A Behavior-Based Approach to Securing Email Systems

Salvatore J. Stolfo, Shlomo Hershkop, Ke Wang,
Olivier Nimeskern, and Chia-Wei Hu

450 Computer Science Building
Fu Foundation School of Engineering & Applied Science
Computer Science Dept., Columbia University, USA
{sal,shlomo,kewang,on2005,charlie}@cs.columbia.eduz

Abstract. The Malicious Email Tracking (MET) system, reported in a prior publication, is a behavior-based security system for email services. The Email Mining Toolkit (EMT) presented in this paper is an offline email archive data mining analysis system that is designed to assist computing models of malicious email behavior for deployment in an online MET system. EMT includes a variety of behavior models for email attachments, user accounts and groups of accounts. Each model computed is used to detect anomalous and errant email behaviors. We report on the set of features implemented in the current version of EMT, and describe tests of the system and our plans for extensions to the set of models.

1 Introduction

The *Email Mining Toolkit* (EMT) is an offline data analysis system designed to assist a security analyst compute, visualize and test models of email behavior for use in a MET system [0]. In this paper, we present the features and architecture of the implemented and operational MET and EMT systems, and illustrate the types of discoveries possible over a set of email data gathered for study.

EMT computes information about email flows from and to email accounts, aggregate statistical information from groups of accounts, and analyzes content fields of emails without revealing those contents. Many previous approaches to "anomaly detection" have been proposed, including research systems that aim to detect masqueraders by modeling command line sequences and keystrokes [0,0].

MET is designed to protect user email accounts by modeling user email flows and behaviors to detect misuses that manifest as abnormal email behavior. These misuses can include malicious email attachments, viral propagations, SPAM email, and email security policy violations. Of special interest is the detection of polymorphic virii that are designed to avoid detection by signature-based methods, but which may likely be detected via their behavior.

The finance, and telecommunications industries have protected their customers from fraudulent misuse of their services (fraud detection for credit card accounts and telephone calls) by profiling the behavior of individual and aggregate groups of customer accounts and detecting deviations from these models. MET provides behavior-based protection to Internet user email accounts, detecting fraudulent misuse and policy violations of email accounts by, for example, malicious viruses.

A behavior-based security system such as MET can be architected to protect a client computer (by auditing email at the client), an enclave of hosts (such as a LAN

V. Gorodetsky et al. (Eds.): MMM-ACNS 2003, LNCS 2776, pp. 57–81, 2003.

with a few mail servers) and an enterprise system (such as a corporate network with many mail servers possibly of different types).

The principle behind MET's operation is to model baseline email flows to and from particular individual email accounts and sub-populations of email accounts (eg., departments within an enclave or corporate division) and to continuously monitor ongoing email behavior to determine whether that behavior conforms to the baseline. The statistics MET gathers to compute its baseline models of behavior includes groups of accounts that typically exchange emails (eg., "social cliques" within an organization), and the frequency of messages and the typical times and days those messages are exchanged. Statistical distributions are computed over periods of time, which serve as a training period for a behavior profile. These models are used to determine typical behaviors that may be used to detect abnormal deviations of interest, such as an unusual burst of email activity indicative of the propagation of an email virus within a population, or violations of email security policies, such as the outbound transmission of Word document attachments at unusual hours of the day.

EMT provides a set of models an analyst may use to understand and glean important information about individual emails, user account behaviors, and abnormal attachment behaviors for a wide range of analysis and detection tasks. The classifier and various profile models are trained by an analyst using EMT's convenient and easy to use GUI to manage the training and learning processes. There is an "alert" function in EMT which provides the means of specifying general conditions that are indicative of abnormal behavior to detect events that may require further inspection and analysis, including potential account misuses, self-propagating viral worms delivered in email attachments, likely inbound SPAM email, bulk outbound SPAM, and email accounts that are being used to launch SPAM.

EMT is also capable of identifying similar user accounts to detect groups of SPAM accounts that may be used by a "SPAMbot", or to identify the group of initial victims of a virus in a large enclave of many hundreds or thousands of users. For example, if a virus victim is discovered, the short term profile behavior of that victim can be used to cluster a set of email accounts that are most similar in their short term behavior, so that a security analyst can more effectively detect whether other victims exist, and to propagate this information via MET to limit the spread and damage caused by a new viral incident.

2 EMT Toolkit

MET, and its associated subsystem MEF (the Malicious Email Filter) was initially conceived and started as a project in the Columbia IDS Lab in 1999. The initial research focused on the means to statistically model the behavior of email attachments, and support the coordinated sharing of information among a wide area of email servers to identify malicious attachments. In order to properly share such information, each attachment must be uniquely identified, which is accomplished through the computation of an MD5 hash of the entire attachment. A new generation of polymorphic virii can easily thwart this strategy by morphing each instance of the attachment that is being propagated. Hence, no unique hash would exist to identify the originating virus and each of its variant progeny. (It is possible to identify the progenitor by analysis of entry points and attachment contents as described in the Malicious Email Filter paper [0].)