DIGITAL INVESTIGATION 3 (2006) 85–88

available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/diin

Digital Investigation

Advances in Onion Routing: Description and backtracing/ investigation problems

*Dario Forte**

Adj Faculty University of Milano at Crema, DFLabs, Italy

ARTICLE INFO

Article history:
Received 27 January 2005
Revised 21 April 2006
Accepted 21 April 2006

Keywords:
Onion routing
Antiforensic
Anonymizers
Privacy
Security
Network investigation
Log files analysis
Traffic analysis
Network intelligence
Digital forensics

ABSTRACT

Onion routers were born about 10 years ago as a sort of blended military/research project. The main goal of it was the avoidance of traffic analysis (TA). TA is used in part to identify the remote IP addresses that a given host seeks to contact. This technique may have various purposes, from simple statistical analysis to illegal interception. In response, to protect personnel whose communications are being monitored by hostile forces, researchers from the US Naval Research Laboratory conceived a system, dubbed "Onion Routing", that eludes the above two operations.

In this author's opinion, the original "owners" of the project lost control of it. Software is freely available that enables anyone to utilize a network of onion routers. The result is a strong evolution (in terms of "privacy") that is very difficult to manage, especially when an international investigation must be performed.

1. Onion Routing: a general description

The objective of Onion Routing is to make it completely impossible for third parties to perform traffic analysis. This goal is achieved by applying cryptographic techniques to networking. The packets transiting the chain of onion routers thus appear anonymous. Practically speaking, there is a group of onion routers distributed around the Internet, each of which has the task of encrypting the socket connections and to act in turn as a proxy. The number and/or type of managed protocols is practically unlimited.

Let's imagine we have to make an http transaction. Here's how it works:

1. The application does not connect directly to the destination web server, but rather to a socket connection with an Onion Routing proxy (at the moment called "OR Rendez Vous").
2. The Onion Routing proxy establishes a direct anonymous connection with its nearest sister. To guarantee the impossibility of interceptions, the first Onion Routing proxy makes another connection with others of its ilk to complete the chain. To avoid hijacking and man-in-the-middle phenomena, the communication between onion routers is forced. Practically speaking, each onion router is only able to identify and dialog with its adjacent kin included in the route.

* Tel.: +39 034 026 242 46; fax: +39 027 004 246 07.
E-mail address: dario.forte@dflabs.com

1742-2876/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.diin.2006.04.007

3. Each time an onion router handles a transaction, it strips away a layer of encryption with respect to the preceding hop. This means that at the end of the route the packet arrives in clear text. This is one of the first problems an investigator may encounter. Because of the encryption and because at each hop the link to the preceding routing point is literally stripped away, trace back becomes impossible. The only way to carry out an effective investigation is to implement a logging function at the proxy level as we will describe in greater detail below.
4. The encryption and transmission of data through the links of the chain are carried out randomly in such a way as to render impossible any sort of "sequence prediction". Furthermore, whenever the connection is interrupted, for any reason, all information relating to a given transaction is deleted from the rest of the chain. It is basically a sort of "no cache" system.

A graphic representation of OR packet flow is shown in Fig. 1, where "A" (TOR client) obtains a list of TOR nodes from a Directory Server "D", and then establishes a connection, via the "+" nodes to the final destinations "C." The nodes are also chosen in a random way. For more info please refer to http://tor.freehaven.net/overview.html.en.

