Anonymity, Privacy and Hidden Services: Improving censorship-resistant publishing

Doctoral Dissertation by

Lasse Øverlier

Submitted to the Faculty of Mathematics and Natural Sciences at the University of Oslo in partial fulfillment of the requirements for the degree Philosophiae Doctor (PhD) in Computer Science

August 2007

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo.
No. 658

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AiT e-dit AS, Oslo, 2007.

Produced in co-operation with Unipub AS.
The thesis is produced by Unipub AS merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Unipub AS is owned by
The University Foundation for Student Life (SiO)

"I disapprove of what you say, but I will defend to the death your right to say it."

- Evelyn Beatrice Hall, writing "The Friends of Voltaire" as S.G. Tallentyre in 1906

Abstract

The request for on-line privacy is rapidly increasing. More and more Internet users realize that information about their on-line activities is highly valuable information for commercial companies and open for potential abuse. Information about who communicates with whom, and who accesses which services, is already used to improve on-line services, e.g. by serving more relevant on-line advertisements which many appreciate. But the problem of letting large commercial companies know your entire surfing history does not seem to be of major concern to the average Internet user. Future services may look into how to prevent this type of information leakage, but this will not help the users of today. In addition, anonymous publication of information, e.g. by dissidents and whistle-blowers, is made nearly impossible for today's Internet users. There exists a need for censorship-resistant Internet services, where anonymous publishing of information can be made. These types of services are already starting to appear. They are combined with anonymizing technologies, and designed to be attack-resistant, accessible from anywhere, have a hidden physical location, and therefore they will be more censorship-resistant.

The overall goal of the research work was to address vulnerabilities in, and to develop new or enhance existing anonymizing network technologies and censorship-resistant services. This thesis presents both analyses and new principles to enhance the anonymizing technology existing today.

The first phase of the research work consisted of an analysis of traffic flow confidentiality in a future military network setting, and an analysis of how to securely anonymize traffic data logs at high-speed interconnections. The thesis presents a new method for securing these logs by creating transaction specific pseudonyms without increasing the amount of logged data. The thesis also presents solutions to allow some elements of the traffic data to be used for statistical analysis and therefore be available for search, while

other parts of the data could be kept anonymous and unlinked to the searchable data.

The second phase of the research work focuses on technologies inside anonymizing networks, their vulnerabilities, and proposes methods to increase security to the existing techniques. The work demonstrates how the predecessor attack works in a live anonymizing network and can be used to locate a so-called hidden service within minutes with only a single compromised node in the network. An analysis of various countermeasures is also presented together with a recommendation on how to best resist this attack by using nodes protecting the initial connection to the anonymizing network.

The thesis presents a method of reducing a hidden service's vulnerability to denial-of-service attacks by using so-called valet nodes to protect the contact points of the hidden service. In addition the valet nodes solution enables the use of completely hidden services, where even the very existence of the service is hidden from the other users and from the network itself. The use of valet nodes also supports a method of obtaining flexible quality of service for both authenticated and anonymous users of a hidden service.

The research work also presents a general improvement of the authenticated Diffie-Hellman key exchange used in building anonymous connections. The solution eliminates the need for the RSA encryption by using predistributed Diffie-Hellman values when setting up session keys for the anonymous connections. This reduces the number of encryptions and the number of messages necessary for constructing an anonymous connection while maintaining forward secrecy. The solution is also easily adaptable to the valet nodes design which will benefit from the use of public Diffie-Hellman values and thereby also avoid the use of RSA. In addition the thesis presents a method to reduce the latency in a hidden service connection by utilizing the extra protection within the valet nodes extension.

Acknowledgments

This thesis is dedicated to my family. This research work could not have been completed without their support. First of all my wife Monica who has stood by me through all challenges and even agreed to spend one year with me and the children in the US while studying. Thanks to my parents, Tine and Svein, for teaching me never to give up and that anything can be accomplished. And lots and lots of thanks to my children, Anine, Kristine, Lars Magnus, Eirin, and Selma, for being the greatest kids anyone could have, and for being patient with me working odd hours. *They are the true meaning of life.*

The research has been carried out mainly at Gjøvik University College (HiG) and at the Norwegian Defence Research Establishment (FFI), in addition to a one year period at the U.S. Naval Research Laboratory (NRL) in Washington DC. The research period has been extended after Selma was born since I have chosen to work half time from August 2006 to the end of 2007. The funding for the research work has been provided by both HiG and FFI.

Many thanks to Gjøvik University College for providing me with a great and highly expansive working environment. This thesis had not been possible if it was not for the huge effort of Dr. Erik Hjelmås who has assisted me with all those little questions that consumes lots of time, in addition to his initial help to locate funding for my research period. Many thanks to my supervisors, Professor Einar Snekkenes at HiG, and Professor Chunming Rong at the University of Stavanger, for helping me and for contributing to a highly interesting research period. Lots of thanks also to my other colleagues at HiG, Professor Chik How Tan, Professor Slobodan Petrovic, Professor Stephen D. Wolthusen, and the others, who have always assisted me whenever I have had research problems. Many thanks to the other research fellows at HiG, Geir Olav Dyrkolbotn, Nils Kalstad Svendsen, Hanno Langweg, Kirsi Helkala, Davrondzhon Gafurov, and Janne Hagen, for many discussions both within and on the outside of the research areas during this

research period.

I would also like to express my gratitude to FFI for letting me pursue a PhD within such an interesting research area. Many thanks to Ronny Windvik who always has taken time to discuss the many research questions of anonymizing networks. Thanks also to Tore J. Berg, Torgeir Broen, Erlend A. Garberg, Lars Hornfelt, Kjetil Mosesen, Camilla Olsen, Tormod Sivertsen, Aasmund Thuv, Ane Daae Weng, and other colleagues at FFI, for their continuous effort in providing a great and fun working environment. I would also like to thank Vidar S. Andersen for help with funding for the research period in general, and for the extra funding enabling the one year stay at NRL in Washington DC.

Many thanks also to my co-authors, Dr. Tønnes Brekne, Dr. André Årnes and Geir Hallingstad for fun and interesting periods of research work.

Last, but definitely not least, I would like to thank Dr. Paul Syverson at the Naval Research Laboratory for his huge effort in both guiding me inside the field of anonymity research, and especially for completing all the paperwork needed to allow me to have a one year research period at NRL. A truely great year with open feedback and discussion on all research questions I, and others, came up with. The discussions at the lunch table was greatly appreciated and solved (and tore down) many research challenges. The year at NRL turned out to be very fruitful in terms of research topics and areas, and has left several research questions still to be completed. Lots of thanks also to the other researchers at NRL, Dr. Catherine Meadows for letting me work in her research group, Dr. Ira Moskowitz for his great sense of humor and countless practical jokes, Dr. Keye Martin for many great off-topic discussions, Dr. Gerard Allwein for his views on society, and Dr. LiWu Chang for interesting research discussions. A great thanks to the Naval Research Laboratory for providing me with a place to work for that year and for supporting a scientist exchange program that many more should take advantage of.

Introduction to the papers

The five research papers that constitutes Part II of this thesis are:

Paper A Lasse Øverlier, Tønnes Brekne and André Årnes. **Non-expanding Transaction Specific Pseudonymization for IP Traffic Monitoring.** In Yvo G. Desmedt, Huaxiong Wang, Yi Mu, and Yongqing Li, editors, *Cryptology and Network Security: 4th International Conference (CANS 2005)*, pages 261–273. Springer-Verlag, LNCS 3810, December 2005.

Paper B Geir Hallingstad and Lasse Øverlier. **Traffic Flow Confidentiality in a Future Network Enabled Capability Environment.** In *Proceedings of the 2007 IEEE Information Assurance and Security Workshop.*, pages 325–332. IEEE, June 2007.

Paper C Lasse Øverlier and Paul Syverson. **Locating Hidden Servers.** In *Proceedings of the 2006 IEEE Symposium on Security and Privacy (S&P'06)*, pages 100–114, May 2006. IEEE Computer Society.

Paper D Lasse Øverlier and Paul Syverson. **Valet Services: Improving Hidden Servers with a Personal Touch.** In George Danezis and Philippe Golle, editors, *Proceedings of the Sixth Workshop on Privacy Enhancing Technologies (PET 2006)*, pages 223–244, Cambridge, UK, June 2006. Springer-Verlag, LNCS 4258.

Paper E Lasse Øverlier and Paul Syverson. **Improving efficiency and simplicity of Tor circuit establishment and hidden services.** In *Proceedings of the Seventh Privacy Enhancing Technologies Symposium (PETS 2007)*, pages 134–152, Ottawa, Canada, June 2007. Springer-Verlag, LNCS 4776.

Related papers:

- Lasse Øverlier and Paul Syverson. **Location Hidden Services and Valet Nodes.** *Teletronikk 2.2007.* `http://telektronikk.no/`, Telenor, 2007
- Lasse Øverlier. **Tunnel direction hiding.** *FFI Notat.* Norwegian Defence Research Establishment, 2007.

Contents

Part I

Introduction

"Anonymity is a shield from the tyranny of the majority."

- U.S. Supreme Court decision No. 93-986 April 19, 1995

1 Introduction

This chapter gives a brief introduction into the background and motivation for the thesis and describes the thesis' outline.

1.1 Background and motivation

As the use of the Internet is continuing to increase rapidly, people leave more and more traces of their on-line activities without being aware of the potential for abuse of this information, or by simply ignoring them. There are many commercial interests in (ab)using this information, e.g. why did you join the on-line "chat-room for depressed", and why are you looking for information about short time credit card loans?

Besides this obviously private and personal information, there exists areas where people are in need of publishing information without being identified. This could be political dissidents or corporate whistle blowers in need of making information publicly (or "corporately") available without having their identities revealed. This so-called censorship-resistant publishing is an important part of anonymity services, and has until now received limited attention within the various research communities. Censorship resistance may be achieved by using *hidden services*. These services are constructed to make general Internet services, like publishing services, available from anywhere at any time, without exposing the IP address and thereby its physical location. Hidden services thereby makes direct denial-of-service attacks and even physical attacks impossible.

As privacy (cf. Section 2.1) is an important and large part of our offline society, there has always been a challenge to define how privacy will be a part of the on-line community in the use of data communication networks. In cooperation with Gjøvik University College there were identified common areas of interest within anonymity research, like privacy, traffic flow confidentiality and censorship-resistant publishing to

be the basis in the main research areas of the thesis. There had already been identified some weaknesses in the location hidden services and this was early identified as an interesting area for deeper privacy research.

The challenges identified during the initial research period have made the research focus rapidly evolve from traffic flow confidentiality, anonymizing network security and anonymity in high speed data logs, towards attacks and improvements on so-called hidden services, as will be described in Chapter 3.

1.2 Structure of the thesis

The thesis is separated into two parts. *Part I* is an introduction to the field of anonymity, privacy and censorship resistant publishing. *Part II* contains the published articles describing the research work of the thesis.

Part I After a brief introduction describing the background, motivation and structure of the thesis, Chapter 2 gives an overview of the research area and related fields, in addition to related work both existing before and completed during the research period. Chapter 3 describes the research work by summarizing the contributions from the individual papers, making a brief discussion of the research work, and describe suggested areas of further research.

Part II This part consists of the following five research papers:

- Non-expanding Transaction Specific Pseudonymization for IP Traffic Monitoring.
- Traffic Flow Confidentiality in a Future Network Enabled Capability Environment.
- Locating Hidden Servers.
- Valet Services: Improving Hidden Servers with a Personal Touch.
- Improving efficiency and simplicity of Tor circuit establishment and hidden services.

