Introducing Traffic Analysis

Attacks, Defences and Public Policy Issues... (Invited Talk)

George Danezis

K.U. Leuven, ESAT/COSIC,
Kasteelpark Arenberg 10,
B-3001 Leuven-Heverlee, Belgium.
George.Danezis@esat.kuleuven.be

Abstract. A lot of traditional computer security has focused on protecting the content of communications by insuring confidentiality, integrity or availability. Yet the meta data associated with it – the sender, the receiver, the time and length of messages – also contains important information in itself. It can also be used to quickly select targets for further surveillance, and extract information about communications content. Such traffic analysis techniques have been used in the closed military communities for a while but their systematic study is an emerging field in the open security community. This talk will present an overview of traffic analysis techniques, and how they can be used to extract data from 'secure' systems.

> "*Traffic analysis, not cryptanalysis, is the backbone of communications intelligence.*"
>
> — Susan Landau and Whitfield Diffie.

1 Introduction

The field of computer security was first studied, often in secretive organizations, to guarantee properties of interest to the military. Since then the open research community has made astounding advances, focusing more and more on the security needs of commercial circles and, since the advent of computers and networks in the home, private individuals and civil society. Still, there is a field, or better described as a set of tools and techniques that are largely underrepresented in the open security research community: the field of *traffic analysis*. While a rich literature exists about securing the confidentiality, integrity and availability of communication content, very little has been done to look at the information leaked, and minimizing this leak, from communication *traffic data*. Traffic data comprises the time and duration of a communication, the detailed shape of the communication streams, the identities of the parties communicating, and their location. The knowledge of what 'typical' communication patterns might look like can also be used to infer information about an observed communication.

The civilian infrastructures, on which state and economic actors are increasingly reliant, is more and more vulnerable to such attacks: wireless and GSM telephony are replacing traditional systems, routing is transparent and protocols are overlayed over others – giving the attackers plenty of opportunity to observe, and take advantage of such traffic data. Concretely attackers can make use of this information to gather strategic intelligence, or to attack particular security protocols, and violate traditional security properties.

In this short introduction we shall highlight the key issues around traffic analysis. We shall start with its military roots and present defenses the military have used against it. Then we shall provide an overview of the research literature on attacks and defenses in contemporary networks. Finally we shall discuss some policy issues relating to the retention of traffic data.

2 Military Roots

Traffic analysis is a key part of signal intelligence and electronic warfare. Michael Hermann, who has served as chair of the UK Joint Intelligence Committee, in his book 'Intelligence Power in Peace and War' [16] describes the value of extracting data from non-textual (to be understood as 'not content') sources:

> These non-textual techniques can establish targets' locations, order-of-battle and movement. Even when messages are not being deciphered, traffic analysis of the target's C3I system and its patterns of behavior provides indications of his intentions and states of mind, in rather the same way as a neurologist develops insights about a silent patient by studying EEG traces from the brain.

Traffic analysis was used in military circles even before the invention of wireless communications. Anderson in his book [3] mentions that in the trench warfare of World War I, the earth returns of the telegraph communication of the enemy was used to extract information up to a few hundred yards away from the transmitting station. Traffic analysis though became an extremely potent source of intelligence when wireless communication became popular, particularly in naval and air operations. Ships at see had to balance the value of communicating against the threat of being detected via direction finding if they transmit. When transmitting strict standards, governing call-signs and communication, had to be adhered too in order to minimize the information that traffic analysis could provide.

Another example of traffic analysis providing valuable intelligence (by Herman [16]) is the reconstruction of the structure of the network structure of the German Air Force radio in 1941 by the British, confirming that a unit was composed of nine and not twelve planes. This allowed a more accurate estimate of the total strength of their opponent. Identification of radio equipment can also be used to detect accurate movements of units: each transmitter has characteristics such as the unintentional frequency modulations, the shape of the transmitter turn-on signal transient, the precise center of frequency modulation, etc that

provide a fingerprint, that can be detected and used to track the device (Similar techniques can be used to identify GSM phones [3]). Back in World War Two radio operators became vary skilled at recognizing the 'hand' of other operators, i.e. the characteristic way in which they type Morse code, which in turn was used as a crude unit identification method (until operators are swapped around!).