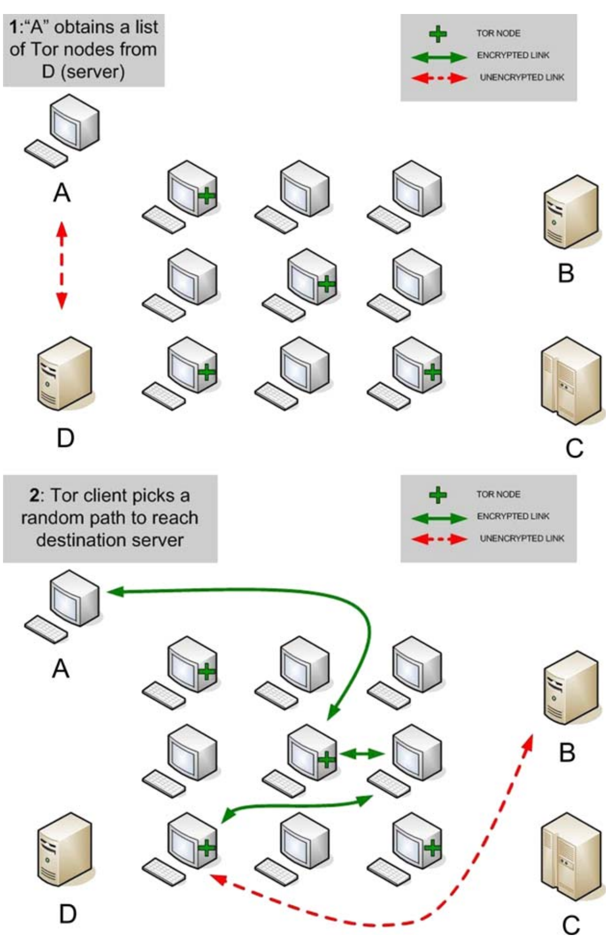

Fig. 1 – How the TOR system works. After the login onto the Rendez Vous Point, the encrypted/anonymized path is totally random.

2. How onion routers can be used (by criminals)

The latest generation of Onion Routers Chain, called TOR [http://tor.freehaven.net/], is used by criminals for 3 main purposes: Hacking, File Exchange and "Terrorism Operation Management".

Using TOR as a Hacking Facilitator. Of course such a routing can be combined with using multiple stepping stones. In both cases the investigator's work will be very difficult, all the more so because the deployment of the TOR chain is on an international scale. The version using multiple stepping stones is illustrated in Fig. 2.

If an attacker combines the use of stepping stones and TOR, he/she can obtain a very high level of "security" in hiding his/her traces. In this case, it is possible to hide the communication between compromised machines and the real attacker.

File Exchange. So far, it is not difficult to understand the danger. But the part above it is just the start. The real issue caused by the project evolution is given by the fact that TOR also makes it possible for users to hide their locations while offering various kinds of services, such as web publishing or an instant messaging server, and also a secure peer to peer, used to exchange child pornography.

"Terrorism Operation Management". Using TOR "rendezvous points," other TOR users can connect to these hidden services each without knowing the other's network identity. This hidden service functionality could allow TOR users to set up a website where people publish material without worrying about censorship. Nobody would be able to determine who was offering the site, and nobody who offered the site would know who was posting to it.

One example of such a service is *Privoxy*, a web proxy with advanced filtering capabilities for protecting privacy, filtering web page content, managing cookies, controlling access, and removing ads, banners, pop-ups and other obnoxious junk. It is based on the Internet Junkbuster. The combination of TOR and Privoxy is lethal for investigation.

3. Challenges for investigators

More than any other type of anonymizer, TOR represents a very big issue for investigators. Starting from the standpoint that, basically, it would be difficult to perform a standard digital investigation in this case, it sounds very difficult to perform a forensic examination on the nodes. However, if found, an examination of a Privoxy based machine could be very useful. Results could be, for example: configuration files, logs, cache/cookies management, etc.

Such examination may be also performed on the TOR Client. Like every Client, a TOR Client could keep interesting information, like the list of Rendez Vous Points, the protocols supported, the services used and so on. But this is not enough. A correlation (when possible) between files found on the machine and traffic generated by the chain could be necessary. However, it is not so easy. As it is now known, without the possibility to intercept the traffic or the payload, the only way to successfully complete a traceback is to make