2 Anonymity background

This chapter will present background information on anonymity, hidden services and censorship resistance to set the thesis' publications in a relevant context. First some definitions are presented in Section 2.1, while Section 2.2 sets the use of anonymity in a wider perspective. Section 2.3 presents high-latency anonymity, and Section 2.4 presents a classification of low-latency anonymity systems and some relevant low-latency anonymity schemes. Section 2.5 describes different methods of achieving censorship-resistant publishing of information, and Section 2.6 briefly describes the the Tor anonymizing network and the hidden services principle.

2.1 Definitions

Anonymity originates from the Greek "anonymia" meaning "without a name". Anonymity is used in many different settings like common social situations, e.g. story telling and using cash, to uttering less popular political views, e.g. by the use of public flyers. This thesis focuses on anonymity in a computer network communication setting which will be described in this chapter.

The informal use of anonymity simply means that one cannot tell who did what. More formally the common and most widespread definition of anonymity is made by Pfitzmann and Hansen [80].

> Anonymity is the state of being not identifiable within a set of all possible acting subjects, called the anonymity set.

Anonymity is often evaluated as an absolute value; either you are anonymous, or you are not. But based on the above definition, anonymity is a probability based on the

size of the anonymity set one is a part of (cf. Section 2.2.1). Attempting to identify who is communicating with whom will also divide the definition into *sender anonymity* - for the originator of the message, *receiver anonymity* - for the receiver of the message, and *relationship anonymity* - against the linking of senders and receivers. Most anonymity systems focus on sender anonymity, e.g. a user is sending an anonymous message or requesting information anonymously from a public website. Anonymity is not *cryptography*, as cryptography only hides the content of the communication channel and not those involved. But cryptography often plays a strong part in the construction of anonymity systems. Neither is anonymity the same as *steganography* [56]. The main objective of steganography is to hide the very existence of the communication, creating *unobservability*.

Strong anonymity will prevent the linking of two transactions (or separable actions) to the same identity.

Pseudonymity can be viewed as the use of roles instead of personal identities. The role acts as a representation (pseudonym) of the person using this role over time, but without connecting this role to the real persons identity, comparable to a nickname.

The term anonymity suffers under the attention it gets from the abuse of the freedom it provides. Therefore other more acceptable terms have arisen in different areas which basically means the same. In the technical definition of computer network anonymity, businesses use the term *network security* when trying to protect their resources, military networks often discuss *traffic flow confidentiality* and *traffic analysis resistance* to counter information leakage, while private citizens relates mostly to the term *privacy*.

Traffic flow confidentiality and *traffic analysis resistance* defines how a network is able to hide the communication patterns of the network, e.g. who communicates with whom.

Privacy is all about an individual being in control of what personal information that is to be distributed to whom. In an Internet setting this information can be "everything" related to this person, e.g. personal identity, home address, email address, private emails, web pages visited last three months, bank account information, etc. Privacy is not to hide this information, but to protect and verify who has access to what information.

Censorship resistant publishing describes methods of disseminating information securely and anonymously, without letting non-authorized users remove or change infor-

mation, and without allowing anyone to make the information unavailable.

2.2 Anonymity and free speech

Article 19 of the UN Universal Declaration of Human Rights expresses:

> Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.

Freedom of the press is a well established principle in democratic countries and viewed as one of the fundamental rights for their citizens. Most people expect that this fundamental right also applies in other areas where expressions can be published, like on the Internet. The technology and evolution of Internet usage is faster than any other previous technological development, and therefore laws are often constructed on the basis of isolated incidents[1, 91] rather than put into a more complete perspective before they are implemented. If the government required every person to wear a GPS-tracker[1] so that anyone could be asked to prove his/her whereabouts upon request, e.g. in abduction cases, terrorist activities, etc., the majority would hopefully *never* accept this even if it could assist in solving many crimes. But by pointing out existing and potential abuse scenarios, the same principle is about to be introduced on Internet usage and on-line activity [36] in several countries.

Existing and potential abuse is an important aspect of Internet anonymity. Unfortunately there are abusers of the different types of anonymity services like there are abusers in all other areas of society, and this will remain a fact also in the future. Cash may be the simplest analogy as cash on the one side allows the public to be anonymous in most ordinary transactions and on the other side can be abused e.g. to avoid taxes. Most illegal activities are still illegal even if they are fulfilled by (ab)using the Internet, and this will unfortunately not disappear by prohibiting new technology. This situation is exactly why we have laws to inform the public of what is deemed legal and illegal behavior, and this is currently also how we define legitimate Internet usage. One problem with this is that a global computer network is only slightly influenced by local (country

[1] A Global Positioning System tracker will store your position at all times. These systems already exists and products implementing these in teenage-model cell phones [92], sneakers [110], hidden car devices and more [26], are all available today.

based) laws and regulations as the services are easily moved from one jurisdiction to another. Some content is illegal in a few countries, like selling Nazi-related material over Internet in France, but legal (even if disputed) in other countries. Similarly it is illegal to drive above the speed limit, but we do not enforce all cars to report continuous speed and position reports of every trip [2] for complete monitoring even if this is now technically possible. The authorities still accepts that it has to provide some freedom to its "users", the citizens, but the government should be prepared to take action if someone abuses that trust.

Currently there is a push for making every user's on-line activity traceable, often with references to serious abuse cases within Internet communication. The *EU data retention directive* [36] will enforce every service provider to store information about Internet (and phone) activity for a period of *"at least six months and not more than two years"*. Typical logged data will be information about the IP address used (location), who the user communicated with, and what they did, e.g. email sent and received and which web sites the user visited. The directive is to be implemented in the EU region by 15 September 2007, but may be delayed by 18 months for the areas of Internet access and usage by individual countries.

But Internet users are also leaving vast amounts of information to commercial companies about their on-line activity. This may be by accident, by lack of knowledge for how this information can be abused, or simply by not knowing how to avoid leaving sensitive data. This is the very core of the *privacy* definition given in Section 2.1; being in control of who learns what about yourself. The value of this information is enormous[3] and most Internet users do not seem to mind giving this information away. However, there may exist times when a user is aware of the privacy risks and would like to be certain of having privacy. This can be in situations where we have taken privacy for granted but, without knowledge, might not have privacy or anonymity after all. Today there exists challenges in not giving away on-line identity, originator, or organization, in many different scenarios, e.g. when:

- sending or receiving a private email (or instant message),
- searching for personal health information,

[2] Not very surprising this is a method currently under development to enforce automatic toll payment on roads. It will not take long before other areas of use are suggested.

[3] Information contained in people's searches and their on-line usage, interests and habits, is the very foundation for companies like Google, Yahoo, Lycos, and many others.

- investigators are accessing open/public information about suspects,
- informants want to give the police on-line anonymous tips,
- journalists try to protect their sources,
- political dissidents attempt to publish information.

These are only a few of the anonymity scenarios the public are familiar with, but where anonymity on-line may be lost in the near future.

At the time of writing there are many products, services and technologies, that can give anonymity of some degree, but mostly only in single usage areas, e.g. The Anonymizer [4] for Internet browsing, Mixmaster [71] for anonymous email, etc. Anonymous publication and dissemination of information have been cumbersome and insecure. Until now anonymous publishing have often been completed using an anonymizing web service or anonymous email tools for accessing public or commercial publishing services and distributing information from there. But as these services are available to all, they are also open for pressure to be shut down, e.g. by denial-of-service attacks, or legal attacks on the publishing service provider. Existing solutions for anonymous publishing will be described in Section 2.5.

Privacy enhancing technologies (PET) have been under development since the early 1990s, and are still undergoing rapid evolution to provide privacy protection for Internet users. Many of these different technologies will be described in this chapter, but a short summary of the early PET systems can be found in Goldberg et al. [44, 42]. Other related and often connected areas like digital cash [95, 54] and e-voting [53] will not be addressed as they are separate areas of research.

2.2.1 Degrees of anonymity

Another challenge in anonymity research is the "level of anonymity", e.g. how can we measure and quantify the anonymity given in an anonymity service or even in a specific situation.

Reiter and Rubin [87] presented a *degree of anonymity* ranging from *absolute privacy*, via degrees of *innocence* and *suspicion*, to *provably exposed*. Goldberg [41] defined *The Nymity Slider* presenting a scale of anonymity ranging from *verinymity*, e.g. proof of identification, to *unlinkable anonymity* where the identity cannot be recovered. Berthold et al. [11] defined the mixer network secure if at least one of the mixer nodes in the

cascade could be trusted. The probability of this is $P = 1 - a^l$, where a is the part of attackers in the network and l is the length of the route. Díaz et al. [28] described the *degree of anonymity after an attack* as the systems current entropy divided by the maximum entropy of the system.

For the Tor network (cf. Section 2.6), the probability of a user connection being compromised is often simplified to the probability of an attacker controlling both the entrance and exit node of the network. If an attacker controls c of the n server nodes in the network, the probability of being secure is $1 - (\frac{c}{n})^2$ if all nodes are selected with equal probability.

The rest of this chapter will give an introduction to the different types of anonymizing technologies - both for personal privacy and for censorship-resistant publishing, and look at some of their weaknesses and strengths.

2.3 High-latency anonymity

Looking at the history of privacy enhancing technologies, there is wide agreement that this expansive area of research was initiated by David Chaum's paper on email mixes [17] in 1981. Mixes are network nodes that accepts a (preferably) large number of messages as inputs and send them out again them with varying new attributes, like new appearance, new/removed encryption layers, and optional random delays giving a new message order in the output. This is the typical functionality for the early remailer services, also called type 0 remailers. These type 0 remailers were services like Helsignius' `anon.penet.fi` that stripped off identifying headers in emails, changed the "From" address to an alias at `anon.penet.fi`, and forwarded the mail to the recipient. The mapping between the originator's email address and the alias was kept in a mapping table at the `anon.penet.fi` service provider, and is one of this service's weak points. Another disadvantage is that the service's construction as a single point of failure makes it quite easy for an attacker monitoring the remailer service to statistically match the input messages to the output messages. The service was later shut down due to legal pressure to retrieve originator identities from the mapping table [48].

Later Cypherpunk remailer services, called type I remailers, are more complex and involves a network of *mixer nodes*. These message based mixer technologies are usually

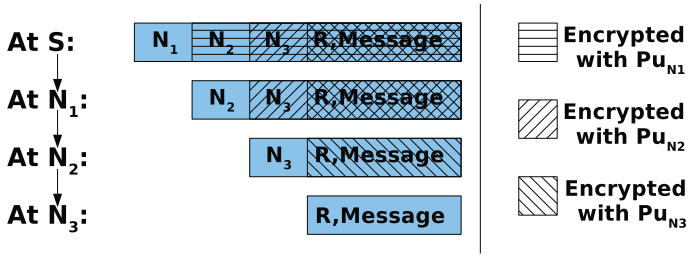

Figure 2.1: Message appearance at the anonymizing network nodes in a sequence of mixes using public key cryptography.

based on public key cryptography, where each consecutive mixer node has a public and a private key (Pu_N and Pr_N). The message, M, is to be sent from a sender, S, to a receiver with address, R, through the mixer nodes N_1, N_2, and N_3. First the message, M, is preceded with the address of the receiver, R. This new message with destination address is now encrypted with the public key of N_3, Pu_{N_3}, and preceded with the address of the node N_3 in the mixer chain. Then this is repeated - the new message is encrypted with the public key of N_2, preceded with the address of N_2, and encrypted with the public key of N_1. The final constructed message is shown on the top of Figure 2.1, where the transformation of the message at the different mixer nodes in the network during sending is illustrated at individual lines. When node N_1 receives the encrypted message, it decrypts the message with its private key, Pr_{N_1}, recovers the address of the next mixer, N_2, and sends the remaining part to this node. N_2 and N_3 does the same, and N_3 is at the end left with the address of the receiver and the message to send there without knowledge of the originator. There have been identified several vulnerabilities to these first types of mixer networks [85, 25, 58].