Why is traffic analysis so valuable? It provides lower quality information compared with cryptanalysis, but it is both easier and cheaper to extract and process. It is easier because ciphers need considerable effort to break (when they break at all). It is cheaper because traffic data can be automatically collected and processed to provide high level intelligence. Computers can clear traffic data and map out structures and locations, while a skilled human operator needs to listen to every radio transmission (often in a foreign language) to extract intelligence. For this reason traffic analysis is often used to perform 'target selection' for further intelligence gathering (such as more intensive and expensive surveillance), jamming or destruction. Given the enormous amount of communication and information in public networks we can expect these 'economics of surveillance' to be ever more relevant and applicable.

Sigint is an arms race, and many *low probability of intercept and position fix* communication methods have been devised by the military to minimize exposure to traffic analysis and jamming (a key reference here is Anderson [3]).Their principles are rather simple: scanning many frequencies can only be done at some maximal rate and a lot of power is necessary to jam a wide part of the frequency spectrum. Therefore the first technique used to evade interception, and foil jamming was *frequency hopping*, now also used in commercial GSM communications. Alice and Bob share a key that determines for each time period the frequency at which they will transmit. Eve on the other hand does not know the key and has to observe or jam the whole chunk of the frequency spectrum that may be used. In practice hopping is cheap and easy to implement, makes it difficult to jam the signal (given that the hop frequency if high enough), but is not very good at hiding the fact that communication is taking place. It is used for tactical battlefield communications, where the adversary is unlikely to have very large jammers at hand.

Direct sequence spread spectrum transforms a high power low bandwidth signal into a high bandwidth low power signal, using a key that has to be shared between Alice and Bob. It is easy for them to extract the signal back, using their key, but an adversary will have to try to extract it from the noise, a difficult task given its low power (that is often under the noise floor). DSSS has also inspired commercial communications and is now used in ADSL and cable modems as CDMA. Its key problem is synchronization, and the availability of a reference signal (like GPS) is of great help when implementing such systems.

The final technique in the arsenal against interception is *burst communication*. The key idea behind these is that the communication is done in a very short burst to minimize the probability the adversary is looking at the particular frequency at the time. A cute variant of this is meteor scatter communications, that use the ionization trail of small meteorites hitting the atmosphere to bounce

transmission between special forces troops in the field and a base station. Meteor scatter can also be used in civilian life when low bandwidth, high latency but very low cost and high availability communications are required.

3 Contemporary Computer and Communications Security

The Internet is no open war, yet there is a lot of potential for conflict in it. We shall see how traffic analysis techniques can be used to attack secured systems, extract potentially useful information, and be used to censor (the equivalent of jamming) or abuse and spam (the equivalent of deception) systems. We shall also outline the key defense strategies one can use on the Internet to foil these attacks – and draw the parallels but also differences with the military world.

The key difference to keep in mind when studying civilian traffic analysis research is that the attackers have fewer means. It is not military powers, with large budgets and the ability to intercept most communications that worry us, but it is commercial entities, local governments, law enforcement, criminal organizations but also terrorist networks that have become the adversary. For that reason research has focused on attacks and solutions that can be deployed at low cost, and provide tangible tactical benefits (a pass phrase, a record of web accesses, ...). Yet lately some work is developing on how traffic analysis can be of use to law enforcement, but also how one can evade from routine surveillance, which integrate a more strategic outlook.

So what can we do if we are not allowed to look at the plaintext contents?

