Research Plan MailVortex

Martin Gwerder

April 7, 2015

It has never been so easy to collect huge amounts of data about people and companies. Todays electronic communication allows not only to quickly reach any point of the world but it also allows to automatically monitoring all kind of interaction that takes place on it. With the huge availability of such monitoring data, dangers of misuse became immanent as [9], [22], [23] and many more do show. Some attempts have been made (such as Mixmaster[13] or Herbivore[11]) to improve privacy again but rarely an attempt has found a broader audience in the wild.

However, none of these attempts ever seriously tried to blend state-of-the-art anonymizing technologies into existing messaging protocols. Moreover, if they did they focused simply on the "encryption" part of the problem completely ignoring meta data and traffic flow (see [16] or [10]). Such an attempt would allow using well-known clients and therefore minimize the amount of change required to move to a new privacy sphere. By blending state-of-the-art anonymizing technologies, it should be impossible to block traffic without causing tremendous collateral damage thus making it very hard to interfere for any adversary.

1 Investigators

This topic has been chosen as main topic for a PhD thesis of Martin Gwerder at the University of Basel. Main supervisor will be Prof. Christian F. Tschudin (Head of Computer Networks Group). A second supervisor has yet to be defined. Being a topic of a PhD thesis the work will be carried out by Martin Gwerder. The supervisors ensure guidance and quality assurance.

2 Research Questions

The following question should be investigated:

> How is it possible to create a secure channel over existing, asynchronous message transfer protocols that is capable of hiding messages and meta information towards any third party.

The information should be at least untraceable for third parties. Optional goals might be (non-conclusive):

- The system may hide information about the sender to the recipient.
- The system may hide information about the recipient to the sender.

As a transport base SMTP as defined in [12] should be used.

Assumption is that an adversary attacking the anonymity of any of the party (sender, recipient or both) has huge founding and almost unlimited capabilities regarding the internet. These capabilities include:

- The capability of monitoring traffic at any point of the internet.
- The capability of controlling DNS.
- The capability of controlling routing.
- The capability of building any kind of infrastructure.

The supervisors ensure guidance and quality assurance.

3 Background and Significance

Almon B. Strowger was the inventor of the first "automatic telephone exchange" which was a major step in information routing automation. He patented it in 1891 [19]. Since then automated routing of information became tremendously important for the western world. Phone calls, internet traffic and even conventional paper mail is routed automatically these days. While automated information routing speeds up the message exchange in our society it enables at the same time to collect automatically data about a persons and their habits. For example are persons more likely to communicate with persons sharing the same interest. Therefore, if a person communicates with several persons known to be passionate philatelists he is more likely to be a philatelist himself.

This is generally known and marketing experts, secret services or research institutes. They try to use available information to be more effective in their job by inventing methods to collect vast amounts of data about persons and then identify interesting individuals or groups for their respective work. While the interest of the data collection owner might be legal, the existence of a database categorizing individuals based on statistical likelihood might be more than questionable. The main problem is that a person in the western world cannot choose in some cases whether he wants to be part of such a data collection or not. Some of the media which are interesting as they offer a lot of data about a person are unavoidable these days. Unavoidable information sources might be public directories, telephone or email.

Ongoing discoveries show that message flows in the internet and our private life are being traced (see [9], [22], [23] and many more). The information obtained through these channels is then being combined with other social streams to obtain a profile of a persons social network. If misused this information may lead to wrong accusations such as being part of illegal activities or socially not accepted groups. It furthermore violates the human rights where Article 19 of the ICCPR states that "everyone shall have the right to hold opinions without interference" and "everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice"[20].

In this thesis, the main topic shall be email. Nowadays, Email is one of the most important business and personal communication media. As such, it contains a lot of individual or secret data that can be easily analysed. As an example just imagine what information could be discovered by analysing the data of a mailbox. Easily identifiable would be who is communicating with the owner of a mailbox ("from" and "to" headers), the topics which are being discussed (keyword analysis of message body and subject) and who else was involved ("cc" and "bcc" headers [whereas the latter is only available in the sent box]). Most of this information (all except the message body) is even available if a message has been encrypted prior to sending. This and the fact that it is easily obtainable makes todayś emails a valuable source for information.

In a business mailbox, we may find respectively business partners, topics of interest to the company, current research and sales data about the companies current products. In the case of a company, the information is even more easily identifiable as the amount of mails is huge. This is because we can analyse all mails of a company and then turn the whole lot into one big result.

