Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 3 (2011) 888–892

Procedia Computer Science

www.elsevier.com/locate/procedia

WCIT 2010

Comparison between normal and TOR-Anonymized Web Client Traffic

Tomáš Liška, Tomáš Sochor, Hana Sochorová

University of Ostrava, 30. dubna 22, Ostrava, Czech Republic

Abstract

TOR is known free tool for traffic anonymization on the Internet. The adverse aspect of using TOR is significant increase of overhead and decrease of the traffic speed. So far no study focusing the quantification of such decrease was published. TOR operation is based on "routing" the traffic through several nodes resulting in difficult and practically impossible direct calculation of the delay caused by TOR application. Therefore the study was performed to quantify the delay associated with using TOR comparing ordinary traffic. Www was used as a model of Internet traffic. The sample of 14 stable www pages and 8-10 files available via http throughout the world was chosen and the round-trip time was measured approx. 10 times both with TOR (at least in two configurations, first automatic and the second manually modified) and without TOR. For manual TOR measurement two onion routers were configured in two locations. Measurements were made in various days of week and various times to eliminate possible fluctuations due to varying traffic target www servers. The main result of the study is that the factor of round-trip time increase varies between 2 and more than 100. In case of configured TOR networks the ratio is lower but still high,1.7 and 34. This is quite rough result that is shown in details in the article. The results show that TOR is useful anonymization tool but the delay caused by its use is enormous. Providing of more exact estimations would require more measurements however.

Selection and/or peer-review under responsibility of the Guest Editor.

Keywords: Traffic anonymization; Onion routing; TOR; TOR network configuration; Www; Latency comparison;

1. Introduction – Internet anonymization

There are lots of situations when user desire to hide details about its activity in the Internet. In such cases various tools and methods for anonymization are used. TOR (The Onion Routing) belongs among the most famous tools for anonymization of traffic through the Internet. Moreover its use is free unlike in case of numerous commercial solutions (usually based on cascade mix) making TOR increasingly popular. Because of its popularity their properties (especially its behavior in practical use) were studied. This paper focuses primarily to the comparison of latency between norman and TOR-anonymized www traffic.

1.1. TOR properties

There are numerous tools for anonymization of Internet traffic. Most of them including TOR rely on concealing the IP address of the originator. The operation of TOR is possible thanks to the fact that many people throughout the world allow using their computers as TOR nodes (called "onion routers"). The anonymized traffic then travels through variety of such TOR nodes instead of direct communication of the originator with the target IP address. The

doi:10.1016/j.procs.2010.12.145

main idea of TOR anonymization consists in the fact that each TOR node has information about its predecessor and successor (if any other than final target) and obviously also information about final target but does not have originator IP adress so as to avoid backtracking of packets (in combination with encryption of packets). More detailed information about TOR can be found in [1].

Decrease in latency and/or bandwidth is an obvious drawback associated with the use of TOR and any similar tool. Significant decrease is obvious not only because adding more nodes into the path between the client and target but primarily due to encryption of traffic between TOR nodes. The main aim of the study described here was to quantify such decrease to allow the assessment of efficiency of using TOR as anonymization tool.

2. Methods – what type of traffic and which way latency decrease was measured

After some initial investigations we decided to focus our measurements on www service. The www service was chosen due to two reasons. First one was the still significant share of www in total Internet traffic (approx. 10% in 2007, see [2]). Another reason for this choice was the nature of www traffic comparing with e-mail also making significant share of Internet traffic; www communication is much more likely to require anonymization.

Parameters like latency (to be more precise "round-trip latency") are quite difficult to measure exactly because its "soft" (i.e. insufficiently exact) definition. The key problem is represented by the issue how to define start and especially the end of communication. This is particularly complicated in case of www communication where the traffic consists of downloading of multiple files consecutively initiated by the www client. Therefore our study assumed certain additional limitations. The main limitation was only to single web client software (Mozilla Firefox, see details in section 2.2) so as to eliminate extrinsic fluctuations in measured data due to different implementation of client functions.

The original idea was to select suitable set of web pages and files available via www service and to compare its downloading with TOR and without it. For better quantification of TOR influence to www traffic several variants of TOR use were suggested. The variants of TOR usage that were used for measurement are described below including brief explanation.

