基于 SN 并行结构的网络取证日志隐匿安全管理*

戴江山, 肖军模
(解放军理工大学通信工程学院, 江苏 南京 210007)

摘 要: 针对目前应用于网络取证的日志结构多样, 存贮分散, 缺乏证据能力保证和难以进行综合取证分析的问题, 提出并建立一种基于无共享资源 SN (shared nothing) 并行结构的安全日志隐匿管理系统. 该系统将分散的异构的日志记录由日志代理和管理网关收集并转发到多个存贮节点. 各存贮节点采用信息分配算法 IDA (information dispersal arithmetic) 将日志记录分散为 n 份并存贮相应分片, 计算所有分片单向散列值并保存. 取证分析时, 管理网关利用任意 m ($m<n$) 个存贮节点的相关信息重建原有日志, 并且通过每个日志记录分片携带的所有分片单向散列值验证日志记录的完整性.

关键词: 网络安全; 网络取证; 无共享资源结构; 日志
中图分类号: TP393.08 文献标识码: A 文章编号: 0529-6579 (2004) S2-0038-04

证据能力, 即证据资格, 是指证据资料在法律上允许其作为证据的资格[1]. 数字证据资料由于其数字化特性, 具有易篡改、易伪造、易删除和修改后易消除痕迹等特点, 因此其证据能力更需要专门的技术手段和严格的取证程序加以保证和证明. 日志包括主机日志、网络日志和边界设备日志是目前网络取证中重要的证据信息来源, 但是其结构多样, 存贮分散, 缺乏安全性和完整性保护, 这使其证据能力很容易遭到质疑, 而且不利于证据的提取和综合分析.

针对以上问题, 本文提出和建立一种基于无共享资源 SN (shared nothing) 并行结构的网络取证安全日志隐匿管理系统. 无共享资源 SN 并行结构就是由多个处理节点构成, 每个处理节点具有自己独立的处理机、内存和存储器, 并由高速通信网络连接. 同分布式系统相比, SN 并行结构最大的不同是不存在全局和局部概念, 各节点具有非独立性只能在数据处理中发挥协同作用[2]. 安全日志隐匿管理系统基于 SN 并行结构原理, 利用日志代理和管理网关将分散的异构的日志记录收集并转发到多个存贮节点. 各存贮节点采用信息分配算法 IDA (information dispersal arithmetic) 将日志记录分散为 n 份并存贮相应分片, 计算所有分片单向散列 Hash 值并保存. 取证分析时, 管理网关利用任意 m ($m<n$) 个存贮节点的相关信息可以重建原有日志, 并且通过每个日志分片携带的所有分片单向散列 Hash 值验证日志记录的完整性.

1 安全日志隐匿管理系统体系结构

基于 SN 并行结构的安全日志隐匿管理系统包括日志代理、管理网关和多个存贮节点 (图 1).

网络取证日志通常包括主机日志、网络日志和边界设备日志. 在日志所在系统安装日志代理, 当日志记录产生后, 除在本地按照原有方式存贮外, 日志代理向管理网关发送连接请求, 并将日志记录加密和数字签名后发送给管理网关.

管理网关包括安全防护、安全监测、管理控制、系统维护、接收转发代理和临时数据库等, 主要完成日志记录正确接收和安全可靠转发到各存贮节点等功能. 安全防护通过防火墙、基于主机的入侵检测系统、访问控制等手段严格限制和过滤进出管理网关的数据包, 以保护管理网关以及存贮节点的安全性. 接收发送代理负责接收和向各存贮节点转发日志代理发送的日志记录. 被接收的日志记录经过安全检测中的数字签名验证后, 被存贮在临时数据库, 然后由接收发送代理负责安全可靠转发到各存贮节点. 为满足网络取证需要, 日志往往需要长期保存, 因此管理系统必须具有系统维护功能, 以随时发现和纠正系统出现的问题, 如存贮节点系统发生故障, 数据完整性遭破坏, 发生拒绝服务攻击等等. 管理网关和各存贮节点均设有系统维护模块, 协作运行保障整个系统安全可靠运行. 管理控

* 收稿日期: 2004-09-11
基金项目: 国家自然科学基金资助项目 (69931040)
作者简介: 戴江山 (1973 年生), 男, 博士研究生; E-mail: daijshan@public.qd.sd.cn

