The Anonymity of Continuous Time Mixes

George Danezis

University of Cambridge, Computer Laboratory,
15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom.

May 25, 2004
Outline

- Continuous time mixes.
  - The anonymity of single mixes.
  - The anonymity of streams going through mixes.
Outline

- Continuous time mixes.
- The anonymity of single mixes.
- The anonymity of streams going through mixes.
Continuous time mixes.

The anonymity of single mixes.

The anonymity of streams going through mixes.
What is a mix?

- A network node **relaying traffic**.
- Bitwise unlinkability between incoming and outgoing traffic (cryptography).
- Destroys the timing correlations, by batching or delaying messages.

Result: Cannot link senders and receivers of messages → anonymity.
What is a mix?

- A network node **relaying traffic**.
- Bitwise unlinkability between incoming and outgoing traffic (cryptography).
- Destroys the timing correlations, by batching or delaying messages.

Result: Cannot link senders and receivers of messages → anonymity.
What is a mix?

- A network node **relaying traffic**.
- Bitwiseunlinkability between incoming and outgoing traffic (cryptography).
- Destroys the timing correlations, by batching or delaying messages.

Result: Cannot link senders and receivers of messages → anonymity.
What is a continuous mix?

- A mix that individually delays each message.
- The delay is selected out of a probability distribution (the delay characteristic).

Obvious questions:
- How much anonymity do continuous time mixes provide?
- Is there an optimal delay characteristic?
What is a continuous mix?

- A mix that **individually delays** each message.
- The delay is selected out of a probability distribution (the **delay characteristic**).

Obvious questions:

- How much anonymity do continuous time mixes provide?
- Is there an optimal delay characteristic?
What is a continuous mix?

- A mix that *individually delays* each message.
- The delay is selected out of a probability distribution (the *delay characteristic*).

Obvious questions:

- How much anonymity do continuous time mixes provide?
- Is there an optimal delay characteristic?
What is a continuous mix?

- A mix that **individually delays** each message.
- The delay is selected out of a probability distribution (the delay characteristic).

Obvious questions:

- How much anonymity do continuous time mixes provide?
- Is there an optimal delay characteristic?
What is a continuous mix?

- A mix that *individually delays* each message.
- The delay is selected out of a probability distribution (the *delay characteristic*).

Obvious questions:
- How much anonymity do continuous time mixes provide?
- Is there an optimal delay characteristic?
Getting formal

- Messages arrive to a single continuous mix according to a **poisson distribution** (uniform distribution over the time line, exponentially distributed delays). Message arrival rate $\lambda_{\alpha}$.
- We denote the delay characteristic as $f(\beta|\alpha)$. The probability a message that arrived at time $\alpha$ leaves the mix at time $\beta$.
- We use the information theoretic metric for anonymity: the **entropy** of the probability distribution relating messages to senders is the **sender anonymity** of the message.

Note: entropy is $\int f(x) \log f(x) dx$.
Getting formal

- Messages arrive to a single continuous mix according to a **poisson distribution** (uniform distribution over the time line, exponentially distributed delays). Message arrival rate $\lambda$. 

- We denote the delay characteristic as $f(\beta|\alpha)$. The probability a message that arrived at time $\alpha$ leaves the mix at time $\beta$. 

- We use the information theoretic metric for anonymity: the **entropy** of the probability distribution relating messages to senders is the **sender anonymity** of the message. 

  Note: entropy is $\mathcal{E}[f(x)] = \int_{-\infty}^{\infty} f(x) \log_2 f(x) \,dx$.
Getting formal

- Messages arrive to a single continuous mix according to a **poisson distribution** (uniform distribution over the time line, exponentially distributed delays). Message arrival rate $\lambda_{\alpha}$.
- We denote the delay characteristic as $f(\beta|\alpha)$. The probability a message that arrived at time $\alpha$ leaves the mix at time $\beta$.
- We use the information theoretic metric for anonymity: the **entropy** of the probability distribution relating messages to senders is the **sender anonymity** of the message.

Note: entropy is $\mathcal{E}[f(x)] = \int_{-\infty}^{\infty} f(x) \log_2 f(x) dx$
Getting formal

- Messages arrive to a single continuous mix according to a **poisson distribution** (uniform distribution over the time line, exponentially distributed delays). Message arrival rate \( \lambda_\alpha \).
- We denote the delay characteristic as \( f(\beta|\alpha) \). The probability a message that arrived at time \( \alpha \) leaves the mix at time \( \beta \).
- We use the information theoretic metric for anonymity: the **entropy** of the probability distribution relating messages to senders is the **sender anonymity** of the message.
  
