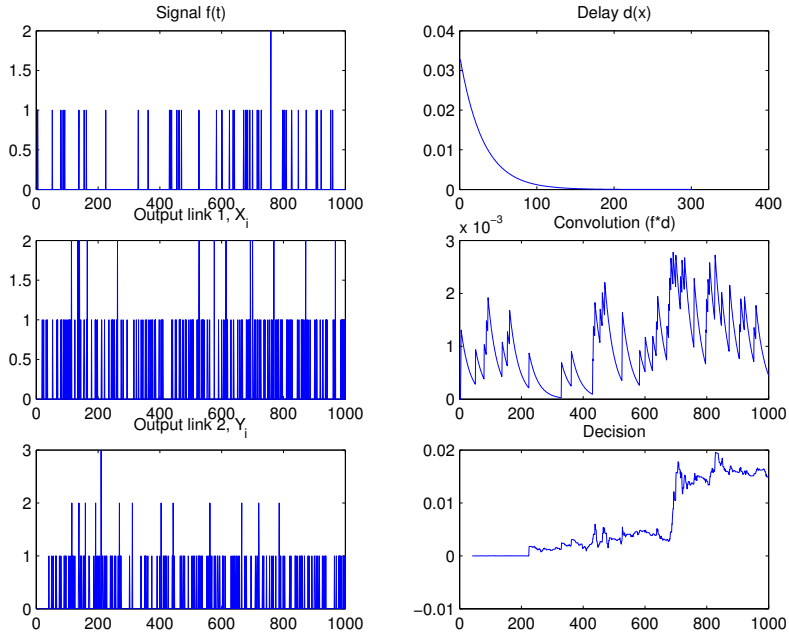


Fig. 2. Final and intermediate results of traffic analysis.

3.3 Performance of the traffic analysis attack

There are two question that need to be answered concerning the traffic analysis attack presented. First the conditions under which it is at all possible must be established. Second the number of observations necessary to get reliable results has to be calculated.

By simple mathematical manipulations with logarithms, we can derive that the likelihood ratio test, applied to select the most appropriate hypothesis can be expressed using sums of random variables:

$$\mathcal{L}_{H_0/H_1} = \frac{\mathcal{L}(H_0|X_i, Y_j)}{\mathcal{L}(H_1|X_i, Y_j)} = \frac{\prod_{i=1}^{n} C_X(X_i) \prod_{j=1}^{m} u}{\prod_{i=1}^{n} u \prod_{j=1}^{m} C_Y(Y_j)} > 1 \tag{27}$$

$$\Rightarrow \log \mathcal{L}_{H_0/H_1} = \sum_{i=1}^{n} \log C_X(X_i) - \sum_{j=1}^{m} \log C_Y(Y_j) + (m-n) \log u > 0 \tag{28}$$

The expression above is equivalent to (26) the rule by which we choose the hypothesis to accept. The condition for which the attack is possible is that the decision rule (28) must not equal zero. This could be the case if both C_X and C_Y were uniform distributions. Even through the inequality might hold it does not give us any measure of confidence in the result. We will therefore attempt to find bounds within which we are confident that the decision is correct.

Note that the two sums will converge to the expectations $nE\left[\log C_X(X) \mid X_i \sim X\right]$ and $mE\left[\log C_Y(Y) \mid Y_j \sim Y\right]$. The notation $X_i \sim X$ means that the samples X_i are sampled from the distribution X, and the samples Y_j from the distribution Y. The two distributions X and Y are different according to which of the two hypothesis is accepted. In case H_0 then $X_i \sim C_X, Y_j \sim U$. Alternatively if H_1 is true then $X_i \sim U$ and $Y_j \sim C_Y$. Without losing generality we will demonstrate when to accept hypothesis H_0. The derivations are the same in the other case.

In case the hypothesis H_0 is correct, $E\left[\log C_X(X) \mid H_0 : X_i \sim C_X\right]$ converges to the entropy of the probability distribution $C_X(t)$, denoted $\mathcal{E}[C_X(t)]$, since the probabilities assigned to each value of the random variable $\log C_X(X)$ follow the distribution C_X.

$$E\left[\log C_X(X) \mid H_0 : X_i \sim C_X\right] = \int_0^T C_X(t) \log C_X(t) dt = \mathcal{E}[C_X(t)] \tag{29}$$

On the other hand $E\left[\log C_Y(Y) \mid H_0 : Y_j \sim U\right]$ converges to the expectation of C_Y namely $E\left[\log C_Y(t)\right]$.

