perpass non-WG
Internet-Draft
Intended status: Informational
Expires: March 08, 2014

B. Trammell
ETH Zurich

September 04, 2013

The Perfect Passive Adversary: A Threat Model for the Evaluation of Protocols under Pervasive Surveillance
draft-trammell-perpass-ppa-00.txt

Abstract

This document elaborates a threat model for the Perfect Passive Adversary (PPA): an adversary with an interest in eavesdropping that can passively observe network traffic at every layer at every point in the network between the endpoints. It is intended to demonstrate to protocol designers and implementors the observability and inferability of information and metainformation transported over their respective protocols, to assist in the evaluation of the performance of these protocols and the effectiveness of their protection mechanisms under pervasive passive surveillance.

Table of Contents

1. Open Issues

1. Lots of things need citations that don't have them yet.

2. Threat analysis and protocol design guidelines need to be completed, which will require them to be started too.

2. Introduction

Surveillance is defined in [RFC6973], Section 5.1.1, as "the observation or monitoring of an individual's communications or activities". Pervasive passive surveillance is the practice of surveillance at widespread observation points, without any modification of network traffic, and without any particular surveillance target in mind. Pervasive passive surveillance allows subsequent analysis and inference to be applied to the collected data to achieve surveillance aims on a target to be identified later, or to analyze general communications patterns and/or behaviors without a specified target individual or group.

An analysis of the costs and benefits of pervasive passive surveillance is explicitly out of scope of this work; we presume a priori that communications systems should aim to provide appropriate privacy guarantees to their users, and that such pervasive surveillance is therefore a bad thing. Therefore, susceptibility to pervasive surveillance should avoided as a design goal in protocol

design. From these assumptions we take the very act of pervasive surveillance to be adversarial by definition.

This document outlines a threat model for an entity performing pervasive passive surveillance, termed the Perfect Passive Adversary (PPA), and explores how to apply this model to the evaluation of protocols. As the primary threat posed by pervasive surveillance is a threat to the privacy of the parties to a given communication, this document is heavily based on [RFC6973].

3. Terminology

The terms Anonymity, Anonymity Set, Anonymous, Attacker, Eavesdropper, Fingerprint, Fingerprinting, Identifier, Identity, Individual, Initiator, Intermediary, Observer, Pseudonym, Pseudonymity, Pseudonymous, Recipient, and Traffic Analysis are used in this document as defined by Section 3, Terminology, of [RFC6973]. In addition, this document defines the following terms:

Observation: Information collected directly from communications by an eavesdropper or observer. For example, the knowledge that <alice@example.com> sent a message to <bob@example.com> via SMTP taken from the headers of an observed SMTP message would be an observation.

Inference: Information extracted from analysis of information collected directly from communications by an eavesdropper or observer. For example, the knowledge that a given web page was accessed by a given IP address, by comparing the size in octets of measured network flow records to fingerprints derived from known sizes of linked resources on the web servers involved would be an inference.

4. The Perfect Passive Adversary

The perfect passive adversary (PPA) is an eavesdropper that can potentially observe every packet of all communications at any or every hop in a network path between the outward-facing network interface of the last trusted machine in the initiator's administrative domain and the recipient, but can take no other action with respect to these communications. Limiting the adversary to being completely passive may under-represent the threat to communications privacy posed especially by well-resourced adversaries, but represents well the maximum capability of a single entity whose surveillance is undetectable without physically securing the entire network path. We also assume that the PPA does not have unlimited resources, i.e., that it will attempt to eavesdrop at the most efficient observation point available to it, and will collect as little raw data as necessary to support its aims.

We explicitly assume the PPA does not have the ability to compromise trusted systems at either the initiator or a recipient of a communication. Indeed, if the adversary is cooperating with one of the communications endpoints, there is no guidance to give to protocol designers that would improve the privacy and security of the individual at the other endpoint: the compromise is then truly out of the scope of the communication enabled by the protocol.

We further assume the PPA does not have priviliged information allowing the reversal of encryption, e.g. compromised key material or knowledge of weaknesses in the design or implementation of cryptographic algorithms at the initiator, recipient, and/or intermediaries. While these risks do exist in the real world, the threat model is simplified if we presume that a given cryptographic protection for a protocol works as advertised.

