MailVortex

A extension to traditional transport protocols to offer anonymity towards third parties

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von
Martin Gwerder (06-073-787)
von Glarus GL
July 7, 2016

Genehmigt von der Philosophisch-Naturwissenschaftlichen FakultÃ‘t
Auf Antrag von
Prof. Dr. Christian F. Tschudin
Prof. Dr. Heiko Schuldt

Basel, FIXME datum der fakultÃ‘tsversammlung FIXME Signature line with Name of Dean of Faculty below

::

Contents

List of Tables

List of Figures

1. Introduction

Numerous events in present and past have shown that data is broadly collected in the internet. Wether this is a problem or not may be a disputable fact. Undisputed is however that if data is not handled with care peoples are accused with numerous “facts” that are more than questionable. To show that this may happen even under complete democratic control we might refer to events such as the “secret files scandal?? (or “Fichenskandal”) in Switzerland. In the years from 1900 to 1990 Swiss government collected 900’000 files in a secret archive (covering roughly 10% of the natural and juristic entities within Switzerland at that time).

A series of similar attempts to attack privacy on a global scale have been discovered by whistle blower Edward Snowden. The documents leaked in 2009 by him claim that there was a data collection starting in 2010. Since these documents are not publicly available it is hard to to prove claims based on these documents. However – due to the fact that the documents were screened by a significant number of journalists spanning multiple countries, the information seems credible.

According to these documents (verified by NRC) NSA infiltrated more than 50k computers with malware to collect classified information. They furthermore infiltrated Telecom-Operators (executed by british GCHQ) such as Belgacom to collect data and targeted high member of governments even in associated states (such as the germans president mobile phone). A later published shortened list of “selectors” showed 68 telephone and fax numbers targeting enconomy, finance and agricultural parts of the german government.

This list of events shows that big players are collecting and storing vast amounts of data for future use. The list of events shows also that the use of this data has in the past been at least partially questionable. As a part of possible counter measures this work analyses the possibility of using state of the art technology to minimize the information footprint of a person on the internet.

We leave a vast information footprint in our daily communication. On a regular email we disclose everything in an “postcard” to any entity on its way. Even when encrypting a message perfectly with todays technology (S/MIME[4] or PGP[3]) leaves at least the originating and the receiving entity disclosed. Most likely other relevant information such as “message subject”, “frequency of exchanged messages”, “size of messages”, or “client beeing used”. A good anonymity protocol has therefore far more attributes to cover than the message itself. Furthermore a protocol anonymising messages should not rely on the trust of infrastructure other than the infrastructure under control of the sending or receiving entity.

Any central infrastructure is bound to be of special interest to anyone gathering data concerning the using entities of such a protocol. So central infrastructure has to be avoided.

In this work a new protocol is designed to allow message transfer through existing communication channels. These messages should be unobservable to any third party. This unobservability does not only cover the message itself but all metadata associated with it.

2. Main Research Question

The main topic of this thesis was defined as follows:

- Is it possible to have specialized asynchronous messaging protocol based on “state of the Science” technologies offering a high level of anonymity (sender and receiver anonymity) towards an advisor with a high budget and privileged access to infrastructure?

Based on this main question there are several sub questions grouped around various topics:

- What technologies and methods can be used to provide anonymity against a potential adverser?
- How can entities utilizing these technologies and methods be attacked?
- How can attacks targeting anonymity of a sending or receiving entity be mitigated by design?

Part I.

Methodes

In this part of the dissertation we collect definitions, methods, and existing research relevant to the topic of this thesis

3. Existing Research and Implementations on the Topic

3.1. Anonymity

As Anonymity we take the definition as specified in [8].

> Anonymity of a subject means that the subject is not identifiable within a set of subjects, the anonymity set. [1]

and

> Anonymity of a subject from an attacker's perspective means that the attacker cannot sufficiently identify the subject within a set of subjects, the anonymity set.[1]

Whereas the anonymity set is defined as the set of all possible subjects.

Especially the second quote is very important to this paper.

3.1.1. k-Anonymity

k-anonymity is a term introduced in [1]. This work claims that no one might be held responsible for an action if the action itself can only be identified as an action which has been taken by one unidentifiable entity out of k entities.

The Document distinguishes between *Sender k-anonymity* where the sending entity can only be narrowed down to a set of k entities and *Receiver k-anonymity*

3.1.2. ℓ-Diversity

In [7] an extended model of k-anonymity. According to the authors it is possible to break a k-anonymity set if there is additional Information available which may be merged into a data set so that a special entity can be filtered from the k-anonymity set. In other words if an anonymity set is to tightly specified a single additional background information might be sufficient to identify a specific entity in an anonymity set.

Their approach is to introduce an amount of invisible diversity into k-anonymous sets so that simple background knowledge is no longer sufficient to isolate a single member.

3.1.3. t-Closeness

While ℓ-diversity protects the identity of an entity it does not prevent informatino gain. A subject which is in a class has the same attributes. This is where t-closeness[6] comes into play. t-closeness is defined as follows:

> An equivalence class is said to have t-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole table is no more than a threshold t. A table is said to have t-closeness if all equivalence classes have t-closeness.

