文章编号:1001 - 9081(2008)10 - 2485 - 03

分布式自治型计算机取证系统研究

鄢喜爱[1,2],杨金民[1],常卫东[2]
(1. 湖南大学 软件学院,长沙 410082; 2. 湖南公安高等专科学校 计算机系,长沙 410138)
(yanxiai222@yahoo.com.cn)

摘 要:针对目前取证系统的时效性不足和通信瓶颈等问题,提出了一种分布式自治型计算机取证系统。该系统利用自治取证节点对所有可能的入侵行为进行实时动态取证,并采用了安全有效的方式对证据及时保存。由于取证节点具有自治取证能力,系统的整体性能得到了优化。实验表明:该系统能实时取到真实有效的电子证据,并具有很强的容错能力。

关键词:计算机取证;分布式;自治型
中图分类号: TP393.08 **文献标志码**:A

Research of distributed and autonomous computer forensics system

YAN Xi-ai[1,2], YANG Jin-min[1], CHANG Wei-dong[2]
(1. *Software College, Hunan University, Changsha Hunan* 410082, *China*;
2. *Computer Department, Hunan Public Security College, Changsha Hunan* 410138, *China*)

Abstract: Currently, most of computer forensics systems are not real-time, and often cause communicational bottleneck. In order to overcome the shortages, a distributed and autonomous computer forensics system was presented. By using the autonomous forensics node, the system could obtain real-time evidence dynamically as soon as network intrusions took place, in which the evidence could be saved in a safe way in time. This way of autonomous forensics could optimize system performance. Experimental result shows that the system can capture the authentic and valid electronic evidence, and has high capability of fault tolerance.

Key words: computer forensics; distributed; autonomous model

0 引言

随着计算机技术的发展和 Internet 的普及,计算机已经成为人们生活中必不可少的日常工具。与此同时,由于 TCP/IP 协议本身的脆弱性,与计算机相关的犯罪案件急剧上升,如何预防和打击计算机犯罪已经成为人们普遍关心的问题[1]。目前的主流的信息安全技术如防火墙、数据加密、入侵检测等可以从一定程度上预防计算机犯罪的发生,但是,这些安全技术和手段并不能够防止所有入侵。因此,借助法律对入侵犯罪者实施惩罚和威慑已是一个非常重要的手段,此手段的关键是找到真实、可靠、具有法律效力的电子证据,计算机取证技术也就应运而生。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[2]。

1 相关研究的观察和启发

目前,国外在取证的理论研究基础上,已进入了实用化的阶段,许多国家相继成立了计算机取证机构、实验室和咨询服务公司,并相继有些产品问世。如美国最大的计算机取证公司 NTI 研制了 Net Threat Analyzer 软件[3];美国 Guidance software 公司研制了基于 Windows 的 Encase 产品[4];英国 Vogon 公司开发了基于 PC、Mac 和 Unix 等系统的数据收集和分析系统 Flight Server;美国的 Sandstorm 公司开发了 Netintercept 网络取证系统[5]等。

国内也已在计算机取证技术研究方面作出了积极反应,开发出了一些应用系统,并且运用计算机取证技术侦破了一些实际犯罪案例[6]。如厦门公安局计算机安全监察处与厦门美亚柏科资讯公司开发的计算机犯罪取证勘察箱;上海金诺网安依托十五攻关项目"存储介质中残缺数据的勘察取证技术"和 863 项目"应急响应中的数据取证和恢复技术"开发的取证系统 DiskForen 产品以及中软公司的网络信息监控分析与取证系统等。

通过仔细分析现有的电子取证系统和产品,我们主要有以下两点观察和启发:

观察 1 计算机取证技术的研究多为静态分析方法,取证工具软件的功能主要集中在磁盘分析上,证据的获取过程不能够自动进行,人工干预比较多,不能够在入侵的同时或者一定的时限内自动获取合法的证据。这种静态的取证模式对于新兴的反取证技术(如数据隐藏技术、数据擦除技术等)显得无能为力。

