ComEX

IEICE Communications Express, Vol.4, No.12, 370–375

Countermeasure against fingerprinting attack in Tor by separated contents retrieval

Naonobu Okazaki[1a)], **Kentaroh Toyoda**[2], **Emiri Yokoyama**[1], **Hirofumi So**[1], **Tetsuro Katayama**[1], **and Mirang Park**[2]

[1] *University of Miyazaki, Japan*

[2] *Kanagawa Institute of Technology, Japan*

a) *oka@cs.miyazaki-u.ac.jp*

Abstract: Tor (The Onion Router) realizes that anonymous web surfing without revealing the user's identity. However, A. Panachenko et al. reveals that an onion router that directly communicates with a user can infer which website a user accesses by leveraging site-specific traffic features, e.g., volume and time, and this attack is called the fingerprinting attack. In this paper, we propose a countermeasure against the fingerprinting attack by obfuscating site-specific traffic features. The idea is to establish two distinct Tor connections and to separately request text-based contents and image-based one through them. We show the effectiveness of our scheme with experiments.

Keywords: anonymous communication, Tor, fingerprinting attack

Classification: Internet

References

[1] C. Shields and B. N. Levine, "A protocol for anonymous communication over the Internet," ACM Conference on Computer and Communications Security (CCS), pp. 33–42, ACM, 2000. DOI:10.1145/352600.352607

[2] R. Sherwood, B. Bhattacharjee, and A. Srinivasan, "P5: a protocol for scalable anonymous communication," IEEE Symposium on Security and Privacy (SP), pp. 58–70, 2002. DOI:10.1109/SECPRI.2002.1004362

[3] N. Mathewson, P. Syverson, and R. Dingledine, "Tor: the secondgeneration onion router," USENIX Security Symposium (SS), 2004.

[4] A. Hintz, "Fingerprinting websites using traffic analysis," in *Privacy Enhancing Technologies*, LNCS 2482, pp. 171–178, Springer Berlin Heidelberg, 2003. DOI:10.1007/3-540-36467-6_13

[5] V. Shmatikov and M.-H. Wang, "Timing analysis in low-latency mix networks: Attacks and defenses," in *European Symposium on Research in Computer Security (ESORICS)*, pp. 18–33, Springer, 2006.

[6] A. Panchenko, L. Niessen, A. Zinnen, and T. Engel, "Website fingerprinting in onion routing based anonymization networks," ACM Workshop on Privacy in the Electronic Society (WPES), pp. 103–114, 2011. DOI:10.1145/2046556.2046570

DOI: 10.1587/comex.2015XBL0131
Received October 4, 2015
Accepted November 5, 2015
Published December 11, 2015

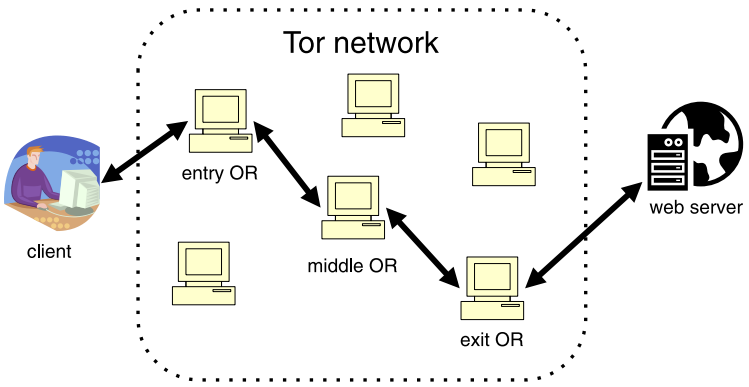


Fig. 1. An example of anonymous communication with Tor

1 Introduction

Safe browsing on the Internet is crucial to preserve privacy for us. In order to realize it, anonymous communication techniques have been extensively studied [1, 2]. Anonymous communication realizes that a user retrieves web contents without revealing his/her identity to a recipient. Especially, Tor is one of the measure implementation of anonymous communication in the Internet [3]. Fig. 1 shows an example that a user communicates with a web server through Tor. A client, who wants to anonymously communicate with a web server, chooses multiple ORs (Onion Router), exchanges a symmetric key with each OR, and connects with a web server via each OR. By sequentially encrypting packets by each symmetric key, any OR cannot identify a website that a client accesses and it realizes anonymous communication.