$$E\left[\log C_Y(Y) \mid H_0 : Y_j \sim U\right] = \int_0^T u \log C_Y(t) dt = E\left[\log C_Y(t)\right] \tag{30}$$

Therefore in case we accept hypothesis H_0 the expected value of the decision rule $\log \mathcal{L}_{H_0/H_1}$ (28) is μ_{H_0}:

$$\begin{aligned} \mu_{H_0} &= E\big[\textstyle\sum_{i=1}^{n} \log C_X(X_i) - \sum_{j=1}^{m} \log C_Y(Y_j) + (m-n)\log u \,\big|\, H_0\big] \\ &= nE\left[\log C_X(X) \mid H_0\right] - mE\left[\log C_Y(Y) \mid H_0\right] + (m-n)\log u \\ &= n\mathcal{E}[C_X(t)] - mE\left[\log C_Y(t)\right] + (m-n)\log u \end{aligned} \tag{31}$$

The variance can be calculated using the above observations:

$$V\left[\log C_X(X) \mid H_0\right] = \int_0^T C_X(t)(\log C_X(t) - \mathcal{E}\left[C_X(X)\right])^2 dt \tag{32}$$

$$V\left[\log C_Y(Y) \mid H_0\right] = \int_0^T u(\log C_Y(t) - E\left[\log C_Y(Y)\right])^2 dt \tag{33}$$

Using these we will calculate the variance $\sigma^2_{H_0}$ of the decision rule $\log \mathcal{L}_{H_0/H_1}$ (28) which is:

$$\begin{aligned} \sigma^2_{H_0} &= V\big[\textstyle\sum_{i=1}^{n} \log C_X(X_i) - \sum_{j=1}^{m} \log C_Y(Y_j) + (m-n)\log u \,\big|\, H_0\big] \\ &= nV[\log C_X(X) \,\big|\, H_0] + mV\left[\log C_Y(Y) \,\big|\, H_0\right] \end{aligned} \tag{34}$$

Using Chebyshev's inequality[2] we can derive the condition necessary in order to accept hypothesis H_0 with confidence p. We require the log-likelihood not to

[2] If a random variable x has a finite mean μ and finite variance σ^2, then $\forall\, k \geq 0 \quad \Pr[|x-\mu| \geq k] \leq \frac{\sigma^2}{k^2}$.

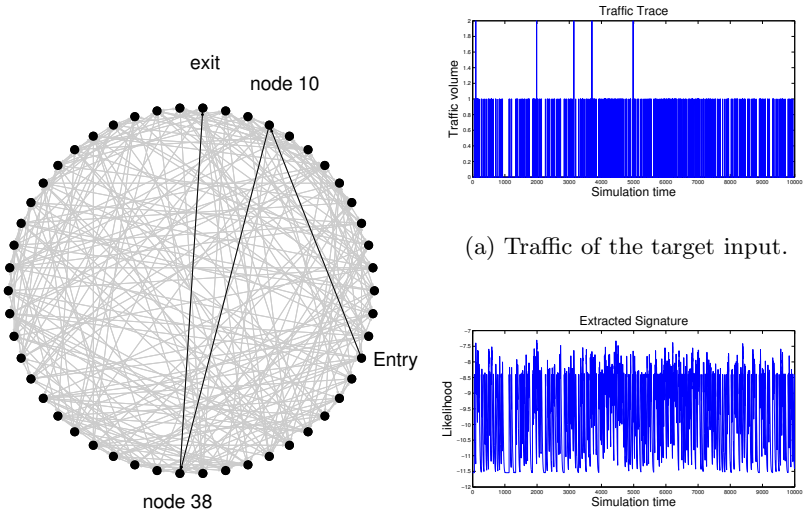
deviate, with probability greater than p, from its expected value (the mean) more than its mean (which would invalidate our decision rule (28)).

$$p = \Pr\left[\left|\log \mathcal{L}_{H_0/H_1} - \mu_{H_0}\right| \geq \mu_{H_0}\right] \leq \frac{\sigma^2_{H_0}}{\mu^2_{H_0}} \quad \Rightarrow \quad p \leq \frac{\sigma^2_{H_0}}{\mu^2_{H_0}} \tag{35}$$

An equivalent test can be derived to assess our confidence when accepting hypothesis H_1.

3.4 Traffic analysis of networks

We modify slightly the simple techniques described above to perform traffic analysis against a mix network composed of continuous-time mixes. Instead of performing a hypothesis test on two links, we compare all the links in the network with the pattern extracted from the input stream that we want to trace. This way each link is assigned a degree of similarity with the traced input. This can be used to infer some information about the intermediate and final nodes on the path.



(a) Traffic of the target input.

