Applicability of Website Fingerprinting Attack on Tor Encrypted Traffic

Mohamad Amar Irsyad Mohd Aminuddin, ZarulFitri Zaaba, Azham Hussain

Abstract: Tor is a famous anonymity tools that provide Internet user with capability of being anonymous in the Internet. By using the Tor network, a user can browser without anyone know the truth of the communication information. Numerous studies have been performed worldwide on deanonymizing the Tor user. One of popular study is the Website Fingerprinting (WF) attack, a subset of passive traffic analysis attack. WF consists of complex traffic analytical process with several limitations and assumptions on the Tor network. In this paper, we will discuss the fundamental principal of WF on Tor network, its assumptions and discussion on whether WF is considered as applicable on attacking the Tor user anonymity(especially in real-world scenario).As a result, the applicability discussion and establishment are presented. This study had found that with the advancement of WF attack, it is applicable to be utilized on Tor encrypted traffic and might become a serious threat to Tor's user anonymity if no proper defense being proposed to prevent the improved WF attack.

Index Terms: website fingerprinting; Tor; machine learning; security; survey;

I. INTRODUCTION

Tor (The Onion Router)[1] is a famous Privacy Enhancement Technology (PET) that provide Internet user with capability of being anonymous in the Internet. Its primary goal is to provide the Internet user with privacy-protected environment by providing anonymousInternet access to anyone that use their software. By using the Tor software, a user can browse (or do other activity) without anyone know the truth of the communication information (such as source and destination IP Address). In the Clearnet (generalInternet), if user open a Google or Facebook webpage, adversary can easily know that the user accessing those websites by simple monitoring the network of the user. Even though the communication to those servers are encrypted making the content of the communication is unreadable, the adversary still knows the user accessing Google or Facebook by simply learn the destination of the traffic that user sent to the Internet.

By using Tor, not only it behaves like a middle person or VPN (Virtual Private Network), it also provides strong anonymity by requiring the message of the communication to travel several locations worldwide before it arrives at its intended recipient. Therefore, the adversary could no longer know the recipient of the message that the user sent.

The popularity of Tor is undeniable as it reaches up to four million users in early 2018[2]. Due to massive popularity, researcher around the world has put considerable effort on analyzing and evaluating the Tor especially in security perspective as whether Tor are really providing strong anonymity to its users. Multiple research had been carried out since its inception to prove its otherwise (proving methods to deanonymize the user)and multiple research had been performed on how to improve Tor to became more efficient and secure.

Until recently, Website Fingerprinting (WF) attack [3]which is a subset of traffic analysis attack had gain in popularity as means of deanonymize the user on the Tor network. WF attack is an attack where an adversary monitor and capture traffic from the client side (the sender/originator of the communication message), then analyses the captured traffic for any known fingerprint(pattern or characteristic).Even though applying WF attack on Tor's user will only give a very basic knowledge on the user browsing activity (such as user open a Google webpage), it could be deeming as serious threats to the Tor's user as the real destination of the message which supposed to be secret are known by adversary. Therefore, WF attack had been thought as dangerous attack to the Tor system that require attention on how to remediate the attack from continuing to deanonymize the tor user[4].

Despite being a serious threats to Tor anonymity, there are still a lot of research need to be carried out to develop WF attack that able to work on world-wide scale as at current stage there are still several critical factors that being ignored by researcher [5]caused the attack might not applicable or effective in the real-world scenario as in the laboratory experiment or controlled environment. There are various factors and approaches need to be carefully examined and tested to ensure better outcome of the WF attack .With that being said, studies on technique and approach to defending WF attack also had been carried causing many traditional WF attack does not applicable on Tor anymore.

Revised Manuscript Received on May 22, 2019.

Mohamad Amar IrsyadMohd Aminuddin, School of Computer Sciences, Universiti Sains Malaysia,11800 USM, Pulau Pinang, Malaysia.

ZarulFitri Zaaba, School of Computer Sciences, Universiti Sains Malaysia,11800 USM, Pulau Pinang, Malaysia.

Azham Hussain, School of Computing, Universiti Utara Malaysia 06010 Sintok Kedah Darul Aman, Malaysia

The main objective of this paper is to survey the applicability of Website Fingerprinting attack on Tor encrypted traffic as the approach of deanonymizing Tor's user. This paper will focus WF attack that specifically focus on learning user browsing behavior on the Tor network.