However, A. Panachenko et al. reveals that an entry OR, that directly communicates with a user, can infer which website a user accesses with high accuracy. They collect site-specific traffic features, e.g., volume, time, and direction of the traffic, and use SVM (Support Vector Machine) to identify browsed websites. This attack is called the fingerprinting attack [4, 5, 6].

The conventional countermeasures try to request dummy requests to obfuscate the traffic features [4, 5]. However, they increase traffic volumes which are not desired from the network aspect. Therefore, it is necessary to propose a countermeasure against a fingerprinting attack without increasing unnecessary traffic.

In this paper, we propose a countermeasure against the fingerprinting attack by obfuscating site-specific traffic features. The idea is to establish two Tor connections and to separately request text-based contents and image-based contents through them. We show the effectiveness of our scheme with experiments.

2 Attacker model

We define the attacker model dealt in this paper. The aim of an attacker is to identify websites that a victim browses. The attacker's procedure consists of two phases: (1) collecting fingerprints of websites and (2) identifying a website. In the following, we describe each phase in detail.

DOI: 10.1587/comex.2015XBL0131
Received October 4, 2015
Accepted November 5, 2015
Published December 11, 2015

2.1 Fingerprints collection phase

An attacker accesses websites that a user likely to access to record traffic features by each website. Popular websites can be obtained with Alexa[1], which discloses the top accessed websites. The collected traffic features are as follows:

- $S_{total}^{p\to q}$: the total amount of packets from p to q (Bytes)
- $N_{total}^{p\to q}$: the total number of packets from p to q
- $S_{avg}^{p\to q}$: the average packet size from p to q (Bytes)
- $S_{var}^{p\to q}$: the variance of packet size from p to q (Bytes)
- $C_{avg}^{p\to q}$: the average chunk size from p to q (Bytes)
- $C_{var}^{p\to q}$: the variance chunk size from p to q (Bytes)

p and q denote either a client c or a web server w, respectively. Therefore, totally 12 ($= 6 \times 2$) features are used and we call a set of them ($S_{total}^{c\to w}, S_{total}^{w\to c}, \cdots, C_{var}^{w\to c}$) as a fingerprint. The chunk size denotes cumulative amount of packets until a client (or a website) receives a packet from the other side.

2.2 Identification phase

After collecting fingerprints for popular websites, an attacker setups an OR, which we call a malicious OR. When the malicious OR is chosen as an entry OR of a client, the attacker tries to identify which site the client browses. In order to do that, the attacker records a fingerprint in the same way as the previous phase. Then, the attacker calculates the similarity between the recorded fingerprint and each fingerprint collected in the previous phase. Finally, the attacker judges the highest similar website as the website that the client browses.

