文章编号:1001-9081(2005)06-1290-04

# 分布式计算机动态取证模型

梁昌宇,吴 强,曾庆凯

(南京大学 计算机科学与技术系,江苏 南京 210093)

(cyliangzy@bbs. nju. edu. cn)

摘要:提出一个分布式计算机动态取证模型,在被保护系统中进行实时动态的证据采集,将证据及时、完整地存储到安全的证据中心,为证据分析和提取工作提供可信的原始证据数据。

关键词: 计算机取证; 动态取证; 证据完整性中图分类号: TP309.2 文献标识码: A

# Distributed and dynamic computer forensic model

LIANG Chang-Yu, WU Qiang, ZENG Qing-Kai

( Department of Computer Science and Technology, Nanjing University, Nanjing Jiangsu 210093, China)

Abstract: The new forensic model was proposed here. Camparing with traditional computer forensic model, the major differenced between these two models lies on the distributed structure and the mechanism of dynamical data gathering. With this two characteristics, forensics system based on the new model could gather real-time evidences dynamically in a distributed system, and save this evidences in a safe place in time. So unauthorised deletion, change to evidences could be detected and prevented. Then the stored evidences could be used for further analysis and review.

Key words: computer forensics; dynamic forensics; integrality of evidence

# 0 引言

随着信息技术的发展与深入应用,与计算机相关的案件,如电子商务纠纷、计算机犯罪等也不断出现,计算机犯罪已成为刑事案件的一个新方向<sup>[1]</sup>。目前的信息安全技术,如防火墙<sup>[2]</sup>、数据加密<sup>[3]</sup>、入侵检测<sup>[4,5]</sup>等大多着眼于对入侵的防范。但是,这些安全技术和手段并不能够防止所有入侵。计算机取证是将计算机调查和分析技术应用于对潜在的、有法律效力的证据的确定与获取之上。当系统遭受入侵时,计算机取证技术能够及时发现,并提取入侵证据以追查入侵来源,清除入侵行为对计算机系统产生的负面影响,尽量恢复受破坏的系统。

计算机取证的手段、方法与传统的物证取证方法不同,它涉及到新的证据形式,即电子证据。与传统证据一样,电子证据也必须是可信的、准确的、完整的。随着计算机犯罪日益增多,电子证据正逐渐成为新的诉讼证据之一<sup>[6]</sup>。电子证据存在于计算机及相关外围设备(包括网络介质)中,对电子证据的获取需要提取存在于计算机系统中的资料,甚至需要从已被删除、加密或破坏的文件中重新获得信息。如何及时地将这些电子证据保护起来,保证其正确性和完整性,提供有效的方法呈现和分析数字证据,找出入侵者,解释入侵过程,是计算机取证的主要工作。

### 1 相关研究工作

目前,计算机取证技术的研究大部分为静态分析方法,即 事件发生后对目标系统的静态分析。这种静态分析的对象主 要为目标系统中的各种文件,包括现存的正常文件、已经被删 除但仍存在于磁盘上(即还没有被新文件覆盖)的文件、隐藏文件、受到密码保护的文件和加密文件等等。这些文件包含了系统日志信息与应用日志信息。静态分析也对位于磁盘特殊区域中的所有相关数据进行分析、提取,包括未分配的磁盘空间以及分配给文件的最后一簇中未被文件使用的剩余空间,其中可能包含先前磁盘操作或者文件系统遗留下来的信息。国内外已经开发了不少静态取证的工具,包括数据硬拷贝和恢复工具,如 Safe Back、SnapBack DatArrest、PowerQuest、DriveCopy等;数据分析取证工具,如 Encase、Mycroft等;日志分析工具,如 Webtrends Tools、Tcpshow等。但这些静态取证技术的共同不足之处在于没有针对攻击行为进行实时监视和实时记录。

随着计算机人侵攻击技术的发展变化,静态取证方法已无法满足要求。因为人侵者在人侵成功后,完全可能将人侵系统的相关数据销毁,以此来隐匿行踪。人侵相关的证据数据可能已经被人侵者恶意地删除或者篡改了。这样,事后的静态取证就不可能取得完整准确的人侵相关数据。因此,需要对静态取证技术进行改进,以适应计算机犯罪技术的这种发展。本文提出了一种分布式动态计算机取证模型,采用了动态取证、分布式证据收集和组织技术,具有更强的证据获取能力,可有效保障证据的安全性。

