SOURCE HIDING PROPERTY OF AN OBSERVABLE BLACK-BOX CHANNEL

Gergely TÓTH
Advisors: Zoltán HORNÁK, Ferenc VAJDA

I. Introduction

Anonymous transport protocols (such as onion-routing [1] [2]) revolutionize anonymous messaging. The design of such protocols is advancing, however their theoretical analysis is incomplete. In this paper I will introduce the model of an observable black-box channel and deduce the extent of conclusions that an eavesdropper observer can draw. Source hiding property that can be used as a theoretical degree of anonymity will also be introduced.

II. Model of an Observable Black-box Channel

In the following the model of an observable black-box channel will be introduced. An observable channel means that an observer can see all messages entering and leaving the channel [3]. However the channel is a black-box, thus the observer cannot see what happens with the messages inside the channel and how these are transmitted and encoded. Since the messages are assumed to be encoded and uniform in length an observer can only gain information from the timing of events.

A. The environment

Let S denote the set of senders, R the set of recipients and M the set of messages. Let $S(m_i)$ denote the sender of message m_i, $R(m_i)$ the recipient of message m_i, $t_S(m_i)$ the time of sending of message m_i and $t_R(m_i)$ the time of receipt of message m_i.

B. The channel

Inside the channel no messages will be born or dropped. The channel is expected to delay the transmission of messages according to the following rules:

- the delay is a probability variable δ with a given density function $f(\delta)$, $\delta = t_R - t_S$, where δ is both message and time invariant;
- the channel delivers all messages before a predefined constant $\delta_{\max}$ (time-to-live) and after a predefined constant $\delta_{\min}$ (minimal delay) thus $\forall_{m_i}[\delta_{\min} < t_R(m_i) - t_S(m_i) < \delta_{\max}]$.

So a channel C can be given by $f(\delta)$, $\delta_{\min}$, $\delta_{\max}$.

C. Message transmission

Assume that $s_a \in S$ sends a message $m_i \in M$ to recipient $r_b \in R$. Message m_i arrives to the channel in an encrypted form $\alpha_i := E_S(r_b, m_i)$ at time $t_S(m_i) = t_S(\alpha_i)$ and will be delivered from the channel in a differently encrypted form $\beta_i := E_R(r_b, m_i)$ at time $t_R(m_i) = t_R(\beta_i)$. It is assumed that E_S and E_R are ideal encryption functions thus α_i can only be decrypted by the channel, β_i only by r_b.

III. The Observer

In this paper only a passive observer will be handled, thus the observer can only eavesdrop all encrypted messages but it cannot change or delay them or otherwise modify any parameters in the system. The aim of the observer is to map or at least guess with good probability who sends messages to whom, thus tracking all communication circuits.

A. *Knowledge of the observer*

The observer knows solely the following:

- the environment (S, R) and the attributes ($f(\delta)$, $\delta_{\min}$, $\delta_{\max}$) of the channel;
- $\varepsilon_S := \{\alpha_i := E_S[S(m_i), m_i]\}$, $\vartheta_S := \{t_S(\alpha_i)\}$, $\varepsilon_R := \{\beta_i := E_R[R(m_i), m_i]\}$, $\vartheta_R := \{t_R(\beta_i)\}$.

B. *History* Ψ *of the System*

Let Ψ denote the history of messages and system parameters given as a tuple C, S, R, ε_S, ε_R, ϑ_S, ϑ_R.

C. *Confidence of the observer*

To evaluate in a channel C^* which sender sends which message, for each delivered encrypted message β_k^* and for each sender s_l^* a probability $P_{\beta_k^*, s_l^*, \Psi^*}$ can be determined. If the observer knows the history Ψ^*, it can determine that β_k^* has been sent by s_l^* with the probability $P_{\beta_k^*, s_l^*, \Psi^*}$.

With the introduction of the sets $\mu_{\beta_i^*, \Psi^*}$ and $\eta_{\beta_i^*, s_l^*, \Psi^*}$ equation (1) yields $P_{\beta_k^*, s_l^*, \Psi^*}$:

$\mu_{\beta_i^*, \Psi^*} := \{\alpha_j^* \mid [t_R(\beta_i^*) - \delta_{\max}^*] \le t_S(\alpha_j^*) \le [t_R(\beta_i^*) - \delta_{\min}^*]\}$ – the sent messages in the relevant time frame;

$\eta_{\beta_i^*, s_l^*, \Psi^*} := \{\alpha_j^* \mid [\alpha_j^* \in \mu_{\beta_i^*, \Psi^*}] \wedge [S(\alpha_j^*) = s_l^*]\}$ – subset of $\mu_{\beta_i^*, \Psi^*}$ respectively for each sender;

$$P_{\beta_k^*, s_l^*, \Psi^*} = P[S(\beta_k^*) = s_l^* \mid \Psi = \Psi^*] = \frac{\sum\limits_{\forall_{\alpha_i^*}[\alpha_i^* \in \eta_{\beta_k^*, s_l^*, \Psi^*}]} f^*[t_R(\beta_k^*) - t_S(\alpha_i^*)]}{\sum\limits_{\forall_{\alpha_j^*}[\alpha_j^* \in \mu_{\beta_k^*, \Psi^*}]} f^*[t_R(\beta_k^*) - t_S(\alpha_j^*)]} \tag{1}$$

The observer naturally looks for the most probable source of a message where $P_{\beta_k^*, \Psi^*} := \max\limits_{s_l^*} P_{\beta_k^*, s_l^*, \Psi^*}$.

IV. Source Hiding Property

A history $\Psi^* := (C^*, S^*, R^*, \varepsilon_S^*, \varepsilon_R^*, \vartheta_S^*, \vartheta_R^*)$ is source hiding with parameter Θ, if $\forall\limits_{\beta_k^* \in \varepsilon_R^*} P_{\beta_k^*, \Psi^*} \le \Theta$.

In order to best limit $P_{\beta_k^*, s_l^*, \Psi^*}$, the $f^*(\delta)$ density function should be chose to have constant $f_{\max}^*$ value between $\delta_{\min}^*$ and $\delta_{\max}^*$ and to be 0 elsewhere. In this case statement (2) can be given for Θ as:

$$\Theta_{\beta_k^*, \Psi^*} = P_{\beta_k^*, \Psi^*} = \frac{\max\limits_{s_l^*} \left| \eta_{\beta_k^*, s_l^*, \Psi^*} \right|}{\left| \mu_{\beta_k^*, \Psi^*} \right|} . \tag{2}$$

V. Conclusions

In this paper a model of an observable black-box channel has been introduced. It was shown what conclusions a passive observer can draw based only on the timing of events and properties of the channel. The source hiding property was introduced, which represents the degree of anonymity. To provide a theoretical background for anonymity the black-box channel should be opened and the inner working of the channel should be evaluated, while possibilities of an active observer (e.g. maliciously delaying messages) should also be analyzed.

References

[1] M. Reed, P. Syverson, D. Goldschlag: "Anonymous Connections and Onion Routing", in *IEEE Journal on Selected Areas in Communication Special Issue on Copyright and Privacy Protection*, 1998.

[2] D. Goldschlag, M. Reed, P. Syverson: "Hiding Routing Information", in *Information Hiding*, R. Anderson (editor), Springer-Verlag LLNCS 1174, pp. 137–150, 1996

[3] A. Pfitzman, M. Kohntopp: "Anonymity, Unobservability, and Pseudonymity – A Proposal for Terminology", in *Designing Privacy Enhancing Technologies*, H. Federrath (editor), Springer-Verlag LNCS 2009, pp. 1–9, 2001