

基于涉密网的计算机取证模型研究

王素芳¹, 高美真¹, 苏继斌²

(1 焦作师范高等专科学校 计算机与信息工程系, 河南 焦作 454000; 2 71391 部队司令部, 河南 开封 475003)

摘 要: 设计一种针对涉密网的基于 Agent 的动态远程控制取证模型, 采取 Agent 的分布式数据采集策略, 引入 syslog 协议实时高效传输相关日志, 将计算机取证与入侵检测技术相结合, 动态获取入侵证据, 提高证据的可信性、有效性及证明能力。

关键词: 网络安全; 计算机取证; 涉密网

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2010)05-0202-04

Study on the Model of Computer Forensics Based on the Secret-Related Network

WANG Su-fang¹, GAO Mei-zhen¹, SU Ji-bin²

(1 Department of Computer & Information Engineering, Jiaozuo Teachers College, Jiaozuo 454001, China;

2 Headquarters of Unit No. 71391, Kaifeng 475003, China)

Abstract: Combining with computer forensic technique and intrusion detection technique, a distributed dynamic computer forensics model based on multi-Agent is presented. The distributed data collection policy is adopted, and introduced the protocol of syslog to translate the related log real-time and efficiently, so that the range of data collection is extended. The dynamic intrusion detection system provides real-time evidences of high legal stringency. The evidence-combined data analysis technique is adopted to decrease the rate of false alarm and enhance the validity of the evidences.

Key words: network security; computer forensics; security-network

1 引言

随着计算机的应用与互联网的发展, 涉密网因其特殊性不可避免地与 Internet 互联. 计算机取证技术^[1]是指对存储在计算机系统或网络设备中潜在电子证据的识别、收集、保护、检查和分析以及法庭出示的过程, 通常是对存储介质、日志的检查和与分析. 与传统证据一样, 电子证据也必须是可信的、准确的和完整的^[2]. 计算机取证目前研究成果有 NTI 的 Net Threat Analyzer^[3]、Guidance software 的 Encase^[4]、Vogon 的 Flight Server, Sandstorm 的 Net intercept^[5]、金诺网安的 DiskForen 等, 但通过研究发现, 在取证工具研发方面, 静态取证工具较多, 动态实时取证系统较少; 在取证理论技术方面, 对于电子

证据的获取, 主要是与现有入侵检测、蜜罐(网)等网络安全技术的结合, 而这些安全技术的局限也就成为了相应取证系统的局限, 如取证秘密性不强、可靠性不高; 在证据获取的途径方面, 主要是被动获取方式, 少有主动获取方法; 在取证系统应用方面, 主要是对本地系统的保护式取证, 少有对它系统的监控式的网络取证.

针对以上问题, 在文献[6-7]的研究基础上, 设计一种针对涉密网的基于 Agent 的动态远程控制取证模型, 将 IDS、SYSLOG 协议等运用于计算机取证, 通过 IDS 检测入侵, 一方面可及时获取入侵者的犯罪证据, 另一方面可以通过 IDS 的入侵响应模块, 对入侵行为采取相应的响应措施, 从而降低入侵损失; 通过 SYSLOG 协议, 实时高效地传输各种日