# 2 系统模型

### 2.1 取证模型结构

根据计算机取证的特点和要求,分布式计算机取证系统 应包括实时证据收集、安全证据保护、证据处理和分析提取等 功能,以构建包括分布式数据采集、集中式处理和管理的证据

收稿日期;2004-11-10;修订日期;2005-02-23 基金项目;863 计划项目(2004AA147070);国家自然科学基金资助项目(60473053) 作者简介:梁昌宇(1980-),男,贵州遵义人,硕士研究生,主要研究方向:计算机安全、操作系统; 吴强(1979-),男,江苏常州人,硕士研究生,主要研究方向:计算机安全、操作系统; 曾庆凯(1963-),男,安徽来安人,教授,主要研究方向:网络人侵检测系统、安全操作系统、安全体系结构等。

收集平台。而且要保障数据安全传输、安全存储和数据完整性。分布式动态计算机取证模型采用 C/S 结构,包括了以下的几个组成部分:

- 1) 基于分布式多 Agent 机制的证据采集器,进行证据的 分布式收集。
  - 2) 集中的证据存储中心,进行安全的证据集中存储。
- 3)专用的通信协议,以保证证据传输的安全性和完整性。
- 4)证据分析中心,向用户提供最终的经过分析推理的有效的证据。
- 5) 用户界面以及用户配置管理接口,以便用户对系统进行管理。

分布式动态计算机取证模型如图 1 所示。



#### 2.2 功能模块和结构设计

### 2.2.1 分布式证据收集器

证据收集器置于被保护系统中,实时采集各种可能的证据信息。收集的证据信息可分为系统安全信息和文件信息两大类。系统安全信息包括入侵检测信息、用户权限操作、系统安全操作和重要数据的操作等;文件信息包括文件保护信息、系统日志、应用日志、重要删除文件和用户定制信息等。不同的证据信息,采集的方式与途径都不一样,所以针对收集的各类信息,必须设计各自的证据收集器。这些证据收集器包括入侵检测证据收集器、用户登录证据收集器、应用日志证据收集器等等。

分布式证据收集器收集的取证信息多种多样,证据信息各有其特征。如果不加处理就把这些取证信息发送到证据存储中心,将导致证据信息的杂乱不堪,这会对后续的存储、分析工作带来困难。因此,按照取证信息的来源,证据被分成了若干类,即不同的证据收集器将产生不同的证据类,如入侵检测证据类、用户登录证据类、应用日志证据类等等。

分布式证据收集器的结构如图 2 所示。



图 2 分布式证据收集器结构图

当证据发送到证据存储中心时,发送的除了证据本身,也包括了证据的所属分类信息,这样,后续的证据存储,证据分析等工作也可以依照分类来进行,从而更加有利于证据的管理与操作。对上述的证据的分类采集与分类存储工作的具体操作,可以示例如下:

- 1) 分布式证据采集器采集到证据信息 Evi。
- 2) 根据证据来源,将分类信息加入证据后将得到(Evi+Cls)。
  - 3) 将(Evi+Cls)交给通信模块,发送到证据存储中心。
  - 4) 证据存储中心收到(Evi + Cls)。
  - 5) 根据 Cls,证据存储中心对证据进行分类存储。

#### 2.2.2 证据通信协议

证据信息是由分布式证据收集器在被保护系统中收集起来的,然后再发送到证据存储中心,而证据通信协议就是规定证据信息如何在证据收集器和证据存储中心之间进行传输的一种规范。在证据由证据收集器到证据存储中心的传输过程中,将会面临证据信息的安全性和完整性威胁。因此,通信协议必须提供一种安全机制,保证证据信息在传输的过程中不会受到非法的窃取和篡改。

#### 2.2.3 证据存储中心

证据存储中心接受并保存大量来自多台目标主机的证据信息,为证据进一步的分析取证作好准备。因此证据的组织、存储和管理是非常重要的。证据的保存处理包括对证据的存储、查询、导出和导入等。

对证据信息这样的结构型数据,采用数据库技术对它进行组织与管理是非常合适的。数据库提供了一个功能强大的接口,可以快捷,高效的对存储的结构化数据进行查询,删除等操作。而且,数据库本身还提供了存取控制技术,以保证证据访问的合法性。

同样,在证据存储中心,涉及到了证据的存储、查询,也会有证据信息分类的问题。证据的分类是在分布式证据收集器中完成的,证据的分类信息会与证据本身一起从证据收集器发送过来。因此可以利用该分类信息对证据进行分类存储及查询。