4 Research Method, Design and Analysis

The following working structure will be used to try to achieve the goals listed above:

1. The current standard has to be analysed. Known strength and weaknesses have to be collected and the basic capabilities of the current transport channels have to be elaborated.
2. The current state of theory regarding transmission of anonymous messages has to be worked out. It has to be distinguished between three types of anonymity at this point. The "sender anonymity" is given if a sender cannot be determined given a message that has been delivered or is being delivered. "Recipient anonymity" is considered as given when the recipient of a message cannot be determined even if the message content is completely known either at the beginning or the end of a transmission. "Third party anonymity" is given if a message cannot be traced by an observer not involved in sending or receiving the message. This means the observer would be unable to determine neither the content of the message nor its source or its destination.
3. By recombining strengths and weaknesses of previous works, the research should lead to the next step. A protocol should be developed that allows a controllable degree of anonymity while using SMTP as transport technology and thus blending into the regular transfer of the current media.
4. Next step is writing a RFC quality document defining a protocol based on the findings of the previous phase.
5. Based on the protocol definition a prototype has to be built. The prototype must be able to run in an isolated environment simulating hundreds of mail servers. The prototype furthermore needs the capability to run as proxy to a mail client (IMAPv4 or POP3 or EWS).
6. Last step is the verification of the newly designed protocol. It should be analysed based on the prototype and attacks that are already known to be more or less successful on other anonymity systems. Emphasis should be laid on the following type of attacks:

   - Traffic analysis
     Can tuples of mail participant be identified?
   - Tag analysis
     How to tag messages in order to follow them thru forwarding nodes.
   - $(n-1)$ analysis
     is it possible to attack the anonymity with the $(n-1)$ attack?
   - Evil nodes
     Is it possible to break anonymity while controlling a certain amount of nodes involved in the delivery process?
   - Bugging
     Is it possible to break anonymity by using side channel attacks or bugging technologies?

5 Potential Risks

It is possible that despite careful design of a target solution no acceptable solution is found to the problem of anonymous message transfer over SMTP. If so the thesis should outline why this is so and what is needed in order to fill the existing gaps. It should list conclusively what has been tried to elaborate anonymity and what countermeasures have been found to break effectivity of these measures.

Another risk might be that if the thesis leads to a fully or partially effective result the system might be misused for illegal activities (such as black mailing, planning terrorist attacks or sending UBE). If possible precausive actions should be taken to avoid such situations (without violating the main goal).

All anonymizing technologies introduce some kind of "noise" information in that the true message content is hidden. This "noise" is an additional load that has to be handled

by the existing mail infrastructure. If successful the reliability of the infrastructure might be endangered due to the additional load opposed by the new protocol. Nodes that are already on the brink of their capabilities might become overloaded and, therefore, unable to handle regular mail traffic. If possible precausive actions should be taken to avoid such situations (without violating the main goal).

As with all encryption technologies antivirus and antispam measurements are rendered ineffective as an antispam or antivirus program located on a server is unable to scan the terminal message content.

6 Potential Benefits

If successful, this thesis will allow normal users without the help of a provider to set up an anonymous communication channel by using well-known, elaborated technology. It will enable a mail user to control the level of anonymity he would like to have when communicating over the internet. This would fill an important gap of the current western information society.

Another benefit would be that the use of emails, its addresses and its clients are well-known to the community. Acceptance should be far higher if those instead of a new message infrastructure may be used.

7 Related Work

In order to create this plan a minimum research has been carried out. The following is a brief excerpt of related papers and solutions that have been found:

