Traffic Analysis Versus Private Browsing on the Web

George Tsikkos
School Of Electronics and Computer Science,
University Of Southampton
gt304@ecs.soton.ac.uk

1. Introduction

The literature search presents a summary of the research performed on the issue of traffic analysis and some of the ways found to resist it. Sections two and three give an overview of the problem by presenting the key aspects of these attacks and how they can be performed. A summary of the key solutions is then presented in section four followed by a brief description of topics related to traffic analysis and Internet privacy.

Detailed descriptions of the solutions are outside the scope of the literature search as these will be described in more detail in the technical report.

2. Overview

Traffic analysis is a fairly recent type of attacks which attempts to create breaches in users' confidentiality. It is a type of Man-In-The-Middle attack that aims at obtaining non-content critical information such as users' identities, bandwidth consumption, length, time and duration of messages exchanged.

Traffic analysis is opposed to cryptanalysis where the goal is to decode messages sent and received and read their content. These two types of attacks can work together in order to choose the victim, by traffic analysis, and read their content using cryptanalysis.

The roots of traffic analysis can be traced back to the military where it was used to retrieve mission critical information by obtaining not the actual content of messages but rather the origins, targets and other non content data, as described by Michael Hermann [1].

Over the years a number of such attacks have been reported and some of them are summarized in a paper of John D. Howard in [2].

3. Traffic Analysis Attacks

One of the key researchers in the field of traffic analysis, Raymond [3] gives a good overview of the problem, describes the main methods of attacks and how they can be performed, along with a comparison of the defence techniques at the time.

Since traffic analysis has only recently appeared on the Internet, the various protocols and tools used for secure interactions have as their main concern defending against cryptanalysis without really "caring" about giving away metadata. Some of the attacks on important "secure" protocols are described below.

3.1 Attacks on SSH

SSH (Secure Shell) is a protocol that enables users to join remote terminals by authenticating using a password and a public keyring. Then all the information transmitted and received is encrypted to ensure confidentiality. Song et al. [4] showed that by analyzing the time between keystrokes the length of passwords can be revealed. In another paper by Monrose et al. [5] proved that passwords can be obtained by observing users' typing patterns. This means that the actual content of users' communications can be extracted as well as their identity even under SSH.

3.2 Attacks on SSL and TLS

SSL (Secure Socket Layer) and TLS (Transport Layer Security) are protocols designed to ensure secure communications on the Web over HTTP. George Danezis [6], another key researcher in the field, as well as Sun et al. [7] and Hintz [8] have proved that even these protocols leak information (such as traffic shape) which allows attackers to "guess" the websites visited as well as the browsing path taken to get to them.

3.3 Attacks on Web Privacy

Caching techniques, used by web browsers to improve efficiency and speed in browsing, might also give away important information which enable "observers" to retrieve previously visited pages by measuring the time taken to load up a page. Felten et al. [9] has shown that the above can be possible with the current structure of web browsers.

3.4 Attacks on Peer-to-Peer Systems

Clauthia and Chatzikokolakis [10] have presented a comparison of a number of anonymous P2P file sharing systems and showed that attackers are able to determine a user's IP address by linking it to the pseudo address, used by the file sharing applications.

4. Defending Against Traffic Analysis

4.1 Analysis Detection

In recent years researchers have designed techniques that attempt to detect whether traffic analysis is taking place in a network and which machines it uses to perform the attack. This detection is performed by observing incoming and outgoing connections and checking whether the stream they carry is the same. If it is, then, it is possible that the machine is used by maliciously to attack another machine. These techniques can either be passive, only observing streams, or active, where the stream can be modulated. Blum et al. [11] and Wang et al. [12] have presented some of these detection mechanisms in their papers.

4.2 Anonymous Browsing

Many cryptographers have attempted to tackle the issue of anonymous communications, even before traffic analysis appeared on the Web. Some of the main solutions are presented below.

In 1981 David L. Chaum, the father of anonymous communications, presented one of the key ideas in the field called the Mix [13] used to avoid traceability in emails. In later years other techniques were developed: some using MIXes and others using different mechanisms. The main solutions for anonymous communications are Onion routing [14-16], Crowds [17], DC-Net [18], WebMIXes [19] and Hordes [20]. Other less famous solutions are Mixminion [21] for email anonymity and Freenet [22] for peer-to-peer.

