The APROB Channel: Adaptive Semi-Real-Time Anonymous Communication

Gergely Tóth and Zoltán Hornák

Budapest University of Technology and Economics
Department of Measurement and Information Systems
H-1117 Budapest, XI., Magyar tudósok krt. 2.
{tgm,hornak}@mit.bme.hu

Abstract. Anonymous communication has become a building block of network services. Besides providing anonymity, speed (and thus *real-time guarantees*) are becoming crucial as well. In this paper we will introduce the *global delaying adversary (GDA)*, an active attacker who is capable of arbitrarily delaying messages, while eavesdropping on all communication channels. This type of foe is particularly relevant for inter-mix relationships, where communication between the partners is secured (by authentication and integrity protection), and delaying remains the only effective external active attacking possibility. To counter GDA, the *adaptive semi-real-time APROB Channel* will be introduced. It will be shown that the APROB Channel can provide a guaranteed level of anonymity under semi-real-time[1] conditions considering that the adversary cannot obtain any additional information by delaying messages, thus this type of attack *will not be reasonable.*

1 Introduction

In this paper we will present and analyze systems for anonymous communication. Sending messages in communication networks anonymously is gaining more and more importance nowadays: this technique is used as the foundation for several privacy enhancing technologies (PETs), and has found adaptations in a wide area ranging from web-browsing to e-payments.

For anonymity we identify two factors as crucial: users want *quality of service* (QoS), i.e. an understandable, modifiable and guaranteed level of anonymity; on the other hand, message delivery times should be limited – in fact *real-time* systems (where a maximal delay is ensured) are required.

The foundation on which this paper is built was presented in [1]: senders used *common QoS parameters while using the MIN/MAX technique* and a *global passive observer* (GPO) was assumed. For this environment the *non-adaptive real-time channel* (PROB Channel) was shown to be effective.

The extension presented in this paper assumes *different QoS parameters for each message*, and takes a stronger adversary – the *global delaying adversary* (GDA) – into

[1] Naturally, against an adversary who can arbitrarily delay messages, no hard real-time guarantee can be given. The notion *semi-real-time* refers to the property that enables the channel to deliver the messages with the real-time requirements, if the adversary does not delay them.

Please use the following format when citing this chapter:
Author(s) [insert Last name, First-name initial(s)], 2006, in IFIP International Federation for Information Processing, Volume 201, Security and Privacy in Dynamic Environments, eds. Fischer-Hubner, S., Rannenberg, K., Yngstrom, L., Lindskog, S., (Boston: Springer), pp. [insert page numbers].

account. In order to provide anonymity under these circumstances, the *adaptive semi-real-time channel* (APROB Channel) will be introduced.

With the evolution to the APROB Channel presented here our aim was to continue the work aimed at providing anonymous communication systems metting higher and higher requirements. We made our assumed adversary stronger, gave more freedom to the users and developed an anonymity system that could guarantee quality of service under these circumstances as well.

In this paper we will still consider single, black-box channels. Our reason for this is simple: first the basic building blocks have to be analyzed on their own and only after their properties are well understood can we analyze more complex networks built of them.

2 Anonymity background

In this paper techniques for anonymous communication will be evaluated. In order to establish a common understanding, first the model of an anonymous message transmission system (AMTS) will be introduced. The section closes with a description of the foundations for measuring anonymity.

2.1 Notations for the anonymous communication scenario

For the purpose of anonymous electronic communication, several anonymous message transmission systems have been proposed. Their structures and modes of operation differ in various aspects, but some common properties are true for most of them. This common basic framework will be defined in the following.