(b) Pattern extracted from target.

Fig. 3. Network, route and stream traced.

To illustrate our techniques we use a mix network made of 50 nodes with 10 links each. The network is sparse, which is consistent with quite a few fielded

systems, such as Freedom [4]. Five hundred streams (500) are routed through this network, using a random path of 4 nodes (the same node cannot appear twice in the path). Each stream contains 400 packets during the period the network is under observation, which is 10000 simulation ticks. Mixes delay packets individually using an exponential mix with mean 10 simulation ticks. Figure 3 presents a view of the network, along with the route that the stream under observation takes. The attacker's objective is to uncover the route of this stream, knowing only its input pattern and entry point, and the traffic on the network links.

As before a pattern (figure 3(b)) is extracted for the input under observation (figure 3(a)) that is compared with each link in the network. The convolution of the input traffic with the exponential delay characteristic, has been used to compute the pattern, but there has been no attempt to model the noise on each channel.

The pattern is compared to the traffic on each link of the network. This returns a measure of similarity of the link to the input traced. This in turn can be used to classify the link, as containing the target input on its way to the second mix (hop 1), the third mix (hop 2) or the final mix (hop 3). Alternatively the link might be unrelated to the target input, and simply contain noise. We choose the decision rule in such a way that we avoid false negatives. Figure 4 shows the classification curves that have been compiled after simulations.

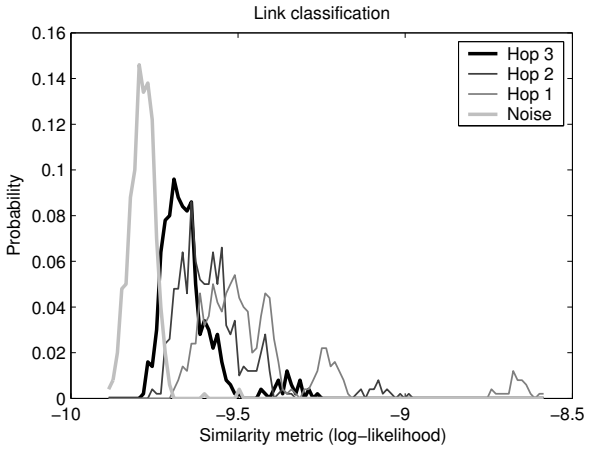
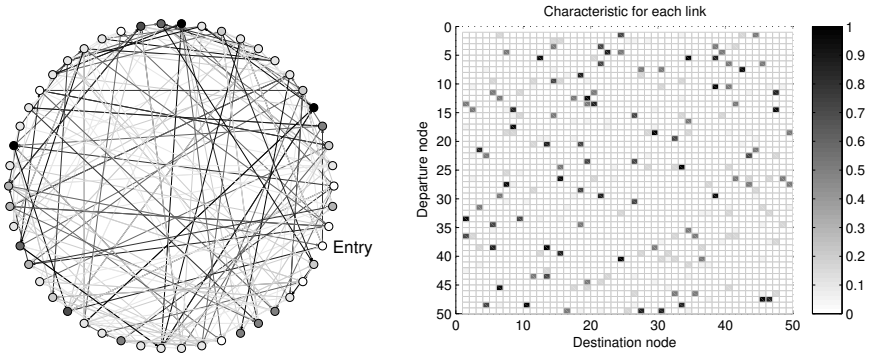


Fig. 4. Classification of the link (false positive and false negative curves)

The classification of each link as 'noise' or 'candidate link' allows us to simplify the graph of the network. Information can also be extracted relating to how likely the link is to contain the signal traced, and therefore a weighted graph (figure 5(a)) and its corresponding matrix (figure 5(b)) can be extracted. The

intensity of the links or the entries in the matrix represents the likelihood a link contains the stream under observation.



(a) Simplified graph

(b) Simplified transition matrix

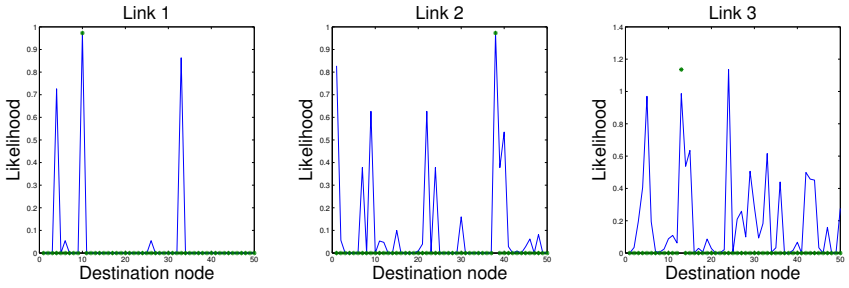
Fig. 5. Simplified network (intensity is the likelihood of containing the target input)

A random walk is performed for one to three hops on the resulting graph, starting that the entry point of the observed stream. This provides us with the likely second, third or final nodes of the path (figures 6(a), 6(b) and 6(c) respectively). The stars on the graphs indicate the actual nodes that relay the target stream. In the example shown the final node is not guessed correctly, but is within the three nodes with highest probability. In the presence of longer delays or more traffic the correct nodes might not be the ones with highest likelihood but the attack still yields a lot of information and significantly reduces the effective anonymity provided to the users.