The rest of the paper is organized as follows. In Section 2, will be brief discussion on how Tor network works and its technical background. In Section 3, we will discuss the fundamental operation of WF attack. Section 4 contain the discussion of WF attack assumption on previous studies. Section 5 the detailed discussion on the applicability of WF attack on Tor encrypted traffic. Finally, the paper is concluded in Section 6.

Tor Background

Tor is run by group of volunteer-operated servers worldwide which currently running more than six thousand relay servers[6]. Relay server (Tor node) is the backbone architecture of the Tor since it functions as the middle entity between communication message originator (client) to the intended recipient (server). The main different between tor and VPN (Virtual Private Network) or other proxy services is that rather than only has one middle entity, Tor provide three middle entity that the message will be transferred through before it reaches its intended recipient. Thus, it will be harder for adversary to monitor the message (to know its sender and recipient) due to it has multiple middle entity and these entities located worldwide which is a difficult task to monitoring altogether. As example, when user connected to Tor and want to open a Google webpage, the user will send request message to first Tor node (entry guard), the first node will forward it to the second node, then the second node will forward to the third node (exit relay) and lastly the third node will forward the message to the Google server. The responses message from the Google server also will travel back to the users in reverse manner. In the event that the adversary managed to compromise one of the Tor Node, it is still impossible to learn the sender and recipient of the message since none of the Tor Node has knowledge one both original source and destination information of the message as shown in Table.

Fig shows the example of Tor operation diagram. The endpoint is the destination of the message communication send by the user. The green-line (protected) means the communication message is protected by Tor encryption. Any traffic captured on green-line traffic will retrieved Tor-based traffic which has significant different from non-Tor traffic. The red-line (unprotected)means the communication message are not protected by Tor. The endpoint itself does not know the true identity of the user since it believed that the exit relay is the message originator who send the request.

Table. 1 Tor node knowledge

Tor Node	Know message source	Know message destination
Entry Guard	TRUE	FALSE
Middle Node	FALSE	FALSE
Exit Relay	FALSE	TRUE

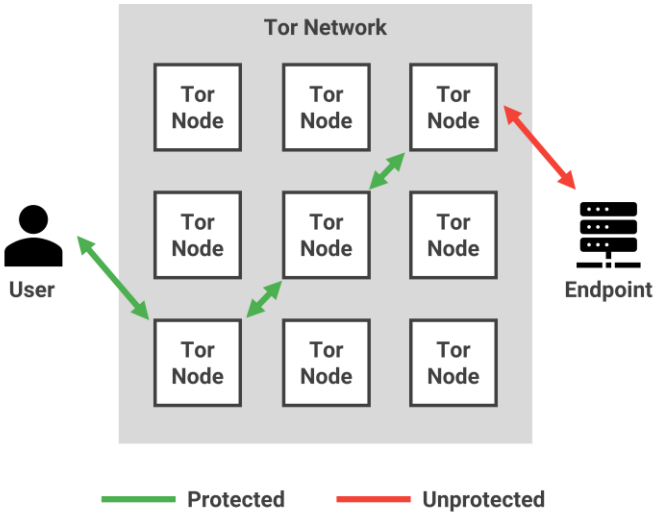

Fig. 1Tor operation diagram

II. WEBSITE FINGERPRINTING

In this section there will explanation on the WF fundamental operation including the basic principle of WF attack, type of WF target, accuracy measurement, attack versus defense and type of study.

WF attack principle

When user browsing the typical webpage, there commonly contains multiple different files that required to be transferred to the user before a complete webpage appeared on the user browser[7]. Therefore, each webpage that user browse generally has specific pattern or characteristic (on the transferring process) that could be analyzed and defined[8]. Those pattern or characteristics are called as fingerprint. In WF attack, the adversary is required to defined fingerprint on webpage that is he has interest on (to detect whether victim open the webpage or not)and store it into database before launching any attack. This defining process involving training a machine learning algorithm on specific sets of fingerprints. The output of the training is a set of machine learning classifier that able to classify traffic based on the defined fingerprint. Then the adversary needs to monitor the traffic between the client to the webpage server. In Tor-based environment, the adversary needs to monitor and captured the victim's traffic anywhere between the user system to the entry guard. The red-line in Fig denoted the position where the adversary will monitor and capture victims' traffic.