The goal of an anonymous message transmission system is to deliver *messages* from *senders* to *recipients* so that it becomes algorithmically hard for an *adversary* to link these messages to senders or recipients. Let us look at the formal model:

- Senders ($s_i \in S$) send encrypted messages ($\alpha_j \in \varepsilon_S$) at times $t_S(\alpha_j)$ through the AMTS. These messages have a fixed size and are encrypted for the AMTS.
- The AMTS receives the sent messages and performs cryptographic operations on them to obtain a different representation. In order to further confuse the adversary, the AMTS delays and reorders messages. How messages are actually encoded and transformed is irrelevant for the purposes of this paper, the main assumption being that the adversary is not able to break the cryptographic functions used.
- After the delay, the AMTS delivers the re-encoded messages ($\beta_k \in \varepsilon_R$) to the recipients ($r_l \in R$) at times $t_R(\beta_k)$. These delivered messages also have a common, fixed size, and are encrypted for the recipient.
- The adversary's aim is to either match the delivered messages to senders ($\beta_k \rightarrow s_i = S(\beta_k)$), or the sent messages to recipients ($\alpha_j \rightarrow r_l = R(\alpha_j)$). In order to do this, adversaries may eavesdrop on communication channels and see when messages are being sent or delivered (*passive* observer), or even influence the network traffic by delaying messages or creating new ones (*active* adversary). Furthermore, it is assumed that each message is independent, so both the channel and the adversary operate on single messages – they do not consider message streams.

Finally, the adversary's aim is to break *sender anonymity* [2] and thus link delivered messages to their senders ($\beta_k \rightarrow S(\beta_k)$).

2.2 Measuring anonymity

In order to provide guaranteed anonymity with quality of service parameters, first we have to define how anonymity is measured. For this reason different metrics have been proposed.

In this paper we will use the source-hiding property as defined in [1] for measuring sender anonymity[2]. For its definition we have to consider that the adversary assigns probabilities to each delivered message – sender pair and then chooses accordingly. In order to model this, let the notion $P_{s_i,\beta_k} = P(S(\beta_k) = s_i)$ indicate the probability that shows, according to the knowledge of the adversary, what the chance is that s_i was the sender of β_k, the message delivered. With this the level of anonymity is defined as the maximal probability with which the adversary may back-trace messages to their senders.

Thus – according to the definition in [1] – an AMTS is *source-hiding with parameter* Θ, if the adversary cannot assign a sender to a delivered message with a probability greater than Θ[3]:

$$\forall_{\beta_k} \wedge \forall_{s_l} : P_{\beta_k,s_l} \leq \Theta \tag{1}$$

3 The non-adaptive, real-time channel for the global passive observer

After having introduced the foundations of anonymous communication, in this section the PROB Channel [1] will be introduced. The main goal of this paper is to present the APROB Channel, an efficient adaptive extension of the PROB Channel.

3.1 The PROB Channel

This paper presents an extension to the PROB Channel. In order to better understand the new features, let us introduce the relevant aspects of the PROB Channel. The PROB Channel is: (1) *non-adaptive* – the delay of an incoming message in the channel does not depend on properties or the actual distribution of incoming messages; (2) for each message the delay is calculated independently; and (3) it is *real-time* – the delay (δ) in the channel has a guaranteed maximum ($\delta_{\max}$), i.e. the PROB Channel ensures that every incoming message leaves the channel within $\delta_{\max}$ time.

[2] In [3] a detailed analysis of anonymity metrics is presented. The most popularly used measure currently are the simple entropy-based [4] and the normalized entropy-based [5]. Due to the disadvatnages outlined in [3] ones, we will stick to the source-hiding property anonymity metric as it is well understandable by the end user and can be interpreted as the *local* view, what the user wants from an anonymity system.

[3] Note, this definition is yet the *global* view of anonymity with a general quality of service parameter. The extension for local QoS is given as a refinement of this equation is (3).

3.2 The global passive observer (GPO)

As a first step in the analysis of characteristics of anonymity, a *global passive observer* was evaluated. The main properties of the GPO are the following: (1) she has knowledge of the *environment*, thus she knows of the potential senders and recipients and of the generally known parameters; (2) she *eavesdrops* on all the traffic to and from the AMTS, thus she knows of message dispatches and deliveries and their actors and timings; (3) however, GPO *cannot decrypt* the messages entering or leaving the channel and since messages have a common, fixed size, size-based decisions are ruled out as well; (4) GPO *cannot alter* the traffic in any way; and finally (5) GPO is *external*, meaning that she cannot gain information from within the channel, she sees the system as a black-box.