# 2.2.4 证据的分析工具

证据存储中心中包含的信息量是巨大的,要完全依靠人工进行分析,其工作量将不可接受,所以还需要一种对计算机取证获取的信息进行自动分析的方法。利用大量各类原始证据数据进行联合推理得到有效的证据是系统的最终目标。不同的证据数据来源不同,所包含证据信息也不同。在证据的计算机表示中组成该证据信息的各个域也是不一样的,所以,针对不同类型的证据信息,需要设计不同的分析工具。分析工具提供了基于时间段、用户(或 IP 地址)等信息的推理方法,使得用户可以在数量庞大的证据信息中,进行快速的定位。这种分析可以是在同类证据中展开,也可以在不同的证据类之间展开。

#### 2.2.5 用户界面以及用户配置管理

用户界面向用户提供一个友好的人机接口。通过用户界面,用户可以方便对取证系统进行相应的操作与配置,使得系统能够更好地适应特定用户的计算机环境,也使得系统可以满足用户的不同的需求。

## 3 系统实现

根据上述系统模型,基于 Linux 系统环境实现了一个分布式动态取证系统。

### 3.1 分布式证据收集器的实现

表1 证据收集器

证据收集器	产生的证据类		
FTP 证据采集器	EVIDENCE_FTPSERVER		
HTTP 证据采集器	EVIDENCE_HTTPSERVER		
WTMP 证据采集器	EVIDENCE_WTMP		
防火墙证据采集器	EVIDENCE_FIREWALL		

证据收集器收集的事件信息非常广泛,其中包含了 Linux 系统的各种日志信息,还包含了人侵检测系统产生的事件信 息。而且,用户还可以对需要收集的信息的种类进行定制。 列举一部分已经实现的证据收集器如表1。

# 3.2 专用的证据通信协议的实现

在通信协议的实现中,采用了循环冗余校验和 MD5 加密算法<sup>[7]</sup>。该通信协议的具体操作步骤描述如下:

在证据的收集以及发送方,进行以下步骤:

- 1)证据通信模块从证据采集器接收到将要传输的证据 信息 Evi + Cls。
  - 2) 为该证据计算校验信息,记为 VER(Evi+Cls)。
- 3)将证据及其校验信息作为一个整体进行加密处理,得到最终将会发送到证据存储中心的数据,记为 ENCRY(VER (Evi + Cla) + (Evi + Cla))。
  - 4) 发送数据 ENCRY(VER(Evi + Cls) + (Evi + Cls))。 在证据的接收方即证据存储中心,进行以下步骤:
- 1) 收到加密数据 ENCRY (VER (Evi + Cls) + (Evi + Cls))。
- 2) 对数据进行解密,将会得到 VER(Evi + Cls) + (Evi + Cls)。
- 3) 用与分布式证据采集器相同的方式为(Evi + Cla) 计算校验值,然后再与已经接收到的 VER(Evi + Cla) 进行比较。如果这两个校验值不一致,就可以确定证据在传输的过程中发生了错误或者已经被非法篡改,因此可以采取相应的措施,如要求证据收集器重传、报告管理员等。如果两个校验值一致,就可以将接收到的证据保存在永久介质中。

从以上的证据通信协议的设计中可以看出:一方面,证据信息在网络中传输时采用的是密文方式,因此,即使被非法窃取,由于窃取者无法对窃取到的密文解密,因此也不会知道证据的内容,不会导致泄密,从而保证了证据的安全性;另一方面,如果窃取者篡改了证据信息,将会导致证据存储中心根据接收到的证据计算的校验值与接收到的校验值不一致,使得篡改证据的行为不会达到目的,从而保证了证据的完整性。

### 3.3 证据的存储中心开发

在证据存储中心,采用了 mysql 数据库对最终的证据进行存储与组织。Mysql 中存储的证据是一个包含了多个域的记录,不同的证据类具有不同的域成员。例如下面就是HTTP证据采集器产生的证据的各个组成域及其描述:

表 2

各组成域	描述		
evidence_class	证据类,整型		
pid	进程号,整型		
status	状态,字符型		
bytes_served	已服务字节数,无符号整型		
client	客户, 2Byte 字符数组		
request	客户请求, 64Byte 字符数组		
vhost	服务主机,32Byte 字符数组		

由于采用了 mysql 进行证据的存储组织与管理,可以很方便高效地实现证据的各种删除,插人,修改操作,为了完成这些操作,我们所要做的仅仅是构造不同的 sql 查询语句,然后发送给 mysql 的服务器。

#### 3.4 证据的分析工具的实现

证据的分析工具被集成进了用户界面,在对证据中心的证据进行分析时,主要包括两种取证分析的方法:

