

Research Plan MailVortex

Martin Gwerder

February 5, 2014

It has never been so easy to collect huge amounts of data about people and companies. Today's electronic communication allows not only to quickly reach any point of the world but it also allows to automatically monitor all kind of interaction which takes place on it. With the huge availability of such monitoring data dangers of misuse became imminent as [ECHELON], [wiki:prism], [wiki:tempora] and many more do show. Some attempts have been made (such as Mixmaster[mixmaster-spec] or Herbivore[herbivore:tr]) 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. And if they did they focused simply on the "encryption" part of the problem completely ignoring meta data and traffic flow (see [RFC3851] or [RFC2015]). Such an attempt would allow to use 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. Guidance and quality assurance is ensured 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 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 [RFC5321] 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.

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 patented it in 1891 [strowger]. 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. So 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 try to use available information to be more effective in their job. They invent methods to collect vast amounts of data about persons and then try to 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 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 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 [ECHELON], [wiki:prism], [wiki:tempora] 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”[iccpr].

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 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 analysis of message body and subject) and who else was involved (“cc” and “bcc” headers [whereas the later 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 today’s 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 due to the fact that we can analyze 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 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 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 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 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 which the true message content is hidden. This “noise” is an additional load which has to be han-

dled 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 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).

As with all encryption technologies anti virus and anti spam measurements are rendered ineffective as an anti spam or anti virus program 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 which have been found:

- In **k-anonymous:ccs2003**[**k-anonymous:ccs2003**] **k-anonymous:ccs2003** 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 **chaum-dc** **chaum-dc** writes an excellent article[**chaum-dc**] about a method to share information while keeping the source of the information anonymous.
- In **chaum-mix**[**chaum-mix**] **chaum-mix** 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:weis2006** focused in their work [**danezis:weis2006**] on the economics of anonymity.
- In [**xor-trees**] **xor-trees** 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 [**herbivore:tr**] 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 [**mixmaster-spec**] specifies the protocol for the Mixmaster infrastructure.
- In [**remailer-history**] **remailer-history** 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 [**RP03-1**] **RP03-1** conclude that privacy on current (static) mixer networks is endangered and dynamic networks are still in their infancy.
- **morphmix:wpes2002** describe in [**morphmix:wpes2002**] 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 **buses03**[**buses03**] the authors **buses03** propose to use a protocol which is clustering multiple messages in “buses” to make traffic analysis harder.
- In [**danezis:wpes2004**] **danezis:wpes2004** propose a message format which should be used as transport protocol for chaumian mixes. This format should be lightweight and efficient.
- **tor-design** describe in [**tor-design**] the design of the Tor-Protocol. Tor is a very common, generic onion routing protocol and [**foci12-winter**] of **foci12-winter** focuses on the Chinese firewall and its dedicated blocking methods for Tor.

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 which 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.

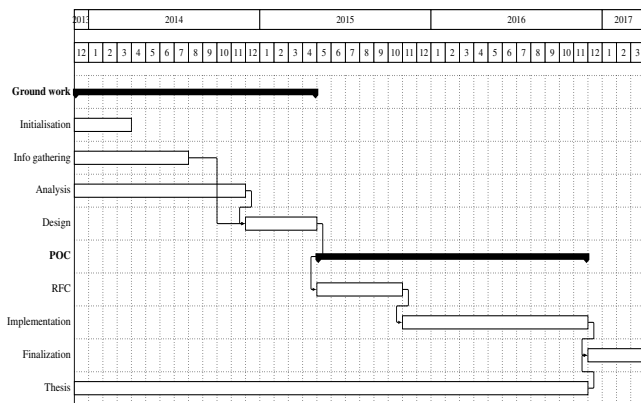


Figure 1: Planed timetable

9 Bibliography