With the above properties, the aim of GPO is to *break sender anonymity* and thus guess who could have been the sender of a delivered message.

3.3 Anonymity in the PROB Channel

Assuming a GPO, the strong anonymity of the PROB Channel could only be enforced with the introduction of the *MIN/MAX property*. The definition was the following: a system possesses the MIN/MAX property with parameters $\tau_{\min}$ and $\tau_{\max}$ ($\tau_{\min} < \tau_{\max}$), if it holds that no sender sends more than one message within any $\tau_{\min}$ time interval, and all senders send at least one message within every $\tau_{\max}$ time interval.

With this strict limitation on the frequency of message sending, a guaranteed level of anonymity could be provided with the PROB Channel, where the source-hiding property could be easily tuned by setting $\tau_{\min}$ and $\tau_{\max}$ appropriately (2):

$$\Theta_{\text{PROB}} \approx \frac{\tau_{\max}}{|S| \cdot \tau_{\min}} \tag{2}$$

As the first step, the PROB Channel provided only a *global* quality of service, the anonymity guarantee being the same for all messages. However, despite of being non-adaptive, the channel could ensure a guaranteed level of anonymity assuming a global passive observer, if the senders conformed to the MIN/MAX property.

4 Environment for the APROB Channel

In the previous sections a basic framework for anonymous communication has been introduced together with the relevant previous work. In this section the environment for the APROB Channel will now be defined: a guaranteed level of anonymity has to be provided even if a global delaying adversary is assumed and users may specify different QoS levels for their messages.

4.1 The global delaying adversary

In our earlier paper [1] a global passive observer was assumed as the adversary for the system. In this paper we wanted the take a stronger opponent into consideration, so the opportunities open to a *global delaying adversary (GDA)* will be evaluated.

Since we are analyzing a static scenario, senders and recipients do not leave the system and their number is known to all participants. In this scenario, GDA has the following properties: (1) she is *global*, thus sees all messages sent into the channel and delivered from the channel, however since message are encrypted, she cannot read the contents or the QoS parameters[4]; (2) she may *delay* any number of messages for arbitrary time; but (3) she may *not create* new messages, nor alter existing ones; (4) finally, she is also *external* and sees the channel as a black-box, being unable to gain information from within the channel.

With this attacker model the question may arise, why not consider a fully active attacker, who is capable of creating or altering messages – and anyway, in what situations is such a GDA realistic? GDA was defined as the adversary model for *inter-mix communication*, i.e. to evaluate characteristics for internal nodes of an anonymity network. Since most current anonymity network implementations [6] establish authenticated channels between the nodes (e.g. with the help of TLS), the packet creation/modification opportunities opent to an external attacker are eliminated. The only active intervention possible is to *delay* messages, and so GDA comes into the picture.

To show that GDA is a real upgrade relative to a global passive observer, the following theorem can be formulated:

Theorem 1. *Anonymity provided for a particular sender by any non-adaptive channel can be completely compromized by a global delaying adversary.*

Proof. For the proof let us recapitulate the definition of the non-adaptive channel: its operation is not affected by the properties and distribution of the actual incoming messages. Thus, the simplest way a GDA could compromize a non-adaptive channel would be if the GDA acted as a bottleneck between the senders and the channel. If she buffered sent messages before reaching the channel and let one message at a time through, she could wait till the non-adaptive channel delivered that message to its recipient. By observing this delivery, GDA could match it to the one sent message and thus to the one sender. After one message has been compromized in this way, the adversary could feed the next message from its buffer into the channel and carry out the same procedure. With this simple technique GDA could break the anonymity provided by any non-adaptive channel. □

From this it follows that in order to cope with a global delaying adversary, an adaptive channel is required which internally monitors the distribution and properties of incoming messages and takes corrective action if necessary.