Type II remailer services like Mixmaster [71] and Babel [46] strengthen the relationship anonymity, improve reply possibilities and address potential attacks like replay and message length matching. But still they have weaknesses like the *n-1 attack* [11] and *trickle attack* [97]. Type III remailers like Mixminion [25] attempts to address these problems by adding long term pseudonyms, replay protection, and *forward anonymity*[4].

An improvement to the mixer networks was proposed by Kesdogan et al. in Stop-And-Go-MIXes [60]. Here the sender precalculates a delay with exponential distribution for each packet at every mixer node and also sets a time window on each packet's arrival

[4] *Forward anonymity* describes the situation where compromise of a long term encryption key does not expose the anonymity in earlier communication. Analogous to (perfect) forward secrecy (PFS) [31].

at the individual nodes. If the packet arrives within this time window it is delayed by the precalculated value before sent to the next mix and is therefore more resistant to active attacks such as deliberate delaying of packets. Other mixer technologies related more to message anonymity than low-latency traffic are Ohkubo and Abe's Hybrid Mix [75, 52], Markus Jacobsson's Flash Mix [51, 70] and George Danezis' FS-mix [22]. But even high-latency mixer networks are vulnerable to some types of traffic analysis attacks [68, 24].

2.4 Low-latency anonymity

The delays involved in the above mentioned technologies are not suitable for low-latency interactive traffic, like web-browsing, where significant amounts of the privacy related information is revealed.

Low-latency anonymity were first proposed by Pfitzmann et al. for ISDN communication [81], but the users had to use fixed and equal bandwidth to a local telephone switch. The proposed system and the scalability was unsuitable to scale towards an Internet sized anonymity network. Another telephony based mixer system was proposed by Jerichow et al. [55], but this thesis will only address technologies for anonymous Internet communication from now on.

Anonymity networks are mainly using three anonymizing technology principles [40] *DC-networks*, *broadcast systems*, and *source rewriting systems*, as shown in Figure 2.2. These anonymizing technologies and some related protocols will be addressed individually in this section.

In addition to anonymity networks, there is the simplest anonymizing technology for low-latency communication, *the anonymizing proxy*. Anonymizing proxies have been commercially available since 1995 [4] and such services are still operative [88]. Anonymizing proxy services can be compared to one-node mixers similar to `anon.penet.fi` keeping the location, i.e. IP-address of the originator, away from the accessed Internet service and replacing it with the address of the anonymizing proxy. *SafeWeb* was a similar commercial-but-free[5] service enabling its users to anonymously access the web using plain HTTPS encryption [29] to reach the SafeWeb anonymizing proxy. SafeWeb later

[5]Using banner ads instead of charging the users directly.

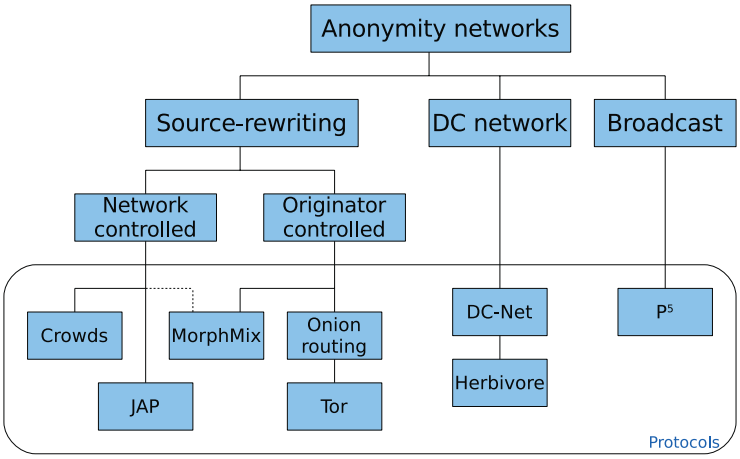

Figure 2.2: Simple classification of anonymizing technologies

added the software solution *TriangleBoy* [50] which enabled SafeWeb users to allow their computers to be used as a one hop forwarding proxy for other SafeWeb users. TriangleBoy would then allow people within restrictive firewalls to more easily find available and non-blocked service points. Both SafeWeb and TriangleBoy disappeared as services after SafeWeb Inc. was acquired by Symantec in 2004.

Feamster et al. presented *Infranet* [37], a service that uses steganography inside plain web content to transmit data retrieved by its servers back to the clients. The Infranet servers are built to be indistinguishable from normal web servers so that traffic to and from these unidentifiable servers appears like plain HTTP [38] traffic. One of the goals for Infranet is to be an option bundled with standard web servers and thereby enabling anyone to easily assist in preventing censorship and surveillance.

Many vulnerabilities have been located and demonstrated [49, 67] against proxy services in general, but their main weaknesses are being a *single point of failure*, a *single point of compromise*, and a *single point of attack*.

2.4.1 DC network

The *dining cryptographers protocol, DC-net,* was introduced by David Chaum [18] in 1988. A user of this network can achieve absolute anonymity within a group of users cooperating at sending anonymous messages. The DC-net principle is that all n users in an anonymity set (i.e. potential actors) share a bit-long secret with at least two other

users. Each host then transmits the xor of all shared bits. The sum of all transmitted bits will be divisible by two (xor all bits equals *zero*). If one user wants to send information, it transmits the inverse of the actual value. This will not be noticeable individually by the others, but the total xor will now be *one* and a proven anonymous transfer from the group can take place. Later discussions on security and proposed improvements of the DC-net protocol can be found in [106, 107].

Another protocol *Herbivore* [40], under development at Cornell University, will use the DC-net principle by having multiple groups of users organized in *cliques*, and transmitting information through one of the members in the clique. Each clique uses an extended DC-net technology with reservation and transmission phases, and the clique is self-controlled with regards to size. Herbivore will provide strong anonymity within one clique.

The bandwidth overhead required in the DC-net protocol has made practical use of DC-nets challenging and almost non-existent.

2.4.2 Broadcast protocols

Broadcast protocols have received less attention than the other technologies as they have too much traffic overhead. A broadcast protocol will typically need all possible senders to send a message to all potential receivers, which gives a huge extra network load in a switched topology as this often is implemented using constant rate transmission to all participants.

One of the few published anonymizing broadcast protocol is *Peer-to-Peer Personal Privacy Protocol*, P^5 [100]. P^5 tries to enhance performance by dividing the members into a hierarchy of broadcast groups, but still requires massive overhead traffic in addition to having the maximum available bandwidth limited by the constant transmission rate.

2.4.3 Source-rewriting networks

The third and most mature anonymizing technology is source-rewriting networks. These networks use many of the principles described in Section 2.3 on high-latency anonymity, but have very low, if any, added delays during the traffic mixing at the network nodes.

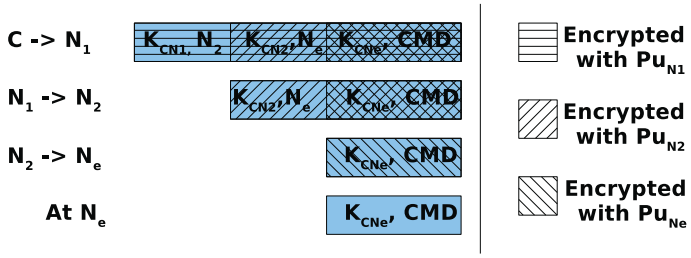

Figure 2.3: Onion routing setup of session keys using public key cryptography.

The first distributed low-latency system for anonymous Internet communication was *onion routing* [45, 86]. An onion routing network consists of several hops which proxy the communication and each hop changes the appearance of the communication by adding or removing an encryption layer. An anonymized communication channel through the onion routing network, called *a circuit*, is initiated by the client using public key cryptography to distribute session keys along the circuit. This initiating *onion*, shown on top of Figure 2.3, is used to create a circuit from the client, C, through the onion router nodes, N_1 and N_2, to N_e, the proxy node (called *the exit node* in later onion-routing based protocols). Each node in the circuit "peels off" one layer from the onion and when the onion has reached its destination, the client shares a session key K_{CN_1},K_{CN_2},K_{CN_e}, with each node in the circuit. These session keys are now used on the data passing through the circuit. One of the major problems with the first onion-router protocol was the vulnerability for replay of the onions. A more complete security analysis for onion routing is presented by Syverson et al. [104].

Freedom Network [12] from Zero-Knowledge Systems Inc., the first commercial anonymizing network, allowed the use of pseudonyms which also enabled them to counter some of the potential abuse. For more information on the security of the Freedom Network see [6, 5]. Rennhard and Plattner introduced *MorphMix* [89, 90], which is a peer-to-peer based system using a *witness node* during the setup of the anonymizing tunnel to counter collusion attacks, but had limited success [105]. *Tarzan* [39] and *Cebolla* [15] uses the UDP protocol to construct an anonymity layer that is able to tunnel IP traffic similar to a router. Tarzan also adds a scheme for cover traffic to improve traffic flow confidentiality. A discussion on the effects of cover traffic in mixer network can be found in [10, 27, 65]. Goldberg and Wagner's *Rewebber* [43] uses a network of en-

crypting proxies for retrieving from and publishing information to the web. Using public key cryptography and accessing the server software through HTTP requests, enables the Rewebber network to be used to interact anonymously with normal web services.

All these anonymizing networks are *originator controlled* (Figure 2.2), meaning that the originator (client) selects which nodes in the mixer network that is to be used. In *network controlled* source rewriting systems the client only passes information to the network and lets the network do the anonymization. Examples of these are *Crowds* [87], *Hordes* [66, 101] and *JAP* [9].

Crowds, introduced by Reiter and Rubin in 1998, is a mixer network where every node in the anonymizing network, *the crowd*, can ask another node in the network to retrieve information on its behalf. The node throws a biased coin and evaluates to fetch the information itself, or send the request on to another randomly selected node in the crowd. When the coin results in retrieving information from the outside, the node completes the request, e.g. downloading a web page, and sends the answer back to the originator in the same (reversed) path. Hordes is an extension of Crowds that improves the sign-on, the distribution of the *hordes* list, and reduces response times by using multicast to anonymize the replies. Crowds and Hordes suffer from a number of vulnerabilities [113, 104, 111, 112].

A network controlled mixer network used by many and still under development is the *Java Anon Proxy* (JAP) [9]. JAP uses a local client side proxy to connect to the first mix in a cascade[6] of mixes, where the last mixer is connected to a web cache proxy. JAP has a large user base, but due to its functionality it is not amenable to the hidden service design (cf. Section 2.6.1) and was therefore not used in the research work.

Several timing and traffic analysis vulnerabilities in these networks have been described; Raymond [85], Back et al. [6], Zhu et al. [115], Kesdogan et al. [58, 59], Serjantov and Sewell [98], and Danezis [23].