1)根据证据的分类进行纵向分析,这种纵向分析主要在同种取证信息之内进行,包括根据时间相关信息进行分析,根

据用户相关信息进行分析,根据主机信息进行分析等等。一般来说,同种取证信息是由同一个应用程序或者同一类系统日志产生的。

2) 一种横向分析方法,主要根据时间、用户、主机等信息在各种分类取证信息之间进行横向分析。这种横向分析将综合两类或者多类取证信息进行分析,考虑到当前计算机犯罪手段多样化的趋势——即一次计算机犯罪过程包含了多种手段因而将在被攻击计算机中产生多种取证信息,这种横向分析将更为有效。

不同的证据提供的信息不同,要在这些不同的证据类之间进行横向分析,关键就是要充分的理解不同的证据类提供的语义信息,只有这样,才能在这些语义信息的基础上更进一步的发掘这些证据类之间的关联。但是,对证据的这种语义信息的理解却是证据分析工具中最为艰难的一步。

#### 3.5 用户界面的实现

该部分的实现包括一个图形用户界面,该界面的顶层菜单包括:

- 1)证据查询菜单。包含了已经实现的证据类的查询操作,如 FTP 证据查询、用户登录证据查询等。
- 2)证据分析菜单。包含了已经实现的证据类的分析操作,这样的分析可以在同种证据类内进行,也可以在多种证据 类之间进行,这取决于分析工具的具体实现。
- 3)数据库导人导出菜单。包含了证据数据库的导人与 导出功能。
  - 4) 选项菜单。包括了取证系统的配置功能。

### 4 系统的测试和分析

为了测试分布式取证系统的性能,我们在以下环境下进行了测试:

CPU	硬盘	内存	网络带宽	操作系统
PIII 1.7G	Segate 40G	256M	100M LAN	Redhat 8.0

其中证据存储中心所在的保护主机与运行分布式证据收集器的被保护系统的软硬件环境相同。由于缺少同类的取证系统,将对比对象设为 Redhat 本身的日志信息系统。进行下列测试:

### 1)被保护系统上的取证信息范围

证据收集器能够采集各种可能的证据信息。在分布式取证系统运行一段时间后,可以在证据存储中心看到人侵检测信息、用户权限操作信息、文件保护信息、系统日志信息、应用日志信息、重要删除文件信息等。而且这些取证信息的种类还可以通过证据收集器的添加动态扩展。而 Redhat 的取证能力局限于操作系统本身提供的 syslog 机制和一些应用级日志,且不能动态扩展。

#### 2) 同时对多台被保护主机的取证能力

在分布式取证系统运行一段时间后,可以在证据存储中心看到多台被保护系统的取证信息。这种同时对多台被保护主机进行取证的能力极大的提高了系统的适应性。Redhat 并没有提供这样的机制。

# 3) 对证据进行的完整性与安全性保护的能力

在采用 Redhat 本身的日志系统和本文中所实现分布式 取证系统两种情况下,我们模仿人侵者侵人被保护系统,首先 删除被保护系统上的一个重要文件,然后再将用户登录信息, 进程记帐信息从系统相关日志中删掉。在第一种情况下,系 统中的重要文件被删除了,而且,也无从知道是哪个用户删除了这个文件,以及在什么时候删除了这个文件,因为相关信息已经被入侵者一起从系统中删掉了。而在采用了分布式取证系统后,在取证系统的证据存储中心,却可以看到一系列证据信息,明确的显示系统的一个重要文件已经被删除了,这些证据信息中还包含了文件被删除的时间,以及是由哪个用户执行了什么命令删除了这个文件。

可见,由于采用了动态的取证方法,入侵者试图通过删除 取证信息以达到隐匿行踪的目的变得不可能,证据信息的完 整性得到了保证。Redhat 不具备这样的保护能力。

### 5 结语

由于采用了新的系统架构,分布式计算机取证模型具有如下优点:

- 1) 网络动态实时取证能力。将人侵检测、实时系统监控等引入计算机取证系统,在被保护主机上对系统的网络流量和系统中的用户行为进行实时监视,并进行判别及时发现对网络和系统的攻击,将这些信息收集起来作为取证的证据。这就弥补静态取证技术缺乏取证及时性的缺点。同时,这种动态实时的取证也保证了证据的安全性和完整性,因为在被保护主机上的相关证据受到可能的人侵者的篡改、删除以前,分布于该主机上的证据收集器就已经通过安全的途径将该证据发送到了证据存储中心。
- 2) 取证信息的易扩展性和灵活的证据处理和分析方式。 通过调整证据收集器、证据分析的方式和工具,使系统取证范

