# Research Plan MailVortex

Martin Gwerder

January 2014

## Abstract

Ongoing discoveries show that message flows in the internet and our private life are being traced (see [1], [2], [3] 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 missused this information may lead to wrong accusations such as being part of illegal activities or socially not acceptable 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"'[4].

*It has never been so easy to collect huge amounts of data about people. The internet allows not only to quickly reach any point of the world but it also allows to automatically monitor all kind interaction which takes place on it. Mobile connection data, Instant messenger data, data from mail accounts and similar kind of data is available in masses and easy to collect and analyse.*

*Ongoing discoveries show that message flows in the internet and our private life are being traced (see [1], [2], [3] 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 missused this information may lead to wrong accusations such as being part of illegal activities or socially not acceptable 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"'[4].*

## 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. Guidance and quality assurance is assured by the supervisors.

## 2 Research Questions

The following question should be investigated:

> How is it possible to create a secure channel over existing, asynchronous message transfer protocols which is capable to hide Messages and Meta-Information towards any third party.

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

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

As a transport base SMTP as defined in [5] should be used.

Assumption is that an adversary attacking the anonymity of any of the party (sender or recipient) 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.

## 3 Background and Significance

Almon B. Strowger was the inventor of the first "'automatic telephone exchange"' which was a mayor step in information routing automation. He patended it in 1891 [6]. Since then automated routing of information became tremendously important for the western world. Phone calls, internet packets and even conventional 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 habits. For example are persons more likely to communicate with persons sharing the same interest. So if a person communicates with several persons known to be passionate pony lovers he is more likely to be a pony lover himself.

This is generally known and marketing experts, secret services or research institutes try to use available information to be more effective. They invent methods to collect vast amount of data about persons and

then try to identify interesting individuals 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 can not choose in some cases whether he wants to be part of such a data collection or not. Some of the media which are interesting are unavoidable. Unavoidable information sources might be public directories, telephone or email.

In this thesis the main topic shall be email. Email is today one of the most important business communication media. As such it contains a lot of individual data which 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 analyse of message body and subject) and who else was involved ("'cc"' and "'bcc"' headers). Most of this information (all except the message body) is even available if a message is being encrypted prior to sending. This and the fact that it is easily capturable makes todays emails a valuable source for information.

# 4 Research Method, Design and Analysis

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

First the current standard has to be analysed. Known strength and weaknesses have to be evaluated and the basic capabilities of the current transport channels have to be elaborated.

Next 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 can not be determined given a message which has been delivered or is being delivered. "'Recipient anonymity"' is considered as given when the recipient of a message can not 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 can not 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 nor its destination.

By recombining strength and weaknesses of previous works the research should lead to the next step. A protocol definition should be written. This protocol should have the following which allows a controllable degree of anonymity while using already existing transport technology and thus blending into the regular transfer of the targeted media.

Next step is writing a RFC quality document defining a protocol based on the findings of the previous phase.

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).

Last step is the verification of the newly designed protocol. It should be analysed based on the prototype and attacks which 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 is it possible to tag messages in order to follow them.
- $(n-1)$ analysis
  is it possible to attack the anonymity with the $(n-1)$ attack?
- Evil nodes
  Is it possible to break anonymity when 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 counter measure has 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 action (such as black mailing, planing terrorist attacks or sending UBE). If possible precausive actions should be taken to avoid such situations (without violating the main goal).

All anonymising technologies introduce some kind of "'noise"' information in which the true message content is hidden. This noise is an additional load which has to be handled by the existing mail infrastructure. If successful the reliability of the infrastructure might be endangered due to additional load. Nodes which 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).

# 6 Potential Benefits

If successful this thesis will allow normal users without the help of a provider to set up an anonymous communication channel which is 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.

# 7 Bibliography

## References


[1] 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.

[2] Wikipedia entry on PRISM 2013. url: https://en.wikipedia.org/wiki/PRISM_%28surveillance_program%29

[3] Wikipedia entry on TEMPORA 2013. url: https://en.wikipedia.org/wiki/Tempora

[4] International Covenant on Civil and Political Rights. UNHR, 1966. url: http://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx

[5] Klensin. RFC5321 Simple Mail Transfer Protocol. IETF, 2008. url: http://tools.ietf.org/pdf/rfc5321.pdf

[6] Almon B. Strowger. Patent 447918: Automatic Telephone Exchange. USPTO, 1891. url: http://patft.uspto.gov/netacgi/nph-Parser?patentnumber=447918