构建基于蜜罐技术的入侵检测系统

The Construction of Intrusion Detection System Based on Honeypot

冯　嵩　（中南大学湘雅学院 湖南长沙 410008）
张　洁　（广东东莞理工学院 523808）
王振力　（南京通信工程学院 江苏南京 210007）

摘要：传统意义上的网络入侵检测只能检测到已知类型的攻击和入侵，对未知类型的攻击则无能为力。蜜罐技术是入侵检测技术的一个重要发展方向，已经发展成为诱骗攻击者的一种非常有效而实用的方法，不仅可以转移入侵者的攻击，保护主机和网络不受入侵，而且可以为入侵的取证提供重要的线索和信息。设计了基于蜜罐技术的网络入侵检测系统，成功的实现了对网络入侵的跟踪与分析。

关键词：网络安全 入侵检测 蜜罐 构建

1　引言

入侵检测（Intrusion Detection）是对入侵行为的发觉，它通过对计算机网络或计算机系统中若干关键点收集信息并对其进行分析，从中发现网络或系统中是否有违反安全策略的行为和被攻击的迹象。进行入侵检测的软件与硬件的组合便是入侵检测系统（Intrusion Detection System，简称 IDS）。但是传统意义上的网络入侵检测比如基于数据挖掘的入侵检测[1]-[3]等只能检测到已知类型的攻击和入侵，对未知类型的攻击则无能为力。

蜜罐[4]（Honeypot）技术目前已经成为入侵检测技术的一个重要发展方向，它不仅可以转移入侵者的攻击，保护主机和网络不受入侵，而且可以为入侵的取证提供重要的线索和信息。本文设计了一种基于蜜罐技术的网络入侵检测系统，介绍了系统组成与关键技术，成功的实现了对网络入侵的跟踪与分析，具有一定的实用价值。

2　蜜罐

蜜罐简单的说就是一种在互联网上运行的计算机系统，专门为吸引并“诱骗”那些试图非法闯入他人计算机系统的人（如电脑黑客）而设计的。蜜罐系统则是一个包含漏洞的诱骗系统，它通过模拟一个或多个易受攻击的主机，给攻击者提供一个容易攻击的目标。由于蜜罐并没有向外界提供真正有价值的服务，因此所有对其链接的尝试都将被视为可疑的。蜜罐的另一个用途是拖延攻击者对真正目标的攻击，让攻击者在蜜罐上浪费时间。这样，最初的攻击目标得到了保护，真正有价值的内容没有受到侵犯，因此蜜罐也可以为追踪攻击者提供有用的线索，为起诉攻击者搜集有力的证据，从这个意义上说，蜜罐就是“诱捕”攻击者的一个陷阱。经过多年的研究与发展，蜜罐技术已经发展成为诱骗攻击者的一种非常有效而实用的方法。

3　系统组成与功能

系统主要由入侵检测系统、取证代理和取证中心组成，如图 1 所示。入侵检测系统配置在各个网段，对该网段的网络运行情况进行监测，并将报警信息发送给取证中心。取证代理根据安全需要配置在需要监控的主机系统内，监测主机系统运行情况、收集相关信息、将相关信息进行完整性处理后发送到取证中心。取证中心负责接收多数据源发送来的数字证据，实现数字证据的完整性验证、处理、存贮和取证分析等功能。

3.1 取证代理

取证代理就是一个配以相应数据捕捉工具的蜜罐，主要用来捕捉发生在蜜罐中所有有关入侵的数据，帮助准确重建攻击者侵入系统后的行为。在入侵者看来，取证代理所在的系统是一个包含漏洞的系统，非常容易受到攻击，因为入侵者无法察觉取证代理已经安装了数据捕捉工具。取证代理在 Linux 系统内核以可调用内核模式(LKM)实现，包括数据捕捉和数据发送。取证客户端可以记录用户通过 read( ) 系统调用访问的所有数据[7]，然后以标准格式表示，并采用 UDP 方式隐蔽发送给取证服务器。由于捕捉的数据以自定义标准格式的表示，因此服务器可以收集运行在不同操作系统上的 honeypot 发送的数据。