3.1 The Traffic Analysis of SSH

The secure shell protocol allows users to log in remote terminals in a secure fashion. It does this by performing authentication using a pass-phrase and a public keyring, and subsequently encrypts all information transmitted or received, guaranteeing its confidentiality and integrity. One would think that any subsequent password entry (that might be required to log in to further remote services), over an SSH connection, should be safe. Song et al. [29] show that there is a lot of information still leaking. In interactive mode SSH transmits every key stroke as a packet. The timing between the key strokes can be used to trivially reveal information about the password lengths. More advanced techniques, using hidden Markov models, can be used to extract further information from inter-packet timing and lower the entropy of the passwords, to make guessing them easier. Some further details include the extraction of a user's password using another user to build a profile, showing that there are similarities that can be exploited between users.

The information one can extract using another user's profile link in with Monrose and Rubin's [23] research on identifying and authenticating users using keystroke dynamics. Although their focus was more on biometrics and authentication their results have a clear relevance to the traffic analysis of SSH. They

show that there is enough variability in typing patterns between users to be able to identity them, particularly after a long sequence has been observed. As a result not only the content of your communications may be leaked but also your identity despite using SSH.

3.2 The Traffic Analysis of SSL

The Secure Socket Layer (SSL, also known as TLS for Transport Layer Security) was introduced primarily to provide private web access. HTTP requests and replies are encrypted and authenticated between clients and servers, to prevent information from leaking out. Yet there is plenty of research [9, 6, 31, 2, 17] to suggest that information is leaking out of this shell.

The key weaknesses come down to the shape of traffic that is inadequately padded and concealed. Browsers request resources, often HTML pages, that are also associated with additional resources (images, stylesheets, ...). These are downloaded through an encrypted link, yet their size is apparent to an observer, and can be used to infer which pages are accessed (the difference between accessing a report on two different companies might leak information if you work in an investment bank). There are many variants of this attack: some attempt to build a profile of the web-site pages and guess for that which pages are being accessed while others use these techniques to beat naive anonymizing SSL proxies. In the later case the attacker has access to the cleartext input streams and he tries to match them to encrypted connections made to the proxy.

Note that latent structure and contextual knowledge are again of great use to extract information from traffic analysis: in Danezis [9] it is assumed that users will mostly follow links between different web resources. A hidden Markov model is then used to trace the most likely browsing paths a user may have taken given only the lengths of the resources that can be observed. This provides much faster and more reliable results than considering users that browse at random, or web-sites that have no structure at all.

3.3 Web Privacy

Can a remote web server, you are accessing, tell if you have also been browsing another site? If you were looking at a competitor maybe giving you a better price might be in order!

Felten et al. [13] show that it is possible to use the caching features of modern web browsers to infer information about the web-sites that they have been previously browsing. The key intuition is that recently accessed resources are cached, and therefore will load much more quickly than if they had to be downloaded from the remote site. Therefore by embedding in a served page some foreign resources, the attacker's web-server can perform some timing measurements, and infer your previous browsing patterns.

Note that this attack can be performed even if the communication medium is anonymous and unlinkable. Most anonymization techniques work at the network layer, making it difficult to observe network identities, but perform only minimal

filtering in higher layers. Being forced to do away with caching would also be a major problem for anonymous communication designers since any efficiency improvement has to be used to make the, already slow, browsing more usable.

3.4 Network Device Identification and Mapping

Can you tell if two different addresses on the Internet are in fact the same physical computer? Kohno et al. at CAIDA [20] have devised a technique that allows an attacker to determine if two apparently different machines are the same device. They note that the clock skew, the amount by which the clock drifts per unit of time, is characteristic of the hardware, and the physical conditions in which the crystal is maintained (heat, light, etc). Therefore if the clock drift of the remote machines seems to match for a long time, it is very likely that the machine is in fact the same. The technique they use is resistant to latency, and can be applied remotely if the target machine implements NTP, SNMP or a web server that echos the time. The technique can be used in forensics to detect target machines, but it can also be used by hackers to detect if they are in a vitalized honey-pot machine, and to determine if two web-sites are hosted on the same consolidated server.