3.2. Zero Trust

Zero trust is not a truly researched model in systems engineering. It is however widely adopted.

[1] footnotes omitted in quote

We refer in this work to the zero trust model when denying the trust in any infrastructure not directly controlled by the sending or receiving entity. This distrust extends especially but not exclusively to the network transporting the message, the nodes storing and forwarding messages, the backup taken from any system, and software, hardware and operators of all systems not explicitly trusted.

3.3. Pseuidonymity

As Pseudonymity we take the definition as specified in [8].

> A pseudonym is an identifier of a subject other than one of the subject's real names. The subject which the pseudonym refers to is the holder of the pseudonym. A subject is pseudonymous if a pseudonym is used as identifier instead of one of its real names.[2]

3.4. undetectability

3.5. unobservability

3.5.1. Ephemeral Identity

3.6. Single Use Reply Blocks

3.7. Censorship

3.7.1. Censorship Resistant

3.7.2. Parrot Coircumvention

3.7.3. Censorship Circumvention

3.7.3.1. Covert Channel

3.7.3.2. Spread Spectrum

3.8. Cryptography

3.8.1. Symmetric Encryption

3.8.1.1. RSA

3.8.1.2. Elliptic Curve Cryptogaphy

3.8.2. Asymmetric Encryption

3.9. System Implementations

The following sections describe

[2] footnotes omitted in quote

3.9.1. Concepts

3.9.1.1. DC Networks

3.9.1.2. MIX Networks

3.9.1.3. Onion Routing

3.9.1.4. Remailer

3.9.2. Implementations

The following sections emphasize on implementations of anonymising (and related) protocols regardless of their usage in the domain of messaging. It si a list of system classes or their specific implementations together with a short analysis of strength and weaknesses. Wherever possible we try to refer to original sources.

3.9.2.1. Pseudonymous Remailer

3.9.2.2. Babel

3.9.2.3. Cypherpunk-Remailer

3.9.2.4. Mixmaster-Remailer

3.9.2.5. Mixminion-Remailer

3.9.2.6. Crowds

3.9.2.7. Herbivore

3.9.2.8. Dissent

3.9.2.9. P5

3.9.2.10. Gnutella

3.9.2.11. Gnutella2

3.9.2.12. Freenet

3.9.2.13. Darknet

3.9.2.14. Sneakernet

3.9.2.15. Hordes

3.9.2.16. Salsa

3.9.2.17. Hydra-Onion

3.10. Known Attacks

In the following sections we emphasize on possible attacks to an anonymity preserving protocols. In the following sections we describe classes of attacks. These attacks may be used to attack the anonymity of any

entity involved in the message channel. In a later stage we test the protocol for immunity against these classes of attacks.

3.10.1. Broken Encryption Algorithms

3.10.2. Attacks Targeting Anonymity

3.10.2.1. Hotspot Attacks

3.10.2.2. Message Tagging and Tracing

3.10.2.3. Side Channel Attacks

3.10.2.4. Bugging Attacks

3.10.3. Denial of Service Attacks

3.10.3.1. Censorship

3.10.3.2. Credibility Attack

4. Applied Methodes

Part II.

Results

5. MessageVortex - Transport Independent Messaging anonymous to 3rd Parties

6. Security Analysis

7. Additional Considerations

7.1. Storage of Messages

The storage of messages sent though MessageVortex should be handled with great care. It seems on the first sight a good idea to merge all messages in a globally available storage such as the mail account of the receiving entity. However – In doing so we would discover the message content to the providing party of a mail account. Since we handled the message with great care and tremendous costs up until this point it would be careless doing so.

Storing them in a localized and receiving entity controlled storage is definitely a good idea but leaves security considerations like a backup possibly to an end user. This might be better but in effect a questionable decision. There is however a third option. By leaving the message unhandled on the last entity of the MessageVortex chain we may safely backup the data without disclosing the message content. Merging the content then dynamically through a specialized proxy would allow the user tu have a unified view on his without compromising the security.

Part III.

Discussion

8. Anonymity

Glossary

adverser FIXME

Agent FIXME

EWS FIXME

IMAP IMAP (currently IMAPv4) is a typical protocol to be used between a Client MRA and a Remote MDA. It has been specified in its current version in [2]. The protocol is capable of fully maintaining a server based message store. This includes the capability of adding, modifying and deleting messages and folders of a mailstore. It does not include however sening mails to other destinations outside the server based store.

LMTP FIXME

Local Mail Store A Local Mail Store offers a persistent store on a local non volatile memory in which messages are beeing stored. A store may be flat or structured (eg. supports folders). A Local Mail Store may be an authoritative store for mails or a "‘Cache Only"’ copy. It is typically not a queue.

mail server admin FIXME

MDA An MDA provides an uniform access to a Local Mail Store.

Remote MDA A Remote MDA is typically supporting a specific access protocol to access the data stored within a Local Mail Store .