In this paper we focus on real-time communication, i.e. where message delay in the channel has a guaranteed upper limit. Naturally, if we assume a GDA, this real-time delivery cannot be guaranteed any more, since the adversary may delay the messages in addition to the delay introduced by the channel. Therefore the notion of *semi-real-time* is proposed: a semi-real-time channel guarantees real-time message delivery if the messages are not delayed by the adversary; should the opposite happen, then the channel operates on a best-effort basis (considering message delay) by conforming to the anonymity QoS parameters of the messages. The main requirement is that the channel

[4] In this regard she has all the capabilities of the global passive observer.

must not allow the adversary to compromize the anonymity provided – if the messages cannot be delivered in real-time with the necessary anonymity guarantees, then the real-time criterium (but not the anonymity requirements) can be dropped.

4.2 QoS diversity

In [1] quality of service was the same for all the senders: they obeyed the MIN/MAX rules and the PROB Channel guaranteed respective source-hiding property ($\frac{\tau_{max}}{|S|\cdot\tau_{min}}$) and guaranteed maximal message delivery (within δ_{max} time).

First we want to loosen the common parameters so that each sender may ask for different source-hiding properties for their sent messages. This way the source-hiding property has to be reconsidered: it is no longer a global requirement (i.e. for all senders and for all messages), but rather a *local* parameter conforming to the needs of each user.

To formalize the above, let the notation $I(\beta_k)$ (*input*) mean the sent message α_j, which was transformed by the AMTS into β_k. Similarly $O(\alpha_j)$ (*output*) means the delivered message β_k corresponding to α_j. Furthermore, QoS diversity is supported by attaching the requested local source-hiding property to each sent message, which will be denoted by $\theta(\alpha_j)$. With these, the following (3) should be ensured by the AMTS:

$$\forall\beta_k \wedge \forall s_l : P_{\beta_k,s_l} \leq \theta(I(\beta_k)) \tag{3}$$

In other words, in the rest of the paper QoS diversity is understood to mean the following: the sender specifies for each sent message α_j a parameter $\theta(\alpha_j)$, which is the requested local source-hiding property of that particular message. The task of the channel is to service these different requests and ensure real-time delivery with the requested QoS parameters.

5 The adaptive, semi-real-time channel

Having defined the environment of the APROB Channel, it is time for the specification of the channel itself. The approach will be the following: first we will construct the channel to provide guaranteed anonymity, then we will extend it to provide real-time guarantees and handle the different QoS requirements. Finally, we will analyze the chances of the adversary.

5.1 Guaranteed anonymity

Withing the context of the local source-hiding property (3) let us construct the APROB Channel in the following way: first it buffers incoming messages until the QoS requirements of *all* messages in the buffer can be fulfilled, and then it flushes the buffer after the messages have been reordered randomly.

This way its operation is similar to that of a simple mix [7]. To finalize the specification, the probabilities that the adversary may calculate for tracing delivered messages β_k back to their senders s_l have to be determined. The following equation defines the adversary's guess for the different probabilities P_{β_k,s_l}:

$$P_{\beta_k,s_l} = \frac{|\{\alpha_j|(\alpha_j \in X_{\beta_k}) \wedge (S(\alpha_j) = s_l)\}|}{|X_{\beta_k}|} \tag{4}$$

The above equation (4) defines the probabily as the fraction of the number of messages sent by a particular sender and the total number of messages in the batch. Since the APROB Channel is basically an adaptive mix, this is the best guess the adversary can make. Having the adversary's guess, the condition under which the buffer of the channel can be flushed, can be formulated properly:

$$\forall(\alpha_j \text{ in the buffer}) \wedge \forall(s_l)\colon P_{O(\alpha_j),s_l} \leq \theta(\alpha_j) \tag{5}$$

With this it is ensured that the buffer is only flushed when the local source-hiding property of *every* message in the buffer is fulfilled, thus the APROB Channel provides *guaranteed anonymity*.

5.2 Getting real-time

The next task of the APROB Channel is to provide *real-time* anonymity. Since the source-hiding property considers the worst case that could happen, we have to formulate the following theorem bearing the results from the PROB Channel [1] in mind:

Theorem 2. *If the senders do not conform to the MAX property, a guaranteed reasonable local source-hiding property (i.e. smaller than $\frac{1}{2}$) cannot be provided with a real-time channel.*

Proof. The MAX property demands that each sender has to send at least one message within each $\tau_{\max}$ time interval. If senders do not conform to the MAX property, then the channel may have fewer messages with which to form the anonymity set than would be required for the QoS parameters.