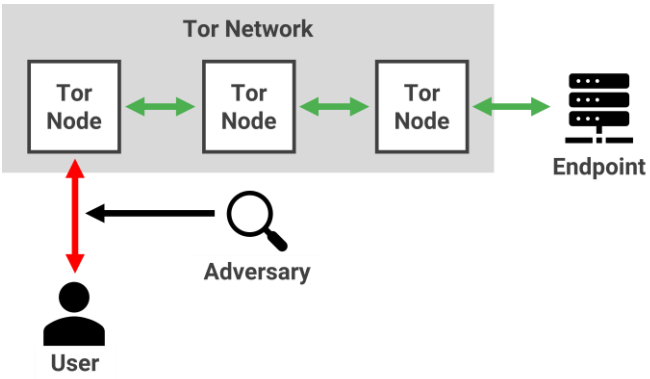

Fig. 2 Adversary monitor traffic at client side

After the capturing process had been made, the captured traffic will be analyzed for its fingerprint. Then, it will be compared with the defined/known fingerprint in the database[9]. The comparison will utilize machine learning classifier (that had been previously trained) in order to process all the pattern or characteristics found of the traffic. There also might be some filtering or sanitizing applied to the fingerprint. The outcome of the machine learning process will determine whether the captured traffic has any match with the fingerprint on the database.

Type of target

To launch WF attack, there are two type of victim target that need to be consider[10]. **Error! Reference source not found.** shows the targets comparison. Both target type has pro and cons and are very dependable on situation and purpose of an attack. If WF attack aim is to detect how many users browsed a certain known webpage or website, non-targeted is the best of choice rather than targeted as the classifier does not trained for a specific victim. If the aim is to learn whether the victim browsed certain webpage, then targeted type is much more appropriate as the classifier does not trained for a specific victim.

Table. 2 Type of Target

Criteria	Targeted	Non-targeted
Definition	Adversary targets specific victim to discover the victim browsing activity.	Adversary target group of victims to discover their browsing activity without focusing on specific victim.
Training	Required classifier that have similar training condition to the victim environment.	Required classifier that use specific condition environment which will applicable to the group of victims.
Location	Adversary located in the same network with victim where the adversary able to monitor directly victim network traffic.	Adversary located anywhere from victim to the first entry node. Could be ISP-level or malicious entry node.

Accuracy Measurement

Study on algorithms require certain type of measurement for comparison and benchmarking purposed. In WF, there are four common measurement type as shown in Table. .

Table. 3 Accuracy measurement

Measurement Type	Description
True Positive (TP)	The frequency that the WF model able to correctly match the fingerprint. Researchers attempt to maximize this measure.
True Negative (TN)	The frequency where WF model able correctly to detect unknown fingerprint that had not been predefined in the database. Researchers attempt to maximize this measure.
False Positive (FP)	The frequency of WF model falsely reported two different fingerprints as matches fingerprint. This will give critical impact on the accuracy and applicability of the WF model. Researchers attempt to minimize this measure.
FN (FN)	The frequency of WF model falsely reported unknown fingerprint which is actually had been predefined in the database. Researchers attempt to minimize this measure.

Although TP is important in measuring the successfulness of WF attack on detecting fingerprint, FP is also considered as important factor on determining the effectiveness of the WF attack. Higher FP means the WF model has higher degree of detecting incorrect fingerprint making the model is highly unusable and ineffective to be utilized. Hence, TP and FP are two measurements that essentially need to be carefully studied during developing WF model[11].

Type of website

There are two common type of website that are applied in the WF attack researches. The first one is the top rank website in the Alexa Top 100 list [12] as has been carried out by [13] and [3]. This type of research not only applicable to Tor encrypted traffic analysis, but it also applicable on WF study on the Clear net or other Anonymity-services such as JAP (Jon Donym Anonymous Proxy)[14] and I2P (Invisible Internet Project) [15]since these websites are accessible on those networks.

