Lane Gramling
EECS 700
Reading 17

Summary - Deep Fingerprinting: Undermining Website Fingerprinting Defenses with Deep Learning

This paper focuses on exploring the application of deep learning techniques to build an improved system for website fingerprinting, capable of maintaining reasonable accuracy even when faced with active countermeasures attempting to mitigate its effectiveness. In this case, the paper focuses on countering proposed defenses against website fingerprinting techniques that could be employed in identifying Tor traffic.

Contributions

After exploring the background of attacks and defenses against website fingerprinting and explaining the most popular defense proposals in the context of their work, the authors segue into their proposed counterattack centered around using convolutional neural networks to form the core of their approach to fingerprinting a large dataset of packet traces captured by running tcpdump on an array of university computers. Data for the packets is represented as a sequence of tuples pairing the timestamp with the size of the packet, using signage to differentiate incoming and outgoing packets. By normalizing their values and observing merely whether the packet was incoming or outgoing, this provided enough differentiating information in terms of features for the learning algorithms to accurately learn the fingerprints of trained websites.

In tests on undefended packet data, the authors' technique for deep fingerprinting achieved 98.3% in testing on communications over Tor. This is the highest among compared tests, with the next most successful attacks ranging from 92.3% and 97.3% accuracy. Likewise, when tested against packet bursts using various defense mechanisms while flowing through Tor, their results performed adequately against the competition in all scenarios. Notably, their DF technique achieved 90.7% against WTF-PAD, effectively eliminating it as a means of protection against website fingerprinting; despite performance against WTF-PAD, the performance against other defenses were not significantly higher than the competition.

Limitations

Despite the success in their testing, I personally felt that they could have put more effort into specifically trying to find situations that reduce the accuracy of their technique. Additionally, the observed results only showed significantly improved performance against one of the four attacks they tested as compared to competition.

It is also worth mentioning that due to the only features fed to the neural network being limited to timestamps paired with a value in {-1,1}, their deep fingerprinting technique feels on the surface as if it's quite simple to defend against, because the narrow nature of the observed data seems simple enough to target with countermeasures obfuscating the traits observed by DF.

Future Work

It would be worthwhile to see what kind of results their algorithm shows under an expanded testing environment. Additionally, as with any case where a new attack shows results and potential for use, there is a door opened for further inquiry into defenses against the attack.