Modeling User Search Behavior for Masquerade Detection*

Malek Ben Salem and Salvatore J. Stolfo

Computer Science Department
Columbia University
New York, USA
{malek,sal}@cs.columbia.edu

Abstract. Masquerade attacks are a common security problem that is a consequence of identity theft. This paper extends prior work by modeling user search behavior to detect deviations indicating a masquerade attack. We hypothesize that each individual user knows their own file system well enough to search in a limited, targeted and unique fashion in order to find information germane to their current task. Masqueraders, on the other hand, will likely not know the file system and layout of another user's desktop, and would likely search more extensively and broadly in a manner that is different than the victim user being impersonated. We identify actions linked to search and information access activities, and use them to build user models. The experimental results show that modeling search behavior reliably detects all masqueraders with a very low false positive rate of 1.1%, far better than prior published results. The limited set of features used for search behavior modeling also results in large performance gains over the same modeling techniques that use larger sets of features.

Keywords: masquerade detection, user profiling, search behavior, svm.

1 Introduction

The *masquerade attack* is a class of attacks, in which a user of a system illegitimately poses as, or assumes the identity of another legitimate user. Identity theft in financial transaction systems is perhaps the best known example of this type of attack. Masquerade attacks are extremely serious, especially in the case of an insider who can cause considerable damage to an organization. Their detection remains one of the more important research areas requiring new insights to mitigate against this threat.

A common approach to counter this type of attack, which has been the subject of prior research, is to apply machine learning (ML) algorithms that produce classifiers which can identify suspicious behaviors that may indicate malfeasance of an impostor. We do not focus on whether an access by some user is authorized

* Support for this work has been partially provided by a DARPA grant ADAMS No. W911NF-11-1-0140.

R. Sommer, D. Balzarotti, and G. Maier (Eds.): RAID 2011, LNCS 6961, pp. 181–200, 2011.
© Springer-Verlag Berlin Heidelberg 2011

since we assume that the masquerader does not attempt to escalate the privileges of the stolen identity, rather the masquerader simply accesses whatever the victim can access. However, we conjecture that the masquerader is unlikely to know the victim's search behavior when using their own system which complicates their task to mimic the user. It is this key assumption that we rely upon in order to detect a masquerader. The conjecture is backed up with real user studies. Eighteen users were monitored for four days on average to produce more than 10 GBytes of data that we analyzed and modeled. The results show that indeed normal users display different search behavior, and that that behavior is an effective tool to detect masqueraders. After all, a user will search within an environment they have created. For example, a user searches for a file within a specific directory, or a programmer searches for a symbol within a specific source code file. We assume the attacker has little to no knowledge of that environment and that lack of knowledge will be revealed by the masquerader's abnormal search behavior. Thus, our focus in this paper is on monitoring a user's behavior in real time to determine whether current user actions are consistent with the user's historical behavior, primarily focused on their unique search behavior. The far more challenging problems of thwarting mimicry attacks and other obfuscation techniques are beyond the scope of this paper.

Masquerade attacks can occur in several different ways. In general terms, a masquerader may get access to a legitimate user's account either by stealing a victim's credentials, or through a break in and installation of a rootkit or key logger. In either case, the user's identity is illegitimately acquired. Another perhaps more common case is laziness and misplaced trust by a user, such as the case when a user leaves his or her terminal or client open and logged in allowing any nearby coworker to pose as a masquerader.

In this paper we extend prior work on modeling user command sequences for masquerade detection. Previous work has focused on auditing and modeling sequences of user commands including work on enriching command sequences with information about arguments of commands [15,10,18]. We propose an approach to profile a user's search behavior by auditing search-related applications and accesses to index files, such as the index file of the Google Desktop Search application. We conjecture that a masquerader is unlikely to have the depth of knowledge of the victim's machine (files, locations of important directories, available applications, etc.), and hence, a masquerader would likely first engage in information gathering and search activities before initiating specific actions. To this extent, we conduct a set of experiments using a home-gathered Windows data. We model search behavior in Windows and test our modeling approach using our own data, which we claim is more suitable for evaluating masquerade attack detection methods.

The contributions of this work are:

- A **small set of search-related features** used for effective masquerade attack detection: The limited number of features reduces the amount of sampling required to collect training data. Reducing the high-dimensional modeling space to a low-dimensional one allows for the improvement of