2.1. WWW test set

The important part of our comparison was the set of www pages and files available via www. The set was built rather in subjective-casual way with trial&error refinement. The most important factors for selecting pages and files into set (in fact two separate sets, one of www pages and the other one of files were built) were their stability during the testing period (both in availability and in the size of page or file) and the worldwide spread of the set (regarding the nature of Internet and limited availability and determinacy of geographical data in the Internet). In case of files the size exceeding 10 MB was pre-postulated. Resulting lists of 14 www pages and 10 files (only with 8 the complete set of measurements were concluded) are listed in brief in the Tables 1 and 2.

Table 1. WWW pages test set (Elements columns contains the number of all files – HTML, images, scripts – forming the page, Hops column shows the number of hops excluding ISP network of the test set, in two cases the tracing up to the the destination was unable to determine)

Item No.	URL	Size (kB)	Elements	Hops	Country
1	http://www.sydneyaustralia.com/en/	189	59	>7	Australia
2	http://english.buenosaires.com/	536	110	9	Argentina
3	https://stag1.osu.cz/wps/portal/	455	20	4	Czech Republic
4	http://santos.globo.com/index_idioma.php?idioma=2	205	101	15	Brazil
5	http://slovnik.cz/	75	N/A	2	Czech Republic
6	http://www.u-tokyo.ac.jp/index_e.html	233	26	15	Japan
7	http://web.up.ac.za/	369	37	>10	South Africa
8	http://www.novyj.ru/	336	13	8	Russia
9	http://www.banik-ostrava.cz/	594	140	3	Czech Republic

10	http://www.lignano.it/en/	127	84	7	Italy
11	http://www.umax.cz/	147	87	4	Czech Republic
12	http://www.zar.mn/index.php	1324	99	12	Mongolia
13	http://www.matrix-se.com/matrix/	225	57	9	Sweden
14	http://www.loxon.de/	415	28	8	Germany

Table 2. WWW files test set (first and last files listed in italics did not allow to complete the test set)

Item No.	URL – description	Size (kB)	Country
1	*Packardbell.de ATI driver*	13,430	Germany
2	K-Lite Codec Pack 5.83 – http://www.slunecnice.cz/sw/k-lite-codec-pack/full/stahnout/	15,591	Czech Rep.
3	http://freesoft.ru/getit_eu.html?file=mm-mm/K- Lite_Codec_Pack_583_Full.exe	15,591	Russia
4	http://http.download.nvidia.com/Windows/81.98/81.98_forceware _win9x_international.exe	20,402	Bulgaria
5	Adobe Reader 9.3 – http://www.slunecnice.cz/sw/acrobat-reader/stahnout/	26,666	Czech Rep.
6	http://ardownload.adobe.com/pub/adobe/reader/win/9.x/9.3/enu/A dbeRdr930_en_US.exe	27,386	Netherlands
7	http://ftp22.nero.com/Nero9/79a1617ac1f8b22196e2c6ed2724df7 4/Nero-9.4.12.708b_lite.exe	33,178	Netherlands
8	Nero 9.4 – http://www.slunecnice.cz/sw/nero-9-free/stahnout/	33,364	Czech Rep.
9	http://game.amd.com/us-en/drivers_catalyst.aspx?p=xp/radeonx- xp	47,595	USA
10	*Packardbell.de ATI Catalyst*	52,352	Germany

2.2. Measurement techniques

All measurements were made using the same www client Mozilla Firefox version 3.6.3 running under OS Windows XP Professional SP3. The use of client cache was eliminated both its deactivation and manual erasing the cache after every measurement. This was necessary due to the fact that despite cache deactivation it was observed that some data are still cached somehow.

The measurement consisted in measuring the time necessary for complete downloading the webpage or saving the file to the hard disk. The measurement itself was made by two Firefox plugins: Fasterfox (ver. 3.8.4) for webpage download and Download Panel (ver. 2009.09.02) for file download. No other applications affecting network communication (excluding antivirus system) were active during any measurement.