  Note: entropy is \( \mathcal{E}[f(x)] = \int_{-\infty}^{\infty} f(x) \log_2 f(x) \, dx \)
The anonymity of a single mix

- Messages arrive at times $X_1 \ldots K$ each distributed according to a uniform distribution $U(t)$ over the time interval of length $T$
- A single message comes out at time $\bar{\beta}$.
- The sender anonymity of this message is:

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

$$\rightarrow \mathcal{E}[f'(\alpha | \beta)] - \log \lambda_\alpha \rightarrow (2)$$

- Interpretation: delay characteristic and volume of traffic increase anonymity.
The anonymity of a single mix

- Messages arrive at times $X_1...K$ each distributed according to a uniform distribution $U(t)$ over the time interval of length $T$
- A single message comes out at time $\beta$.
- The sender anonymity of this message is:

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

$$\rightarrow E[f'(\alpha|\beta)] - \log \lambda_\alpha \quad (2)$$

- Interpretation: delay characteristic and volume of traffic increase anonymity.
The anonymity of a single mix

Messages arrive at times $X_1\ldots K$ each distributed according to a uniform distribution $U(t)$ over the time interval of length $T$. A single message comes out at time $\beta$. The sender anonymity of this message is:

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

$$\rightarrow \mathcal{E}[f'(|\beta|)] - \log \lambda_\alpha \left(2\right)$$

Interpretation: delay characteristic and volume of traffic increase anonymity.
The anonymity of a single mix

- Messages arrive at times $X_1...K$ each distributed according to a uniform distribution $U(t)$ over the time interval of length $T$.
- A single message comes out at time $\beta$.
- The sender anonymity of this message is:

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

$$\rightarrow \mathcal{E}[f'(\alpha|\beta)] - \log \lambda_\alpha \rightarrow (2)$$

- Interpretation: delay characteristic and volume of traffic increase anonymity.
Optimal delay strategy

- What is the optimal delay to maximize anonymity?
  - Infinite
- Given a particular expected latency?
- Answer: The exponential delay ($sg$-mix)

\[ f(\beta|\alpha) = \mu e^{-\mu(\beta-\alpha)}. \]  

(3)
What is the optimal delay to maximize anonymity? Infinite

Given a particular expected latency?

Answer: The exponential delay ($sg$-$mix$)

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

(3)
Optimal delay strategy

- What is the optimal delay to maximize anonymity?
  Infinite
- Given a particular expected latency?
  Answer: The exponential delay (sg-mix)

\[ f(\beta|\alpha) = \mu e^{-\mu(\beta-\alpha)}. \] (3)

Anonymity: \( \mathcal{A} = -\log \frac{\xi \varrho}{\mu} \)
Optimal delay strategy

- What is the optimal delay to maximize anonymity?
  Infinite
- Given a particular expected latency?
- Answer: The exponential delay (sg-mix)

\[
f(\beta|\alpha) = \mu e^{-\mu(\beta-\alpha)}.\tag{3}\]

Anonymity: \( A = -\log \frac{\lambda \alpha e}{\mu} \)
Optimal delay strategy

- What is the optimal delay to maximize anonymity? Infinite
- Given a particular expected latency?
- Answer: The exponential delay (sg-mix)

\[ f(\beta|\alpha) = \mu e^{-\mu(\beta-\alpha)} \]  

(3)

Anonymity: \( \mathcal{A} = -\log \frac{\lambda e}{\mu} \)
Stream based traffic analysis of continuous mixes

Characteristics of stream based systems

- Many smaller packets travel over the same route.
- Minimal batching to achieve low-latency (approximated by a delay characteristic function).
- Used for web-browsing or ssh: some clear patterns of traffic.

Attacker objectives

- Trace a stream from a sender, through the network of mixes, to the receiver.
- Possible because more information is available (than single packet anonymous email).
Stream based traffic analysis of continuous mixes

Characteristics of stream based systems

- Many smaller packets travel over the same route.
- Minimal batching to achieve low-latency (approximated by a delay characteristic function).
- Used for web-browsing or ssh: some clear patterns of traffic.