The tools available to the PPA are therefore direct observation and inference. Direct observation involves taking information directly from eavesdropped communications - e.g., URLs identifying content or email addresses identifying individuals from application-layer headers. Inference, on the other involves analyzing eavesdropped information to derive new information from it; e.g., searching for application or behavioral fingerprints in observed traffic to derive information about the observed individual from them, in absence of directly-observed sources of the same information.

5. Threat analysis

On initial examination, the PPA would appear to be trivially impossible to defend against. If the PPA has access to every byte of every packet of a communication, then full application payload and content is available. Guidance to protocol designers to provide cryptographic protection of confidentiality in their protocols (e.g.,

through the use of TLS [RFC5246] at the transport layer and S/MIME [RFC3851] end-to-end) improves this situation somewhat, but metadata such as source and destination IP addresses and ports are still available to allow correlation and association of communications. Protocols that route messages based on recipient identifier or pseudonym, such as SMTP [RFC2821] and XMPP [RFC6120], still require intermediate systems to handle these. If each hop of the communication is not secured, these identifiers may be available to an eavesdropper.

Assuming that the PPA's resources are not unlimited allows us to back away from this worst-case scenario. Storing full packet information for a fully-loaded 10 Gigabit Ethernet link will fill one 4TB hard disk (the largest commodity hard disk available as of this writing) in less than an hour; storing network flow data from the same link, e.g. as IPFIX Files [RFC5655], requires on the order of 1/1000 the storage (i.e., 4GB an hour). Flow-based surveillance approaches, which store only communications metadata, are therefore more scalable for pervasive surveillance, so it is worthwhile to analyze information which can be inferred from various network traffic capture and analysis techniques other than full packet capture.

In the remainder of this analysis, we list kinds of information which can be directly observed and those which can be used for inference through e.g. fingerprinting. The former group may seem somewhat obvious, but are included for completeness.

5.1. Information subject to direct observation

[EDITOR'S NOTE: list includes but not limited to communications content, application-layer identifiers, network- and transport-layer identifiers, association of DNS queries with subsequent usage of information in the answers.]

5.2. Metainformation useful for inference

[EDITOR'S NOTE: list includes but not limited to interpacket timing; packet sizes; flow packet and octet counts; presence of options which could lead to OS fingerprinting for deNATting, etc.]

6. Guidelines for protocol evaluation

[EDITOR'S NOTE: How to look at a protocol and evaluate the observability of the information it transports?]

[EDITOR'S NOTE: General guidance: end-to-end encryption when possible. Apply unlinked pseudonyms for message routing on envelopes around end-to-end encrypted content.]

[EDITOR'S NOTE: General guidance: Fingerprinting can rely on packet and flow size information; the inclusion of null information in packets, or grouping information into more/fewer packets can reduce this risk at the expense of usable bandwidth; though this is implementation guidance, protocols should make it possible do to dhis. Similarly, fingerprinting can rely on inter-packet timing information: injecting delay into packet transmission can reduce this risk at the expense of latency.]

7. Acknowledgments

Thanks to Dilip Many, Daniel Borkmann, and Stephan Neuhaus, who contributed to an initial version of this work.

8. References

8.1. Normative References

[RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J., Morris, J., Hansen, M., and R. Smith, "Privacy Considerations for Internet Protocols", RFC 6973, July 2013.

8.2. Informative References

[RFC2821] Klensin, J., "Simple Mail Transfer Protocol", RFC 2821, April 2001.

[RFC3851] Ramsdell, B., "Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 3.1 Message Specification", RFC 3851, July 2004.

[RFC5246] Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, August 2008.

[RFC5655] Trammell, B., Boschi, E., Mark, L., Zseby, T., and A. Wagner, "Specification of the IP Flow Information Export (IPFIX) File Format", RFC 5655, October 2009.

[RFC6120] Saint-Andre, P., "Extensible Messaging and Presence Protocol (XMPP): Core", RFC 6120, March 2011.

Author's Address

Brian Trammell
Swiss Federal Institute of Technology Zurich
Gloriastrasse 35
8092 Zurich
Switzerland

Phone: +41 44 632 70 13
Email: trammell@tik.ee.ethz.ch