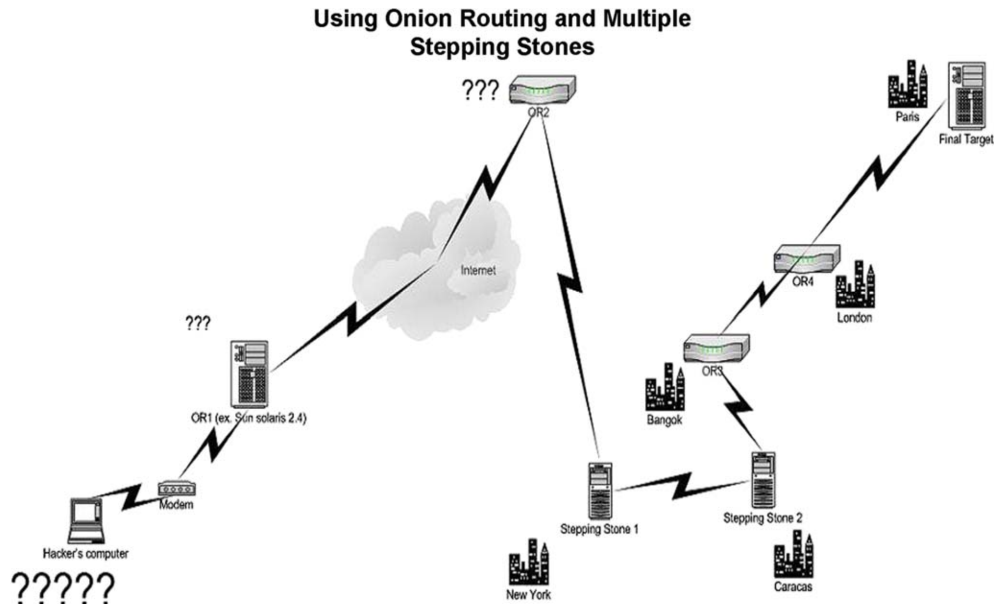

Fig. 2 – In this particular period, we are assisting to a blended use of TOR and stepping stones based attack. This scheme illustrates how it is possible to chain the two methods.

a correlation among packets and/or connections. Not being able to monitor the flow of packets, partially due to the complete lack of control over hop randomization and over the "no cache" setup of ORs, it may become extremely difficult (almost impossible) to conduct an investigation.

However, there is some experiment, recently presented during the 2005 IEEE Symposium on Security and Privacy, which explores the possibility of combating Onion Routing from the inside, for example, being part of the chain, meaning setting up a node with sniffing capabilities. Basically speaking, it is a form of traffic analysis (TA). The experiment mentioned above (Murdoch and Danezis, 2005) is theoretically feasible and is of low cost. It also had some practical implementation. However, this method has one important problem: the more the traffic that passes through TOR, the poorest analysis performances are obtained. Furthermore, as we said before, this possibility implies that the TA Machine is a trusted part of the chain, which, in this moment, is not always possible.

We should also keep in mind that, in this moment, hackers use TOR in two ways. The first is that if they own the chain they can obviously set it up so there will be no activity logging at the proxy level and also perhaps set up the ACP with some user restrictions, but certainly not regarding the protocols that can be used. This ultimately means it is extremely difficult to backtrace the evildoer who used the chain for illicit purposes. The second possibility that might raise its ugly head is if an attacker uses an OR chain and attempts to compromise a router or wages a Denial of Service before or after a specific attack. Here the routers are hit both with Denial of Service and with a bona fide attack against a specifically targeted vulnerability.

The last available chance is, just in case, exploiting the real weakness of the entire chain, which seems to be the link between the final point and the destination of the session. The second part of Fig. 1 can help that understanding.

As readers can see in the mentioned figure, there is an unencrypted link that, regardless from the path randomly chosen, could be exploited by the investigators, at least to start with the backtracing.

In fact, If the "B" computer is a web server the last TOR point could reveal (theoretically) some trace about the origin of the transaction. For this reason, the development team suggests to adopt a proxy for every single component, like a sort of anonymizer. On the other way, investigators can try to perform a backtracing.

4. Can TOR be investigated?

It is not easy. As readers can understand reading this article, there are a lot of factors, technical and "criminal" which could generate some issues during the investigation. It is the author's opinion that one of the few chances given to investigators is the analysis of the Proxy and TOR Client Machines. It is also true that traffic analysis could become more easy to perform in the future, but, in this moment, even in presence of a possible implementation, this possibility is fairly remote.