启发 1 取证系统必须采用实时的动态方式对电子证据进行提取和保存,能在网络攻击行为发生的时候自动地收集

收稿日期:2008 - 04 - 15;**修回日期**:2008 - 06 - 15。
基金项目:国家自然科学基金资助项目(60473031);公安部应用创新计划项目(2006YYCXHNST024)。
作者简介:鄢喜爱(1972 -),男,湖南长沙人,副教授,硕士,主要研究方向:信息安全; 杨金民(1967 -),男,湖南长沙人,副教授,博士,主要研究方向:系统容错; 常卫东(1967 -),男,湖南长沙人,副教授、硕士,主要研究方向:信息安全。

网络攻击证据并对攻击行为作出反应。

观察 2 取证系统的证据收集平台大多采用集中式处理和管理的证据收集方式,系统遵循自顶向下的控制流程,层次结构复杂。这样容易产生主机性能瓶颈、单点失效、网络带宽不足、负载不均衡等问题。

启发 2 取证系统可以采用分布式自治型的证据收集平台,各取证子节点均能独立地完成证据的收集,不过分依赖主机,各节点之间仅传递必要的信息,并能相互协调地完成工作。

基于以上两点观察和启发,我们设计了一种分布式自治型计算机取证系统,该系统具有自治、实时和通信低开销等特性,能及时主动地获取计算机犯罪的电子证据。

2 分布式自治型计算机取证系统

分布式自治型计算机取证系统的拓扑结构如图 1 所示,采用具有自治性的取证节点组成三层分布式体系结构。证据服务器是全局事务的控制中心,负责取证任务的发起和取证结果的收集,若某网段自治取证节点发生故障,能通过服务器的协调产生一个新的自治取证节点;自治取证节点层是本系统的关键层,负责某逻辑网段中被取证端证据的实时采集、分析、保存,并将取证结果通过专用的通信协议提交给服务器。

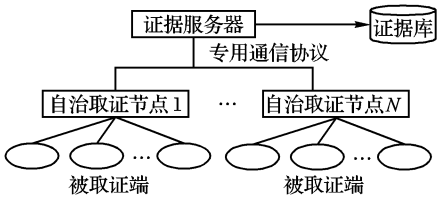

图 1 取证系统的整体拓扑结构

2.1 自治取证节点

每个自治取证节点均能单独实现其自治区域被取证端证据的检测、提取、保存,其模块的组成和工作流程如图 2 所示。证据检测模块完成对网络数据的检测,若发现有可疑行为,则激活证据提取模块;证据提取模块分析初步的检测结果,依据给定的规则对证据进行提取;本地保存模块按一定的策略存储生成的证据,供以后查询和进一步分析;证据传输模块负责将证据记录安全地提交至证据服务器。

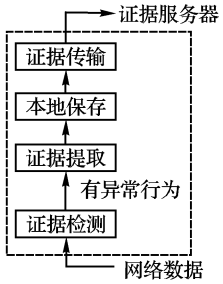

图 2 自治取证节点的功能模块及流程

2.1.1 证据的检测

检测模块通过对网络数据流的捕捉和分析,对网络的运行情况进行实时监控,当发现入侵企图或入侵事件时,立即发送一个取证请求消息给取证模块开始取证。为了方便检测,我们在证据服务器上设置了一个初始的特征知识库,初始知识库收集了常见攻击行为的特征记录,并通过系统的自学习不断地更新,实时发布到各自治取证节点。检测模块将捕获的数据流放入自己的检测队列,与知识库中的记录进行匹配,一旦发现可疑的数据,就立即向证据提取模块发出取证请求消息,并由证据提取模块执行分析取证程序,开始对攻击现场进行证据的提取。