The opposite question is often of interest. Given two connection originating from the same network address, have they actually been initiated by one or multiple machines? This is of particular relevance to count the number of machines behind NAT (Network Address Translation) gateways and firewalls. Bellovin [4] noted that the TCP/IP stack of many operating systems provides a host specific signature that can be detected, and used to estimate the number of hosts behind a gateway. To be exact the IPID field, used as a unique number for each IP packet, is in the windows operating system a simple counter that is incremented every time a packet is transmitted. By plotting the IPID packets over time, and fitting lines through the graph, one can estimate the number of unique Windows hosts.

Finally a lot of network mapping techniques have been introduced in the applied security world, and included in tools such as *nmap* [14]. The key operation that such tools perform is scanning for network hosts, open network ports on hosts, and identifying the operating systems and services running on them to assess whether they might be vulnerable to attack. The degree of sophistication of these tools has increased as more and more people started using network intrusion detection tools, such as the open source snort [32], to detect them. Nmap now can be configured to detect hosts and open ports using a variety of techniques including straight forward ping, TCP connect, TCP syn packet, but also indirect scans. For example the FTP protocol allows the client to specify to the server that it should connect to a third machine. The client can therefore use this feature to scan a third host by requesting the server the open connections to the remote ports, and observing the type of failure that occurs. The full nmap documentation is well worth a read [15].

3.5 Detecting Stepping Stones

A lot of work has been done by the intrusion detection community to establish if a host is being used as an attack platform [34, 7]. The usual scenario involves a firewall that sees incoming and outgoing connection, and tries to establish if a pair of them may be carrying the same stream. This might mean that the internal machine is compromised and used to attack another host, i.e. it is a stepping stone for the attacker to hide their identity.

The two main classes of techniques for detecting stepping stones are *passive*, where the firewall only observes the streams, and *active*, where the stream of data is modulated (often called watermarked). Since an adversary is controlling the content of the stream, and maybe encrypting it, both types of detection rely on traffic data, usually the correlation between packet inter arrival times, to match incoming and outgoing streams. The family of traffic analysis techniques that arise are similar with those used to attack anonymous communication channels.

The key result in this area is that if the maximum latency of the communication is bounded there is no way of escaping detection in the long run. This result is of course tied to a particular model (the adversary can match packet for packet, which is not obvious if the streams are encrypted under different keys or mixed with other streams), and cover channels out of its scope may prove it wrong and escape detection. Note that arbitrary active detectors are extremely difficult (maybe even impossible) to defeat.

4 Exploiting Location Data

Wireless communication equipment is often leaking location data to third parties, or wireless operators. The extent to which these can be used to degrade security properties is still to be seen but some experiments have already been performed, and their results may be a precursor to a rich set of attacks to come.

Escudero-Pascual [26] describes an experiment he set up at the 'Hacker's at Large' (HAL) summer camp. The camp had multiple wireless LAN access points, that recoded the wireless MAC address of users that were using them. This provided a time-map of user's movements throughout the event, including clues about which talks they were attending (the access points were related to the venues). Even more striking were the inferences that could be drawn about the relationship between users: random pairs of users would expect to have a low probability of using the same access point at any time. Furthermore access point usage between them over time should be uncorrelated. As a result any correlation between two users that is above average, is indicative of a relationship between the users, i.e. they are consistently moving together at the same time around the camp.

Intel research at Cambridge, also designed a similar experiment. Members of staff were issued with bluetooth devices that would record when another transmitting bluetooth device was in range. The idea was to measure the ambient bluetooth activity, to tune ad-hoc routing protocols for real world conditions, but

also to establish how often a random pair of devices meet to establish how the ad-hoc communication infrastructure could be used for two way communications. To the surprise of the researchers analyzing the data, the devices of two members of staff were found to be meeting each other rather often at night – which led them to draw conclusions about their, otherwise undisclosed, relationship.