Attacker objectives

- Trace a stream from a sender, through the network of mixes, to the receiver.
- Possible because more information is available (than single packet anonymous email).
Stream based traffic analysis of continuous mixes

Characteristics of stream based systems

- Many smaller packets travel over the same route.
- Minimal batching to achieve low-latency (approximated by a delay characteristic function).
- Used for web-browsing or ssh: some clear patterns of traffic.

Attacker objectives

- Trace a stream from a sender, through the network of mixes, to the receiver.
- Possible because more information is available (than single packet anonymous email).
Stream based traffic analysis of continuous mixes

Characteristics of stream based systems

- Many smaller packets travel over the same route.
- Minimal batching to achieve low-latency (approximated by a delay characteristic function).
- Used for web-browsing or ssh: some clear patterns of traffic.

Attacker objectives

- Trace a stream from a sender, through the network of mixes, to the receiver.
- Possible because more information is available (than single packet anonymous email).
Stream based traffic analysis of continuous mixes

Characteristics of stream based systems

- Many smaller packets travel over the same route.
- Minimal batching to achieve low-latency (approximated by a delay characteristic function).
- Used for web-browsing or ssh: some clear patterns of traffic.

Attacker objectives

- Trace a stream from a sender, through the network of mixes, to the receiver.
- Possible because more information is available (than single packet anonymous email).
A simple case

We use a **single exponential mix**:

- The target stream of data $f(t)$ goes into a mix.
- The mix has two outputs, padded with random messages up to a certain volume.
- The mix delays each input message according to an exponential distribution $d(t)$.
- The attacker observes the messages output at times $X_i$ on the first link and $Y_j$ on the second link.
- From these he will try to guess which link contains the target data.
A simple case

We use a **single exponential mix**:

- The target stream of data $f(t)$ goes into a mix.
- The mix has two outputs, padded with random messages up to a certain volume.
- The mix delays each input message according to an exponential distribution $d(t)$.
- The attacker observes the messages output at times $X_i$ on the first link and $Y_j$ on the second link.
- From these he will try to guess which link contains the target data.
A simple case

We use a **single exponential mix**:

- The target stream of data $f(t)$ goes into a mix.
- The mix has two outputs, padded with random messages up to a certain volume.
- The mix delays each input message according to an exponential distribution $d(t)$.
- The attacker observes the messages output at times $X_i$ on the first link and $Y_j$ on the second link.
- From these he will try to guess which link contains the target data.
A simple case

We use a **single exponential mix**:  
- The target stream of data $f(t)$ goes into a mix.  
- The mix has two outputs, padded with random messages up to a certain volume.  
- The mix delays each input message according to an exponential distribution $d(t)$.  
- The attacker observes the messages output at times $X_i$ on the first link and $Y_j$ on the second link.  
- From these he will try to guess which link contains the target data.
A simple case

We use a **single exponential mix**:

- The target stream of data \( f(t) \) goes into a mix.
- The mix has two outputs, padded with random messages up to a certain volume.
- The mix delays each input message according to an exponential distribution \( d(t) \).
- The attacker observes the messages output at times \( X_i \) on the first link and \( Y_j \) on the second link.
- From these he will try to guess which link contains the target data.
Model the continuous mix operation

- We pretend that the timings of output packets are random samples of a function of the input target stream.
- The mix delays the stream \( f(t) \) according to the exponential distribution \( d(t) \). We convolve them to get an estimate of the where packets are likely to come out.

\[
C(t) = (d * f)(t) = \int d(x)f(t - x)dx
\] (4)

- We see if link 1 or link 2 are most likely generated by \( C(t) \). We do this using the likelihood ratio:

\[
\frac{\mathcal{L}(H_0|X_i, Y_j)}{\mathcal{L}(H_1|X_i, Y_j)} = \frac{\prod_{i=1}^{n} C(X_i) \prod_{j=1}^{m} u}{\prod_{i=1}^{n} u \prod_{j=1}^{m} C(Y_j)} > 1
\] (5)
Model the continuous mix operation

- We pretend that the timings of output packets are random samples of a function of the input target stream.
- The mix delays the stream $f(t)$ according to the exponential distribution $d(t)$. We convolve them to get an estimate of the where packets are likely to come out.

$$C(t) = (d * f)(t) = \int d(x)f(t - x)dx$$  \hspace{1cm} (4)