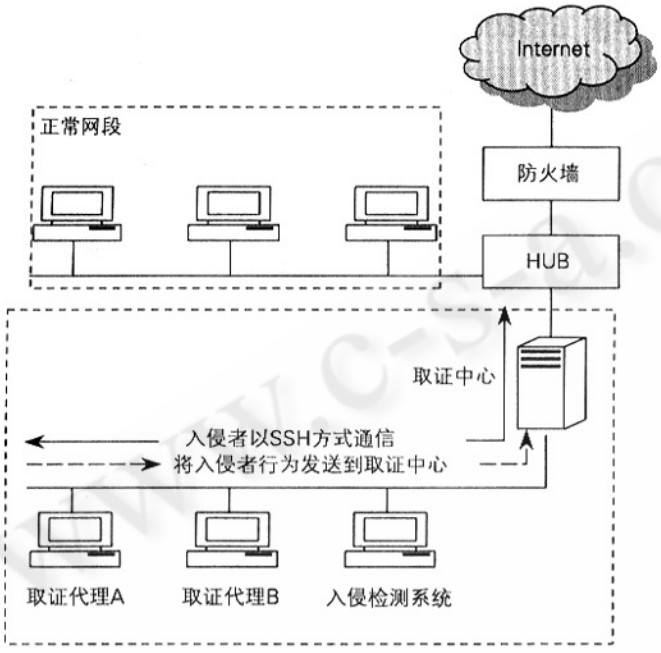

图 1　基于蜜罐的入侵检测系统总体结构

3.2 取证中心

取证中心是系统控制和管理的中心，主要用来收集攻击者潜在的入侵犯罪证据，重建攻击者的入侵过程，并防止攻击者成功侵入 honeypot 后，以 honeypot 为"跳板"攻击其它正常主机。取证服务器采用桥(bridge)模式。桥式工作在数据链路层，对从端口接收到的 MAC 帧根据目的地址进行转发和过滤。桥这种模式使取证服务器没有 IP 地址，没有 MAC 地址，没有数据报路由以及数据报的 TTL 消耗，这使取证服务器被构建成为一个对攻击者来说"不可见"的过滤控制设备，使攻击者更难以检测和觉察它的存在。通常取证服务器建立在 Linux 环境下，本系统使用 Rad HatLinux9.0。而大多数版本的 Linux 在默认安装情况下支持桥的功能。另外，取证服务器具有网关功能，它将入侵检测系统同网络其它部分隔离开来，任何进出入侵检测系统的数据包必须经过取证服务器，这样就可以对数据包进行过滤，实现对无论是来自内网还是外网攻击的控制和取证。

3.3 入侵检测系统

入侵检测系统采用基于特征检测的 Snort 系统[6]。Snort 系统根据规则定义检测网络中的数据报，当规则被触发后产生告警信息，发送到取证中心。取证中心开放固定端口不断监听，当接收到数据主动采集端的连接请求后，启动线程实现相应信息的接收。当取证中心接收到信息后，首先进行 MD5 算法验证，然后从数据库表中提取相关记录建立潜在证据关联，最后根据潜在证据类型提取字段内容，并分类规范存贮在数据库相应表中。

4 系统关键技术与方法

4.1 数字证据完整性算法

数字证据完整性算法包括完整性处理算法和完整性验证算法两部分，目的是保证潜在证据在长期存贮过程中的完整性，以及取证分析时，对于数字证据被篡改、伪造、删除和添加等情况能够正确检测和对存在完整性问题的记录实现定位。

以取证中心接收到取证代理发送的主机进程信息为例说明完整性处理算法，如图 2 所示。

(1) 取证中心接收到新的进程信息 D_i，从数据库主机进程信息表中读取最后一条记录 $i-1$ 的 Hash 字段值 $Hash_{i-1}$；

(2) D_i 内属于各字段的信息以空格为间隔标志。将 D_i 同 Step1 中读取的 $Hash_{i-1}$ 以空格为间隔连接为字符串 S_i = (D_i□ $HASH_{i-1}$) (□表示空格)，采用 MD5 算法计算 S_i 的 128 位消息摘要 $Hash_i$。

