Online social networks leak personal information to tracking sites, new study shows

August 24 2009

More than a half billion people use online social networks, posting vast amounts of information about themselves to share with online friends and colleagues. A new study co-authored by a researcher at Worcester Polytechnic Institute (WPI) has found that the practices of many popular social networking sites typically make that personal information available to companies that track Web users' browsing habits and allow them to link anonymous browsing habits to specific people.

The study, presented recently in Barcelona at the Workshop on Online Social Networks, part of the annual conference of the Association for Computing Machinery's Special Interest Group on Data Communications, is the first to describe a mechanism that tracking sites could use to directly link browsing habits to specific individuals.

"When you sign up with a social networking site, you are assigned a unique identifier," says Craig Wills, professor of computer science at WPI, who conducted the study with an industry colleague. "This is a string of numbers or characters that points to your profile. We found that when social networking sites pass information to tracking sites about your activities, they often include this unique identifier. So now a tracking site not only has a profile of your Web browsing activities, it can link that profile to the personal information you post on the social networking site. Now your browsing profile is not just of somebody, it is of you."

Like most commercial websites, online social networks use third-party tracking sites, called aggregators, to learn about the browsing habits of their visitors. Cookies are maintained by a [Web browser] and contain information that enable tracking sites to build profiles of the websites visited by a user. Each time the user visits a new website, the tracking site can review those cookies and serve up ads that might appeal to the user. For example, if the user frequently visits food sites, he or she might see an ad for a new cookbook.

Online networking sites have gone a step further by allowing for transmission of unique identifiers. It is a particularly troubling practice for two reasons, Wills says. "First," he notes. "users put a lot of information about themselves on social networking sites. Second, a lot of that information can be seen by other users, by default. There are mechanisms users can use to limit access to their information, but we found through previous research that most users don't take advantage of them." With a unique identifier, a tracking site could gain access to a user's name, physical address, email address, gender, birth date, educational and employment information, and much more.

With the "leakage" of this type personal information, there is a significant risk of having one's identity linked to an inaccurate or misleading browsing profile. Browsing profiles record the websites a particular computer has accessed, not who was using the computer at the time or why particular sites were chosen. According to Wills, this leaves room for inaccurate profiling by tracking sites, a situation that has the potential to lead to serious problems. When a computer is used by more than one person, or a person browses for curiosity rather than intent, it leaves room for misinterpretation, he notes. "Tracking sites don't have the ability to know if, for example, a site about cancer was visited out of curiosity, or because the user actually has cancer. Profiling is worrisome on its own, but inaccurate profiling could potentially lead to issues with employment, health care coverage, or other areas of our personal lives."

Wills says the researchers do not know what, if anything, tracking sites do with the unique identifiers that social networks transmit to them. They say they have communicated with all of the sites they studied to inform them about the privacy leakage, but have not heard back officially from any. "We are not saying that they are necessarily trying to leak private information," he says. "But once someone is in possession of your unique identifier, there is so much they can learn about you. And even if tracking sites do not use the information themselves, can they guarantee that it will never find its way into other hands? For these reasons, we feel this issue is something that we should to be concerned about."

The researchers also note that while users of social networking sites can protect themselves to some degree by limiting the amount of information they post and using the protections the sites make available to them to limit access to their information, the easiest way to prevent privacy leakage would be for social networking sites to stop making unique identifiers visible.

More information: View the full study here: conferences.sigcomm.org/sigcom … s/wosn/papers/p7.pdf

Source: Worcester Polytechnic Institute (news : web)