Several of the above techniques have been used to implement applications and systems for users. Some of the popular applications are Tor [23, 24], Ants [25], Anonymizer [26] and Tarzan [27].

A lot of research has been performed to prove vulnerabilities in systems like the above. One of the best papers on this was written by Diaz et al. [28] which compared key defence mechanisms and attempted to show the level of anonymity provided by each of them. Other main papers on this are [29-33].

5. Further Reading

In recent years a fair amount of publicity has been given to the issues related to the Internet and the privacy of its users. Several organizations exist that aim at protecting users' privacy rights [34, 35].

On top of all others in recent years traffic analysis has been made legal! The European Union in 2006 has enacted a directive [36] that requires member states to "force" their communication providers to capture and retain information like telephone calls, email messages and other communications.

Even though traffic analysis can be used maliciously there is also a good side to it. Systems like Slingbox Pro, Microsoft Zune and Nike iPod [37] use traffic analysis with "pure" intentions. Other such systems include Google PageRank, credit card fraud detection as well as systems used for analysis of social networks and fight against crime and terrorism.

6. References

[1] M. Herman, *Intelligence Power in Peace and War*. Cambridge University Press, 1996.

[2] John D. Howard, "An Analysis Of Security Incidents On The Internet 1989 – 1995," PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, 1997.

[3] J.F. Raymond, "Traffic analysis: protocols, design issues, and open problems," *International workshop on Designing privacy enhancing technologies: design issues in anonymity and unobservability",* Springer-Verlag New York, 2001, pp. 10-29

[4] D. X. Song, D. Wagner, and X. Tian, "Timing analysis of keystrokes and SSH timing attacks," in *Proceedings of 10th USENIX Security Symposium*, 2001, pp. 337–352.

[5] F. Monrose, M. K. Reiter and S. Witzel, "Password hardening based on keystroke dynamics," *International Journal of Information Security*, Volume 1/Number 2, Springer Berlin, 2002, pp. 69-83

[6] G. Danezis, "Traffic Analysis of the HTTP Protocol over TLS", [Online] Available: http://research.microsoft.com/users/gdane/papers/TLSanon.pdf, 2007.

[7] Q. Sun, D. R. Simon, Y.Wang,W. Russell, V. N. Padmanabhan, and L. Qiu. "Statistical identification of encrypted web browsing traffic," in *Proceedings of IEEE Symposium on Security and Privacy*, 2002, pp. 19-30.

[8] A. Hintz, "Fingerprinting Websites Using Traffic Analysis," *Privacy Enhancing Technologies*, Volume 2482, Springer Berlin, 2003, pp. 229-233.

[9] E. Felten and M. Schneider, "Timing attacks on web privacy," in *Proceedings of the 7th ACM Conference Computer and Communications Security*, 2000, pp. 25–32.

[10] T. Clauthia, and K. Chatzikokolakis, "A Survey of Anonymous Peer-to-Peer File Sharing," *Embedded and Ubiquitous Computing*, Volume 3828, Springer Berlin, 2005, pp.744-755.

[11] A. Blum, D. X. Song, and S. Venkataraman. "Detection of interactive stepping stones: Algorithms and confidence bounds," *Recent Advances in Intrusion Detection*, Volume 3224, Springer Berlin, 2004, pp. 258–277.

[12] X. Wang and D. S. Reeves, "Robust correlation of encrypted attack traffic through stepping stones by manipulation of interpacket delays," in *Proceedings of the 10th ACM Conference Computer and Communications Security*, 2003, pp. 20–29.

[13] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, Volume 4/Number 2, February 1981, pp. 84-90.

[14] M. G. Reed, P. F. Syverson, and D. M. Goldschlag, "Anonymous connections and onion routing," *Selected Areas in Communications*, Volume 16/Number 4, IEEE, May 1998, pp. 482-494.

[15] M. G. Reed, P. F. Syverson, and D. M. Goldschlag, "Onion routing," *Communications of the ACM*, Volume 42/Number 2, ACM New York, February 1999, pp. 39-41.