志信息;模型通过多 Agent 协作,实现分布式环境下,电子证据的动态收集、动态获取;通过对对证据的融合分析,有效提高证据的准确性与可信性。

2 模型设计与分析

2.1 模型原型

文中提出的基于 Agent 的动态远程控制取证模型原型如图 1 所示。在逻辑结构上,模型分为证据收集部分与证据分析两大部分。

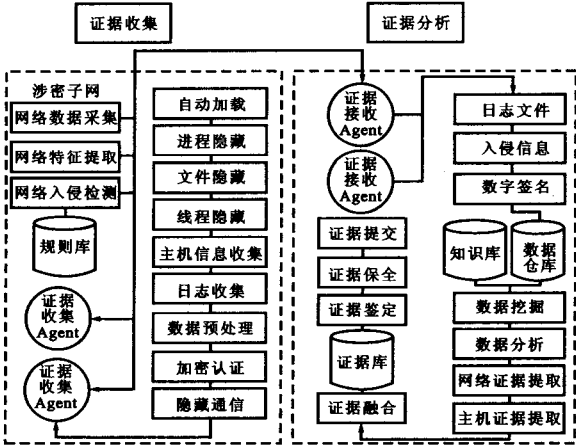


图 1 基于 Agent 的动态远程控制取证模型

2.2 系统结构

针对涉密网取证,文中给出取证的系统结构如图 2 所示。

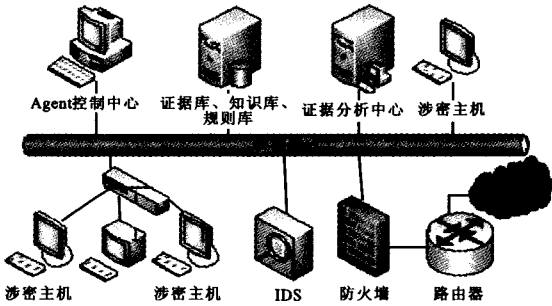


图 2 针对模型的取证系统结构图

Agent 控制中心主要功能包括取证控制、远程触发、身份认证、隐蔽通信等子模块。分别完成对取证内容(取什么、取多少)的控制,对被取证端命令的发布;对被取证端取证程序的触发;取证端与被取证端的身份认证和隐蔽通信等。

数据库服务器含有规则库、知识库、证据仓库、证据库等。规则库存放 SNORT 检测规则;知识库存放与入侵规则相关的专家知识;证据仓库存放 Agent 采集的 SYSLOG 相关日志;证据库则存放经数

据挖掘与融合后可以作为司法鉴定的证据。

证据分析中心主要将 Agent 传递过来的原始取证证据进行证据过滤、分类、融合、认证、保全等。

2.3 证据收集模块

(1) 自动加载. Agent 一旦接到 Agent 控制中心指令或检测到设备异常,便通过修改被取证端注册表和注册表隐藏技术保证取证程序的自动启动和运行. Agent 终端以系统服务的方式长期运行一个守护程序. 该程序不间断地采集被取证机相关日志文件,键盘操作记录以及其他现场证据的采集。

(2) 进程隐藏. 为防止入侵者警觉,通过对象隐藏技术实现取证程序进程的隐藏。

(3) 文件隐藏. 通过调用 API 特定函数,修改输出缓冲区记录方式实现程序文件隐藏。

(4) 线程监控. 通过线程监控技术实现程序的抗查杀功能。

(5) 主机信息收集. 利用 Agent 功能对主机的相关参数及信息进行收集。

(6) 日志收集. 利用 SYSLOG 协议将网络设备相关日志收集存储及传输至证据分析中心的原始证据数据库. 关键技术分析中的 3.1 节将详细探讨本日志收集子模块。

(7) 数据预处理. 由于日志文件非常庞大,且日志文件记录中的一些数据属性与安全性无关,为了提高数据传输速度和主机入侵检测的效率,数据预处理 Agent 对收集到的日志文件进行精简,按统一的格式提交给主机入侵检测 Agent。

(8) 加密认证. 由于计算机取证的整个过程必须具有不可篡改性,因此,数据在传输过程中要具有抵抗黑客非法入侵的能力,同时还要提高对远程数据传输的保密性,以避免在传输中遭受非法窃取。

(9) 网络入侵检测与网络特征参数提取. 网络数据采集 Agent 每截取一个数据包后,对其进行解包,然后送给网络特征提取 Agent. 对于每一个数据包,网络入侵检测 Agent 根据其特征,与规则库的相应规则进行匹配,如果满足条件,则认为有入侵事件发生,将该数据传送到证据分析中心保存,以便作进一步的证据分析提取,否则丢弃。

(10) 隐藏通信. 通过隧道技术和 DES 加密技术,与证据接收端商议密钥并传递,实行取证数据的安全通信。

以上诸模块的结合不仅保证了被取证端取证程序的秘密生存与运行,还保证了取证端能实时与被取证端的连接与通信;并且取证端能够根据自身要

求进行取证控制。

2.4 证据分析模块

收集 Agent 自动收集网络和主机日志及相关信息,对收集到的信息进行预处理后,将收集到的证据经 SYSLOG 协议加密传输至证据分析中心的证据数据仓库。通过专家知识库进行数据挖掘与数据分析,提取网络证据与主机证据,后经证据融合,存至证据库中。可根据司法需要,可进一步进行证据鉴定、证据保全和证据提交。根据需要,取证证据可做进一步的保全、检索、备案等处理,也可根据入侵部位对网络有针对性地加固。

(1) 日志文件

由 Agent 收集到的日志包括 OS 日志、网络设备日志、数据库日志及典型应用日志等。日志信息经格式化处理后^[8]存储到数据仓库中。为保证原始证据的不可抵赖性,可对日志进行数字签名。