3 Proposed scheme

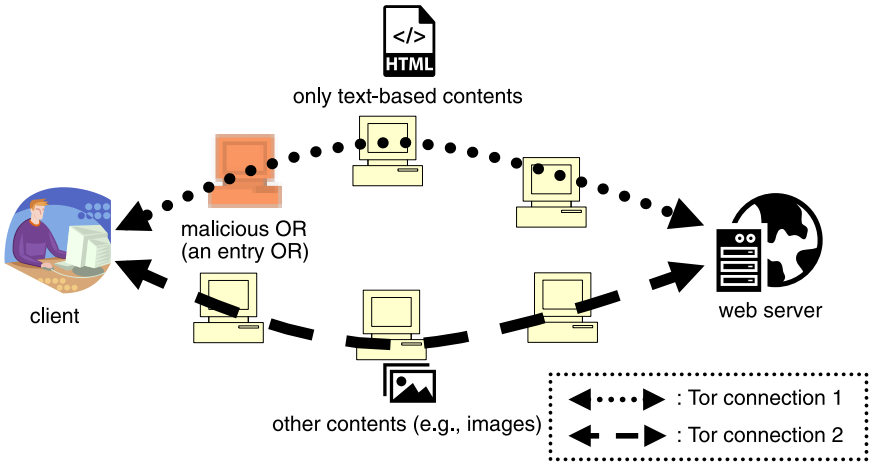
Here, we propose a countermeasure against the fingerprinting attack by connecting with two distinct Tors and separately retrieving a text-based content (such as an HTML file) and other contents (such as an image file) with them. Fig. 2(a) shows a concept of our scheme. At first, a web browser only retrieves and shows text-based HTML contents through a Tor connection. Then a user clicks desired non-text contents, i.e., images and videos, and a web browser retrieves them through another Tor connection. By doing this, our scheme makes a malicious OR infeasible to identify an accessed website from collected traffic features, since it may only contain only text-based contents or non-text-based contents and timing to retrieve contents is also obfuscated. We argue that it is unlikely for a client to choose two different entry ORs controlled by an attacker.

3.1 Implementation

We explain a way to implement the proposal in a client system. In order to realize that a client can retrieve images on demand, when a text-based HTML is retrieved, let a web browser attach a HTML code to show a button to retrieve an image above the image location. Fig. 2(b) and 2(c) show an implementation of an image retrieval button.

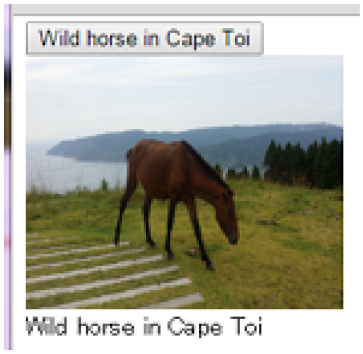
[1]http://www.alexa.com/topsites

DOI: 10.1587/comex.2015XBL0131
Received October 4, 2015
Accepted November 5, 2015
Published December 11, 2015



(a) Separated contents retrieval with two distinct Tor connections.

(b) Initially loaded webpage with an image retrieval button.



(c) A webpage after an image button is clicked.

Fig. 2. An implementation of an image retrieval button.

In the current release of Tor[2], a client cannot simultaneously boot up multiple Tor clients. Therefore, we boot up a virtual machine on a client and establish another Tor connection to retrieve image contents. Retrieved images are saved in a shared directory that is accessible by the host OS (Operating System). Finally, a web browser on a host OS displays a web page by combining text-based contents and images on the shared directory.

4 Evaluation

In order to show the effectiveness of our scheme, we compare attacker's detection accuracy against the real website traffic information between (1) with our scheme and (2) with no countermeasure. We evaluate the attacker's success rate r_i for a website i and the overall success rate R averaged over i and they are defined as follows:

$$r_i = \frac{N_{success}(i)}{N_{trial}(i)}, \quad R = \frac{1}{N_{sites}} \sum_{i}^{N_{sites}} r_i, \tag{1}$$

where $N_{success}(i)$, $N_{trial}(i)$, N_{sites} denote the number of websites that successfully identified by attackers, the number of trials that an attacker tries to identify, and the

[2]https://www.torproject.org/

DOI: 10.1587/comex.2015XBL0131
Received October 4, 2015
Accepted November 5, 2015
Published December 11, 2015