- In "k-Anonymous Message Transmission"[1] Ahn, Bortz, and Hopper emphasised on the level of anonymity introducing a measure of anonymity depending on a constant number of people to whom a transmission-tuple may be traced back.
- In 1988 Chaum writes an excellent article[4] about a method to share information while keeping the source of the information anonymous.
- In "Untraceable Electronic Mail, Return, Addresses, and Digital Pseudonyms"[3] Chaum describes mechanisms addressing anonymity. While the main principles of cloaking a sender (or recipient) by encrypting information "onion like" are still valid, the approach for using mixes showed considerable weaknesses.
- Danezis and Wittneben focused in their work [6] on the economics of anonymity.
- In [8] Dolev and Ostrobsky introduced a method broadcasting a message to multiple recipients while keeping the source secret. The introduced method has a $O(1)$ complexity degree scoring over most of other solutions considerably.
- Herbivore as described in [11] introduces a proprietary protocol implementing Chaums's DC-Nets and a routing layer. This network is based on a centrally known topology of servers.
- The document [13] specifies the protocol for the Mixmaster infrastructure.
- In [14] Parekh focuses on the history of remailers and mixers. His bottom line was that these infrastructures are now well known and anonymity is at danger when using these technologies. In [18] Rennhard and Plattner conclude that privacy on current (static) mixer networks is endangered and dynamic networks are still in their infancy.
- Rennhard and Plattner describe in [17] a method for blending serveral message streams in a way that traffic analysis of encrypted messages is no longer successful. The authors claim that this methode is much more efficient than working with fixed block sizes.
- In "Buses for Anonymous Message Delivery"[2] the authors Beimel and Dolev propose to use a protocol that is clustering multiple messages in "buses" to make traffic analysis harder.
- In [5] Danezis and Laurie propose a message format which should be used as transport protocol for chaumian mixes. This format should be lightweight and efficient.
- Dingledine, Mathewson, and Syverson describe in [7] the design of the Tor-Protocol. Tor is a very common, generic onion routing protocol and [24] of Winter and Lindskog focuses on the Chinese firewall and its dedicated blocking methods for Tor.
- Few steganographic work has been published on research level that has reached a broader audience and all of them (eg. JSteg, OutGuess, invisible secrets, appendX and F5[21]) seem to be broken (see *An Introduction to Steganography*[15] or Hiding in Plain sight FIXME cite)

8 Planed Timeline

It is expected for this thesis to have a duration until 2017. Two mayor phases will be carried out during this time. First, I will do in a "ground work" phase all the information collection, classification and analysis. Furthermore, a design will be created on that practical work should be based.

Second, a prototype is being described and built. Based on the prototype and the analysis in the previous phase the design will be analyzed for weaknesses and design flaws.

All this information is continuously collected in a thesis document and finalized at the end of the "POC" phase.

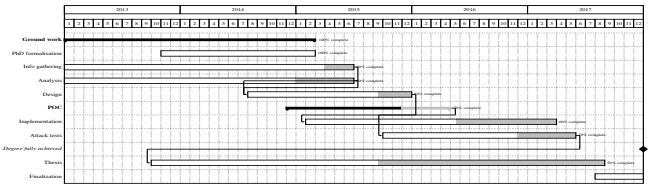

Figure 1: Planed timetable

9 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. 3).

[2] Amos Beimel and Shlomi Dolev. "Buses for Anonymous Message Delivery". In: *Journal of Cryptology* 16.1 (2003), pp. 25–39. DOI: 10.1007/s00145-002-0128-6. URL: ftp://ftp.cs.bgu.ac.il/pub/people/dolev/37.ps (cit. on p. 3).

[3] David Chaum. "Untraceable Electronic Mail, Return, Addresses, and Digital Pseudonyms". In: *Communications of the ACM* (1981). URL: http://www.cs.utexas.edu/~shmat/courses/cs395t_fall04/chaum81.pdf (cit. on p. 3).

[4] David Chaum. "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability". In: *Journal of Cryptology* 1 (1988), pp. 65–75. URL: http://www.cs.ucsb.edu/~ravenben/classes/595n-s07/papers/dcnet-jcrypt88.pdf (cit. on p. 3).

[5] George Danezis and Ben Laurie. "Minx: A simple and efficient anonymous packet format". In: *Proceedings of the Workshop on Privacy in the Electronic Society (WPES 2004)*. Washington, DC, USA, Oct. 2004. URL: http://apache-ssl.securehost.com/minx.pdf (cit. on p. 3).

[6] George Danezis and Bettina Wittneben. "The Economics of Mass Surveillance and the Questionable Value of Anonymous Communications". In: *Proceedings of the Fifth Workshop on the Economics of Information Security (WEIS 2006)*. Ed. by Ross Anderson. Cambridge, UK, June 2006. URL: http://www.cosic.esat.kuleuven.be/publications/article-788.pdf (cit. on p. 3).

[7] Roger Dingledine, Nick Mathewson, and Paul Syverson. "Tor: The Second-Generation Onion Router". In: *Proceedings of the 13th USENIX Security Symposium*. Aug. 2004. URL: http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA465464 (cit. on p. 3).