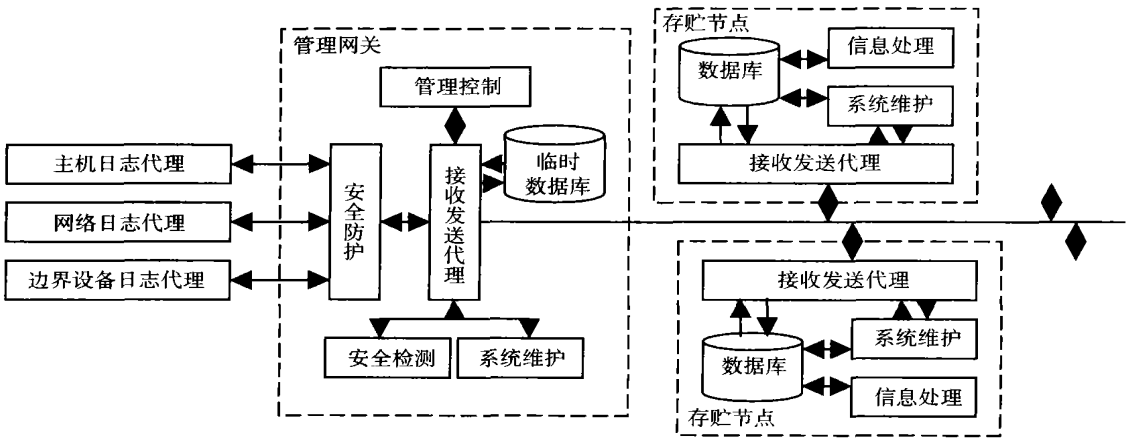

图 1　基于 SN 并行结构的安全日志隐匿管理系统体系结构

Fig. 1　The structure of the security log anonymous management system based on the SN parallel structure

制实现整个系统检测、设置以及利用各种方式对日志进行综合分析实现网络取证和完整性验证等。

存贮节点包括信息处理、系统维护、接收发送代理和数据库等，主要实现日志的隐匿安全存贮。存贮节点上的接收发送代理接收到日志记录后，采用信息分配算法 IDA 对日志记录进行分散并计算所有分片的单向散列 Hash 值，然后存贮相应日志记录分片和所有分片 Hash 值。由于不同类别日志内容结构差别较大，为减少存贮空间的冗余度，这里采用多数据库存贮方式，即在同一存贮节点根据日志类型将数据库分为面向主机日志、面向边界设备日志和面向网络日志数据库。

2　安全日志隐匿管理系统运行

安全日志隐匿管理系统运行主要包括日志传输、日志存贮、日志分析和系统维护。

2.1　日志传输

日志传输分为 2 个阶段（图 2），包括日志代理同管理网关间的传输和管理网关同存贮节点间的传输，其中前者传输空间为非安全传输域，后者传输空间为安全传输域，运行优先级前者高于后者。

日志代理同管理网关间的日志传输主要实现在非安全传输域内日志记录的安全可靠传输和被管理网关正确接收。日志代理首先对待发送日志记录进行内容加密（如对称加密算法 DES[3]）和数字签名（如公开密钥数字签名算法 DSA[3]），然后采用 TCP 协议向管理网关发送连接请求并完成可靠传输。管理网关内建有日志代理数字签名公钥表，对接收到的日志记录经过数字签名验证后存贮到临时数据库。

管理网关同存贮节点间日志传输主要实现在安全传输域内日志记录的可靠传输和被各存贮节点正确接收。采用 UDP 协议，管理网关将临时数据库内日志记录同时发送给各存贮节点，传输数据包在 UDP 数据包头后增加固定标识头。各存贮节点根据数据包固定标识头识别和正确接收到日志记录后发送接收确认。管理网关在规定时间内接收到所有存贮节点发送的接收确认后，删除该日志记录并发送下一条，否则重发。

2.2　日志存贮

存贮节点采用信息分配算法 IDA[4] 将日志记录分散为 n 份并存贮相应日志记录分片，计算所有

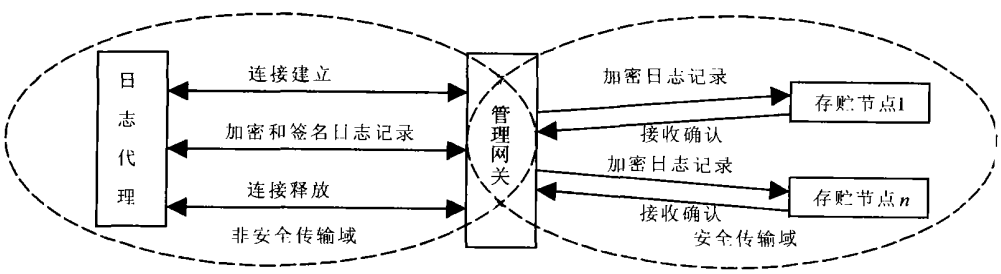

图 2　日志传输过程图

Fig. 2　The transmission process of the log

分片的单向散列 Hash 值（如安全散列算法 SHA[3]）并存贮。

信息分配算法 IDA 基本思想是将长度为 $L=|D|$ 的数据信息 D 分散为 n 份，其中任意 m（$m\leqslant n$）份可以重组原有信息，具体方法如下：