The second type of website is Tor Hidden Service (HS). HS is a website (which have .onion domain) that are similar with website on Clear net but only accessible via the Tor as it located in the Tor network. User that is not in the Tor network are unable to browse the website.

HS true location or IP Address are only known by its operator. Thus, WF studies on Tor be inclined to focus more on HS-based type of website[16][17][18].

Fingerprint characteristics

In encrypted traffic analysis, the content of the traffic is unknown. In Clear net, the destination IP Address and Port are clearly visible. However, in Tor the destination of IP Address and Port are unknown as the only available information are pointing to the entry guard. Therefore, there are needs for other type of information to be recognize as characteristics or patterns that could be analyzed and extracted[19]. Current researches are focusing on three main categories of traffic metadata characteristics. Those categories are circuit-based (such as circuit lifetime, cell inter-arrival times and cells per circuit life time), flow-based (such as flow segment size, round trip time, duration and burst volume) and packet-based (such as packet length, frequency, header).However, due to uniqueness and usefulness of these characteristics, not all feature will be utilize in WF [20]as only useful and effective characteristics will be recognized as useful fingerprint and will be stored on database.

Fingerprint Variations

A webpage might have variation of contents especially on dynamic website such as Facebook and Twitter where the content changed based on the user account and preferences. There is also localization factor where different region or country of user will browse on different language or content such as when opening Google and Amazon. Therefore, the sevariations require separate fingerprinting where individual fingerprint are required for each of the variant to ensure accurate and precise WF outcome. There is also requirement of updating the fingerprint from time to time as a webpage might have updated content.

Assumptions

There are several assumptions being made by researcher in order to make WF attack probable in Tor. Some of the assumption are very important as it will affect the accuracy of the WF model significantly when the assumption is not being applied. Below are the explanations on common assumptions in WF studies.

Closed-world vs Open-world

Close-world assume that the user should only visit a certain number of web pages[21]. This means that the user is only allowed to open specific number of web pages and no other webpage are allowed to be browsed as shown in Figure 1. On the other hand, open-world assumption allowing the user to open any web pages without any restriction as shown in Fig. Close-world originally being assumed for algorithm experiment purpose in controlled environment and generally has higher degree of accuracy but does not suitable to be utilize in the real-world scenario of Tor. Despite the open-world might allowing for a better real-world implementation, it has higher complexity in order to achieve high accuracy and effectiveness WF attack[22].

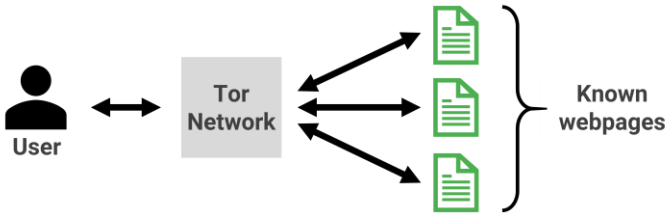

Figure 1: Closed-world scenario.

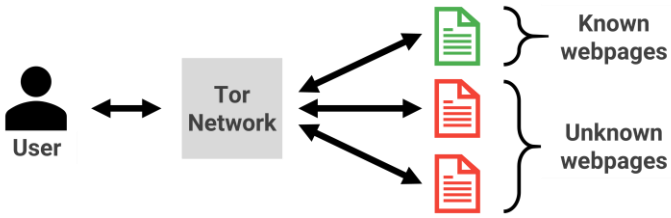

Fig. 3 Open-world scenario

Replicability

In other to define fingerprint and train the machine learning classifier for WF attack, there are certain important environment need to be considered. Replicability is an assumption that the adversary is able to imitate the victim condition (such as network connection and browser version) in order to define and train the classifier. This assumption is crucial especially in targeted-type of victim. Different condition will produce different fingerprint outcome which will greatly impact the WF accuracy and effectiveness. In some situation, this assumption could be unfeasible as the adversary have no method on profiling the victim condition in order to launch WF attack[5].

Browsing Behavior

There are multiple ways of browsing a website. In WF, assumption of browsing behavior means the adversary assume that the user browses the web in sequential manners[21]. This means the user open one page after another and only use single browser tab and window. With this assumption, multi-tab browsing or multi-window browsing is not relevant. This assumption is very difficult to be employed in the real-world scenario since user are likely to browse in multi-tab or multi-window as had been found by [23]. Thus, it is only suitable to be use in the controlled environment.