Download times of www pages were measured in sets having 5 rounds each. Every round comprised gradual displaying every webpage from the list in Tab. 1 and measuring the download time. There were 5 periods of measurement (usually during a single day) and in each of these days 5 sets were measured. From those 25 measured data maximum and minimum values were found and excluded from subsequent processing to avoid extremal fluctuation to affect results. The measurement process of file download time was similar. The client PC was in a small LAN connected to ISP network using home connection with the bandwidth 6 Mbps for download and 2 Mbps for upload during all measurement described here.

Such measurements were made in the following five modes of network operation:

1. normal operation of the client PC without TOR,
2. client like in mode No. 1, onion router (OR) using 0.5 Mbps in LAN,
3. client with TOR in default settings,
4. client with TOR, manually configured TOR network – fixed ingress and exit ORs (see note below),
5. client with TOR, manually configured TOR network – ingress and exit ORs from manually configured list of 15 ORs (this mode was used only for www pages download measurement),
6. clients with TOR, manually configured TOR network – ingress and exit ORs in test network (this mode was used only for file download measurement – see note below).

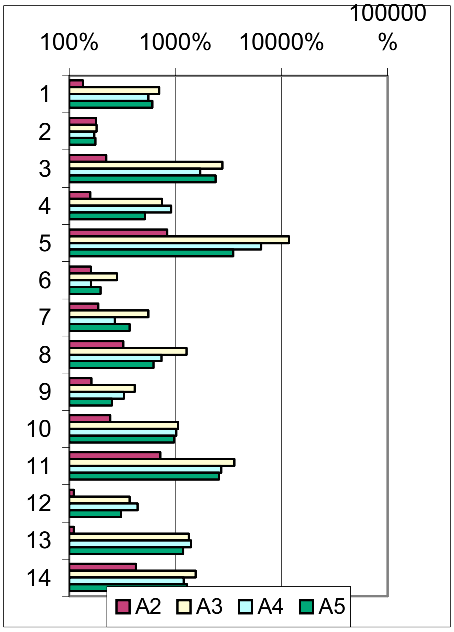
For all modes using TOR for active traffic (modes 3 to 6) the network of ORs consisted 3 ORs. In mode No. 4 the

OR network persistence time was manually set to 30 minutes (default 1 min.) to allow keeping the same conditions (the same network of ORs) for the complete set of measurements. In mode 6 there were two own experimental ORs used: the ingress OR was located in another site than client network, the exit OR was in the same network where the client was located. It means that the only OR that could vary was the second OR. Due to the fact that TOR network is subject to relatively frequent changes this was taken into account in result evaluation. Measurements where significant change occurred during download causing longer delay time were excluded from further processing. Number of such measurements was not significant and the effect to the total results was almost neglectable anyway.

3. Results

The results are split into two separate parts identified with letter A (web page download) and B (file download) respectively. Web page and file download times are show in the Fig. 1 a) and b) respectively. It can be seen that differences between download times without TOR and with TOR are enormous (in extreme case A3 (TOR default) the download time was almost 120 times longer than in case A1 (without TOR)- Such extreme values (occuring only in case of web pages when the basic time is very short – typically 2 seconds or less) could be given partly by the measurement method. The measured round-trip time includes the one-time overhead form setting up ot the TOR that is likely independent on the round-trip time. This time causes the more significant increase of the TOR/non-TOR ratio for shorter basic time values. In more typical cases (basic times A1 above 2 seconds) the TOR/non-TOR ratio is between 2 and 14. In case of file download the ratio is higher, between 16 and 103 (in case of files).

Note: The detailed explanation of the difference in delay ratio is not available at the moment but it seems to be associated with significantly longer time of downloading a single file (files downloaded were much larger than typical files composing the web pages in the test set).