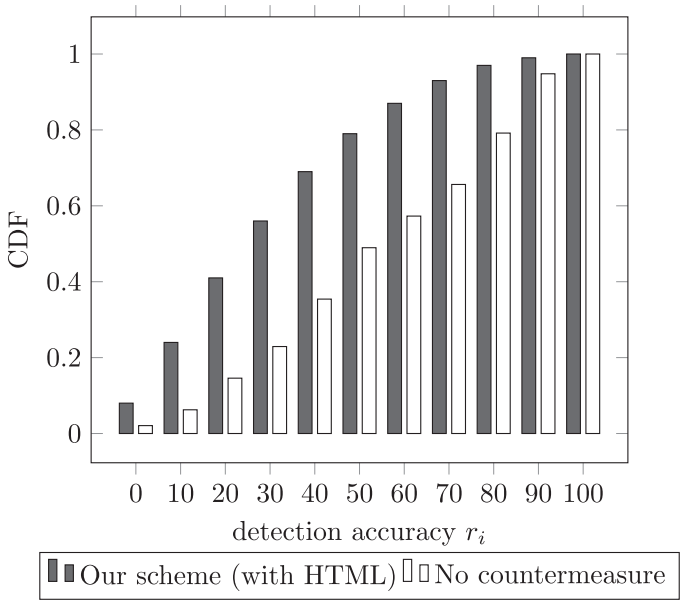


Fig. 3. CDF against success rate r_i.

number of websites, respectively. In this evaluation, we set $N_{sites} = 100$, $N_{trial}(i) = 10$ for $i \in [1, 100]$ if not stated otherwise.

For evaluation, a client accesses the top $N_{sites} = 100$ websites from Alexa by two minutes with and without the proposed system and collects fingerprints. We repeat the same procedures by $N_{trial}(i) = 10$ times for each. An attacker accesses the same 100 websites by ten times for each, collects fingerprints, and identifies browsed websites by the procedure described in Section 2. We assume that an attacker against the proposed scheme knows a client only retrieves text-based contents or image-based contents but cannot retrieve both of them because a client is unlikely to choose two entry ORs from ORs controlled by the same attacker. We evaluate it on a computer that operates Windows 7 Professional, Firefox 28.0 as a web browser, and Tor v0.2.3.25.

At first, we evaluate attacker's success rate when a client only obtains HTML contents. Fig. 3 shows a CDF (Cumulative Density Function) versus the success rate r_i. CDF quickly approaches 1 against a better approach since it means that most of websites are incorrectly identified. As we can see from Fig. 3, our scheme clearly decreases success rate against without any countermeasure. The overall success rate R is 57% for no countermeasure approach while 35% for our scheme (with HTML), respectively. Therefore, by separately retrieving contents, we can decrease the attacker's success rate by 22% on average.

We pay attention to 20 websites whose $r_i \geq 0.9$ if no countermeasure is taken. That is, we focus on websites that an attacker can easily identify. Here we also compare R of (1) no countermeasure, (2) our scheme with only HTML, and (3) our scheme with only images, respectively. Here, 'our scheme with only images' denotes that an attacker knows that a client retrieves only images through his/her OR. Assuming the worst case, a client retrieves the same images as the attacker chooses in the training phase but the order and timing of images retrieval may differ. In collecting image contents, an attacker uses the same web browser that can

DOI: 10.1587/comex.2015XBL0131
Received October 4, 2015
Accepted November 5, 2015
Published December 11, 2015

separately retrieve text contents and images and only 50% of images for a website. Our scheme with HTML and our scheme with images both decrease R from 93% to 31% and 53%, respectively. Again, our schemes much decrease the attacker's success rate by 40%–62%. We can also see that R is higher when an attacker observe fingerprints calculated from only images. This is because the size of images is much larger than that of texts and the fingerprints calculated from images can be more characteristic. In addition, we assume that an attacker can retrieve the same images that a client does and thus it is easier for an attacker to infer websites.

5 Conclusion

We have proposed a countermeasure against the fingerprinting attack by separately retrieving website contents with two Tor connections. The aim is to obfuscate site-specific traffic features measured by an attacker. We implement our scheme with a computer and show that our scheme effectively decreases the attacker's overall detection accuracy by approximately 22%.

DOI: 10.1587/comex.2015XBL0131
Received October 4, 2015
Accepted November 5, 2015
Published December 11, 2015