Background activity

Adversary assume that the user does not performing any other activity on the network. This is to ensure that the WF does not incorrectly analyze traffic that was not come from user browsing activity[24].

An example of background activity such as Operating System Update in the background or user downloading a torrent file .These activities will affect the WF attack accuracy. However, this assumption only could be ensured as true in controlled ecosystem where in real-world environment, there are a lot of background activity happening in user computer system without user consent.

III. DISCUSSION ON APPLICABILITY

Having carefully analyzed the properties of WF, there are several important factors need to be point out before concluding the applicability of WF attack on Tor encrypted traffic .First, Tor as anonymity service really provide strong anonymity towards its user where it became harder on WF to be working accurately and effectively compared to the Clear net. Although there are a lot of successful attack on the Tor's user where it could be considered as dangerous threats to user anonymity in Tor, most of the accurate and effective attack previously are more focusing on controlled environment where a lot of significant assumption had been made.

When dealing on assumptions, there are several important factors need to be taken into account. Closed-word assumption are only suitable for lab experiment rather than real-world application. Thus, any WF attack that have high accuracy in closed-world does not guarantee that it performs the same on open-world. Reliability is another great issue which need to be considered. Retrieving user system profile is generally very hard especially when the adversary does not personally know the targeted victim .However, research by [21] had shown that it able to gather minimal amount of data for replicability while maintaining higher effectiveness on the WF attack .Although browsing behavior is critical factor, there are research that able to tackle the issue[13]. The background activity assumption problem also has been manage to be solved by[25]. It is expected that more effective and efficient improvement on WF attack will be proposed in the future.

Only until recently, progression toward open-world attack closer to real-world scenario being made by researchers with enhanced WF attack technique. There are also variety of advancement had been done to those crucial assumption as had been stated above. Ten years ago, the WF seems not applicable to the Tor network as it has numerous limitation and constraint that are hard to overcome which making only seems viable in controlled-environment. Nevertheless, with advancement research on WF attack, we could see that it starts to slowly become a real threat to Tor user anonymity. This is proven as researchers start to perform elaborate studies on effective and efficient methods of defending Tor network from WF attack[26][13][4][27]. There is also some scenario where WF attack is successfully launch on Tor network. Hence, WF attack currently is applicable on Tor encrypted traffic.

IV. CONCLUSION

To sum up, WF attack has gone a long way and there is lot of improvement had been made over the years as it does applicable to be utilized as attacking tool to Tor encrypted traffic. Despite currently Tor able to withstand several WF attack effectively, progression on WF attack advancement are something need to be look forward into as not only it currently applicable on attacking Tor, it might become a real threat to the Tor user anonymity if no proper defense being proposed to prevent the improved WF attack.

Previous surveys and reviews studies are more focusing on the WF attack applicability in general. Our paper had managed to go deeper and discuss the survey on applicability of WF attack on Tor encrypted traffic. This will hopefully become a helpful guidance for future researches that are related to WF attack on Tor.