Katti et al. [57] recently introduced *information slicing*. This protocol splits a message into multiple parts, *slices*, and sends them to the receiver through different paths of the anonymizing network. Only the receiver of all the slices will have enough information to be able to decrypt the message. One promising thing about this protocol is that it does not require public key cryptography and therefore no distributed key

[6]Chaum defined[17] a *mixer cascade* to be a series of mixes where any of the mixes should be able to provide secrecy of the correspondence between the input and output messages.

management scheme.

Tor [33], the largest low-latency protocol and a protocol supporting hidden services will be presented in Section 2.6.

2.5 Censorship-resistant publishing services

Anonymizing networks may give the anonymity needed to protect privacy and confidentiality from some clients' perspective, but there are situations where the services publishing the information require, need, or wish to remain anonymous. These are so-called censorship-resistant publishing services. Many people suffer from governmental censorship or are afraid of simply loosing their jobs, and are therefore made unable to express or publish their concerns and opinions. But not only dissidents trying to publish information about situations not widely known, need these types of services. Other scenarios likely to exist may be:

- Employees making their board of directors aware that the company is breaking the law (e.g. following the Sarbanes-Oxley Act [96]).
- A blogger on the inside of a firewall, e.g. on a shared IP address and therefore unable to set up a normal web service accessible from the Internet. A hidden service will be available through a most firewalls.
- Publishing a blogg that cannot be traced or shut down by the authorities in your country.

The first service designed to resist denial-of-service attacks was Ross Anderson's Eternity [3] service, distributing the service's storage on many Eternity servers. The service provides long term storage of data and uses payment as incentive for making a large number of cooperating servers store a copy of the data, and thereby make the data extremely hard to delete unless the attacker knows all servers. A proposal for strengthening the Eternity service [7] has also been published.

Other systems that store the entire published document at multiple locations are Freenet [19] and Publius [109]. Freenet uses a peer-to-peer network to resist censorship and sustain availability even in the case where only one of the nodes is available. Freenet's peer-to-peer network is in itself a large storage area where the storage space

is distributed among all the nodes of the network. Freenet is still under active development [82]. Publius by Waldman et al. was designed for publishing content on the web and to guarantee the persistence of stored files. Publius encrypts the stored file and splits the encryption key using Shamir secret sharing [99] and spreads these key shares on different locations. A client must therefore have access to multiple servers for retrieving the entire key and be able to read the content of the file.

Another method for censorship-resistance involves splitting the stored file into many blocks and spread these blocks onto a subset of the system's storage servers. *FreeHaven* [32] uses a reputation system among its nodes involving contracts between the servers for storing data for others. FreeHaven uses Rabin's information dispersal algorithm [83] to split the document into *shares* before distributing them onto the servers. FreeHaven suffers from not defining the underlying anonymous communication channel where many of the anonymity issues exists. Waldman and Mazière's *Tangler* [108] makes newly published documents dependent on previously published documents, and this dependency is what the authors define as *entanglement*. Thereby Tangler creates incentives for the storage and replication of older documents in addition to preventing the servers from being in control of what the other servers may publish.

GNUnet [8] is a framework for peer-to-peer networking designed for anonymous censorship-resistant file sharing. GNUnet is fully decentralized and does not have a central trusted public service, but it has also been found vulnerable to location attacks and to censorship [63]. Several other peer-to-peer storage systems [20, 35, 93] have been developed and many are still in use. More information on these peer-to-peer networks and darknets[7] and current peer-to-peer implementations can be found on-line searching for protocols like BitTorrent, WASTE, KaZaA, FastTrack, and LimeWire.

2.6 Tor and Hidden Services

Tor [33] is the largest public anonymizing network currently in use. Tor builds upon onion routing technology and uses a network of routing nodes (Tor servers) to transport traffic for the users (Tor clients). Tor was deployed in 2003, updated to support hidden services (cf. Section 2.6.1) in 2004, consists currently of approximately 1000 active Tor

[7] A darknet is a private virtual network where its users communicate only with other users they somehow trust.

server nodes and an estimated 200.000+ weekly users, and has until now never been down.

Wei Dai presented in *PipeNet* [21] an anonymizing technique where the client established a connection through an anonymizing network by extending one hop at a time, and exchange an ephemeral encryption key with each node in the connection path. Tor uses the same principle to construct a circuit through the set of Tor servers.

All communication between the Tor nodes (client to server, and server to server) uses Transport Layer Security (TLS) [29] to create forward secrecy (FS) [31] on every communication link. Forward secrecy is important to prevent any attack to compromise and access earlier communication information. If the TLS link is not present when the nodes starts a communication channel, the TLS session is created first. The TLS links are left out of the rest of the description of Tor and its hidden services as they are always present on every communication link between two Tor nodes.

One of the major problems of all anonymizing networks is bootstrapping, i.e. how to locate and start using the anonymous network. This is often the simplest way of blocking an anonymous service [62]. Tor uses a directory service where the directory servers have their identities and public keys hard coded (but configurable) in the client code. The use of the Tor directory service is at the time of writing undergoing significant changes to address vulnerabilities in the original design [33]:

- By stopping access to the directory servers the clients will be unable to download the list of server nodes and thereby not able to connect to the network.
- By forcing the client to download all the server nodes, the network will meet problems scaling the network size.
- One of the directory servers could construct false information and make a large portion of the network believe it.

The directory service distributes a signed list of server nodes, with the nodes' (self-announced) network bandwidth and contact information, i.e. IP address and port numbers.

For the Tor client to use the anonymizing network, it first selects which server nodes to use in the circuit (currently the default number of hops is three) and selects by random three server nodes in the network. This random selection uses the nodes announced bandwidth to weigh the random selection, making a node with 10Mbit bandwidth ten

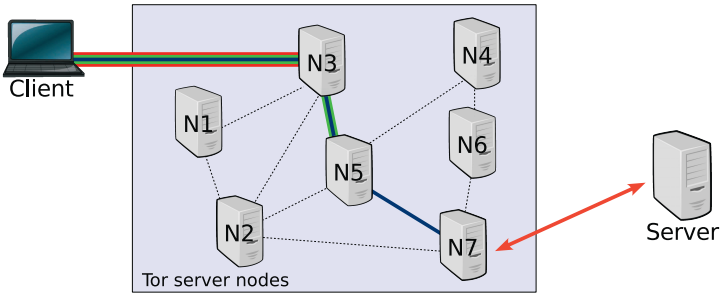

Figure 2.4: Setup of circuit through the Tor network.

times more likely to be chosen than a node with 1Mbit bandwidth. This enables the network to distribute load more equally among the participating server nodes[8]. The abuse of the Tor server nodes' self-announced values is discussed and implemented in *Paper C* [77] and is a known vulnerability in Tor.

The final tunnel, the anonymizing *circuit*, is shown in Figure 2.4 for a client accessing a public server through the nodes $N3$, $N5$, and $N7$. The circuit is established by the client connecting to the first node, $N3$, and using ephemeral Diffie-Hellman[30] to exchange a session key used for encrypting the communication between the two nodes. A classical man-in-the-middle attack on Diffie-Hellman is avoided by encrypting the client's Diffie-Hellman value with the server node's public RSA key. After the secure connection to $N3$ is established, the client sends a message to $N3$ and asks it to extend the circuit to node $N5$, where the client again uses an authenticated Diffie-Hellman to exchange a session key directly with $N5$ without letting $N5$ know at which node the key exchange originates. From the perspective of $N5$, the originator could be $N3$, but it could also be any node, client or server, inside the Tor network. The same extension is completed from $N5$ towards $N7$ and the client may now anonymously communicate with $N7$. The client is then ready to setup anonymous communication sessions to public services on the outside of the anonymizing network using $N7$ as the *exit node*, accessing these services on the client's behalf. Note that for every step of the circuit towards the exit node, one layer of encryption on the traffic data is removed, meaning that the client adds three layers of encryption for each packet traveling out, and each node in the circuit removes one layer before forwarding the packet. This way the packet will appear

[8]As there is a limited and small number of exit nodes available, the selection of nodes also attempts to take this into consideration. This work is currently in active development.

different at every node in the path and avoid being traced. TLS will also provide change of appearance to external adversaries, but if nodes within the anonymizing network are colluding, then a packet without this internal change of appearance will be traceable at two different non-adjacent server nodes, e.g. at $N3$ and $N7$ in this example. When sending reply data from the public service back to the client, this process is reversed and each node instead adds a layer of encryption so the client then has to remove all three layers upon arrival of the packet.

Applications running on the client can now tunnel TCP sessions[9] through this anonymous tunnel by using a SOCKS [61, 64] interface and thereby enabling all TCP client connections to be tunneled over to the exit node and be established as if originating at the exit node. A client can multiplex several connections over the same anonymizing tunnel, but the user must be aware that unencrypted protocols will be visible to the exit server. So if one connection is used for anonymous surfing, it will compromise anonymity to e.g. post a blog or authenticate in another way through the same tunnel.

Since the Tor server network is open for anyone[10] to join, it is vulnerable to the Sybil attack [34], where an attacker inserts (or controls) many nodes of the network without the other users' knowledge. And as long as the communication channels between the servers are over public channels, several other vulnerabilities will also exist [74, 85, 58, 65, 102]

2.6.1 Hidden Services

In 2004 the Tor developers released an upgrade to the anonymizing network that included a method to add so-called *hidden services* inside the network. These services were designed [33] to resist denial-of-service attacks and be unable to locate, i.e. not find the service's IP address and thereby its physical location. So by setting up a hidden service, no one, not even the service's own users, should be able to locate it or prevent the service from being available. The Tor hidden services is is a general service hiding technique that can be used by many anonymizing networks, and is not specific only to Tor.

A hidden service is not a publishing service itself, but simply a method of accessing

[9] Tor supports only TCP sessions as it runs over TLS. Supporting UDP (or IP) over a TCP based channel raises a lot of challenges, and a new design from the ground up is likely to be constructed first.

[10] Anyone with a server accessible at a public IP address.

a hidden Internet service through an anonymizing network. In order to address when to use these hidden services, the Tor network uses a URL with the virtual "top-level domain" *.onion*. Every time a Tor client is requested to access a server name ending in *.onion*, the client knows that this is a connection to a hidden service and downloads the hidden service's contact information from the directory service (anonymously). The principle of hidden services is that the anonymity client and the hidden service agree upon connecting to a *rendezvous point* using plain anonymizing connections. When the rendezvous point connects these two circuits, the client and the hidden service are able to communicate privately without knowing where the other part is located, and without the rendezvous server knowing who is communicating, nor what kind of data is exchanged. More details on Tor hidden services can be found in Part II of the thesis.

Attacks on hidden services have often been related to the different attacks on the Tor anonymity network itself [113, 73]. Others, like Murdoch's clock-skew attack [72] directed specifically towards hidden services have addressed how to reveal the *location* of the hidden service. But finding the location is not the only attack vector against hidden services. Other threats against the current hidden services design have been identified already in the original design paper [33], but have received less academic attention. Denial-of-service attacks without locating the hidden service's IP address is still possible, e.g. by blocking access to the directory service where the contact information is held, or by blocking access to the introduction points where the hidden service is listening for connection requests. Using a separate set of directory servers for hidden services, and combining this with the use of distributed hash tables like CAN [84], Chord [103], Pastry [94], or Tapestry [114], could be implemented to support the storage, lookup, and retrieval of hidden services' contact information, and would increase attack-resistance on the directory servers as mentioned in Part II.

