MŰEGYETEM 1782

Budapest University of Technology and Economics
Department of Measurement and Informaton Systems

Measuring Anonymity in a Passive, Real-time System

Anonymous message transmission should be a key feature in network architectures. It should ensure that delivered messages are impossible or at least infeasible to be traced back to their senders. For this purpose model of the passive, real-time, observable, black-box (PROB) channel was defined. In this model attackers (we call them observers) try to circumvent applied protection measures and to link senders to delivered messages. In order to measure the level of anonymity provided by the system, probability can given, with which observers can determine the senders of delivered messages (source-hiding property) or the recipients of sent messages (destination-hiding property). In order to reduce the confidence of an observer, possible counter-measures can be defined that ensure specified upper limit for the probability with which an observer can mark someone as the sender or recipient of a message. Finally results of simulations demonstrate the strength of the techniques.

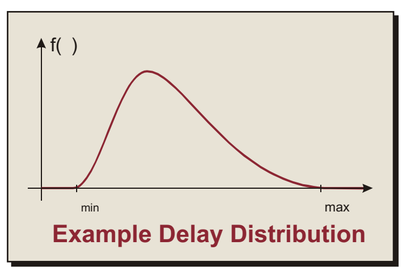
Model of the PROB-channel

Aim of the PROB-channel is to relay messages between senders and recipients while making it unfeasible to link them — thus it implements an anonymous transport protocol. The PROB-channel can be characterized as follows:

- The channel is **passive**, as its operation is not affected by properties and distribution of incoming messages, i.e. network delay has static distribution.
- The channel is **real-time**, thus messages will be delivered before a message-invariant maximal delay.
- All input and output of the channel is **observable**, so one can detect all messages.
- The channel is **black-box**, since it is analyzed as a whole, internal implementation is not considered. The observer cannot see what happens to messages inside the channel and how they are encoded.
- We furthermore assume that messages passing through the channel are equally sized and properly encrypted, thus an observer can only draw conclusions from the timing of the messages, **content does not provide information**.

Specification of the Channel

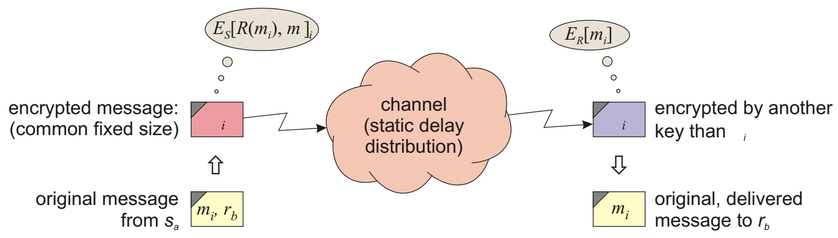
Senders (S) send messages (M) to recipients (R). $S(m_i)$ denotes the sender of a message m_i, $R(m_i)$ is the recipient. Time of sending is $t_S(m_i)$, whereas time of delivery is $t_R(m_i)$. No messages are born or deleted in the channel. **Delay** of the messages in the channel is a probability variable with a public distribution function. This distribution f( ) is **static,** thus message and time invariant and it is before $_{min}$ and after $_{max}$ zero.



Example Delay Distribution

Message Sending

Messages together with the recipients' ID enter the channel in the encrypted form $_i$. After the delay they will be delivered to the recipients as $_i$ (encrypted with different keys).



The Adversary — an Observer

Possibilities of a **passive observer** were analyzed. Such an observer can only eavesdrop on all communication channels, but he cannot alter anything, nor can he decrypt encrypted messages. We assume that the observer knows the **history** of the system, which contains the environment and the parameters of the channel, all sent $_S$={ $_i$} and delivered $_R$={ $_i$} encrypted messages and the times of sending $_S$={t_S($_i$)} and receipt $_R$={t_R($_i$)}.

In order to evaluate, which sender sent which message, for each encrypted delivered message and for each sender a probability (1) can be determined. If the observer knows the history * of the system, he can conclude that a certain message was sent by a certain sender with a certain probability.

$$P_{k,s_l^*,*} \quad P[S(\ _k^*)\quad s_l^* \mid \quad ^*] \qquad (1) \qquad P_{k,*}: \ \max_{s_l^*} P_{k,s_l^*,*} \qquad (2)$$

In order to **back-trace** the message to their senders the observer calculates the probabilities (1) and marks the most probable sender as the potential real sender of the message in question.

Global Back-tracing

In order to compute the probabilities in (1) the **obvious and optimal** solution would be to perform a global back-tracing, thus the observer would try all possibilities and choose the most probable one. In order to do this, one has to generate all possible match combinations of sent and received messages. After having all combinations, based upon their probabilities one can calculate (1).
Unfortunately this approach is **exponential** by the number of messages and thus ineffective for practical use.

Local Back-tracing

Performs the observer the delivered message ⇨ sender matching for each delivered message independently, then (3) gives the probability defined in (1).