首先在有限域 GF（2^8）将日志记录 D 划分为长度为 m 的数据分组序列，不足位由 0 填充。

$$D=D_1,D_2,\cdots=$$
$$(d_{1,1},d_{1,2},\cdots,d_{1,m}),(d_{2,1},d_{2,2},\cdots,d_{2,m}),\cdots$$

其中，$|D_i|=m$。

在有限域 GF（2^8）上利用 $m\times n$ 矩阵 T 实现 D_i 的 m 输入 n 输出的线性变换，即输入

$$D_i=d_{i,1},d_{i,2},\cdots,d_{i,m}$$

输出

$$O_i=D_i\cdot T=o_{i,1},o_{i,2},\cdots,o_{i,n}$$

对 D 中所有数据分组进行同样处理就得到，

$$O=O_1,O_2,\cdots=$$
$$(o_{1,1},o_{1,2},\cdots,o_{1,n}),(o_{2,1},o_{2,2},\cdots,o_{2,n}),\cdots$$

序列 $P_i=o_{1,i},o_{2,i},\cdots(i\leqslant n,|P_i|=L/m)$ 就是日志记录 D 的第 i 个分片，并存贮在第 i 个存贮节点。

转换矩阵 $T=(t_{ij})_{m\times n}$，必须满足任意 m 个列向量是线性无关的。其选取有多种方法如：设 $x_1,x_2,\cdots,x_n\in GF(2^8)$，$y_1,y_2,\cdots,y_m\in GF(2^8)$ 且满足：$x_i+y_i\neq 0$；$x_i\neq x_j(i\neq j)$；$y_i\neq y_j(i\neq j)$，则定义

$$t_i=\left(\frac{1}{x_i+y_1},\cdots,\frac{1}{x_i+y_m}\right),1\leqslant i\leqslant n$$

T 是由 $t_1,t_2,\cdots,t_n$ 行向量组成的 $n\times n$ 矩阵。如果有

$$|T|=\frac{\prod_{i<j}(x_i-x_j)(y_i-y_j)}{\prod_{i,j}x_i+x_j}\neq 0$$

则向量 $t_1,t_2,\cdots,t_n$ 中任意 m 个向量是线性无关的。

2.3 日志分析

日志分析就是根据取证要求采取多种查询方式如 MAC 地址、日志类型、时间段等查询有关日志信息，并完成日志信息完整性检测。管理网关采用 UDP 协议向 n 个存贮节点同时发送查询条件，当接收到 m 个存贮节点的数据返回时，则拒绝继续接收。根据接收数据，管理网关采用 IDA 算法重新构建出原有日志记录，并对每个日志记录分片所携带的该日志记录所有分片 Hash 值相比较，如果相同则表示该日志是完整的，否则发出完整性告警。

为表述方便这里假设接收到编号为 1 至 m 存贮节点的返回数据（如果含有其它编号存贮节点如 $m<i\leqslant n$，则只需使用其存贮的数据分片和转换矩阵 T 中的对应列向量），针对某条日志记录，存贮节点获得分片 $P_1,P_2,\cdots,P_m$，其中 $P_i=o_{1,i},o_{2,i},\cdots(1\leqslant i\leqslant m)$。

则得到

$$O'=O'_1,O'_2,\cdots=$$
$$(o_{1,1},o_{1,2},\cdots,o_{1,m}),(o_{2,1},o_{,2,2},\cdots,o_{2,m}),\cdots$$

T 中相对应的 1 ~ m 列向量组成 $m\times m$ 矩阵 $T'=(t_{ij})_{m\times m}$，由于 T' 中列向量是线性无关的，即 $D_i\cdot T'=O'_i$，因此有 $D_i=O'_i\cdot T'^{-1}$（该计算在有限域 GF(2^8) 上完成）。对 O' 中所有分组 O'_i 进行同样处理，就得到

$$D=D_1,D_2,\cdots=$$
$$(d_{1,1},d_{1,2},\cdots,d_{1,m}),(d_{2,1},d_{2,2},\cdots,d_{2,m})\cdots$$

即恢复出原有日志记录。

2.4 系统维护

系统维护采用存贮节点轮值和时间滑动窗口方法维护存贮节点安全性和日志的完整性。存贮节点内建有所有存贮节点地址表，每间隔某个固定时间段 t（如每天），轮值存贮节点 i 计算本节点最近固定时间窗口（如一个月）内存贮的日志记录分片的单向散列 Hash 值，然后同存贮的相应 Hash 值相比对。如果全部相同，则采用 UDP 协议向所有节点发送此时间窗口内日志记录所有分片 Hash 值。如果存在不同，则通知管理网关，管理网关利用 m 个有效节点数据检测或重新建立 Hash 值存在不同的日志记录，并发送给节点 i，节点 i 采用 IDA 和 SHA 重新计算和存贮相关日志信息。存贮节点 j 接收到轮值存贮节点 i 发送来的时间窗口内日志记录所有分片 Hash 值，则同本节点存贮的相应值相比较，如果存在不同则通知管理网关，并采用同样方法检测或重新建立、计算和存贮相关日志信息。此外由于系统维护是以固定时间为间隔，因此在经过时间段 t 存贮节点没有接收到相应轮值节点发送来的信息，并且管理网关也没有接收到日志信息维护要求，则轮值存贮节点系统可能发生故障，通知管理网关。