3 Contribution and Summary

The thesis consists of the following five research papers:

Paper A Lasse Øverlier, Tønnes Brekne and André Årnes. **Non-expanding Transaction Specific Pseudonymization for IP Traffic Monitoring.** In Yvo G. Desmedt, Huaxiong Wang, Yi Mu, and Yongqing Li, editors, *Cryptology and Network Security: 4th International Conference (CANS 2005)*, pages 261–273. Springer-Verlag, LNCS 3810, December 2005.

Paper B Geir Hallingstad and Lasse Øverlier. **Traffic Flow Confidentiality in a Future Network Enabled Capability Environment.** In *Proceedings of the 2007 IEEE Information Assurance and Security Workshop.*, pages 325–332. IEEE, June 2007.

Paper C Lasse Øverlier and Paul Syverson. **Locating Hidden Servers.** In *Proceedings of the 2006 IEEE Symposium on Security and Privacy (S&P'06)*, pages 100–114, May 2006. IEEE Computer Society.

Paper D Lasse Øverlier and Paul Syverson. **Valet Services: Improving Hidden Servers with a Personal Touch.** In George Danezis and Philippe Golle, editors, *Proceedings of the Sixth Workshop on Privacy Enhancing Technologies (PET 2006)*, pages 223–244, Cambridge, UK, June 2006. Springer-Verlag, LNCS 4258.

Paper E Lasse Øverlier and Paul Syverson. **Improving efficiency and simplicity of Tor circuit establishment and hidden services.** In *Proceedings of the Seventh Privacy Enhancing Technologies Symposium (PETS 2007)*, pages 134–152, Ottawa, Canada, June 2007. Springer-Verlag, LNCS 4776.

The time line in Figure 3.1 shows how the different papers are interconnected through the research period - improving the security and speed of anonymous communication

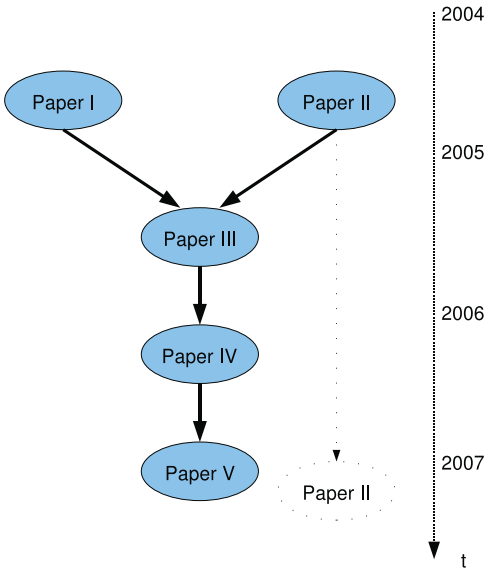

Figure 3.1: Papers and time line

and hidden services. The first two papers were completed in parallel works in cooperation with different research teams while initializing the hidden services research. The following three papers address the vulnerabilities, challenges and improvements around anonymous communication and location hidden services.

3.1 Contribution of Paper A

Non-expanding Transaction Specific Pseudonymization for IP Traffic Monitoring

This paper [76] presents a solution to securely pseudonymize IP addresses in high speed, large scale traffic data collections, while still maintaining a secure, flexible and configurable method of searching for data in these logs. As the security of anonymous communications is directly related to the possibility, availability and searchability of traffic data logs, common interests with researchers from the LOBSTER and SCAMPI EU-projects were identified. This paper was a result from cooperation with researchers working at the Centre for Quantifiable Quality of Service in Communication Systems in Trondheim.

One problem of earlier pseudonymization schemes for IP address logs is the narrow span of possible IP addresses. For IPv4 even a complete address span with combined IP_{from}, IP_{to} is only 64-bit wide, which regardless of earlier methods of pseudonymization will be vulnerable to different types of attacks [14, 13]. Another important aspect is to avoid expanding the logs, as the logs are to be implemented on high speed interconnections of the Internet which already carry traffic in the multi-gigabit range.

A new secure logging scheme is suggested in this paper, describing how to construct a non-expanding transaction specific pseudonymization by using stream ciphers. Individual strong stream ciphers are applied to each bit column of the n-bit traffic data. This way n stream cipher keys will protect one bit column each of traffic data, and searching inside individual columns can be enabled by sharing only the necessary keys. The logging scheme is non-expanding as it adds no extra data and is therefore able to keep the logs to a minimum which is highly relevant in these interconnections. The technique also enables transaction specific pseudonyms to be constructed for each row of data which will protect the logs from injection attacks. To further reduce the potential for abuse it is suggested to rotate encryption keys[1] after each block of k rows has been pseudonymized, and thereby limiting the amount of traffic data available to an attacker. The pseudonymization scheme presented can also be used to cover logging and searching of traffic data (i.e. content), not only IP addresses. The scheme is suitable for parallelization and is therefore also eligible for an efficient hardware implementation.

3.2 Contribution of Paper B

Traffic Flow Confidentiality in a Future Network Enabled Capability Environment

This paper [47] is an analysis of how traffic flow confidentiality will become a challenge in military network enabled capability (NEC) environments [16]. These future networks require a high degree of flexibility for efficient exchange of information. This will likely move information protection closer to the edge of the network i.e. towards the highest layers in the standard network model. On the other side, high demand for availability will likely move integrity protection towards lower levels to eliminate rogue traffic already

[1] Actually it is the initialization vectors (IVs) that are rotated for each block of data.

at the source. Now both the integrity and confidentiality of NEC environments will secure the network, and the paper makes an analysis of how this enhanced flexibility influences the adversary's potential for traffic flow analysis.

A Friendly Force Tracing Scenario using satellite communication to create situation reports and allow the exchange of messages, is used as an example to analyze the problem of traffic flow confidentiality in the scenarios of encryption applied at the link level, IP level (IPsec), and at the object level.

The paper shows how this future scenario leaks traffic flow information at the different network levels, how the existing countermeasures will not effectively help this without compromising the wanted flexibility and availability. Not even anonymous communication using hidden services can accomplish this without introducing some key management scheme and lower layer confidentiality which would terminate this flexibility.

3.3 Contribution of Paper C

Locating Hidden Servers

The main contribution of this paper [77] is the demonstration of effective intersection attacks in a live anonymizing network, and the introduction and analysis of different countermeasures against these attacks.

The research shows how an attacker can locate the IP address of a hidden server in a matter of minutes by controlling only one compromised/evil node in the Tor server network. Using only one node the location attack can be performed within a couple of minutes or at most a couple of hours, and by using two nodes the attack will always succeed within a few minutes. As shown in the paper, a connection to a hidden service is completed when the hidden service connects back to the rendezvous point. By opening connections to a hidden service again and again, thereby forcing the hidden server to connect back to the rendezvous point through different random circuits, the attacker can use statistical methods to locate the hidden server's IP address. First the evil node uses timing analysis to determine whether it has been made part of the circuit from the hidden server to the rendezvous point. If this is confirmed and the IP address of the previous nodes are stored in a list, the hidden server's IP address will be over-represented

in the list. This is more commonly known as an *intersection attack*, or the *predecessor attack*.

Countermeasures discussed are *dummy traffic*, *extended circuit length*, and *entry guard nodes*. Of these only entry guard nodes, a small set of preselected permanent nodes used as first nodes for all anonymous connections, is shown to be a countermeasure that significantly reduces the success rate of the attack. The paper makes an analysis of the possible variations of entry guard nodes and completes an experiment using the same attack when entry guard nodes are implemented. The paper shows that by using entry guard nodes an attacker will be able to identify the location of these entry guard nodes, but *not* the location of the hidden server. Using *backup guard nodes* - a list of preselected spare nodes, or *layered guard nodes* - where each guard node has its own list of second level guard nodes for the next hop, will further slow down the attack.

As a result from the attention the research work received, the report and a live demonstration of the attack was presented at two other conferences[2] in addition to its publication release. Recent work by Abbot et al. [2] has already extended this attack into locating Tor clients using the same principles and this paper's traffic pattern matching algorithms.

3.4 Contribution of Paper D

Valet Services: Improving Hidden Servers with a Personal Touch

Until now most published work on hidden services have focused on the vulnerability of locating the hidden servers and almost no work have focused on another important design goal for the hidden services, censorship-resistance. There were known problems with the existing hidden service design making it possible for an attacker to stop a hidden service by launching a DoS attack on the introduction points or on the directory servers.

The main contribution of this paper [78] is the introduction of the *valet nodes*, created to reduce a hidden service's vulnerability to denial-of-service (DoS) attacks and add quality of service (QoS) as a service option to both anonymous and authenticated users of a hidden service. Additionally valet nodes not only hides the introduction points

[2]BlackHat Federal `http://blackhat.com/`, and ShmooCon `http://shmoocon.org/`.

from being located, but the research shows how to hide the very existence of a hidden service from everyone but the users knowing the exact service address.

Recalling that the introduction points are vulnerable to attacks, the *valet nodes* protect the introduction points by hiding the introduction point's identity from the clients. In addition neither the valet nodes nor the introduction points knows which service they are being used for. The information for connecting to a hidden service is located in *contact information tickets* (CIT) containing a description of the valet nodes and an encrypted extension message for the valet node identifying which introduction point the valet node should extend the circuit to. The client will not at any time know which introduction points are being used, and cannot target them for attacks. By having more than one valet node per introduction point, and reducing the probability of a client knowing all valet nodes, the probability of a successful denial-of-service attack on the service is significantly reduced.

To hide the very existence of a hidden service the network has to restrict access to the hidden service's CITs. This is accomplished by encrypting both the CITs and the CIT identifiers with keys derived from the hidden service's public key. The consequence of this is that the client must have access to the public key, which is the hidden service's unified resource locator, in order to both access and decrypt the contact information ticket. No one else will be able to identify the CIT nor the address (URL) of a hidden service. The dynamics of these descriptors can be high, involving valid time periods, client authentication tokens, and other types of cookies. The descriptors can always be verified as the CITs are signed with the (already known) public key of the hidden service. Updates of these CITs are made possible and verifiable by using a reverse hash chain scheme.

A deeper analysis of the security in locating all introduction points is completed using varying numbers of introduction points and valet nodes per introduction point. E.g. by using three valet nodes for each of the hidden service's three introduction points, an attacker must control 100 nodes in a 500 node anonymizing network in order to have a 12% chance of locating all three introduction points.

The paper also describes how quality of service for both authenticated and anonymous users can be added through the use of valet nodes and CITs, and how the valet nodes scheme is not influenced if the anonymizing network starts to use distributed hash

tables as a replacement for directory services.

3.5 Contribution of Paper E

Improving efficiency and simplicity of Tor circuit establishment and hidden services

This paper [79] proposes a protocol for Tor circuit establishment that eliminates the need for RSA encryption and decryption, and it also suggests how to let the new protocol improve the valet node design by eliminating the need for an external rendezvous point.

In Tor circuit setup a client must interact and setup encrypted tunnels with each node in the path towards the exit node. As the clients know the identity of the server nodes through a signed certificate and a public RSA key, the client can only confirm the server node's identity by either encrypting a message or confirming a signature with the node's public key. This is in the current implementation of anonymous circuits involving three (the length of the anonymizing tunnel) RSA encryptions/decryptions on the client side, and one encryption/decryption on each of the nodes in the path. Since the current Tor implementation uses RSA *in addition to* the Diffie-Hellman ephemeral key exchange, this paper proposes a new protocol combining authentication and key exchange by using predistributed Diffie-Hellman (DH) values for each of the server nodes in the Tor network.