(2) 数据库

在模型中涉及证据相关的日志仓库、规则库、知识库和证据库。规则库主要是检测入侵检测匹配数据的,这里用 2.8.3 版本的 SNORT^[7]提供的。日志仓库存储采集到的原始日志数据,日志仓库经知识库的专家智能分析系统进行数据挖掘与融合成一定法律效力的证据库。

(3) 数据挖掘

采用数据挖掘中的聚类方法对日志数据进行分析,详见关键技术分析中的 3.3 节。

(4) 证据提取与融合

将上一步挖掘出的日志证据结合网络与主机相关参数证据的提取与融合。

(5) 证据保全与提交

根据司法需要,可对证据库中的证据进行保全与提交。保全可采取加密、异地存储等手段。提交可在由于收集到的信息来自于多台主机和多个子网,分析中心通过证据融合,可以发现单台主机无法发现的协作式攻击等更复杂的攻击行为,并提取其入侵证据。

3 关键技术分析

3.1 证据信息提取

基于 Agent 动态远程控制取证模型中的证据信息包含系统日志、可选的蜜罐日志、IDS 日志等。日志的收集由 Syslog 协议实现。

Syslog 协议是加里福尼亚大学伯克立软件分布研究中心(BSD)开发的,现已形成工业标准。在路由

器、交换机、服务器等网络设备中,syslog 记录着系统中的任何事件,管理者可以通过查看系统记录,随时掌握系统状况。它能够接收远程系统的日志记录,在一个日志中按时间顺序处理包含多个系统的记录,并以文件形式存盘。同时不需要连接多个系统,就可以在一个位置查看所有的记录。syslog 使用 UDP 作为传输协议,通过目的端口 514(也可以是其他定义的端口号),将所有安全设备的日志管理配置发送到安装了 syslog 软件系统的日志服务器,syslog 日志服务器自动接收日志数据并写到日志文件中。

在 Agent 控制中心中配置 syslog 日志服务器,Agent 的部分功能由 syslog 实现。网络设备将日志信息以用户数据报协议(UDP)方式传送到远端服务器(Agent 控制中心),远端接收日志服务器必须通过 syslogd 监听 UDP 端口 514,并根据 syslog.conf 配置文件中的配置处理本机,接收访问系统的日志信息,把指定的事件写入特定文件中,供后台数据库管理和响应之用。意味着可以让任何事件都登录到一台或多台服务器上,以备后台数据库用 off-line(离线)方法分析远端设备的事件。

3.2 证据数据库创建

日志仓库可存放在服务器的 MySQL 数据库中。这里重点给出模型中引入的 SYSLOG 协议相关方法。修改 sqlsyslogd.sql 与 config 文件方法如下:

① 修改 sqlsyslogd.sql 文件

```
# vi sqlsyslogd.sql
create database sqlsyslogd;
use sqlsyslogd;
create table logs (
  Id int(10) NOT NULL auto_increment,
  Timestamp varchar(16),
  Host varchar(50),
  Prog varchar(50),
  Mesg text,
  PRIMARY KEY (id)
);
use mysql;
create user sqlsyslogd@localhost identified by 'foo';
grant all on sqlsyslogd.* to sqlsyslogd;
flush privileges;
# sql -u root -p < sqlsyslogd.sql
# vi /usr/local/etc/sqlsyslogd.conf
foo
```

② 修改 syslog-ng.conf 文件,添加下面几行

```
destination mysql {
program("/usr/local/sbin/sqlsyslogd -u sqlsyslogd \
```

```

    - t logs sqlsyslogd - p");
};
log {
source(all);
destination(mysql);
}.

```

3.3 证据分析

模型中的证据分析首先将采集到的原始证据存储证据仓库中,然后经知识库中的专家知识经人工智能方法识别出可能的证据存储在证据库中.文中采用数据挖掘中的聚类方法对日志数据进行分析.聚类就是将物理或抽象对象的集合分组成由类似的对象组成的多个类的过程.由聚类生成的簇是一组数据对象的集合,这些对象与同一簇中的对象彼此相似,与其他簇中的对象相异.聚类分析通过获得数据分布的情况,将数据对象分组成多个簇,观察每个簇的特点,集中对特定的某些簇做出进一步的分析.文中通过相似性和距离函数对日志数据进行聚类分析,主要包括以下两个方面:

