

Stalking Online: on User Privacy in Social Networks

ABSTRACT

With the extreme popularity of Web and online social networks, a large amount of personal information has been made available over the Internet. On the other hand, advances in information retrieval, data mining and knowledge discovery technologies have enabled users to efficiently satisfy their information needs over the Internet or from large-scale data sets. However, such technologies also help the adversaries such as web stalkers to discover private information about their victims from mass data.

In this paper, we study privacy-sensitive information that are accessible from the Web, and how these information could be utilized to discover personal identities. In the proposed scenario, an adversary is assumed to possess a small piece of "seed" information about a targeted user, and conduct extensive and intelligent search to identify the target over both the Web and an information repository collected from the Web. In particular, two types of attackers are modeled, namely tireless attackers and resourceful attackers. We then analyze detailed attacking mechanisms that could be performed by these attackers, and quantify the threats of both types of attacks to general Web users. With extensive experiments and sophisticated analysis, we show that a large portion of users with online presence are highly identifiable, even when only a small piece of (possibly inaccurate) seed information is known to the attackers.