All nodes publish a list of individual descriptors, e.g. IP address, TCP ports, RSA public key(s), nickname, etc. This list is signed with a private RSA key and used for authentication when connecting to a node. By adding a public DH-value[3] to this list of signed identifiers we can use this value as the node's public DH-value during the initial handshake. Security is maintained by rotating this value regularly. Then a half-authenticated setup can be completed if the client constructs a message like "$DH_c, E_K\{data\}$", where DH_c is the clients ephemeral public DH-value for this connection, E_K is the DH-key derived from the clients private DH-value, and the server node's public DH-value. Only the server node with access to the associated private DH-value will be able to derive the correct key and decrypt *data*. This is more commonly known as an ElGamal key agreement [69, p. 517], or a half-certified Diffie-Hellman. The sim-

[3]The paper proposes to add a list of these DH-values with different validity periods for each of the anonymizing servers as these descriptor lists are updated regularly anyway.

plification eliminates the RSA encryption and saves three exponentiations on the client, and one on each of the other nodes in the circuit.

The paper also demonstrates how to reduce the number of exchanged messages in various circuit setup scenarios, and makes an analysis on the security of these proposed protocols. In addition the research shows how an improved protocol extending the ElGamal key agreement can utilize forward secrecy immediately after the key exchange has been completed. This increases the number of exponentiations again from the earlier proposed low limit, but it is still fewer than the original RSA based protocol as many of the exponentiations may be processed when idle and not during circuit setup.

The hidden service's circuit setup will also gain from these predistributed DH-values as valet nodes are more easily implemented, and because the rendezvous point may be eliminated. The new protocol may use ephemeral introduction points where the communication continues to use the initial introduction point circuit. Another possibility discussed is to use the last circuit node *in front of* the valet node as a rendezvous point, thereby eliminating the need for a separate client-to-rendezvous point connection.

3.6 Summary of thesis contribution

Network anonymity is a wide area of research including anonymous email, anonymous browsing and access of services, and censorship-resistant publishing. This thesis has contributed with an analysis of traffic flow confidentiality and anonymizing networks, and found methods to make sure that it is possible to have secure and flexible transaction specific pseudonymous logging of traffic data.

As the EU data retention directive [36] currently is being implemented, the transaction specific pseudonymous logging technique is well suited for securing the many high speed communication logs that will be created. In addition, the solution enables the search of some data areas inside the logged data without revealing all data, meaning that it is possible to find out *if* an activity has been committed without giving away the identity (location) of the user until the assumed activity is confirmed.

The thesis has demonstrated that intersection attacks do work in live networks, by implementing the predecessor attack in the Tor network, and that a location attack on hidden services can be completed using multiple methods with only one or two malicious

nodes. The time to locate a hidden service using only one evil node was shown to vary from a couple of minutes to, at the highest, a couple of hours. After the entry guard node countermeasures were implemented the research confirmed the assumption that these guard nodes could be found, but not the hidden service itself. The hidden service has thereby been added an extra layer of protection and more direct attacks on the guard nodes must be completed in order to attempt to locate the hidden service.

The research has shown how the use of valet nodes enhances the hidden services' resistance to denial-of-service attacks by protecting their introduction points. In addition the valet nodes technique enables the possibility of completely hidden services, methods for individual or group based quality of service for the users, and methods to avoid using the rendezvous point in the hidden service connection setup. The number of valet nodes used for each introduction point has also been shown to decrease the probability of locating all introduction points from one (existing solution today lists all introduction points) to almost zero unless the attacker controls a major portion of the anonymizing network.

A general improvement of the authenticated Diffie-Hellman key exchange has been presented eliminating the need for the RSA encryption and decryption by using pre-distributed Diffie-Hellman values. This has reduced the number of encryptions and the number of messages necessary for setting up an anonymous circuit while maintaining forward secrecy. The solution is also easily adaptable to the valet nodes design which will benefit from the use of public DH-values and also avoid the use of RSA. In addition the latency in connecting to hidden services may be reduced without setting up new connections to external rendezvous points.

3.7 Further research

Inside the main area of the research work there have been identified multiple fields for further research. First there is testing and/or simulating the various extended guard nodes schemes, where the layered guard node scheme might be the most interesting case. By simulating a variable number of guard nodes for each layer and constructing scenarios with a variable number of colluding nodes, it should be easier to estimate the change in anonymity protection on both hidden services and clients of the anonymizing

network.

Second, there continues to be more research into location threats of hidden services than accessibility threats, so this area of research is only in its beginning. Implementing valet nodes and contact information tickets to look at performance and security issues are the first natural extensions of this work. In addition many research questions arises when moving from directory servers, e.g. towards a distributed hash table lookup service. Update, trust, reliability and synchronization, are only a few of the problems a new replacement for the directory servers must look into. More general problems with anonymizing networks is ongoing research, like adding blocking resistance and how to make the anonymizing traffic to look like normal Internet traffic, etc.

Many anonymity issues are also related to the contact information tickets and the use of valet nodes. By allowing a client accessing a hidden service to get different quality-of-service based on e.g. previous behavior, there is the question on whether the client can remain anonymous and have this previous behavior remain as detached events and therefore unlinkable. In addition a better analysis of the situation where a completely hidden service (secret hidden key with optional authorization) has its public key exposed. There exists at the moment no better solution but to "re-hide" the service by redistributing a new public key, which is both cumbersome and has several vulnerabilities.

The proposed Diffie-Hellman enhancement should also be followed up by a more formal analysis of the protocol, and an evaluation of the different cryptographic methods that can be used to achieve the half-certified key exchange. Before changing the connection to the hidden service, a more formal analysis of the security in the proposed protocols should be completed.

In addition the research performed on pseudonymous logging should be followed up by looking at more effective search and statistical algorithms for the proposed protection scheme. And it should be constructed hardware tests and implementations of the secure logging and search algorithms to see how hardware performance of the pseudonymization scheme will be. This research might reveal how the proposed pseudonymization scheme can keep up with the continuously increasing bandwidth and log capacity which service providers will be required to provide, keep secure, and made searchable upon request.

Bibliography

[1] 107th U.S. Congress. Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT ACT) Act of 2001. Retrieved July 10, 2007, from U.S. Government Printing Office: `http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_public_laws&docid=f:publ056.107.pdf`.

[2] Timothy G. Abbott, Katherine J. Lai, Michael R. Lieberman, and Eric C. Price. Browser-based attacks on tor. In *Proceedings of the Seventh Privacy Enhancing Technologies Symposium (PETS 2007)*, Ottawa, CA, June 2007. Springer.

[3] Ross J. Anderson. The eternity service. In *Proceedings of Pragocrypt '96*, 1996.

[4] The Anonymizer. `http://www.anonymizer.com/`.

[5] Adam Back, Ian Goldberg, and Adam Shostack. Freedom systems 2.1 security issues and analysis. White paper, Zero Knowledge Systems, Inc., May 2001. Retrieved June 30, 2007 from `http://osiris.978.org/~brianr/crypto-research/anon/www.freedom.net/products/whitepapers/Freedom_Security2-1.pdf`.

[6] Adam Back, Ulf Möller, and Anton Stiglic. Traffic analysis attacks and trade-offs in anonymity providing systems. In Ira S. Moskowitz, editor, *Proceedings of Information Hiding Workshop (IH 2001)*, pages 245–257. Springer-Verlag, LNCS 2137, April 2001.

[7] Tonda Benes. The strong eternity service. In Ira S. Moskowitz, editor, *Proceedings of Information Hiding Workshop (IH 2001)*, pages 215–229. Springer-Verlag, LNCS 2137, April 2001.

[8] Krista Bennett and Christian Grothoff. GAP – practical anonymous networking. In Roger Dingledine, editor, *Proceedings of Privacy Enhancing Technologies Workshop (PET 2003)*, pages 141–160. Springer-Verlag, LNCS 2760, March 2003.

[9] Oliver Berthold, Hannes Federrath, and Stefan Köpsell. Web MIXes: A system for anonymous and unobservable Internet access. In H. Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 115–129. Springer-Verlag, LNCS 2009, July 2000.

[10] Oliver Berthold and Heinrich Langos. Dummy traffic against long term intersection attacks. In Roger Dingledine and Paul Syverson, editors, *Proceedings of Privacy Enhancing Technologies workshop (PET 2002)*, pages 110–128. Springer-Verlag, LNCS 2482, April 2002.

[11] Oliver Berthold, Andreas Pfitzmann, and Ronny Standtke. The disadvantages of free MIX routes and how to overcome them. In H. Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 30–45. Springer-Verlag, LNCS 2009, July 2000.

[12] Philippe Boucher, Adam Shostack, and Ian Goldberg. Freedom systems 2.0 architecture. White paper, Zero Knowledge Systems, Inc., December 2000. Retrieved June 30, 2007 from `http://osiris.978.org/~brianr/crypto-research/anon/www.freedom.net/products/whitepapers/Freedom_System_2_Architecture.pdf`.

[13] Tønnes Brekne and André Årnes. Circumventing IP-Address Pseudonymization in $O(N^2)$ Time. In M.Y. Sanadidi, editor, *Proceedings of IASTED Communication and Computer Networks (CCN 2005)*, October 2005.

[14] Tønnes Brekne, André Årnes, and Arne Øslebø. Anonymization of IP Traffic Monitoring Data: Attacks on Two Prefix-preserving Anonymization Schemes and Some Proposed Remedies. In George Danezis and David Martin, editors, *Proceedings of Privacy Enhancing Technologies workshop (PET 2005)*, pages 179–196. Springer-Verlag, LNCS 3424, May 2005.

[15] Zach Brown. Cebolla: Pragmatic IP Anonymity. In *Proceedings of the 2002 Ottawa Linux Symposium*, June 2002.

[16] T. Buckman, M. Booth, J. Busch, B. Caplan, B. Christiansen, R. van Engelshoven, K. Eckstein, G. Hallingstad, T. Halmai, P. Howland, V. Rodriguez-Herola, D. Kallgren, S. Onganer, R. Porta, C. Shawcross, P. Szczucki, and K. Veum. Nato network-enabled capability feasibility study volume ii: Detailed report covering

a strategy and roadmap for realizing an nnec networking and information infrastructure (nii), version 2.0. NATO Consultation, Command and Control Agency document (NATO Unclassified), October 2005.

[17] David L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–90, February 1981.

[18] David L. Chaum. The dining cryptographers problem: Unconditional sender and recipient untraceability. *Journal of Cryptology*, 1(1):65–75, January 1988.

[19] Ian Clarke, Oskar Sandberg, Brandon Wiley, and Theodore W. Hong. Freenet: A distributed anonymous information storage and retrieval system. In H. Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 46–66. Springer-Verlag, LNCS 2009, July 2000.

[20] Frank Dabek, M. Frans Kaashoek, David Karger, Robert Morris, and Ion Stoica. Wide-area cooperative storage with CFS. In *SOSP '01: Proceedings of the eighteenth ACM symposium on Operating systems principles*, pages 202–215, Banff, Alberta, Canada, October 2001. ACM Press.

[21] Wei Dai. Pipenet 1.1. Usenet post, August 1996. Retrieved July 10, 2007 from `http://www.weidai.com/pipenet.txt`.

[22] George Danezis. Forward secure mixes. In Jonsson Fisher-Hubner, editor, *Proceedings of 7th Nordic Workshop on Secure IT Systems*, pages 195–207, Karlstad, Sweden, November 2002.

