De-Anonymizing Social Network Users

Dixita Sharegar

Sunday, February 7, 2016

Millions of users use social networking sites to share personal and business related information on websites like Facebook, LinkedIn and Xing. Sites like Amazon and eBay also use the concepts of groups and are prone to such attacks. Although these websites try to protect the users data, there are always new ways to attack and extract sensitive information. The paper A Practical Attack to De-Anonymize Social Network Users presents a technique of obtaining the identity of a user from his group membership. If a user visits the attackers website, then the attacker can get enough information to uniquely identify the user. If not, then the attacker at least gets a smaller subset of data to work with.

Every social network model is a graph of V users and E edges representing friendship between them. Social networks also have groups G where many users with common interests or demographics join. A user can belong to multiple groups where every individual group is represented by g. To represent this data we use a vector :

$$T(v) := (T_g(v))_{g \epsilon G} \tag{1}$$

$$T_g(v) = \left\{ \begin{array}{c} \text{1 if v is a member of a group g} \\ \text{0 if v is not a member of a group g} \end{array} \right\}$$

$T(v)$ is called the group fingerprint of the user. This information will be used in the deanonymizing attack.

Web applications use HTTP GET parameters to communicate with the users. These hyperlinks contain user IDs or group IDs that help the attacker to identify a user or group. The first step to de-anonymizing attack is that the attacker sends a list of hyperlinks to the users browser. Next, by using client side scripting, he forces the users browser to check the list with the users browsing history. At the end, the attacker gets a list of visited URLs. This paper discusses 2 models, the basic attack and the improved attack.

1. Basic Attack: The attacker would first study the social network, the dynamic URL pattern, the numbering pattern of IDs etc. and then perform the attack. This attack is not feasible as every potential victims browser has to perform a lot of data processing and is time consuming.

2. Improved Attack: The attacker studies the social network like above. Then he steals the users browsing history and checks if group pages were recently accessed. If we find a matching page we assume that the user is a member of that group. We can now generate the candidte set and de-anonymize the user.

Web crawling techniques can also be used to get user data. A custom crawler and a third party crawler were used to test this attack on social networking sites like Facebook, LinkedIn and Xing .Data for 1.8 million unique users was obtained from 6,446 public and 108 private groups in Xing. The data gathered by the attacker can be inconsistent as users could have deleted their browsing history, members are added and deleted to groups, so data needs to be collected by the attacker on a regular basis.

We can mitigate the attacks either at the server side or client side. At the server side we should try using HTTP POST instead of GET as GET parameters are stored in browsing history. Users should disable browser history or use private browsing mode. At Xing they have added random numeric characters that represent current date and user number to their URLs to solve the problem.

References

[1] Wondracek, G.;Holz, T.;Kirda, E.;Kruegel, C., *A Practical Attack to De-anonymize Social Network Users*, 2010 IEEE Symposium on,vol., no., pp.223-238, 16-19 May 2010.