[8] Shlomi Dolev and Rafail Ostrobsky. "Xor-trees for efficient anonymous multicast and reception". In: *ACM Trans. Inf. Syst. Secur* 3.2 (2000), pp. 63–84. URL: ftp://ftp.cs.bgu.ac.il/pub/people/dolev/31.ps (cit. on p. 3).

[9] Temporary Committee on the ECHELON Interception System. *REPORT on the existence of a global system for the interception of private and commercial communications (ECHELON interception system)*. 2001. URL: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A5-2001-0264+0+DOC+PDF+V0//EN&language=EN (cit. on pp. 1, 2).

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

[11] Sharad Goel, Mark Robson, Milo Polte, and Emin Gun Sirer. *Herbivore: A Scalable and Efficient Protocol for Anonymous Communication*. Tech. rep. 2003-1890. Ithaca, NY: Cornell University, Feb. 2003. URL: http://www.cs.cornell.edu/People/egs/papers/herbivore-tr.pdf (cit. on pp. 1, 3).

[12] J. Klensin. *RFC5321 Simple Mail Transfer Protocol*. IETF, 2008. URL: http://tools.ietf.org/pdf/rfc5321.pdf (cit. on p. 1).

[13] Ulf Möller, Lance Cottrell, Peter Palfrader, and Len Sassaman. *Mixmaster Protocol — Version 2*. IETF Internet Draft. July 2003. URL: http://tools.ietf.org/pdf/draft-sassaman-mixmaster-03.pdf (cit. on pp. 1, 3).

[14] Sameer Parekh. "Prospects for Remailers". In: *First Monday* 1.2 (Aug. 1996), p. 5. URL: https://www.gnunet.org/sites/default/files/Prospects%20for%20Remailers.pdf (cit. on p. 3).

[15] Niels Provos and Peter Honeyman. *An Introduction to Steganography*. IEEE, 2003. URL: http://www.citi.umich.edu/u/provos/papers/practical.pdf (cit. on p. 3).

[16] B. Ramsdell. *RFC3851 Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 3.1 Message Specification*. IETF, 2004. URL: http://tools.ietf.org/pdf/rfc3851.pdf (cit. on p. 1).

[17] Marc Rennhard and Bernhard Plattner. "Introducing MorphMix: Peer-to-Peer based Anonymous Internet Usage with Collusion Detection". In: *Proceedings of the Workshop on Privacy in the Electronic Society (WPES 2002)*. Washington, DC, USA, Nov. 2002. URL: http://cecid.sourceforge.net/morphmix.pdf (cit. on p. 3).

[18] Marc Rennhard and Bernhard Plattner. "Practical Anonymity for the Masses with Mix-Networks". In: *Proceedings of the IEEE 8th Intl. Workshop on Enterprise Security (WET ICE 2003)*. Linz, Austria, June 2003. URL: https://gnunet.org/sites/default/files/RP03-1.pdf (cit. on p. 3).

[19] Almon B. Strowger. *Patent 447918: Automatic Telephone Exchange*. 1891. URL: http://patft.uspto.gov/netacgi/nph-Parser?patentnumber=447918 (cit. on p. 1).

[20] UNHR. *International Covenant on Civil and Political Rights*. 1966. URL: http://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx (cit. on p. 2).

[21] Andreas Westfeld. "F5 - A Steganographic Algorithm". In: *none* none (2002). URL: http://www.ws.binghamton.edu/fridrich/research/f5.pdf (cit. on p. 3).

[22] Wikipedia. *Prism*. 2013. URL: https://en.wikipedia.org/wiki/PRISM_%28surveillance_program%29 (cit. on pp. 1, 2).

[23] Wikipedia. *Tempora*. 2013. URL: https://en.wikipedia.org/wiki/Tempora (cit. on pp. 1, 2).

[24] Philipp Winter and Stefan Lindskog. "How the Great Firewall of China is blocking Tor". In: *Proceedings of the USENIX Workshop on Free and Open Communications on the Internet (FOCI 2012)*. Aug. 2012. URL: https://www.usenix.org/system/files/conference/foci12/foci12-final2.pdf (cit. on p. 3).

10 Changelog

Date	Version	Changes
7.4.2015	1.0	Initial version of this document
2.5.2016	1.1	Updated timeline to reflect new progress
25.11.2016	1.2	Adapted planed end and progress to reality