Reference

Murdoch Steven J, Danezis G. Low-cost traffic analysis of TOR. In: IEEE symposium on security and privacy; May 2005.

Further readings

Albitz Paul, Liu Cricket. DNS and BIND. 4th ed. O'Reilly; 2001.
Alhambra, daemon9. Project Loki: ICMP tunneling. Phrack Magazine(49), <http://www.phrack.org/phrack/49/P49-06>; 1996;6.

Bejtlich Richard. The TAO of network security monitoring. Addison Wesley; 2005. p. 505–17.
Carrillo Juan F, Ospina Carlos, Rangel Mauricio, Rojas Jaime A, Vergara Camilo. Covert channels sobre HTTP, <http://www.criptored.upm.es/guiateoria/gt_m142m.htm>; 2004.
Comer DE. Internetworking with TCP/IP. Prentice Hall; 1995.
Dyatlov Alex, Castro Simon. Exploitation of data streams authorized by a network access control system for arbitrary data transfers: tunneling and covert channels over the HTTP protocol, <http://www.gray-world.net/projects/papers/html/covert_paper.html>; 2003.
Forte Dario. Analyzing the difficulties in backtracing onion router traffic. The International Journal of Digital Evidence(3). Utica College, <http://www.ijde.org/archives/02_fall_art3.html>, Fall 2002;1.
Forte Dario. The art of log correlation – tools and techniques for log analysis. In: Proceedings of the ISSA conference 2004, Johannesburg, South Africa; 2004.
Forte Dario, Vetturi Michele, Zambelli Michele, Maruti Cristiano. SecSyslog: an approach to secure logging based on covert channels. In: Proceedings of the 1 SADFE conference, IEEE CS Taiwan. Gives a possible implementation of a log system at OR level.
Forte Dario, et al. Forensic computer crime investigation. In: Johnson Thomas Alfred, Johnson Thomas A, editors. Forensic science. CRC Press; November 2005. gives a more deep perspective of the problem, even from a legal and investigative standpoint.
Kaminsky Dan. Black Ops of DNS, <http://www.doxpara.com/dns_bh>; 2004.
Murdoch Steven, Lewis Stephen. Embedding covert channels into TCP/IP, <http://www.cl.cam.ac.uk/users/sjm217/papers/ih05-coverttcp.pdf7>; 2005.
Owens Mark. A discussion of covert channels and steganography, <http://www.sans.org/rr/whitepapers/covert/678.php>; 2002.
RFC 2136: dynamic updates in the domain name system. IETF, <http://www.ietf.org/rfc/rfc2136.txt>; 2002.
RFC 791: Internet protocol. IETF, <http://www.ietf.org/rfc/rfc0791.txt>; 2002.
RFC 793: transmission control protocol. IETF, <http://www.ietf.org/rfc/rfc0793.txt>; 2002.
RFC 2535: domain name system security extensions. IETF, <http://www.ietf.org/rfc/rfc2535.txt>; 2002.
Rowland Craig H. Covert channels in the TCP/IP protocol suite. First Monday, <http://www.firstmonday.org/issues/issue2_5/rowland/>; 1997.
Simple Nomad. README for the ncovert2 tool, <http://ncovert.sourceforge.net/>; 2003.
Szor Peter. The art of computer virus research and defense. Addison Wesley; 2005.
The onion routing second generation EFF page, <http://tor.eff.org/overview.html.en>; <http://tor.eff.org/tor-design.pdf>; 2005. Gives an in depth overview of the project.
U.S. Department Of Defense. Trusted computer system evaluation criteria; 1985.

Dario Forte, CFE, CISM, is adjunct Professor of Digital Forensic and Incident Response at University of Milano, Crema Research Center, where he is also responsible of the Italian chapter of the Honeynet Research Alliance. Former Police Officer, Dario is in the editorial board of advisors of Elsevier's Network Security and DI. He is also founder and CEO of DFLabs, an Italian company specialized in Incident Response and Digital Forensic (www.dflabs.com). The author acknowledges Eoghan Casey for his support during the editing/review phase of this article.