Considering the least reasonable $\frac{1}{2}$ local source-hiding property, if within the parameters of the real-time delivery only one sender sends a message into the channel (which without the MAX property may happen), the real-time channel can only choose from the following options: (1) if the channel is non-adaptive, it will deliver the messages regardless of the small anonymity set and thus break the anonymity guarantee; (2) if the channel is adaptive, then it has an alternative: it can drop messages for which the anonymity *and* real-time guarantees cannot be fulfilled at the same time or (3) it can wait until the anonymity requirements can be ensured.

From this it can be seen that without the MAX property, if senders do not send enough messages, the real-time *and* anonymity (i.e. local source-hiding property) guarantees cannot be ensured at the same time. □

On the other hand, with the MAX property, if every sender asks for a local source-hiding property θ equal or greater to $\frac{1}{|S|}$, and every sender sends periodically with $\tau_{\max}$, then after $\tau_{\max}$ time the buffer can always be flushed. This means a guaranteed $\delta_{\max} = \tau_{\max}$ delay and thus a real-time guarantee.

Naturally, if the APROB Channel functioned this way (i.e. each sender has to periodically send exactly one message), the system would not provide any flexibility. The

aim of this section has been to introduce the real-time property *together* with the already guaranteed anonymity. In the next section we will loosen the restrictions and let the APROB Channel unfold.

5.3 Handling the different QoS requirements

Up until now only the MAX property has been described. With that in mind every sender periodically sent one message and thus the real-time requirement *and* the anonymity guarantee could be maintained.

On the other hand, in order to provide flexibility, we have to enable the senders to send faster.

If a particular sender s_l wants to send $\phi(\alpha_j)$ messages within a $\tau_{\max}$ interval after having sent α_j, then he has to accept a larger source-hiding property $\theta(\alpha_j)$ for α_j, namely:

$$\theta(\alpha_j) = \frac{\phi(\alpha_j) + 1}{|S| + \phi(\alpha_j)} \tag{6}$$

The above equation (6) illustrates the case when all the other senders send only one message in the relevant time interval and s_l sends $\phi(\alpha_j)$.

However, because the number of messages sent and the achievable level of anonymity are strongly connected, (6) can be rearranged for senders wishing a particular source-hiding property $\theta(\alpha_j)$ for a certain message:

$$\phi(\alpha_j) = \frac{\theta(\alpha_j) \cdot |S| - 1}{1 - \theta(\alpha_j)} \tag{7}$$

Naturally, it is up to the sender to specify, how much anonymity is needed and what value should be assigned to the messages. With the help of the above equations we will now create the notion of *well-timed* and *ill-timed* messages. For their definition we introduce the following: α_j^k $(k \geq 1)$ indicates the k^{th} sent message after α_j has been sent, where both α_j and α_j^k have the same sender (i.e. $S(\alpha_j) = S(\alpha_j^k)$) and α_j^k follows α_j in a $\tau_{\max}$ window (i.e. $(t_S(\alpha_j^k) - t_S(\alpha_j)) \leq \tau_{\max}$). With these, the sent messages can be divided as follows:

$$\alpha_j\colon \begin{cases} \text{ill-timed} & \exists \alpha_i \wedge \exists k\colon (\alpha_j = \alpha_i^k) \wedge (k > \phi(\alpha_i)) \\ \text{well-timed} & \text{otherwise.} \end{cases} \tag{8}$$

With this categorization of messages the specification of the APROB Channel has to be extended – the condition for flushing the buffer of the channel (5) has to be restricted to only consider well-timed messages:

$$\forall(\text{well-timed } \alpha_j \text{ in the buffer}) \wedge \forall(s_l)\colon P_{O(\alpha_j),s_l} \leq \theta(\alpha_j) \tag{9}$$

It is essential to note that the APROB Channel can decide – based on the number and distribution of messages sent by each sender and the QoS parameters attached to the messages – whether a message is well-timed or not, so the extension of (5) into (9) is valid.