Unfortunately local back-tracing has a great disadvantage. Originating from its local aspect even in a very simple scenario it can produce false results. However since only local back-tracing is feasible especially for larger sets, in the following we assume a locally back-tracing observer for the drawn conclusions.

$$P_{k,s_l^*,*} \quad \frac{f^*[t_R(\ _k^*)\quad t_S(\ _i^*)]}{\sum_{i}^{*}[\ _i^* \quad {}^{*}_{k,s_l^*,*}]\ f^*[t_R(\ _k^*)\quad t_S(\ _j^*)]} \qquad (3)$$
$$\sum_{j}^{*}[\ _j^* \quad {}^{*}_{k,*}]$$

Measure for Anonymity

Source-hiding Property

History of a system is **source-hiding with parameter** if the observer cannot assign any sender to any delivered message $_k$ with a probability greater than (4). In the model this property can be seen as a **numerical measure for sender anonymity**.

$$P_{k\ R\ k,} \qquad (4)$$

Destination-hiding Property

Similarly to (1), also the probability that a certain sent message was delivered to a certain recipient can be calculated (5). With that, history of a system is **destination-hiding with parameter** if the observer cannot assign a recipient to any sent message $_j$ with a probability greater than (7). Thus this property can be seen as a **numerical measure for recipient anonymity**.

$$P_{*,r_l^*,*} \quad P[R(\ _j^*)\quad r_l^* \mid \quad ^*] \qquad (5) \qquad P_{*,*}: \ \max_{r_l^*} P_{*,r_l^*,*} \qquad (6) \qquad P_{j\ S\ j,} \qquad (7)$$

Anonymity Ensuring Techniques

MIN/MAX Property

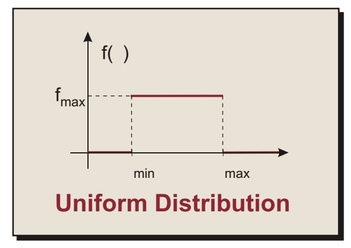
In order to limit the confidence of an observer performing local back-tracing (3) we have to decrease the numerator and increase the denominator. Thus senders have to have to obey the following **frequency rules**: senders cannot send more than one message in a given time interval ($_{max}$) and they have to send at least one message in another given interval ($_{min}$). With these restrictions an **upper limit** can be given for the confidence of the observer (8). Though this approach works for the source-hiding property, for the destination-hiding property it is not the case (recipients cannot be enforced to receive messages).

$$P_{k,} \qquad \hat{P} \quad \frac{\max_{i\ 1\ i\ 1\ \min\ q\ i\ \min}^{\min} f(q)}{|S|\ \min_{i\ 1\ i\ 1\ \max\ q\ i\ \max}^{\max} f(q)} \qquad (8)$$

Uniformly Distributed Delay

The aim is to force the observer to **pick** the sender of a delivered message **randomly** from the potential senders. In order to achieve this, the delay distribution has to be uniform. With this we get (9):

$$P_{k,} \quad \frac{\max_{s_l} \left| \ _{k,s_l,} \right|}{\left| \ _{k,} \right|} \qquad (9)$$



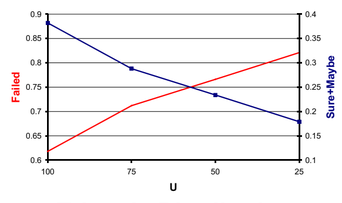
Uniform Distribution

If we also have **MIN/MAX senders**, then the confidence of the observer can be further limited (10). Finally with $_{min}$ = $_{max}$ the **global optimum** can also be reached (11).

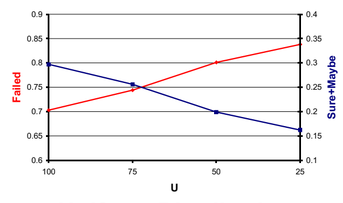
$$P_{k,} \qquad \hat{P} \quad \frac{\min}{|S|_{\ max}} \quad \frac{\max}{|S|_{\ min}} \qquad (10) \qquad P_{k,} \qquad \hat{P} \quad \frac{1}{|S|} \qquad (11)$$

Simulation Results

General Message Sending

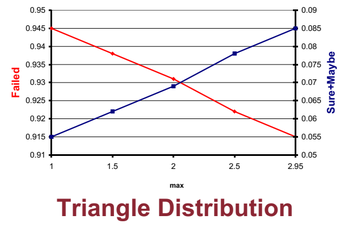


Triangle Distribution

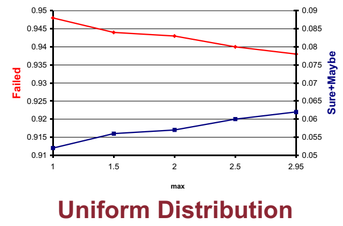


Uniform Distribution

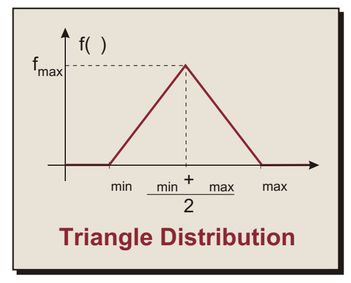
MIN/MAX Message Sending



Triangle Distribution



Uniform Distribution



Triangle Distribution

20 senders and 20 recipients

$_{min}$ = 1 and $_{max}$ = 4

Simulation duration T = 2000 (each sender sends messages between time index 0 ... T)

General senders send messages with a random maximal frequency between 0 ... U

MIN/MAX senders use $_{min}$ = 0.9

Theoretical maximum of 0.05 for the confidence of the observer is almost reached with MIN/MAX sending (for $_{max}$ = 1.0)

Contact: Gergely TÓTH
Fax: +36-1-205-3098 • E-mail: tgm@mit.bme.hu • URL: http://www.mit.bme.hu/~tgm