2.1.2 证据的提取

证据提取模块的主要功能是进行数据分析,并获取真实、有效、不可抵赖的电子证据。

证据提取模块的工作流程如图 3 所示。在得到检测模块的请求消息之后,获取可疑行为的特征数据,如果通过匹配后发现是已知的攻击行为,直接进行证据的收集;如果是未知的行为,将数据处理成符合数据仓库存储格式后存入数据仓库,通过数据挖掘对数据仓库的数据进行消脏、选择、格式转换,使用关联规则分析、分类、联系分析技术方法发现事件之间的时间和空间上的联系,找到用于数据分析的特征、模式、规则与知识[7],再通过数据分析判定是否为入侵行为,如果是入侵行为,就进行证据的收集,并将分析出的新规则存入知识库。

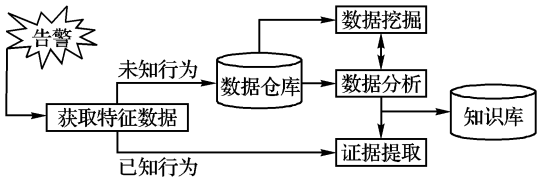

图 3 证据提取模块的工作流程

证据收集的内容主要包括两大部分:一部分是入侵发生时的主机运行参数,例如系统内存信息、CPU 使用率、进程运行状况、缓冲区信息、磁盘文件的读写等,获得这些数据实际上是对整个攻击现场的快照,是还原入侵犯罪场景不可缺少的因素。另外一部分主要指网络通信情况。如网络数据包大小、数据流量、攻击方源地址、目的地址、端口、使用协议、网络连接数量、连接时间、连接类型等。

2.1.3 证据的本地保存策略

为了完全地实现区域自治,我们在每个自治取证节点设有本地证据库,保存本网段收集到的取证结果。证据的收集是一个长期的工作,随着时间的推移,证据库存储的数据记录将会越来越大,这样势必影响系统的运行质量[8],由于取证结果已及时发送到了服务器,所以可以采取一定的存储策略对本地记录保存的时间进行取舍。取舍规则应当遵循安全威胁越大证据记录保存时间越长的规则。

我们在证据收集时,根据入侵发生时的主机运行参数来确定攻击的强度,将证据记录的安全级别定义为低、中、高三种情况,安全威胁越大级别越高。定义时间 T_0 为记录保存的最短时间(如一个月),具体的取舍规则如表 1 所示。

表 1 证据记录的保存规则

规则	安全级别	存储时间
1	低	不少于 T_0
2	中	不少于 $3T_0$
3	高	不少于 $6T_0$

2.1.4 取证结果的传输

证据传输模块主要功能是将最终的取证结果安全地提交给证据服务器。为了防止证据在传输过程中遭到窃听、篡改、重放等攻击手段,我们在证据传输时设置了专用的通信协议对证据进行数字签名,其实现步骤如下:

1)通过增量备份的方式从本地证据库中找到将要传输的证据信息 M。

2)利用 SHA 安全散列函数计算出要传输信息的散列码,记为 $h=H(M)$。

3)利用 RSA 加密算法将散列码进行加密处理,为了提高算法的安全性,采用 1024 位的密钥,得到要发送的密文信息,记为 $C=E_k(H(M))$。

4) 发送密文信息 C 至证据服务器。

2.2　证据服务器

证据服务器作为全局的控制中心,向各个自治取证节点发出控制指令,回收取证节点的取证结果,协调控制整个取证系统。

证据服务器的一个主要功能是接收来自取证节点的取证结果,在进行数字签名验证之后,将取证结果进行筛选、组合和重构,使之能再现整个事件的攻击过程,并生成完整的报告,在加盖时间戳之后保存到证据服务器的证据库中。

证据服务器的另一个主要功能是解决某自治取证节点出现故障时的进程迁移问题。在网络入侵事件中,入侵者一般对网络拓扑结构有一定的了解,他们在进行网络入侵时,为了逃避打击,往往会对自治取证节点首先发起拒绝服务的攻击,自治取证节点出现故障的可能性较大。为了解决这一问题,证据服务器可以通过“心跳”来感知自治取证节点是否存活。所谓“心跳”技术[9],是指证据服务器每间隔一定的时间发送消息给取证节点,这个消息称为心跳。两个心跳消息之间的间隔称为心跳间隔,心跳间隔视故障发生的概率而定。服务器从取证节点处得到心跳,从而知道取证节点的存活。如果服务器间隔一定时间(比如 3 个心跳间隔)没有收到取证节点的心跳,则认为取证节点失败。当服务器发现取证节点失败时,通过选举算法(例如选举存活取证节点中数字标识最大的节点)选举产生新的取证节点来接管原故障节点的工作。