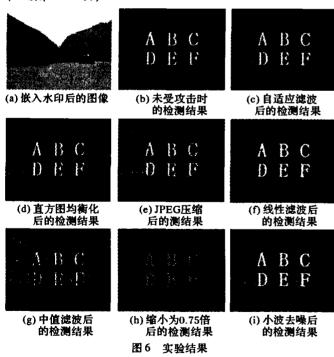
围、分析推理能力可以根据应用环境和需要进行调整,保证取证系统的可扩展性。这样更容易实现对原始证据的可靠性保护,保证证据收集的灵活性和实时性,证据分析推理过程的安全可靠性。

进一步的改进工作包括采用更可靠的加密算法、更多的信息采集技术以扩充取证模型,研究证据信息统一表示、开发通用的证据分析工具等。

#### 参考文献:

- [1] 朱庆华, 邹志仁. 信息系统安全与计算机犯罪[J]. 情报学报, 1999, 18(6).
- [2] RANUM MJ. Thinking about Firewalls [R]. Proceedings of Second International Conference on Systems and Network Security and Management (SANS-II), 1993.
- [3] SCHNEIER B. secrets and lies: digital security in a networked world[M]. chain machine press 2001.
- [4] LUNDIN E, JONSSON E. Survey of Intrusion detection Research [R]. Technical Report nr. 02 - 04. Department of Computer Engineering, Chalmers University of Technology, Goteborg, Sweden 2002.
- [5] LUNT TF. Automated Audit Trail Analysis and Intrusion Detection
  [R]. A Survey Proceedings of the 11th National Computer Security
  Conference, 1988 10.
- [6] CARBER L. computer Forensics: High-tech Law Enforcement[J]. IEEE security. 2001, 34(1).
- [7] RIVEST R. The MD5 Message-Digest Algorithm[S]. Network Working Group, RFC1321, April 1992.

#### (上接第1281页)



### 4 结语

提出了一种新的图像数字盲水印算法,该算法基于相关理论,采用单元模板嵌入信息。实验表明该算法具有较好的水印隐蔽性和鲁棒性。该水印算法提供了一种新的思路,也存在着许多需要改进的地方,作者将在后续的试验中进行进一步完善,包括改进模板结构、应用图像的局部特征、使用动态模板和应用人类视觉模型<sup>[9]</sup>等,进一步提高水印的隐蔽性和鲁棒性,增强水印算法的自适应性。

### 参考文献:

- [1] 李华,朱光喜,朱耀庭. 一种基于人眼视觉感知模型的数字水印 隐藏方法[J]. 电子学报,2000,28(10)、
- [2] LIU J, ZHANG X, SUN JD. a new image watermarking scheme based on DWT and ICA[J]. IEEE Int Conf Neural Networks & Signal Processing, 2003: 1489.
- [3] 张力, 韦岗, 张基宏. --种小波域自适应鲁棒闭环数字盲水印技术[J]. 电子学报, 2003, 31(10).
- [4] WANG SH, LIN YP. wavelet tree quantization for copyright protection watermarking [J]. IEEE transactions on image processing, 2004, 13(2):154-166.
- [5] SHAO YW, WU GW. A wavelet based adaptive watermarking algorithm[J]. IEEE, 2001:384 390.
- [6] TSEKERIDOU S, PITAS I. Wavelet-based self-similar watermarking for still images[J]. IEEE International Symposium on Circuits and Systems, 2000: 220 - 221.
- [7] HONG I, KIM I. A blind watermarking technique using wavelet transform[J]. IEEE, 2001: 1946 - 1951.
- [8] SAGETONG P, ZHOU WS. dynamic wavelet feature based watermarking for copyright tracking in digital movie distribution systems [J]. IEEE, 2002: III - 653.
- [9] PAN R, GAO YX. A new wavelet watermarking technique [J]. Proceedings of 4# World Congress on Intelligent Control and Automation, 2002: 2065 2070.
- [10] GONZALEZ RC, WOODS RE. Digital Image Processing M]. Second Edition, Beijing: Publish House of Electronics Industry (in Chinese with English abstract), 2002
- [11] KANG XG, HUANG JW. A DWT-DFT Composite Watermarking Scheme Robust to Both Affine Transform and JPEG Compression [J]. IEEE transactions on circuits and systems for video technology, August 2003, 13(8):776-786.