Poster: Novel Website Fingerprinting Techniques

Tao Wang
PhD candidate
University of Waterloo

Ian Goldberg
Faculty
University of Waterloo

I. Background

Web-browsing clients may choose to browse the web by using proxies with encryption to secure their privacy. An attacker watching the outgoing connection of such a client will not be able to know the content or destination of their packets, unless the attacker uses website fingerprinting. Website fingerprinting is the process by which an attacker identifies which server a client is communicating with under encryption by observing communication patterns. For example, a client making a search on google and another posting a blog entry on a dissident blog can be observed to receive incoming packets and send outgoing packets in different amounts, ordering, and rates.

Works such as those by Liberatore and Levine (CCS 06) and Herrmann et al. (CCS 09) have demonstrated that website fingerprinting poses a significant threat to plain TLS encryption, such as that used by SSH tunneling and VPNs. They have concluded that users of such technologies cannot expect their destination server to be hidden from an observer; the observer can guess their website with almost absolute certainty. However, it remains unclear if website fingerprinting is powerful enough to pose a threat to Tor. Tor uses padding, suffers from random delays, and performs a number of background activities that intefere with website fingerprinting. Several authors such as Panchenko et al. (WPES 11) and Cai et al. (CCS 12) have proposed attacks on Tor, but they have not reached 90% accuracy. On the other hand, Dyer et al. (IEEE S&P 12) have demonstrated that many defenses are either inefficient or ineffective.

II. Impact

Attackers using website fingerprinting pose a significant threat to the privacy expected from privacy enhancing technologies, such as Tor, the popular anonymity network. We focus on Tor, which is particularly interesting because website fingerprinting on Tor is much more difficult. A number of parties have an interest in conducting such an attack. Totalitarian government agencies, nervous about being unable to conduct surveillance on an anonymity network, may seek to do so with website fingerprinting. They may also seek to observe and classify anonymous traffic in order to justify a ban by demonstrating the potential prevalence of illegal activity. It could be possible to even block specific sites from Tor. On the other hand, attackers may attempt to track, identify, and observe users who expect privacy. We need to fully understand the potential impact of website fingerprinting and what steps a user may take to reduce it. As it stands, anonymity network users, for instance, may be misled into thinking that their privacy is guaranteed and no one can determine their destination server, when it is all but certain when faced with website fingerprinting.

III. Our Contributions

We present a new framework for understanding website fingerprinting attacks and defenses. This framework is based on the particular pattern that the attack exposes and the defense covers. There are roughly three patterns targeted by previous attacks and defenses: resource length (packet size), total transmission size, and packet order. Using this framework, we present novel attacks and defenses.

Website fingerprinting attacks are far from perfection, as they are relatively new. We have identified several ways to improve previous work and by applying those ideas, we have engineered a stronger website fingerprinting attack in terms of accuracy. An attacker who wishes to know if a client is visiting a particular site through Tor may do so at 96% accuracy with our attack. We seek to present this novel attack under our framework in our poster.

From Dyer et al.'s work, previous website fingerprinting defenses are ineffective. Dyer et al. and Cai et al. have proposed their own defenses, but they are not tested against new attacks and there are limitations to the assumptions behind these defenses. Tor developers have been unwilling to adopt such defenses due to their bandwidth cost, and they have implemented their own low-cost defense. With our framework and experiments we demonstrate that the current defensive mechanism of Tor is ineffective, and we propose our own novel webnsite fingerprinting defense (work in progress).

There is no consensus on experimental techniques behind website fingerprinting due to its novelty and the complex practical issues underlying the field. Researchers have presented website fingerprinting attacks and defenses with different experiment sizes, format, assumptions on client behavior or attacker capability, data collection techniques, data processing techniques, and settings. Authors in this field have often neglected to discuss such factors, possibly because they are not considered important. We discuss how variations in these assumptions can significantly affect experimental results. We hope to present our own experimental technique, compare it with other authors, and facilitate discussion on proper experimental methodology.

Website fingerprinting is a relatively new field, and many questions regarding its practicality remain unanswered even in the simplest case. To the authors' understanding, there is no work on how website fingerprinting would function if the client terminates the web loading process prematurely (when a page loads slowly and the client has already found the information or link they need); if the client loads several pages at once (when a page loads slowly); or if a page has significant AJAX content (for dynamic pages). While we have not solved these problems, we seek discussion on these issues.