[16] P. F. Syverson, G. Tsudik, M. G. Reed, and C. E. Landwehr, "Towards an analysis of onion routing security", *Designing Privacy Enhancing Technologies:Workshop on Design Issue in Anonymity and Unobservability*, Springer-Verlag, July 2000, pp. 96–114.

[17] M. Reiter and A. Rubin, "Crowds: Anonymity for web transactions". *ACM Transactions on Information and System Security*, Volume 1/Number 1, 1998, pp. 66-92.

[18] D. Chaum, "The dining cryptographers problem: Unconditional sender and recipient untraceability," *Journal of Cryptology*, Volume 1/Number 1, Springer New York, 1988, pp. 65-75.

[19] O. Berthold, H. Federrath, and S. Kopsell. "Web MIXes: A system for anonymous and unobservable Internet access", *Designing Privacy Enhancing Technologies: Workshop on Design Issue in Anonymity and Unobservability*, Springer Verlag, 2001, pp. 115–129.

[20] C. Shields and B. N. Levine, "A protocol for anonymous communication over the Internet," in *Proceedings of the 7th ACM Conference on Computer and Communications Security*, 2000, pp.33-42.

[21] G. Danezis, R. Dingledine, and N. Mathewson, "Mixminion: Design of a Type III Anonymous Remailer Protocol," in *Proceedings of the 2003 IEEE Symposium on Security and Privacy*, May 2003, pp. 2-15.

[22] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong, "Freenet: A distributed anonymous information storage and retrieval system," in *Proceedings of the International Workshop Design Issues in Anonymity and Unobservability*, Berkeley, 2001, pp. 46–66.

[23] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," in *Proceedings of the 13th USENIX Security Symposium*, 2004.

[24] The Tor Project, http://www.torproject.org/.

[25] ANts P2P, http://antsp2p.sourceforge.net/.

[26] The Anonymizer, http://www.anonymizer.com/.

[27] M. J. Freedman and R. Morris. "Tarzan: A peer-to-peer anonymizing network layer," in *Proceedings of the 9th ACM Conference on Computer and Communications Security*, 2002.

[28] C. Dıaz, S. Seys, J. Claessens, and B. Preneel, "Towards measuring anonymity," in *Proceedings of Privacy Enhancing Technologies Workshop*, 2002, pp. 54–68.

[29] M. Wright, M. Adler, B. N. Levine, and C. Shields, "An analysis of the degradation of anonymous protocols," in *Proceedingsof the Network and Distributed Security Symposium*, IEEE, February 2002.

[30] S. J. Murdoch, G. Danezis, "Low-cost traffic analysis of Tor," *Security and Privacy, 2005 IEEE Symposium*, May 2005, pp. 183-195.

[31] A. Serjantov, R. Dingledine, and P. Syverson. "From a trickle to a flood: active attacks on several mix types," in *Proceedings of Information Hiding Workshop*, 2002.

[32] O. Berthold, A. Pfitzmann, and R. Standtke, "The disadvantages of free MIX routes and how to overcome them," *Designing Privacy Enhancing Technologies: Workshop on Design Issue in Anonymity and Unobservability*, Springer-Verlag, 2000, pp. 30–45.

[33] A. Back, U. Moller, and A. Stiglic, "Traffic analysis attacks and trade-offs in anonymity providing systems," in *Proceedings of the 4th International Workshop on Information Hiding*, 2001, pp. 245-257.

[34] Privacy Rights Clearinghouse, http://www.privacyrights.org/netprivacy.htm/.

[35] Privacy.org, http://www.privacy.org/.

[36] European Union, Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC, http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l_105/l_10520060413en00540063.pdf/.

[37] T. S. Saponas, J. Lester, C. Hartung, S. Agarwal, and T. Kohno, "Devices That Tell On You: Trends in Consumer Ubiquitous Computing," in *Proceedings of the 16th USENIX Security Symposium*, 2007.

Notes:

The main references to be used in the technical report will be mainly taken from sections three and four since these are used to describe the solutions. The rest describe the problem and some guidelines for additional reading mainly on Internet privacy issues.