Local MDA A Local MDA is typically giving local applications access to a server store. This may be done thru an API, a named socket or similar mechanisms.

MRA A Mail receiving Agent. This agent receives mails from a agent. Depending on the used protocol two subtypes of MRAs are available.

Client MRA A client MRA picks up mails in the server mail storage from a remote MDA. Client MRAs usually connect thru a standard protocol which was designed for client access. Examples for such protocols are POP or IMAP

Server MRA Unlike a Client MRA a server MRA listens passively for incomming connections and forwardes received Messages to a MTA for delivery and routing. A typical protocol supported by an Server MRA is SMTP

MS-OXMAPIHTTP FIXME

MSA A Mail Sending Agent. This agent sends mails to a Server MRA.

MTA A Mail Transfer Agent. This transfer agent routes mails between other components. Typically an MTA receives mails from an MRA and forwardes them to a MDA or MSA. The main task of a MTA is to provide reliable queues and solid track of all mails as long as they are not forwarded to another MTA or local storage.

MTS A Mail Transfer Service. This is a set of agents which provide the functionallity tor send and receive Messages and forward them to a local or remote store.

MSS A Mail Storage Service. This is a set of agents providing a reliable store for local mail accounts. It also provides Interfacing which enables clients to access the users mail.

MUA A Mail User Agent. This user agent reads mails from a local storage and allows a user to read existing mails, create and modify mails.

Privacy From the Oxford English Dictionary: “

1. The state or condition of beeing withdrawn from the society of others, or from the public intrest; seclusion. The state or condition of beeing alone, undisturbed, or free from public attention, as a matter of choice or right; freedom from interference or intrusion.
2. Private or retired place; private apartments; places of retreat.
3. Absence or avoidance of publicity or display; a condition approaching to secrecy or concealment. Keeping of a secret.

4. A private matter, a secret; private or personal matters or relations; The private parts.
5. Intimacy, confidential relations.
6. The state of being privy to some act.

”[9, FIXME]
In this work privacy is related to definition two. Mails should be able to be handled as a virtual private place where no one knows who is talking to whom and about what or how frequent (except for directly involved people).

POP POP (currently in version 3) is a typical protocol to be used between a Client MRA and a Remote MDA. Unlike IMAP it is not able to maintain a mail store. Its sole purpose is to fetch and delete mails in a server based store. Modifying Mails or even handling a complex folder structure is not doable with POP

Service FIXME

SMTP SMTP is the most commonly used protocol for sending mails across the internet. In its current version it has been specified in [5].

Storage A store to keep data. It is assumed to be temporary or persistent in its nature.

user FIXME

UBE FIXME

Bibliography

[1] Luis von Ahn, Andrew Bortz, and Nicholas J. Hopper. “k-Anonymous Message Transmission”. In: *Proceedings of the 10th ACM Conference on Computer and Communications Security (CCS 2003)*. Ed. by Vijay Atluri and Peng Liu. ACM Press, Oct. 2003, pp. 122–130. DOI: 10.1145/948109.948128. URL: http://www.abortz.com/papers/k-anon.pdf (cit. on p. 9).

[2] M. Crispin. *RFC3501 INTERNET MESSAGE ACCESS PROTOCOL - VERSION 4rev1*. IETF, 2003. URL: http://tools.ietf.org/pdf/rfc3501.pdf (cit. on p. 27).

[3] M. Elkins. *RFC2015 MIME Security with Pretty Good Privacy (PGP)*. IETF, 1996. URL: http://tools.ietf.org/pdf/rfc2015.pdf (cit. on p. 1).

[4] N. Freed and N. Borenstein. *RFC2045 Multipurpose Internet Mail Extensions; (MIME) Part One: Format of Internet Message Bodies*. IETF, 1996. URL: http://tools.ietf.org/pdf/rfc2045.pdf (cit. on p. 1).

[5] J. Klensin. *RFC5321 Simple Mail Transfer Protocol*. IETF, 2008. URL: http://tools.ietf.org/pdf/rfc5321.pdf (cit. on p. 28).

[6] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. “t-closeness: Privacy beyond k-anonymity and l-diversity”. In: *2007 IEEE 23rd International Conference on Data Engineering*. IEEE. 2007, pp. 106–115. DOI: 10.1109/ICDE.2007.367856. URL: http://www.utdallas.edu/~mxk055100/courses/privacy08f_files/tcloseness.pdf (cit. on p. 9).

[7] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. “l-diversity: Privacy beyond k-anonymity”. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1.1 (2007), p. 3. URL: http://www.cs.cornell.edu/~vmuthu/research/ldiversity.pdf (cit. on p. 9).

[8] Andreas Pfitzmann and Marit Hansen. *A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management*. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf. v0.34. Aug. 2010. URL: http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf (cit. on pp. 9, 10).

[9] A. Stevenson. *Oxford Dictionary of English*. Oxford reference online premium. OUP Oxford, 2010. ISBN: 9780199571123. URL: http://www.oed.com (cit. on p. 28).