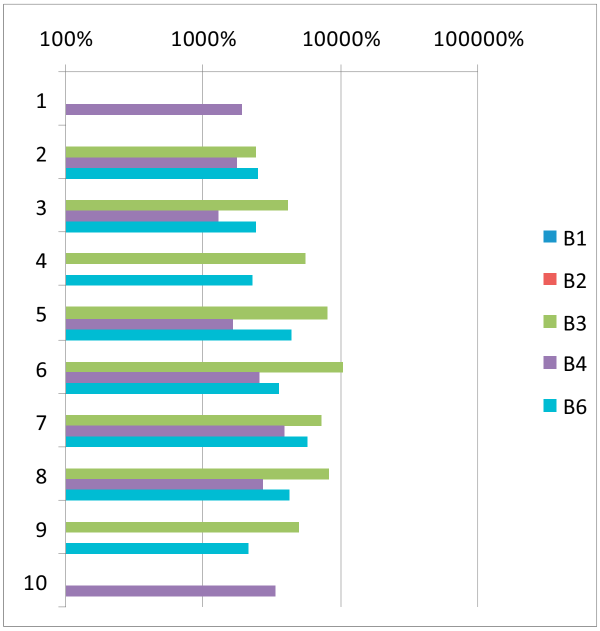


Fig. 1. Relative download times of web pages (left) and files (right). Web pages and files numbering according to the lists in Tab. 1 and 2, description of modes A01 – A05 and B01 – B06 see in the section 2.2. Download times are show in relative values (100% is represented by A1 or B1 values). Diagrams use logarithmic scale.

More detailed comparison of various TOR operating modes is shown in the figure 2. The diagram compares the file download times in TOR default mode (B3) to configured TOR networks (B4 and B6). It can be seen that in most cases the download time B3 (TOR default) is significantly longer than B4 and B6. Thus the assumption that TOR network worth to configure seems to be confirmed. Also the non-neglectable positive difference between B6 and B4 confirms the assumption that the delay in configured TOR network where fast public ORs are chosen would be significantly lower than in TOR network with ingress and exit ORs with relatively slow Internet connection assigned. It is clear that the main purpose of making measurements described as A5 and B6 respectively was not to find the fastest TOR network but rather to test the behavior of ORs in LAN under control.

In configured TOR network the file download delay ratio was between 13 and 34. In case of web pages not shown in the diagram the delay ratio (taking into account only basic times up to 2 seconds) was between 1.7 and 14.1. For the difference between ratios in case of web page and file download see the note above in this section.

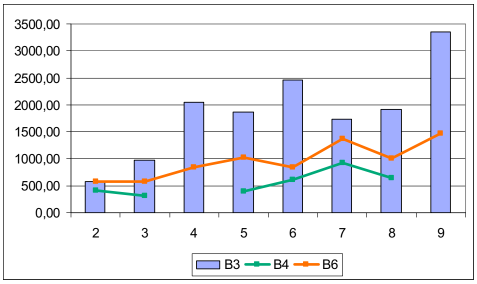


Fig. 2. Absolute file download times - comparison between TOR default mode (B3) and TOR with configured network (B4, B6)

Also the factor of “distance” could play its role in the download times measured and was studied in a limited scope. In networks the distance is usually taken into account using special metrics taking the network infrastructure layout into account. Here only the number of hops is available for web pages. As seen in the Tab. 1 all pages located outside Europe have the number of hops 9 or more (with possible exception of page No. 1 with unfinished tracing and hop count >7) while pages in Europe has the hop count less than or equal to 9. This is not surprising because the measurement was performed in central Europe that has better and more direct connection with the rest of Europe than with other parts of the world. When the web page test set was split according to the geographical criteria then one can observe that the average delay factor for European pages was 21.6 while for non-European pages it was only 4.1. The difference is however given by the significant difference in basic times (European web pages basic times A01 were significantly shorter) but in general the penalty factor for using TOR favors more distant websites.

4. Conclusions,

As expected, the measurements described above confimmed that TOR is significantly slower comparing normal Internet use. There is a way how to decrease the delay singificantly comparing the default TOR behaviour but the delay still remains significant. Therefore one can conclude that TOR offers feasible tool for Internet traffic anonymization but its use should be limited to very special case because the price (paid not in money but in longer time spent by communication) is quite high.

References

1. R. Dingledine, N. Mathewson, P. Syverson, Tor: The Second-Generation Onion Router. [online]. <quot.. 2010-08-10>. Available at WWW: <https://svn.torproject.org/svn/projects/design-paper/tor-design.pdf>

2. Ipoque [online]. 2008 <quot. 2010-08-10>. Internet Study 2007. Available at WWW: <http://www.ipoque.com/resources/internet-studies/internet-study-2007>