[23] George Danezis. Statistical disclosure attacks: Traffic confirmation in open environments. In Gritzalis, Vimercati, Samarati, and Katsikas, editors, *Proceedings of Security and Privacy in the Age of Uncertainty, (SEC2003)*, pages 421–426, Athens, May 2003. IFIP TC11, Kluwer.

[24] George Danezis. The traffic analysis of continuous-time mixes. In David Martin and Andrei Serjantov, editors, *Proceedings of Privacy Enhancing Technologies Workshop (PET 2004)*, pages 35–50. Springer-Verlag, LNCS 3424, May 2004.

[25] George Danezis, Roger Dingledine, and Nick Mathewson. Mixminion: Design of a Type III Anonymous Remailer Protocol. In *Proceedings of the 2003 IEEE Symposium on Security and Privacy (S&P'03*, pages 2–15. IEEE Computer Society, May 2003.

[26] Janine DeFao. Parents turn to tech toys to track teens. *San Fransisco Chronicle*, July 2006. Retrieved June 30, 2007, from `http://sfgate.com/cgi-bin/article.cgi?f=/c/a/2006/07/09/BIGMOTHER.TMP`.

[27] Claudia Díaz and Bart Preneel. Taxonomy of mixes and dummy traffic. In *Proceedings of I-NetSec04: 3rd Working Conference on Privacy and Anonymity in Networked and Distributed Systems*, pages 215–230, Toulouse, France, August 2004.

[28] Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In Roger Dingledine and Paul Syverson, editors, *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)*, pages 54–68. Springer-Verlag, LNCS 2482, April 2002.

[29] T. Dierks and E. Rescorla. The Transport Layer Security (TLS) Protocol Version 1.1. IETF RFC 4346, April 2006. Retrieved June 30, 2007, from `http://ietf.org/rfc/rfc4346.txt`.

[30] Whitfield Diffie and Martin E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, IT-22(6):644–654, November 1976.

[31] Whitfield Diffie, Paul C. van Oorschot, and Michael J. Wiener. Authentication and authenticated key exchanges. *Designs, Codes and Cryptography*, 2(2):107–125, 1992.

[32] Roger Dingledine, Michael J. Freedman, and David Molnar. The Free Haven Project: Distributed anonymous storage service. In H. Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability*, pages 67–95. Springer-Verlag, LNCS 2009, July 2000.

[33] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th USENIX Security Symposium*, pages 303–320, August 2004.

[34] John Douceur. The Sybil Attack. In Peter Druschel, Frans Kaashoek, and Antony Rowstron, editors, *Proceedings of the Peer To Peer Systems: First International Workshop (IPTPS 2002)*, pages 251–260. Springer-Verlag, LNCS 2429, March 2002.

[35] Peter Druschel and Anthony Rowstron. PAST: A large-scale, persistent peer-to-peer storage utility. In *Proceedings of The 8th Workshop on Hot Topics in Operating Systems*, pages 75–80, Schoss Elmau, Germany, May 2001.

[36] European Parliament. Directive on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks. Directive 2006/24/EC. Retrieved July 10, 2007, from `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32006L0024:EN:NOT`.

[37] Nick Feamster, Magdalena Balazinska, Greg Harfst, Hari Balakrishnan, and David Karger. Infranet: Circumventing web censorship and surveillance. In *Proceedings of the 11th USENIX Security Symposium*, pages 247–262, San Francisco, CA, August 2002.

[38] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee. Hypertext Transfer Protocol – HTTP/1.1. IETF RFC 2616, June 1999. Retrieved June 30, 2007, from `http://ietf.org/rfc/rfc2616.txt`.

[39] Michael J. Freedman and Robert Morris. Tarzan: A peer-to-peer anonymizing network layer. In *Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS 2002)*, pages 193–206, Washington, DC, USA, November 2002.

[40] Sharad Goel, Mark Robson, Milo Polte, and Emin Gün Sirer. Herbivore: A scalable and efficient protocol for anonymous communication. Technical Report 2003-1890, Cornell University, Ithaca, NY, USA, February 2003.

[41] Ian Goldberg. *A Pseudonymous Communications Infrastructure for the Internet*. PhD thesis, UC Berkeley, December 2000.

[42] Ian Goldberg. Privacy-enhancing technologies for the Internet, II: Five years later. In Roger Dingledine and Paul Syverson, editors, *Proceedings of Privacy Enhancing Technologies workshop (PET 2002)*, pages 1–12. Springer-Verlag, LNCS 2482, April 2002.

[43] Ian Goldberg and David Wagner. TAZ servers and the rewebber network: Enabling anonymous publishing on the world wide web. *First Monday*, 3(4), August 1998.

[44] Ian Goldberg, David Wagner, and Eric Brewer. Privacy-enhancing technologies for the internet. In *Proceedings of the 42nd IEEE Spring COMPCON*, pages 103–109. IEEE Computer Society Press, February 1997.

[45] David M. Goldschlag, Michael G. Reed, and Paul F. Syverson. Hiding Routing Information. In R. Anderson, editor, *Proceedings of Information Hiding: First International Workshop*, pages 137–150. Springer-Verlag, LNCS 1174, May 1996.

[46] Ceki Gülcü and Gene Tsudik. Mixing E-mail with Babel. In *SNDSS '96: Proceedings of the 1996 Symposium on Network and Distributed System Security (SNDSS '96)*, pages 2–16, Washington, DC, USA, February 1996. IEEE Computer Society.

[47] Geir Hallingstad and Lasse Øverlier. Traffic flow confidentiality in a future network enabled capability environment. In *Proceedings of the 2007 IEEE SMC Information Assurance and Security Workshop*, pages 325–332. IEEE, June 2007.

[48] Johan Helsingius. Press release about the closing of `anon.penet.fi`. Retrieved June 30, 2007, from `http://www.eff.org/Privacy/Anonymity/960830_penet_closure.announce`.

[49] Andrew Hintz. Fingerprinting websites using traffic analysis. In Roger Dingledine and Paul Syverson, editors, *Proceedings of Privacy Enhancing Technologies workshop (PET 2002)*, pages 171–178. Springer-Verlag, LNCS 2482, April 2002.

[50] Stephen Hsu. SafeWeb Inc. - TriangleBoy press release in ZeroPaid archive, March 2001. Retrieved July 10, 2007, from `http://www.zeropaid.com/bbs/archive/index.php/t-1236.html`.

[51] Markus Jakobsson. Flash Mixing. In *Proceedings of 1999 ACM Symposium of Distributed Computing - PODC*, pages 83–89. ACM Press, 1999.

[52] Markus Jakobsson and Ari Juels. An optimally robust hybrid mix network. In *PODC '01: Proceedings of the twentieth annual ACM symposium on Principles of distributed computing*, pages 284–292, Newport, Rhode Island, United States, 2001. ACM Press.

[53] Markus Jakobsson, Ari Juels, and Ronald L. Rivest. Making mix nets robust for electronic voting by randomized partial checking. In *Proceedings of the 11th USENIX Security Symposium*, pages 339–353, San Francisco, CA, August 2002.

[54] Markus Jakobsson and David M'Raïhi. Mix-based electronic payments. In *SAC '98: Proceedings of the Selected Areas in Cryptography*, pages 157–173, London, UK, 1999. Springer-Verlag, LNCS 1556.

[55] Anja Jerichow, Jan Müller, Andreas Pfitzmann, Birgit Pfitzmann, and Michael Waidner. Real-Time MIXes: A Bandwidth-Efficient Anonymity Protocol. *IEEE Journal on Selected Areas in Communications*, 16(4):495–509, May 1998.

[56] Niel F. Johnson and Sushil Jajodia. Exploring steganography: Seeing the unseen. *IEEE Computer*, 31(2):26–34, February 1998.

[57] Sachin Katti, Jeffery Cohen, and Dina Katabi. Information slicing: Anonymity using unreliable overlays. In *Proceedings of the 4th USENIX Symposium on Network Systems Design and Implementation (NSDI)*, pages 43–56, April 2007.

[58] Dogan Kesdogan, Dakshi Agrawal, and Stefan Penz. Limits of anonymity in open environments. In Fabien Petitcolas, editor, *Proceedings of Information Hiding Workshop (IH 2002)*, pages 53–69. Springer-Verlag, LNCS 2578, October 2002.

[59] Dogan Kesdogan, Dakshi Agrawal, Vinh Pham, and Dieter Rautenbach. Fundamental limits on the anonymity provided by the MIX technique. In *Proceedings of the 2006 IEEE Symposium on Security and Privacy (S&P'06)*, pages 86–99. IEEE CS, May 2006.

[60] Dogan Kesdogan, Jan Egner, and Roland Büschkes. Stop-and-go MIXes: Providing probabilistic anonymity in an open system. In David Aucsmith, editor, *Proceedings of Information Hiding Workshop (IH 1998)*, pages 83–98. Springer-Verlag, LNCS 1525, 1998.

[61] David Koblas and Michelle R. Koblas. SOCKS. In *Proceedings of the Third USENIX UNIX Security Symposium*, pages 77–83, Baltimore, MD, September 1992. USENIX Association.

[62] Stefan Köpsell and Ulf Hilling. How to achieve blocking resistance for existing systems enabling anonymous web surfing. In *Proceedings of the 2004 ACM Workshop on Privacy in the electronic society (WPES 2004)*, pages 47–58, Washington, DC, USA, October 2004.

[63] Dennis Kügler. An Analysis of GNUnet and the Implications for Anonymous, Censorship-Resistant Networks. In Roger Dingledine, editor, *Proceedings of Privacy Enhancing Technologies workshop (PET 2003)*, pages 161–176. Springer-Verlag, LNCS 2760, March 2003.

[64] M. Leech, M. Ganis, Y. Lee, R. Kuris, D. Koblas, and L. Jones. SOCKS Protocol Version 5. IETF RFC 1928, March 1996. Retrieved June 30, 2007, from `http://ietf.org/rfc/rfc1928.txt`.

[65] Brian N. Levine, Michael K. Reiter, Chenxi Wang, and Matthew Wright. Timing attacks in low-latency mix-based systems (extended abstract). In Ari Juels, editor, *Financial Cryptography, 8th International Conference, FC 2004*, pages 251–265. Springer-Verlag, LNCS 3110, 2004.

[66] Brian Neil Levine and Clay Shields. Hordes — A Multicast Based Protocol for Anonymity. *Journal of Computer Security*, 10(3):213–240, 2002.

[67] David Martin and Andrew Schulman. Deanonymizing users of the safeweb anonymizing service. In *Proceedings of the 11th USENIX Security Symposium*, pages 123–137, San Francisco, CA, August 2002.

[68] Nick Mathewson and Roger Dingledine. Practical traffic analysis: Extending and resisting statistical disclosure. In David Martin and Andrei Serjantov, editors, *Proceedings of Privacy Enhancing Technologies workshop (PET 2004)*, pages 17–34. Springer-Verlag, LNCS 3424, May 2004.

[69] Alfred J. Menezes, Scott A. Vanstone, and Paul C. Van Oorschot. *Handbook of Applied Cryptography*. CRC Press, Inc., Boca Raton, FL, USA, 1996.

[70] Masashi Mitomo and Kaoru Kurosawa. Attack for Flash MIX. In *Proceedings of the 6th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology*, pages 192–204. Springer-Verlag, LNCS 1976, December 2000.