在上述的容错解决方案中,我们是假设了服务器不会发生故障。当然,证据服务器受攻击的可能性也很大。在条件允许的情况下,可以建一个远程备份服务器,采用双机热备的方式运行。

3　系统测试

根据上述设计的取证模型,我们开发了一个分布式自治型取证的原型系统,并在操作系统为 Redhat Linux 9.0、硬件配置如表 2 的环境下,对典型的 SYNFLOOD 攻击行为进行了测试。实验中将被取证端分成三个逻辑网段,分别由自治节点 1、2、3 进行监控取证。

表 2　实验环境的硬件配置

身份	CPU	硬盘	内存	网络带宽	节点数
证据服务器	P4 3.2GHz	Maxtor 160 GB	1 GB	100 MB	1
自治取证节点	P4 3.2GHz	Maxtor 160 GB	1 GB	100 MB	3
被取证端	P4 3.2GHz	Maxtor 80 GB	512 MB	100 MB	15

1)取证能力测试。首先利用 VC6.0 编写一个 SYNFLOOD 攻击的程序,对逻辑网段 1 中的 1 号被取证机进行攻击,在自治取证节点 1 中,攻击行为的信息可以实时准确地被检测出来,并可以将证据提取、保存。图 4 为进行攻击后得到的相关信息。

图 4　自治取证节点获取的相关信息

2)系统容错能力的测试。采用 DDoS 攻击方法,使自治节点 1 处于瘫痪状态,然后对逻辑网段 1 中的 2 号被取证机进行 SYNFLOOD 攻击。可以在自治节点 3 的监控系统中发现相关的攻击信息。这是由于服务器选择了自治节点 3 来接管自治节点 1 的工作,只是时间上稍有延迟。

4　结语

随着时代的发展,取证与反取证的对抗必将越来越激烈,对取证技术的要求必将越来越高。本文所设计的分布式自治型取证系统层次结构清晰、简单,能准确、实时、有效地收集到无法律疑点的证据;能充分地发挥各自治取证节点的自治性,有效减少网络通信流量和解决服务器的负载瓶颈问题;各自治节点能协同工作,在某一节点出现故障情况下,能由其他节点接管工作,具有很强的容错性能;此外,系统还有良好的可扩展性。进一步的工作是对未知攻击行为的检测和各自治取证节点负载均衡等问题的改善。

参考文献:

[1] MARCUS K, KATE S. The future of computer forensics: A needs analysis survey[J]. Computers & Security, 2004, 23(1):12 - 16.

[2] 梁昌宇, 吴强, 曾庆凯. 分布式计算机动态取证模型[J]. 计算机应用, 2005, 25(6):1290 - 1293.

[3] Frensic Computing Ltd . Computer Forensics & Security Software Tools[EB/OL]. [2002 - 06]. http://www.forensics-intl.com/thetools.html.

[4] WEI W. EnCase Forensic Functionality[EB/OL]. [2003 - 07]. 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]. http://www.sandstorm.net/products/#thetop.

[6] 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.

[7] 钟秀玉. 计算机动态取证系统模型研究[J]. 微计算机信息, 2006, 22(8):43 - 45.

[8] 秦拯, 李建辉, 邹建军, 等. 基于模糊理论的实时取证模型[J]. 湖南大学学报: 自然科学版, 2006, 33(4):115 - 118.

[9] 赵高峰, 胡运发. 基于心跳技术的 3 阶段提交协议[J]. 计算机工程与应用, 2004, 40(11):177 - 179.