

Masquerade attacks based on user's profile

I. Razo¹, C. Mex-Perera¹ and R. Monroy²

¹ Center for Electronics and Telecommunications, ITESM, Campus Monterrey
Av. Eugenio Garza Sada 2501 Sur, Col. Tecnológico
Monterrey, N. L., CP 64849 Mexico

² Computer Science Department, ITESM, Campus Estado de Mexico
Carretera al lago de Guadalupe, Km. 3.5, Estado de Mexico, CP 52926, Mexico
{razozapata, carlosmex, raulm}@itesm.mx

Security in computer networks is an issue of special interest due to the importance of information inside organizations. Besides, a lot of communication actually involves computer networks, so it is necessary to provide security within.

There are several problems related to security in computer networks. In this work only detection of masquerade attacks is explored. A masquerade attack occurs when an illegitimate user tries to impersonate a legitimate user; therefore, the masquerade user gets the privileges from the legitimate user account. The task of detecting masquerade users is not easy since the masquerade user has yielded the name and password of a valid user (probably an administrator). However, detecting illegitimate users could be done if information about the behavior of the impersonate user is taken as a characteristic pattern which is valid only for this user.

Schonlau has developed a dataset to compare various intrusion detection methods [1]. Original Schonlau dataset (SEA) contains UNIX commands for 50 users. These commands are divided in blocks of 100 commands (sessions). Each user has 150 sessions. The first 50 sessions are not contaminated and constitute the training dataset. The last 100 blocks of each user may or may not be masquerade blocks. A block is either totally contaminated or legitimate from the user. SEA dataset comes with a matrix that shows where the masquerade sessions are located for each user.

Masquerade sessions in SEA come from working sessions of other users who had no intention to act as intruders. In order to avoid detection, the intruder must act like the legitimate user, that is, he must know the victim's work profile. This reason motivates the idea of including knowledge of the legitimate user behavior into masquerade sessions.

The present research is focused in the analysis of methods for detecting masquerade users. Performance of methods is studied with masquerade sessions which have local properties of each user, which is the knowledge about frequency of commands and repetitive sequences of commands (scripts).

Methods for detecting masqueraders may deal with frequency of commands typed by users and extraction of grammars for each user [2, 3, 4, 5]. Four methods are analyzed: two of them work with frequency (uniqueness and command frequencies) and the others work with rules obtained from users's behavior (customized grammars and hybrid grammars).

Including knowledge of valid users into masquerade sessions could be per-

formed in several ways, consequently, yielding several datasets. SEA-I is the first modification to SEA. SEA-I is built using statistical properties of users, so, a masquerade session can be built having the same probability distribution of commands than the probability distribution of the training data. In this case scripts are not taken into account when building the masquerade sessions. Scripts are used to build SEA-II. In this process scripts are generated in the following way: firstly extracting rules (grammars) from the training set, secondly rules could be sorted by frequency, length and priority (SEA-II could be based on these three criterions), and finally masquerade sessions are filled with rules.

Another way of building masquerade sessions is hiding attack sequences inside them. SEA-III is built by a method that allows to hide a desired sequence into a legitimate session [6]. In this way the attack is more realistic and with a certain level of intelligence. Once created the masquerade sessions for all users, they are located in exactly the same positions as the original dataset.

As mentioned before, performance of four methods is evaluated with this group of datasets. In order to obtain profiles of users, supervised learning is applied for each method. Afterwards, during the classification process, sessions labelled as legitimates are used to update the profile of the examined user.

Performance of methods is measured with ROC curves and show different properties. First, when sessions use probability distribution, methods based on grammars have better performance. Second, sessions created with scripts are extremely difficult to detect. However, a mixture of methods based on frequency and scripts could be considered in order to overcome this issue since the performance is highly reduced according to masquerader strategies. Third, hiding attack sequences into sessions has an important impact in the detection task, almost as scripts, although attack sequences are too small. Finally, as a general conclusion, sessions created with knowledge of users are very hard to identify and it is necessary to study other kind of techniques, possibly combination methods already mentioned.

References

1. Schonlau, M.: Masquerading used data. (Matthias Schonlaus home page) <http://www.schonlau.net>.
2. Schonlau, M., DuMouchel, W., Ju, W., Karr, A., Theus, M., Vardi, Y.: *Computer Intrusion: Detecting Masquerades*. Statistical Science 16 (2001) 1-17.
3. Schonlau, M., Theus, M.: *Detecting masquerades in intrusion detection based on unpopular commands*. Information Processing Letters 76 (2000) 33-38
4. Latendresse, M.: Masquerade detection via customized grammars. Second International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment. LNCS 3548, pp. 141-159, 2005.
5. Posadas R., Mex-Perera C., Monroy R., and Nolasco-Flores Juan,: *Hybrid Method for Detecting Masqueraders Using Session Folding and Hidden Markov Models*, MICAI: LNCS 4293, pp. 622-631, 2006.
6. Sufatrio and Roland H.C. Yap,: *Improving Host-Based IDS with Argument Abstraction to Prevent Mimicry Attacks*, RAID 2005, LNCS 3858, pp. 146-164, 2006.