# Non-expanding Transaction Specific Pseudonymization for IP Traffic Monitoring

Lasse Øverlier[1,2], Tønnes Brekne[3], and André Årnes[3]

[1] Norwegian Defence Research Establishment, P.B. 25, 2027 Kjeller, Norway
lasse.overlier@ffi.no, http://www.ffi.no/
[2] Gjøvik University College, P.B. 191, 2802 Gjøvik, Norway
lasse@hig.no, http://www.hig.no/
[3] Centre for Quantifiable Quality of Service in Communication Systems,
Norwegian University of Science and Technology,
O.S. Bragstads plass 2E, N-7491 Trondheim, Norway
{tonnes, andrearn}@q2s.ntnu.no, http://www.q2s.ntnu.no/

**Abstract.** This paper presents a scheme for transaction pseudonymization of IP address data in a distributed passive monitoring infrastructure. The approach provides high resistance against traffic analysis and injection attacks, and it provides a technique for gradual release of data through a key management scheme. The scheme is non-expanding, and it should be suitable for hardware implementations for high-bandwidth monitoring systems.

## 1 Introduction

This paper presents a scheme for transaction pseudonymization[1] of IP addresses in traffic data collected from distributed passive network monitoring sensors on high-capacity network links. This work continues our earlier work in evaluating candidate solutions for anonymization of passive monitoring data in the context of the LOBSTER[2] and SCAMPI [3] projects. The motivation for this research is that pseudonymization of network monitoring data becomes challenging when it must simultaneously satisfy the conflicting requirements of privacy and traffic analysis applications. Also, the huge amount of real-time data handled at high-capacity backbone network connections imposes strict resource constraints.

We begin by introducing some terminology, along with the context and motivation for this work. After listing some pivotal assumptions, we give a brief overview of injection attacks, which our work is designed to protect against. Some related work is mentioned, before we proceed with a description of the

[1] We employ this term in the sense of "one-time pseudonyms" as mentioned in [1]. We have previously used the term *instance specific pseudonymization* in our papers.
[2] LOBSTER is a pilot European Infrastructure for large-scale monitoring of broadband Internet infrastructure, see http://www.ist-lobster.org/
[3] SCAMPI is a EU project for creating a scalable and programmable monitoring platform for the Internet, see http://www.ist-scampi.org/

Y.G. Desmedt et al. (Eds.): CANS 2005, LNCS 3810, pp. 261–273, 2005.

scheme and its associated key management scheme. The paper ends with a description of the scheme's capabilities, and an analysis of some of its security properties. Finally, we present the conclusions of this work.

We have previously shown that an active adversary could efficiently attack prefix-preserving pseudonymization of IP addresses gathered using passive network monitors[2]. We have also demonstrated how *any* static pseudonymization scheme fails in the face of injection attacks, where an adversary sends forged IP packets with arbitrary source and destination IP addresses in such a way that they are recognizable in their pseudonymized forms [3].

The term *static pseudonymization*, refers to a scheme where each plaintext value has a unique and unchanging pseudonym. *Transaction pseudonymization* refers to a scheme where each pseudonym for a plaintext value is unlinkable[4] to any other pseudonym of the same plaintext value. In this way, there is no recognizable relationship between different pseudonyms of the same plaintext value.

The scheme presented in this paper is transaction specific, providing protection against injection attacks, while supporting efficient matching of pseudonyms for an authorized user through the use of partial disclosure of address information. The scheme is non-expanding and requires no more storage space than the original plaintext address. It is intended to provide a flexible solution for pseudonymization in high-capacity networks, supporting different applications and user groups with various requirements and trust levels.

### 1.1 Context and Threat Model

In the following, we base our context and threat model assumptions on [2, 3]. A reiteration is given here for the benefit of the reader. We consider only the pseudonymization of IP-addresses, although our methods are applicable to other data types as well. The IP addresses are assumed to be $n$ bits in length.

The context is that of passive sensors monitoring an IP network, and anonymizing captured traffic data. The sensors are programmable network monitoring cards[5] capable of operating on high-capacity links ($\leq$ 10Gbit/s). The IP addresses are anonymized at the sensor node, and a sensor identifier is appended to the data. The data rates involved impose strict performance requirements on all processing tasks. As the network monitoring system is distributed, the pseudonymization scheme has to be consistent across the sensors in order to support distributed analysis applications.

We wish to prevent adversaries from reidentifying IP addresses under the following assumptions:

**Assumption 1.** *The adversary may send forged network traffic with arbitrary source and destination IP addresses.*

**Assumption 2.** *The adversary is capable of ensuring that injected packets are captured by at least one passive sensor.*

[4] Unlinkability means that "two or more items within a system are no more and no less related than they are related concerning a-priori knowledge" [1].

[5] Examples of such cards are SCAMPI cards and Endace DAG cards.