- We see if link 1 or link 2 are most likely generated by $C(t)$. We do this using the likelihood ratio:

$$\frac{L(H_0|X_i, Y_j)}{L(H_1|X_i, Y_j)} = \frac{\prod_{i=1}^{n} C(X_i) \prod_{j=1}^{m} u}{\prod_{i=1}^{n} u \prod_{j=1}^{m} C(Y_j)} > 1$$  \hspace{1cm} (5)
Model the continuous mix operation

- We pretend that the timings of output packets are random samples of a function of the input target stream.
- The mix delays the stream \( f(t) \) according to the exponential distribution \( d(t) \). We convolve them to get an estimate of the where packets are likely to come out.

\[
C(t) = (d * f)(t) = \int d(x)f(t - x)dx
\]  

- We see if link 1 or link 2 are most likely generated by \( C(t) \). We do this using the likelihood ratio:

\[
\frac{\mathcal{L}(H_0|X_i, Y_j)}{\mathcal{L}(H_1|X_i, Y_j)} = \frac{\prod_{i=1}^{n} C(X_i) \prod_{j=1}^{m} u}{\prod_{i=1}^{n} u \prod_{j=1}^{m} C(Y_j)} > 1
\]
Just forget the maths . . .
Analysis

- The attack is computationally cheap but requires a lot of data.
- Given enough messages the stream can be traced.
- We have derived confidence intervals.
- Longer delays, less traffic or more cover traffic make attack slower.
- All of these make systems slower or expensive.

Future work

- Cover traffic is other streams and can be modeled.
- Compress the patterns, and extract features that detect quickly.
- Active attacks that modulate input stream.
Analysis

- The attack is computationally cheap but requires a lot of data.
- Given enough messages the stream can be traced.
- We have derived confidence intervals.
- Longer delays, less traffic or more cover traffic make attack slower.
- All of these make systems slower or expensive.

Future work

- Cover traffic is other streams and can be modeled.
- Compress the patterns, and extract features that detect quickly.
- Active attacks that modulate input stream.
Analysis

- The attack is computationally cheap but requires a lot of data.
- Given enough messages the stream *can* be traced.
- We have derived confidence intervals.
  - Longer delays, less traffic or more cover traffic make attack slower.
  - All of these make systems slower or expensive.

Future work

- Cover traffic is other streams and can be modeled.
- Compress the patterns, and extract features that detect quickly.
- Active attacks that modulate input stream.
Analysis

- The attack is computationally cheap but requires a lot of data.
- Given enough messages the stream *can* be traced.
- We have derived confidence intervals.
- Longer delays, less traffic or more cover traffic make attack slower.
- All of these make systems slower or expensive.

Future work

- Cover traffic is other streams and can be modeled.
- Compress the patterns, and extract features that detect quickly.
- Active attacks that modulate input stream.
Analysis

- The attack is computationally cheap but requires a lot of data.
- Given enough messages the stream can be traced.
- We have derived confidence intervals.
- Longer delays, less traffic or more cover traffic make attack slower.
- All of these make systems slower or expensive.

Future work

- Cover traffic is other streams and can be modeled.
- Compress the patterns, and extract features that detect quickly.
- Active attacks that modulate input stream.
Analysis

- The attack is computationally cheap but requires a lot of data.
- Given enough messages the stream can be traced.
- We have derived confidence intervals.
- Longer delays, less traffic or more cover traffic make attack slower.
- All of these make systems slower or expensive.

Future work

- Cover traffic is other streams and can be modeled.
- Compress the patterns, and extract features that detect quickly.
- Active attacks that modulate input stream.
The objective of the attacker is to trace the route (shown above).
The attacker compares each link with the convolved target input.
Network traffic analysis: step 3

- A random walk is performed for one, two and three steps on the weighted graph to provide the most likely destinations.
- The anonymity of the stream is greatly reduced (green stars indicate actual actual destination)
Network traffic analysis: step 3

- A random walk is performed for one, two and three steps on the weighted graph to provide the most likely destinations.
- The anonymity of the stream is greatly reduced (green stars indicate actual destination).
Network traffic analysis: step 3

- A random walk is performed for one, two and three steps on the weighted graph to provide the most likely destinations.
- The anonymity of the stream is greatly reduced (green stars indicate actual actual destination)
Network traffic analysis: step 3

- A random walk is performed for one, two and three steps on the weighted graph to provide the most likely destinations.
- The anonymity of the stream is greatly reduced (green stars indicate actual destination)

![Graphs showing likelihood of destination nodes for different links]
Conclusions

The anonymity of single continuous mixes:

- We can quantify it (assumption of traffic).
- There is an optimal strategy, the exponential mix.