Thus, senders may send *faster* than $\tau_{\max}$ if they are willing to accept a greater local source-hiding property θ than $\frac{1}{|S|}$ (i.e. you may send more messages with less anonymity). In this case they have to consider the equations (6) and (7) in order to construct well-timed messages. Since the channel only tries to fulfill the anonymity requirements of well-timed messages, with QoS parameters derived from these equations the previous maximal delay of $\delta_{\max} = \tau_{\max}$ can be kept, so this extension still guarantees real-time delivery.

5.4 Resililence against GDA

Finally, let us consider the adversary (GDA): she may delay certain messages and thus create a situation where for well-timed messages in the channel's waiting queue after $\tau_{\max}$ time the QoS requirements cannot be fulfilled. In this case, two solutions can be chosen by the APROB Channel.

- The first approach would be to drop such messages. This way the result of the adversary's actions would be the denial of service for the messages, which she might have also achieved by delaying those messages infinitely.
- The other approach would be to drop the real-time restriction and to let such messages wait until the QoS requirements can be fulfilled. Thus the attacker has succeeded in down-rating the system to a non-real-time one. By simply delaying such messages, the she could have achieved the same.

We have to emphasize that in neither case have the anonymity requirements of messages been broken, the adversary could only either down-rate the system into a non-real-time one or force some to be dropped. GDA could *not gain anything* besides what by definition she had already been capable of. Thus:

Theorem 3. *The APROB Channel (1) provides a guaranteed level of anonymity against a global delaying adversary with a local source-hiding property according to (6) and (7) while delivering messages in a semi-real-time manner, and (2) a GDA cannot compromize the achieved level of anonymity by delaying messages.*

6 Conclusion

In this paper the APROB Channel has been introduced for real-time anonymous communication. It has been proven that the construction could provide guaranteed anonymity while also fulfilling semi-real-time guarantees. The channel even enabled users to ask for different anonymity levels (quality of service). Under these circumstances it has been shown that a global delaying adversary cannot gain any new information by delaying messages, and thus it would not be worth for such an adversary to delay messages.

For future work two main directions present themselves. Should the adversary have *a priori* information about the preferences of senders, i.e. how they choose recipients, then this would introduce a serious breach in the anonymity provided by current

anonymity systems. Thus, adaptive channels need to be constructed that maintain information about the users' preferences and shape the traffic in order to confuse the adversary. Up to now we have only considered black-box channels and thus external adversaries. A big step forward will be to organize *networks* of the analyzed channels and evaluate their properties.

References

1. Tóth, G., Hornák, Z.: Measuring anonymity in a non-adaptive, real-time system. In: Proceedings of Privacy Enhancing Technologies (PET2004). Springer-Verlag, LNCS, Forthcoming (2004)
2. Pfitzmann, A., Köhntopp, M.: Anonymity, unobservability, and pseudonymity – a proposal for terminology. In Federrath, H., ed.: Designing Privacy Enhancing Technologies. Volume 2009 of Springer-Verlag, LNCS., Berkeley, CA (2001) 1–9
3. Tóth, G., Hornák, Z., Vajda, F.: Measuring anonymity revisited. In Liimatainen, S., Virtanen, T., eds.: Proceedings of the Ninth Nordic Workshop on Secure IT Systems, Espoo, Finland (2004) 85–90
4. Serjantov, A., Danezis, G.: Towards an information theoretic metric for anonymity. In Syverson, P., Dingledine, R., eds.: Proceedings of Privacy Enhancing Technologies (PET2002). Volume 2482 of Springer-Verlag, LNCS., San Francisco, CA (2002)
5. Díaz, C., Seys, S., Claessens, J., Preneel, B.: Towards measuring anonymity. In Syverson, P., Dingledine, R., eds.: Proceedings of Privacy Enhancing Technologies (PET2002). Volume 2482 of Springer-Verlag, LNCS., San Francisco, CA (2002) 54–68
6. Dingledine, R., Mathewson, N., Syverson, P.: Tor: The second-generation onion router. In: Proceedings of the 13th USENIX Security Symposium. (2004)
7. Chaum, D.: Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM **4** (1981) 84–88