This is well in line with evidence gathered by the MIT reality mining project [1]. The project distributed about a hundred mobile phones to students and staff of the Media Lab, under the condition that all their traffic data (GSM, bluetooth and location data) would be used for analysis. The users were also asked to fill in forms about themselves and who they consider to be their friends or colleagues. The traffic data and questionnaires were then used to build classifiers: it tuned out that calling or being with someone at 8pm on a Saturday night is a very good indicator of friendship.

They also uncovered location signatures that could differentiate a student from a member of staff. What is even more impressive is that they did not use the physical locations to draw inferences, but instead the frequency at which they are at places designated as 'work' or 'home'. Students tended to have a more uncertain schedule, while members of staff were much more predictable in their habits. This of course led to research about the amount on entropy that location data provides, and as expected for some individuals given a set of locations they are at some moment it is possible to predict with high probability their next moves and locations.

So the evidence from these preliminary studies is suggesting that whatever the wireless medium used, mobile phone, wireless LAN or bluetooth, sensitive information about your identity, your relations to others and your intentions can be inferred merely though traffic analysis.

5 Extracting High Level Intelligence

Contemporary sociology models groups of individuals, not as a mass or a fluid, but in terms of their positions within a 'social network'. The controversial basis for a lot of this research is that the position of an agent in the social network is in many ways more characteristic of them than any of their individual attributes. This position determines their status, but also their capacity to mobilize social resources and act (social capital). This position can also be determined via traffic analysis, yielding a map of the social network, and the position of each actor within it!

Social network Analysis [35], and experimental studies, has recently gained popularity and led to interesting results, that are of use to traffic analysis, but also more generally network engineering. It was first noted by Milgram [33] that typical social networks present a 'small world' property, in that they have a low diameter (experimentally determined to be about 6 hops between any two members) and to be efficiently navigable. In other words there are short paths, i.e. intermediaries between you an anyone else in the world, and you can find them efficiently (think of using hints from location and profession). This work has

been used to build efficient peer-to-peer networks, but so far has been underused in security and trust analysis. Another key finding is that 'weak links' – people you do not know all that well – are instrumental in helping you with activities that are not common but still very important. A well studied example is finding a job, where people using 'far links' are on average more successful, than those who limit themselves to their local contacts.

The first mathematical studies [27] of social networks, or power law networks as they were described because of the degree distribution of their edges, tell us a lot about their resilience to failure. It turns out that they are extremely resistant to random node failures, meaning that they stay connected and maintain a low diameter even when many random nodes have been removed. On the other hand such networks are very sensitive to the targeted removal of the nodes with high degree. After a few nodes have been removed the network will become disconnected, and the diameter increases substantially well before that. An equally effective attack is for an adversary to remove nodes according to their 'betweeness', i.e. how many other nodes in the network they connect. Traffic analysis can be used to select the appropriate targets to maximize communication degradation and disruption.

Recent research by Nagaraja et al. [24] tries to find strategies for a peer-to-peer network of nodes to resist such node deletion attacks. The intuition behind their strategies is that nodes connect to other random nodes in order to get resilience, while connecting according to a power law strategy to get efficient routing. When under attack the network regenerates links to maximize fault tolerance. When things are calm it reconfigures itself to be efficient.

Social network analysis starts being used for criminal intelligence [30, 19]. Investigators try to map, often using traffic analysis techniques on telephone or network traffic and location data, criminal organizations. This can be done to select targets for more intensive surveillance, but also to select appropriate targets for arrest and prosecution. Often these arrests are aim to maximally disrupt the organization targeted. In this case it is not always appropriate to arrest the most central, or the most well connected member – this often merely serves as a promotion opportunity for smaller crooks to take up the position. It was found to be more effective to instead arrest the 'specialists', i.e. those people in the organization that have a unique position or skills, that others would find difficult to fill. Examples include those who can forge papers, or crooked customs officials.