Continuous stream analysis:

- Message based and connection based anonymous communication systems exhibit patterns and can be attacked.
- The attacks presented go beyond proof-of-concept, are well understood, robust and extensible.

The future?

- Attack and defense go hand in hand: new systems must take into account these attacks and provide countermeasures.
- Are secure anonymous communication systems possible at all?
Conclusions

The anonymity of single continuous mixes:

- We can quantify it (assumption of traffic).
- There is an optimal strategy, the exponential mix.

Continuous stream analysis:

- Message based and connection based anonymous communication systems exhibit patterns and can be attacked.
- The attacks presented go beyond proof-of-concept, are well understood, robust and extensible.

The future?

- Attack and defense go hand in hand: new systems must take into account these attacks and provide countermeasures.
- Are secure anonymous communication systems possible at all?
Conclusions

The anonymity of single continuous mixes:

- We can quantify it (assumption of traffic).
- There is an optimal strategy, the exponential mix.

Continuous stream analysis:

- Message based and connection based anonymous communication systems exhibit patterns and can be attacked.
- The attacks presented go beyond proof-of-concept, are well understood, robust and extensible.

The future?

- Attack and defense go hand in hand: new systems must take into account these attacks and provide countermeasures.
- Are secure anonymous communication systems possible at all?
Conclusions

The anonymity of single continuous mixes:
  - We can quantify it (assumption of traffic).
  - There is an optimal strategy, the exponential mix.

Continuous stream analysis:
  - Message based and connection based anonymous communication systems exhibit patterns and can be attacked.
  - The attacks presented go beyond proof-of-concept, are well understood, robust and extensible.

The future?
  - Attack and defense go hand in hand: new systems must take into account these attacks and provide countermeasures.
  - Are secure anonymous communication systems possible at all?
Conclusions

The anonymity of single continuous mixes:

- We can quantify it (assumption of traffic).
- There is an optimal strategy, the exponential mix.

Continuous stream analysis:

- Message based and connection based anonymous communication systems exhibit patterns and can be attacked.
- The attacks presented go beyond proof-of-concept, are well understood, robust and extensible.

The future?

- Attack and defense go hand in hand: new systems must take into account these attacks and provide countermeasures.
- Are secure anonymous communication systems possible at all?
Conclusions

The anonymity of single continuous mixes:

» We can quantify it (assumption of traffic).
» There is an optimal strategy, the exponential mix.

Continuous stream analysis:

» Message based and connection based anonymous communication systems exhibit patterns and can be attacked.
» The attacks presented go beyond proof-of-concept, are well understood, robust and extensible.

The future?

» Attack and defense go hand in hand: new systems must take into account these attacks and provide countermeasures.
» Are secure anonymous communication systems possible at all?
Conclusions

The anonymity of single continuous mixes:
- We can quantify it (assumption of traffic).
- There is an optimal strategy, the exponential mix.

Continuous stream analysis:
- Message based and connection based anonymous communication systems exhibit patterns and can be attacked.
- The attacks presented go beyond proof-of-concept, are well understood, robust and extensible.

The future?
- Attack and defense go hand in hand: new systems must take into account these attacks and provide countermeasures.
- Are secure anonymous communication systems possible at all?
Conclusions

The anonymity of single continuous mixes:

▶ We can quantify it (assumption of traffic).
▶ There is an optimal strategy, the exponential mix.

Continuous stream analysis:

▶ Message based and connection based anonymous communication systems exhibit patterns and can be attacked.
▶ The attacks presented go beyond proof-of-concept, are well understood, robust and extensible.

The future?

▶ Attack and defense go hand in hand: new systems must take into account these attacks and provide countermeasures.
▶ Are secure anonymous communication systems possible at all?
Conclusions

The anonymity of single continuous mixes:
- We can quantify it (assumption of traffic).
- There is an optimal strategy, the exponential mix.

Continuous stream analysis:
- Message based and connection based anonymous communication systems exhibit patterns and can be attacked.
- The attacks presented go beyond proof-of-concept, are well understood, robust and extensible.

The future?
- Attack and defense go hand in hand: new systems must take into account these attacks and provide countermeasures.
- Are secure anonymous communication systems possible at all?