(1) 日志属性独立性分析. 设属性 A 和 B , $A = \{A_i, 1 \leq i \leq n\}$, $B = \{B_i, 1 \leq i \leq n\}$, A 和 B 是属性值,属性相关性度量公式^[8]为 $\gamma_{A,B} = [\sum (A_i - \bar{A})(B_i - \bar{B})] / [(n-1)\sigma_A\sigma_B]$. 其中: $\bar{A} = (\sum A_i) / n$

表示 A 的平均值; $\sigma_A = \sqrt{\sum (A_i - \bar{A})^2 / (n-1)}$ 表示 A 的标准差. 根据 $\gamma_{A,B}$ 的值判断日志属性之间的相关性,若 $\gamma_{A,B} \rightarrow 0$ 就说明属性 A, B 是独立的属性. 根据这个方法提取日志属性四个独立的主要属性,即类别、数字、时间和字符串.

(2) 日志记录聚类分析. 设样本 $x = \{x_1, x_2, \dots, x_m\}$ 和 $y = \{y_1, y_2, \dots, y_m\}$. 其中: $x_i, y_i (1 \leq i \leq m)$ 为样本指标; 明氏距离 $D_q(x, y) = (\sum_i |x_i - y_i|^q)^{1/q}$. 若距离 $D_q(x, y)$ 越近,则样本 x 与 y 越相似; 若距离 $D_q(x, y)$ 越远,则样本 x 与 y 差异越大. 这样就可以通过计算新的日志记录与规则库中已有的聚类 c 中的标准记录 Y 之间的距离 $D_q(x, y)$, 并与阈值 A 进行比较. 若 $D_i(x, y) < \lambda$, 则说明属于聚类 c ; 否则不属于聚类 c . 因此,可以根据聚类 c 的性质来判断日志记录 x 的性质^[9-10].

通过上述方法可对日志数据库中大量日志数据进行分析,找出系统和网络中的异常情况,必要时调用入侵响应模块来保证网络系统的安全,及时跟踪监测与取证,提取关键证据,为下一步展开司法调查提供足够的可信的证据.

4 结束语

通过理论分析与实验评测,与其他取证模型相比,基于 Agent 动态远程控制取证模型引入 syslog 协议、Agent 入侵检测等方法,将代理分布至涉密网中的设备,不仅取证范围广,而且避免了分析中心的性能瓶颈,有效提高了系统的处理速度和能力.模型在对涉密网中系统用户行为及网络流量进行监控的同时,对易丢失数据进行同步记录,实时获取入侵证据,解决了事后证据收集的困难,提高了证据的证明能力.采用证据融合的数据分析技术,增加了证据的可信度,提高了证据的有效性.

参考文献:

- [1] 刘宝旭,马建民,池亚平. 计算机网络安全应急响应技术的分析与研究[J]. 计算机工程, 2007, 33(10): 128-130.
- [2] 丁丽萍,王永吉. 计算机取证的相关法律技术问题研究[J]. 软件学报, 2005, 16(2): 260-274.
- [3] Kruse, Heiser, John R, et al. Computer forensics & security software tools[EB/OL]. (2002-06-10). <http://www.forensics-intl.com/thetools.html>
- [4] Wei W. EnCase forensic functionality[EB/OL]. (2003-07-14). <http://www.guidance-software.com/lawenforcement/ef-index.asp>.
- [5] Corey V, Peterman C, Shearin S, et al. Network forensics analysis[EB/OL]. (2002-06-14). <http://www.sandstorm.net/products/#thetop>.
- [6] 杨卫平,段丹青. 基于多 Agent 的分布式计算机动态取证模型研究[J]. 计算机应用与软件, 2008, 25(3): 81-82.
- [7] 史伟奇,张波云,谢冬青. 基于远程控制技术的动态取证系统[J]. 计算机工程, 2007, 33(16): 117-119.
- [8] 刘必雄,许榕生,杨泽明. 基于多 agent 与 proxy 技术的日志提取分析系统设计[J]. 计算机应用研究, 2008, 25(2): 619-621.
- [9] Snort users manual 2.8.3. the snort project[EB/OL]. (2008-09-17)[2009-03-22]. <http://www.snort.org/docs/snort-htmanuals/htmanual-283>.
- [10] Han Jiawei, Kamber M. Data mining: concept and techniques[M]. 范明,小峰,译. 北京:机械工业出版社, 2001: 223-236.

作者简介:

王素芳 女, (1973-), 硕士, 讲师. 研究方向为网络管理与网络安全.

高美真 女, (1970-), 硕士, 副教授.

苏继斌 男, (1975-), 硕士, 工程师.