On the other hand traffic analysis inspired techniques can be used to protect systems and build trust. Advogato [21] is a social network based system, that provides a community for free software developers. The fact that they introduce each other allows the system to establish whether an author is likely to be a spammer, and filter their messages out. Similarly google's PageRank [25] uses techniques that are very similar to web-page and social network profiling – in that it considers pages that are more central in the network (with more links pointing to them) as more authoritative. Techniques have also been devised [18]

to automatically detect and extract web communities. Their results can be used both to assist or attack users.

6 Resisting Traffic Analysis on the Internet

A relatively old, but only recently mainstream, sub-area of computer security research is concerned with 'anonymous communications' and more generally communications that do not leak any residual information from their meta data. The field was started by David Chaum [8], introducing the mix as a basic building block for anonymity, and has continued since, adapting the techniques to provide private email communications and more recently web-browsing. A thorough overview of the field and key results is available in [10, 28]. Fielded anonymous communication systems, that are the direct products of 20 years of research, include Mixmaster [22] and Mixminion [11] for email, and JAP [5] and Tor [12] for web-browsing. They all increase the latency of communication and its cost in terms of traffic volumes.

A range of traffic analysis attacks have been used to degrade the security of anonymous communications networks. Long term intersection attacks (also referred to as disclosure attacks) reply on long term observations of input and output messages to detect communicating parties. Stream traffic analysis has been used to trace web requests and replies through low-latency networks. Finally the attacker can infiltrate the network or try to influence the way in which honest nodes chose paths to anonymize their traffic. Lately attacks have focused on weaker adversaries, and it has been shown that some forms of traffic analysis can be performed even without any access to the actual data streams to be traced. So little importance has been payed to securing public networks against traffic analysis that the information leaked can be detected and abused even far away from its source...

7 Instead of Conclusions...

Our understanding of the threat that traffic analysis attacks represent on public networks is still fragmented, and research in this growing field is still very active. The results we have presented should act as a warning call against ignoring this threat: traffic analysis not only can be used to collect more information in general but can also be used to bypass security mechanisms in place.

Our study of these techniques should also have some impact on public policy matters. The most relevant of these is the current debate on traffic data retention in the E.U. – plans to store all traffic data for a long time to facilitate law-enforcement investigations. Policy makers must be informed of the wealth of information that could be extracted from such data about every aspect of the networked society. Storing these, in an easily accessible manner, represents a systemic vulnerability that cannot be overstated enough. Allowing even anonymized profiles to be extracted from such data would greatly facilitate privacy violations and routine surveillance. Traffic analysis resistance is a public good – the more an

attacker knows about the habits of your neigbours the more they can tell about you! Similarly our study of jamming resistant communications can shed light on potential means by which criminals might communicate, 'under the radar' of law enforcement.