4 Further considerations and future work

Measuring anonymity. The work presented measures the average anonymity provided by a mix strategy. One of the important assumptions is that the expected number of messages is received in any time interval t, namely $\lambda_\alpha t$. The actual number of messages received in any interval may vary according to the Poisson distribution. Should a mix be flooded by the attacker's messages the rate needs to be adjusted to the level of genuine traffic.

Mix strategies that take into account the number of messages queueing or that adapt their parameters according to the rate of arrival of messages have not been explicitly studied. The metric proposed should still be usable with them, although their delay characteristic function may be dependant of additional factors such as the rate of arrival of messages λ_α. We expect the functions



(a) Second node (b) Third node (c) Final node

Fig. 6. Likely second, third and final nodes on the path.

that depend upon the delay characteristic, such as the mean and variance of the latency, to still be usable.

Traffic analysis. More work needs to be done on how far the traffic analysis attack presented against stream-based anonymity systems can be exploited. Techniques from transient signal detection, as surveyed in [21], can be used as the foundation for a theory of traffic analysis. Some straightforward extensions of the work presented could be to simplify the extracted patterns, by retaining only the parts that are good at discriminating well the target stream, or at making the matching quicker. An experimental evaluation of how the length of the stream, or a more realistic distribution of packets, affects anonymity should also be easy to perform.

The attack assumes that an adversary can observe a "naked" stream somewhere in the network, in order to build a model later used for detection. An attacker might acquire the knowledge that a series of messages belong to the same stream by observing unpadded links at the edges of the mix network or by the means of subverted nodes. This assumption might be invalidated if cover traffic is used on all links, but variants of the attack might still work. Some preliminary results suggest that good models can be created despite this.

The attack can be performed by a passive adversary, without any knowledge of the relationships between packets on the attacked links. When an attacker knows the relationship between packets in the same stream, as a subverted node would, it is much easier to perform the statistical tests since the cover traffic can be discarded. In other words we expect most of the anonymity provided, up to the point where the path goes through a corrupt node, to be easily cancelled if the node applies our attack.

Furthermore the attacks are passive, in the sense that the attacker does not modify in any way the characteristics of the traffic. An active attacker would modulate the input traffic in order to maximise the chances of detecting it. They

could introduce periodicity, allowing for periodic averaging for noise cancellation or injecting patterns of traffic specially designed to be easily detected. Unless the anonymity system takes special steps beyond delaying the traffic to destroy such structure, traffic streams will quickly be traceable.

5 Conclusions

The information theoretic anonymity metric is adapted to describe the properties of mixes that simply delay individual packets. We proved that the optimal delaying strategy is the exponential mix, for which we calculate the anonymity and latency.

A very powerfully attack is then presented that traces streams of messages following the same path through a delaying mix network. We present the conditions under which it is possible, and derive expressions that an adversary can use to assess his confidence. The attack is efficient enough to to be applied against whole networks by a global passive adversary. When performed by an adversary controlling subverted nodes or with the ability to shape traffic on the links, its effects are even more devastating. This attack is applicable to systems that provide real-time anonymous communications and leaves us very sceptical about the possibility of secure and efficient such constructions, in the absence of heavy amounts of cover traffic or delay.

Acknowledgements This work has been substantially improved after discussions with Ross Anderson, Markus Kuhn, Piotr Zielinski and Andrei Serjantov.

