DNS-based User Tracking (Attacks and Defenses) (Abstract)

Authors: Deliang Chang, Joann Qiongna Chen, Zhou Li, and Xing Li
Presenter: Zhou Li, UC Irvine

DNS requests are often coming through recursive resolvers till reaching the authoritative name servers. Due to the trend of centralizing resolvers (e.g., Google Public DNS handles a huge volume of DNS requests from all over the places), it is a valid concern that individual users' privacy can be compromised when the resolvers log and mine the DNS traffic data. One prominent threat is *user tracking*, through which a user's activities across different networks and devices can be correlated, for purposes like personalized advertisement, surveillance, etc. In this talk, I will describe a recent study [1] we have done in studying the problem of DNS-based user tracking with real-world DNS data traces.

In the first part of the talk, I'll formally define the research problem and overview the existing works. Our key finding is that though the prior works have demonstrated the concerns are valid, they all focused on the *closed-world setting*, which means that victim users must be known to the adversary during a training period. To better capture the real-world attacker's capabilities, we propose a new machine-learning based tracking method called DSCORR, which is tailored to the open-world setting, and we show it outperforms the prior works. I'll elaborate on the key techniques of DSCORR, including domain-based word embedding and automatic threshold generation.

In the second part of the talk, I'll describe our efforts in protecting users from the DNS-based user tracking. Our major insight is that the users can add noisy DNS requests to confuse the tracking methods that are based on machine learning or statistical learning. Our defense system is called LDPRESOLVE, which incorporates *local differential privacy* into users' stub resolvers to provide privacy guarantee.. The evaluation result on the same dataset shows the DNS-based user tracking can be effectively curbed, e.g., tracking accuracy degraded from 93% to 10.1%.

Finally, I'll talk about the potential next steps of this work and this research direction, like the feasibility of incorporating other privacy primitives (e.g., Shuffle model) to achieve better privacy and utility tradeoff for defense.

References:

[1] Chang, Deliang, Joann Qiongna Chen, Zhou Li, and Xing Li. "Hide and Seek: Revisiting DNS-based User Tracking." In 2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P), pp. 188-205. IEEE, 2022.