REFERENCES

1. R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The Second-generation Onion Router," in *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13*, 2004, p. 21.
2. The Tor Project, "Users – Tor Metrics." [Online]. Available: https://metrics.torproject.org/userstats-relay-country.html?start=2018-01-01&end=2018-01-30&country=all&events=off. [Accessed: 28-Jan-2019].
3. A. Panchenko, L. Niessen, A. Zinnen, and T. Engel, "Website Fingerprinting in Onion Routing Based Anonymization Networks," in *Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society*, 2011, pp. 103–114.
4. X. Cai, R. Nithyanand, and R. Johnson, "CS-BuFLO: A Congestion Sensitive Website Fingerprinting Defense," in *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, 2014, pp. 121–130.
5. The Tor Project, "A Critique of Website Traffic Fingerprinting Attacks | Tor Blog." [Online]. Available: https://blog.torproject.org/critique-website-traffic-fingerprinting-attacks. [Accessed: 28-Jan-2019].
6. The Tor Project, "Servers – Tor Metrics." [Online]. Available: https://metrics.torproject.org/networksize.html. [Accessed: 28-Jan-2019].
7. Patrick Sexton, "How a webpage is loaded and displayed." [Online]. Available: https://varvy.com/pagespeed/display.html. [Accessed: 28-Jan-2019].
8. Y. Shi and K. Matsuura, "Fingerprinting Attack on the Tor Anonymity System," in *Information and Communications Security*, 2009, pp. 425–438.
9. F. Mercaldo and F. Martinelli, "Tor Traffic Analysis and Identification," in *2017 AEIT International Annual Conference*, 2017, pp. 1–6.
10. M. Juarez, S. Afroz, G. Acar, C. Diaz, and R. Greenstadt, "A Critical Evaluation of Website Fingerprinting Attacks," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. 263–274.
11. X. Cai, R. Nithyanand, T. Wang, R. Johnson, and I. Goldberg, "A Systematic Approach to Developing and Evaluating Website Fingerprinting Defenses," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. 227–238.
12. Alexa Internet, "Alexa Top 500 Global Sites." [Online]. Available: https://www.alexa.com/topsites. [Accessed: 28-Jan-2019].
13. M. Juarez, M. Imani, M. Perry, C. Diaz, and M. Wright, "Toward an Efficient Website Fingerprinting Defense Marc," *Comput. Res. Repos.*, vol. abs/1512.0, 2015.
14. JAP Team, "JAP -- ANONYMITY & PRIVACY." [Online]. Available: https://anon.inf.tu-dresden.de/index_en.html. [Accessed: 28-Jan-2019].
15. I2P, "I2P Anonymous Network." [Online]. Available: https://geti2p.net/en/. [Accessed: 28-Jan-2019].
16. H. Haughey, G. Epiphaniou, H. Al-Khateeb, and A. Dehghantanha, "Adaptive Traffic Fingerprinting for Darknet Threat Intelligence," in *Cyber Threat Intelligence*, vol. 70, A. Dehghantanha, M. Conti, and T. Dargahi, Eds. Cham: Springer International Publishing, 2018, pp. 193–217.

17. A. Panchenko, A. Mitseva, M. Henze, F. Lanze, K. Wehrle, and T. Engel, "Analysis of Fingerprinting Techniques for Tor Hidden Services," in *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society*, 2017, pp. 165–175.
18. A. Kwon, M. AlSabah, D. Lazar, M. Dacier, and S. Devadas, "Circuit Fingerprinting Attacks: Passive Deanonymization of Tor Hidden Services," in *24th USENIX Security Symposium (USENIX Security 15)*, 2015, pp. 287–302.
19. M. A. I. M. Aminuddin, Z. Fitri, M. Kaur, and D. Singh, "A Survey on Tor Encrypted Traffic Monitoring," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, 2018.
20. J. Yan and J. Kaur, "Feature Selection for Website Fingerprinting," *Proc. Priv. Enhancing Technol.*, vol. 2018, no. 4, pp. 200–219.
21. T. Wang and I. Goldberg, "On Realistically Attacking Tor with Website Fingerprinting," *Proc. Priv. Enhancing Technol.*, vol. 2016, no. 4, pp. 21–36, 2016.
22. E. Oh, S. Li, and N. Hopper, "Fingerprinting Keywords in Search Queries over Tor," *Proc. Priv. Enhancing Technol.*, vol. 2017, no. 4, pp. 251–270, 2017.
23. C. von der Weth and M. Hauswirth, "DOBBS: Towards a Comprehensive Dataset to Study the Browsing Behavior of Online Users," in *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, 2013, pp. 51–56.
24. T. Wang, X. Cai, R. Nithyanand, R. Johnson, and I. Goldberg, "Effective Attacks and Provable Defenses for Website Fingerprinting," in *23rd USENIX Security Symposium (USENIX Security 14)*, 2014, pp. 143–157.
25. A. Panchenko *et al.*, "Website Fingerprinting at Internet Scale," in *Network and Distributed System Security Symposium (NDSS)*, 2016.
26. G. Cherubin, "Bayes, not Naïve: Security Bounds on Website Fingerprinting Defenses," *Proc. Priv. Enhancing Technol.*, vol. 2017, no. 4, 2017.
27. R. Nithyanand, X. Cai, and R. Johnson, "Glove: A Bespoke Website Fingerprinting Defense," in *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, 2014, pp. 131–134.