References

1. Adam Back, Ulf Möller, and Anton Stiglic. Traffic analysis attacks and trade-offs in anonymity providing systems. In Ira S. Moskowitz, editor, *Information Hiding workshop (IH 2001)*, volume 2137 of *LNCS*, pages 245–257. Springer-Verlag, April 2001.
2. Oliver Berthold and Heinrich Langos. Dummy traffic against long term intersection attacks. In Roger Dingledine and Paul Syverson, editors, *Privacy Enhancing Technologies workshop (PET 2002)*, volume 2482 of *LNCS*, pages 110–128. Springer-Verlag, 2002.
3. Oliver Berthold, Andreas Pfitzmann, and Ronny Standtke. The disadvantages of free MIX routes and how to overcome them. In H. Federrath, editor, *Designing Privacy Enhancing Technologies*, volume 2009 of *LNCS*, pages 30–45. Springer-Verlag, July 2000.
4. Philippe Boucher, Adam Shostack, and Ian Goldberg. Freedom systems 2.0 architecture. White paper, Zero Knowledge Systems, Inc., December 2000.
5. David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–88, February 1981.
6. Claudia Diaz and Andrei Serjantov. Generalising mixes. In Roger Dingledine, editor, *Privacy Enhancing Technologies workshop (PET 2003)*, volume 2760 of *LNCS*, pages 18–31, Dresden, Germany, March 2003. Springer-Verlag.

7. Michael J. Freedman and Robert Morris. Tarzan: A peer-to-peer anonymizing network layer. In Vijayalakshmi Atluri, editor, *ACM Conference on Computer and Communications Security (CCS 2002)*, pages 193–206, Washington, DC, November 2002. ACM.
8. Ceki Gülcü and Gene Tsudik. Mixing E-mail with Babel. In *Network and Distributed Security Symposium — NDSS '96*, pages 2–16, San Diego, California, February 1996. IEEE.
9. Dogan Kesdogan, Dakshi Agrawal, and Stefan Penz. Limits of anonymity in open environments. In Fabien A. P. Petitcolas, editor, *Information Hiding workshop (IH 2002)*, volume 2578 of *LNCS*, pages 53–69, Noordwijkerhout, The Netherlands, 7-9 October 2002. Springer-Verlag.
10. Dogan Kesdogan, Jan Egner, and Roland Büschkes. Stop-and-Go MIXes: Providing probabilistic anonymity in an open system. In David Aucsmith, editor, *Information Hiding workshop (IH 1998)*, volume 1525 of *LNCS*, pages 83–98, Portland, Oregon, USA, 14-17 April 1998. Springer-Verlag.
11. Brian N. Levine, Michael K. Reiter, Chenxi Wang, and Matthew Wright. Timing attacks in low-latency mix systems. In *Finacial Cryptography (FC'04)*, 2004.
12. Jean-François Raymond. Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems. In Hannes Federrath, editor, *Designing Privacy Enhancing Technologies*, volume 2009 of *LNCS*, pages 10–29. Springer-Verlag, July 2000.
13. Michael G. Reed, Paul F. Syverson, and David M. Goldschlag. Anonymous connections and onion routing. *IEEE Journal on Selected Areas in Communications*, 16(4):482–494, May 1998.
14. Marc Rennhard and Bernhard Plattner. Introducing MorphMix: Peer-to-Peer based Anonymous Internet Usage with Collusion Detection. In *workshop on Privacy in the Electronic Society (WPES 2002)*, Washington, DC, USA, November 2002.
15. Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In Roger Dingledine and Paul Syverson, editors, *Privacy Enhancing Technologies workshop (PET 2002)*, volume 2482 of *LNCS*, pages 41–53, San Francisco, CA, USA, 14-15 April 2002. Springer-Verlag.
16. Andrei Serjantov, Roger Dingledine, and Paul Syverson. From a trickle to a flood: Active attacks on several mix types. In Fabien A. P. Petitcolas, editor, *Information Hiding workshop (IH 2002)*, volume 2578 of *LNCS*, pages 36–52, Noordwijkerhout, The Netherlands, 7-9 October 2002. Springer-Verlag.
17. Andrei Serjantov and Richard E. Newman. On the anonymity of timed pool mixes. In *workshop on Privacy and Anonymity Issues in Networked and Distributed Systems*, pages 427–434, Athens, Greece, May 2003. Kluwer.
18. Andrei Serjantov and Peter Sewell. Passive attack analysis for connection-based anonymity systems. In *European Symposium on Research in Computer Security (ESORICS 2003)*, Gjovik, Norway, 13-15 October 2003.
19. Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.
20. Paul F. Syverson, Gene Tsudik, Michael G. Reed, and Carl E. Landwehr. Towards an analysis of onion routing security. In Hannes Federrath, editor, *Designing Privacy Enhancing Technologies*, volume 2009 of *LNCS*, pages 96–114, Berkeley, CA, USA, 25-26 July 2000. Springer-Verlag.
21. Zhen Wang and Peter Willett. A performance study of some transient detectors. *IEEE transactions on signal processing*, 48(9):2682–2685, September 2000.
22. Robert Weinstock. *Calculus of variations*. Dover publications, 1974. ISBN: 0486630692.