3 安全日志隐匿管理系统性能分析

3.1 安全性

管理网关将传输空间分为安全传输域和非安全传输域。在非安全传输域，系统采用加密和数字签名方式防止日志记录被伪造、篡改和利用 Sniffer 监听和收集。管理网关设有安全防护并且通过严格的安全策略设置，保证系统安全运行。日志被分散存贮在 n 个存贮节点，攻击者要破坏日志必须要侵

入 $n-m+1$ 个存贮节点，而这极大增加了攻击者侵入的难度和其被发现的概率。另外，控制任意 $t<m$ 个节点无法重新构造和获取原有日志，而只有在网络取证被授权时才能重建日志内容，这就保证了日志内容的安全性和隐匿性。

3.2　完整性

日志代理同管理网关间采用 TCP 协议，以及管理网关同存贮节点间采用 UDP 加接收确认的协议保证日志记录被完整可靠传输和存贮。存贮节点存贮有日志记录所有分片的 Hash 值，系统维护利用这些信息可自动维护和保证日志长期存贮的完整性。取证分析时，存贮节点利用任意 m ($m<n$) 个节点 (包括自身节点) 重构原有信息，通过日志记录分片携带的所有分片 Hash 值比较，保证取证分析日志的证据能力。

3.3　实时性

系统采用管理网关从多数据源收集和向多存贮节点转发日志，因此管理网关是系统日志处理瓶颈。但网络取证是对长期收集的日志的综合性分析，因此系统主要应用于日志的长期安全管理，对实时性处理要求并不太高。而且管理网关采用接收优先级高于转发优先级、临时存贮转发等方式保证了日志的正确接收和可靠转发。网络取证日志分析时，存贮节点需要对相关日志信息进行收集、重构和完整性验证，因此所需的系统时间相对较长。但网络取证中数字证据资料的完整性要求往往比取证分析的时间性要求更为重要，而且系统可采用二级查询方式，即将有关信息收集重组和完整性验证后存贮在当地，然后进行更为详细的取证分析。

4　结束语

针对网络取证中日志证据能力保证以及日志综合分析问题，本文提出并设计了一种基于 SN 并行结构的安全日志隐匿管理系统，该系统采用信息分配 IDA 算法将日志分散隐匿存贮在多个存贮节点，保证网络取证分析中日志的安全性、完整性、隐匿性以及便于综合分析查询。证据资料收集和分析是网络取证的主要内容，本系统在保证日志证据能力的情况下对日志进行隐匿安全管理，但在日志取证分析方面还存在一定的局限性。通过对多台每天频率使用的主机进行日志记录统计发现，平均每天每台主机产生日志达到 200 多条，而其中大多是重复的、无价值的。因此必须发展日志综合分析以及显示方法，挖掘日志间的相关性，重构攻击者的入侵过程以及以易理解的方式显示结果，这是也本系统进一步研究和发展的方向。

参考文献：

[1] 齐爱民, 刘颖. 网络法研究[M]. 北京：法律出版社, 2003：322.

[2] 萨师煊, 王珊. 数据库系统概论[M]. 北京：高等教育出版社, 2000：327.

[3] SCHNEIER B. Applied cryptography second edition: protocols, algorithms, and source code in C[M]. New York: John Wiley & Sons, 1996：185－212.

[4] RABIN M. Efficient dispersal of information for security, load balancing, and fault tolerance[J]. J ACM, 1989, 36(2)：335－348.

Network Forensics Log Anonymous Security Management Based on the SN Parallel Structure

DAI Jiang-shan, *XIAO Jun-mo*

(Institute of Communications Engineering, PLA University of Sciences and Technology, Nanjing 210007, China)

Abstract: Logs are the important evidence sources in the network forensics, but their storage mode and format make them lack of protection of evidence ability and effective forensics analysis. A kind of security log anonymous management system based on the SN (shared nothing) parallel structure is proposed and designed. The log agents and management gateway collect and forward the log records to the multi-storage nodes. The log records are respectively dispersed into n shares by the information dispersal arithmetic in the node, and the node stores the corresponding share and the hash values of all shares. The management gateway can reconstruct the log records through log information in m ($m<n$) random nodes and validate the integrity of the log records through the hash values of all shares.

Key words: network security; network forensics; shared nothing structure; log