References

1. Mit media lab: Reality mining. Massachusetts Institute of Technology Media Lab.
2. Heyning Cheng And. Traffic analysis of ssl encrypted web browsing.
3. Ross Anderson. *Security engineering*. Wiley, 2001.
4. Steven M. Bellovin. A technique for counting natted hosts. In *Internet Measurement Workshop*, pages 267–272. ACM, 2002.
5. Oliver Berthold, Hannes Federrath, and Stefan Köpsell. Web MIXes: A system for anonymous and unobservable Internet access. In H. Federrath, editor, *Designing Privacy Enhancing Technologies*, volume 2009 of *LNCS*, pages 115–129. Springer-Verlag, July 2000.
6. George Dean Bissias, Marc Liberatore, , and Brian Neil Levine. Privacy vulnerabilities in encrypted HTTP streams. In *5th Workshop on Privacy Enhancing Technologies (PET2005)*, 2005.
7. Avrim Blum, Dawn Xiaodong Song, and Shobha Venkataraman. Detection of interactive stepping stones: Algorithms and confidence bounds. In Erland Jonsson, Alfonso Valdes, and Magnus Almgren, editors, *RAID*, volume 3224 of *Lecture Notes in Computer Science*, pages 258–277. Springer, 2004.
8. David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–88, February 1981.
9. George danezis. Traffic analysis of the http protocol over tls. `http://www.cl.cam.ac.uk/~gd216/TLSanon.pdf`.
10. George Danezis. *Better Anonymous Communications*. PhD thesis, University of Cambridge, 2004.
11. George Danezis, Roger Dingledine, and Nick Mathewson. Mixminion: Design of a Type III Anonymous Remailer Protocol. In *IEEE Symposium on Security and Privacy*, Berkeley, CA, 11-14 May 2003.
12. Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th USENIX Security Symposium*, August 2004.
13. Edward W. Felten and Michael A. Schneider. Timing attacks on web privacy. In *ACM Conference on Computer and Communications Security*, pages 25–32, 2000.
14. Fyodor. Nmap – free security scanner for network exploitation and security audit. `http://www.insecure.org/nmap/`.
15. Fyodor. Nmap manual. `http://www.insecure.org/nmap/man/`.
16. Michael Herman. *Intelligence Power in Peace and War*. Cambridge University Press, 1996.
17. Andrew Hintz. Fingerprinting websites using traffic analysis. In Roger Dingledine and Paul F. Syverson, editors, *Privacy Enhancing Technologies*, volume 2482 of *Lecture Notes in Computer Science*, pages 171–178. Springer, 2002.
18. Jon M. Kleinberg. Hubs, authorities, and communities. *ACM Comput. Surv.*, 31(4es):5, 1999.
19. Peter Klerks. The network paradigm applied to criminal organisations. In *Connections 24(3)*, 2001.

20. Tadayoshi Kohno, Andre Broido, and Kimberly C. Claffy. Remote physical device fingerprinting. In *IEEE Symposium on Security and Privacy*, pages 211–225. IEEE Computer Society, 2005.
21. Raphael L. Levien. *Attack resistant trust metrics*. PhD thesis, University of California at Berkeley, 1995. Draft Thesis.
22. U. Moeller, L. Cottrell, P. Palfrader, and L. Sassaman. Mixmaster protocol version 2. Technical report, Network Working Group, May 25 2004. Internet-Draft.
23. Fabian Monrose, Michael K. Reiter, and Susanne Wetzel. Password hardening based on keystroke dynamics. In *ACM Conference on Computer and Communications Security*, pages 73–82, 1999.
24. Shishir Nagaraja and Ross Anderson. The topology of covert conflict. Technical report, University of Cambridge, Computer laboratory, 2005.
25. Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.
26. Alberto Escudero Pascual. *Anonymous Untraceable Communications: Location privacy in mobile internetworking*. PhD thesis, Royal Institute of Technology - KTH / IMIT, 2001.
27. William J. Reed. A brief introduction to scale-free networks. Technical report, Department of Mathematics and Statistics, University of Victoria, 2004.
28. Andrei Serjantov. *On the anonymity of anonymity systems*. PhD thesis, University of Cambridge, 2004.
29. Dawn Xiaodong Song, David Wagner, and Xuqing Tian. Timing analysis of keystrokes and timing attacks on SSH. In *Tenth USENIX Security Symposium*, 2001.
30. Malcom K Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. In *Social Networks (13)*, 1991.
31. Qixiang Sun, Daniel R. Simon, Yi-Min Wang, Wilf Russell, Venkata N. Padmanabhan, and Lili Qiu. Statistical identification of encrypted web browsing traffic. In *IEEE Symposium on Security and Privacy*, pages 19–30, 2002.
32. Snort team. Snort. `http://www.snort.org/`.
33. J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32(425), 1969.
34. Xinyuan Wang and Douglas S. Reeves. Robust correlation of encrypted attack traffic through stepping stones by manipulation of interpacket delays. In Sushil Jajodia, Vijayalakshmi Atluri, and Trent Jaeger, editors, *ACM Conference on Computer and Communications Security*, pages 20–29. ACM, 2003.
35. Stanley Wasserman, Katherine Faust, Dawn Iacobucci, and Mark Granovetter. *Social Network Analysis : Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge University Press, 1994.