[71] Ulf Möller, Lance Cottrell, Peter Palfrader, and Len Sassaman. Mixmaster protocol version 2. Internet-Draft, December 2004. Retrieved July 10, 2007, from `http://www.ietf.org/internet-drafts/draft-sassaman-mixmaster-03.txt`.

[72] Steven J. Murdoch. Hot or not: Revealing hidden services by their clock skew. In *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS 2006)*, pages 27–36. ACM Press, November 2006.

[73] Steven J. Murdoch and George Danezis. Low-cost traffic analysis of Tor. In *Proceedings of the 2005 IEEE Symposium on Security and Privacy (S&P'05)*, pages 183–195. IEEE Computer Society, May 2005.

[74] Steven J. Murdoch and Piotr Zieliński. Sampled traffic analysis by internet-exchange-level adversaries. In Nikita Borosov and Philippe Golle, editors, *Proceedings of the Seventh Privacy Enhancing Technologies Symposium (PETS 2007)*, Ottawa, Canada, June 2007. Springer-Verlag, LNCS.

[75] Miyako Ohkubo and Masayuki Abe. A length-invariant hybrid mix. In Tatsuaki Okamoto, editor, *ASIACRYPT '00: Proceedings of the 6th International Conference on the Theory and Application of Cryptology and Information Security*, pages 178–191, Kyoto, Japan, 2000. Springer-Verlag, LNCS 1976.

[76] Lasse Øverlier, Tønnes Brekne, and Andrè Årnes. Non-expanding Transaction Specific Pseudonymization for IP Traffic Monitoring. In Yvo G. Desmedt, Huaxiong Wang, Yi Mu, and Yongqing Li, editors, *Cryptology and Network Security: 4th International Conference (CANS 2005)*, pages 261–273. Springer-Verlag, LNCS 3810, December 2005.

[77] Lasse Øverlier and Paul Syverson. Locating hidden servers. In *Proceedings of the 2006 IEEE Symposium on Security and Privacy (S&P'06)*, pages 100–114. IEEE Computer Society, May 2006.

[78] Lasse Øverlier and Paul Syverson. Valet services: Improving hidden servers with a personal touch. In George Danezis and Philippe Golle, editors, *Proceedings of the Sixth Workshop on Privacy Enhancing Technologies (PET 2006)*, pages 223–244, Cambridge, UK, June 2006. Springer-Verlag, LNCS 4258.

[79] Lasse Øverlier and Paul Syverson. Improving efficiency and simplicity of tor circuit establishment and hidden services. In *Proceedings of the Seventh Privacy Enhancing Technologies Symposium (PETS 2007)*, pages 134–152, Ottawa, Canada, June 2007. Springer-Verlag, LNCS 4776.

[80] Andreas Pfitzmann and Marit Hansen. Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management - a consolidated proposal for terminology. Draft, version 0.29, July 2007. Retrieved July 31, 2007 from `http://dud.inf.tu-dresden.de/Anon_Terminology.shtml`.

[81] Andreas Pfitzmann, Birgit Pfitzmann, and Michael Waidner. ISDN-mixes: Untraceable communication with very small bandwidth overhead. In *Proceedings of the GI/ITG Conference on Communication in Distributed Systems*, pages 451–463. Springer-Verlag, February 1991.

[82] The Free Network Project. `http://freenetproject.org/`.

[83] Michael O. Rabin. Efficient dispersal of information for security, load balancing, and fault tolerance. *Journal of the ACM*, 36(2):335–348, 1989.

[84] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, and Scott Schenker. A scalable content-addressable network. In *SIGCOMM '01: Proceed-*

ings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications, pages 161–172, San Diego, California, United States, 2001. ACM Press.

[85] Jean-François Raymond. Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems. In H. Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 10–29. Springer-Verlag, LNCS 2009, July 2000.

[86] Michael G. Reed, Paul F. Syverson, and David M. Goldschlag. Proxies for Anonymous Routing. In *ACSAC '96: Proceedings of the 12th Annual Computer Security Applications Conference*, pages 95–104, Washington, DC, USA, December 1996. IEEE Computer Society.

[87] Michael Reiter and Aviel Rubin. Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security (TISSEC)*, 1(1):66–92, June 1998.

[88] Relakks. `http://www.relakks.com/`.

[89] Marc Rennhard and Bernhard Plattner. Introducing MorphMix: Peer-to-peer based anonymous internet usage with collusion detection. In *WPES '02: Proceedings of the 2002 ACM workshop on Privacy in the Electronic Society*, pages 91–102, Washington, DC, USA, November 2002. ACM Press.

[90] Marc Rennhard and Bernhard Plattner. Practical anonymity for the masses with MorphMix. In Ari Juels, editor, *Proceedings of Financial Cryptography (FC '04)*, pages 233–250. Springer-Verlag, LNCS 3110, February 2004.

[91] James Risen and Eric Lichtblau. Bush lets U.S. spy on callers without courts. *The New York Times*, December 2005. Retrieved June 30, 2007, from `http://www.nytimes.com/2005/12/21/politics/21nsa.html?ex=1292821200&en=91d434311b0a7ddc&ei=5088&partner=rssnyt&emc=rss`.

[92] Terry Rombeck. Tracking teens: Cell phone features offer mixed benefits. *Lawrence Journal-World*, August 2006. Retrieved June 30, 2007, from `http://www2.ljworld.com/news/2006/aug/01/tracking_teens/`.

[93] Anthony Rowstron and Peter Druschel. Storage management and caching in PAST, a large-scale, persistent peer-to-peer storage utility. In *SOSP '01: Proceedings of the eighteenth ACM symposium on Operating systems principles*, pages 188–201, Banff, Alberta, Canada, October 2001. ACM Press.

[94] Antony Rowstron and Peter Druschel. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In *Proceedings of the 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware 2001)*, pages 329–350, Heidelberg, Germany, November 2001.

[95] K. Sako and J. Killian. Receipt-free mix-type voting scheme. In Louis C. Guillou and Jean-Jacques Quisquater, editors, *EUROCRYPT '95: International Conference on the Theory and Application of Cryptographic Techniques*, pages 393–403, Saint-Malo, France, May 1995. Springer-Verlag, LNCS 921.

[96] Paul Sarbanes and Michael G. Oxley. Public Company Accounting Reform and Investor Protection Act of 2002 (Sarbanes-Oxley Act). Pub. L. No. 107-204, 116 Stat. 745, July 2002. Retrieved June 30, 2007, from U.S. Government Printing Office via GPO Access: `http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_bills&docid=f:h3763enr.tst.pdf`.

[97] Andrei Serjantov, Roger Dingledine, and Paul Syverson. From a trickle to a flood: Active attacks on several mix types. In Fabien Petitcolas, editor, *Proceedings of Information Hiding Workshop (IH 2002)*, pages 36–52. Springer-Verlag, LNCS 2578, October 2002.

[98] Andrei Serjantov and Peter Sewell. Passive attack analysis for connection-based anonymity systems. In Einar Snekkenes and Dieter Gollmann, editors, *Computer Security – ESORICS 2003*, pages 116–131. Springer-Verlag, LNCS 2808, October 2003.

[99] Adi Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, November 1979.

[100] Rob Sherwood, Bobby Bhattacharjee, and Aravind Srinivasan. P5: A protocol for scalable anonymous communication. In *Proceedings of the 2002 IEEE Symposium on Security and Privacy (S&P'02)*, pages 58–70. IEEE Computer Society, May 2002.

[101] Clay Shields and Brian Neil Levine. A protocol for anonymous communication over the Internet. In *Proceedings of the 7th ACM Conference on Computer and Communications Security*, pages 33–42, Athens, Greece, November 2000.

[102] Vitaly Shmatikov and Ming-Hsui Wang. Timing analysis in low-latency mix networks: Attacks and defenses. In Dieter Gollmann, Jan Meier, and Andrei

Sabelfeld, editors, *Computer Security – ESORICS 2006*, pages 18–33. Springer-Verlag, LNCS 4189, September 2006.

[103] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, and Hari Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In *SIGCOMM '01: Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 149–160, San Diego, California, United States, August 2001. ACM Press.

[104] Paul Syverson, Gene Tsudik, Michael Reed, and Carl Landwehr. Towards an Analysis of Onion Routing Security. In H. Federrath, editor, *Designing Privacy Enhancing Technologies: Workshop on Design Issue in Anonymity and Unobservability*, pages 96–114. Springer-Verlag, LNCS 2009, July 2000.

[105] Parisa Tabriz and Nikita Borisov. Breaking the collusion detection mechanism of morphmix. In George Danezis and Philippe Golle, editors, *Proceedings of the Sixth Workshop on Privacy Enhancing Technologies (PET 2006)*, pages 368–384, Cambridge, UK, June 2006. Springer-Verlag, LNCS 4258.

[106] Michael Waidner. Unconditional sender and recipient untraceability in spite of active attacks. In *Proceedings of EUROCRYPT 1989*, pages 302–319, Houthalen, Belgium, April 1990. Springer-Verlag, LNCS 434.

[107] Michael Waidner and Birgit Pfitzmann. The dining cryptographers in the disco: Unconditional sender and recipient untraceability with computationally secure servicability (abstract). In *Proceedings of EUROCRYPT 1989*. Springer-Verlag, LNCS 434, April 1990.

[108] Marc Waldman and David Mazières. Tangler: a censorship-resistant publishing system based on document entanglements. In *Proceedings of the 8th ACM Conference on Computer and Communications Security (CCS '01)*, pages 126–135, Philadelphia, PA, USA, November 2001.

[109] Marc Waldman, Aviel D. Rubin, and Lorrie Faith Cranor. Publius: A robust, tamper-evident, censorship-resistant, web publishing system. In *Proceedings of the 9th USENIX Security Symposium*, pages 59–72, August 2000.

[110] Elizabeth Wolff. Sneakers with seekers: High-tech 'track' shoes locate wearers. *New York Post*, November 2006. Retrieved June 30, 2007, from `http://www.nypost.com/seven/11262006/news/worldnews/sneakers_with_seekers_worldnews_elizabeth_wolff.htm`.

[111] Matthew Wright, Micah Adler, Brian Neil Levine, and Clay Shields. An analysis of the degradation of anonymous protocols. In Clifford Neuman, editor, *Proceedings of the ISOC Network and Distributed Security Symposium - NDSS '02*, San Diego, CA, US, February 2002. IEEE CS.

[112] Matthew Wright, Micah Adler, Brian Neil Levine, and Clay Shields. Defending anonymous communication against passive logging attacks. In *Proceedings of the 2003 IEEE Symposium on Security and Privacy (S&P'03)*, pages 28–41. IEEE Computer Society, May 2003.

[113] Matthew K. Wright, Micah Adler, Brian Neil Levine, and Clay Shields. The predecessor attack: An analysis of a threat to anonymous communications systems. *ACM Trans. Inf. Syst. Secur.*, 7(4):489–522, November 2004. A preliminary version of this paper appeared in [111].

[114] Ben Y. Zhao, Ling Huang, Jeremy Stribling, Sean C. Rhea, Anthony D. Joseph, and John Kubiatowicz. Tapestry: A resilient global-scale overlay for service deployment. *IEEE Journal on Selected Areas in Communications*, 22(1):41–53, 2004.

[115] Ye Zhu, Xinwen Fu, Bryan Graham, Riccardo Bettati, and Wei Zhao. On flow correlation attacks and countermeasures in mix networks. In David Martin and Andrei Serjantov, editors, *Proceedings of Privacy Enhancing Technologies workshop (PET 2004)*. Springer-Verlag, LNCS 3424, May 2004.