(3) 提取 D_i 内相应信息以及消息摘要值 $Hash_i$，存贮到数据库主机进程表相应字段内，生成新的第 i 条记录。

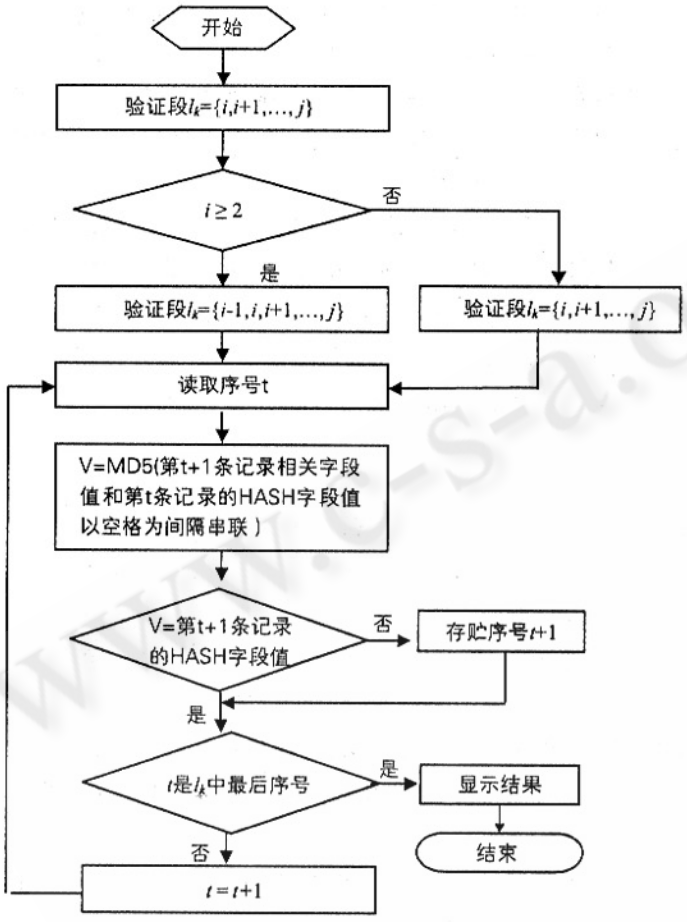

图 2　完整性验证流程图

完整性验证面临的主要问题是数据库表中存贮来自于多数据源的信息，因此满足取证分析条件的相关记录并不完全连续。系统采用验证段划分和分段验证的方法实现完整性验证。

验证段划分：如图 2 所示，根据取证分析条件，取证中心从数据库表中读取满足条件的记录序号，当相邻记录序号间隔数 $d>2$，则记为序列断开点；当相邻记录序号 $d=2$ 时，填加缺少的序号；根据以上规则将取证分析序列 l 划分为以断开点为划分的验证段集合 $l=\{l_1, l_2, \cdots, l_m\}$。

分段验证：如图 2 所示，设验证段 $l_k=\{i, i+1, \cdots, j\} \in l$，$l_k$ 中元素是记录序号。如果 $i \geq 2$，则 l_k 中增加序号 $i-1$，得 $l_k=\{i-1, i, i+1, \cdots, j\}$，否则 l_k 保持不变。读取 l_k 中第一个序号 t 所指记录相关字段内容，以及第 $t-1$ 记录的 Hash 字段值。以空格为间隔将读到内容串联形成字符串 s_i。采用 MD5 算法计算 s_i 的 128 位单向散列，同第 i 条记录存贮的 Hash 字段值比较，如果相等则通过完整性验证，否则存在完整性问题。如此的过程循环，直到遍历验证段的所有记录。

通过以上方法可以检测潜在数字证据存贮过程中的完整性问题。假设某验证段的记录序号序列为 $\{i, i+1, \cdots, k-1, k, k+1, \cdots, j\}$，如果其中序号为 k 的记录被删除，则验证时会有 $s_{k+1}=$（第 $k+1$ 条记录字段间添加空格形成字符串，同原来的第 $k-1$ 条记录 Hash 字段值串联），MD5（s_{k+1}）值同该条记录存贮的 Hash 字段值不同，因此可以判断该记录的完整性遭到破坏。存贮记录被添加、篡改或伪造，其验证方法同理。

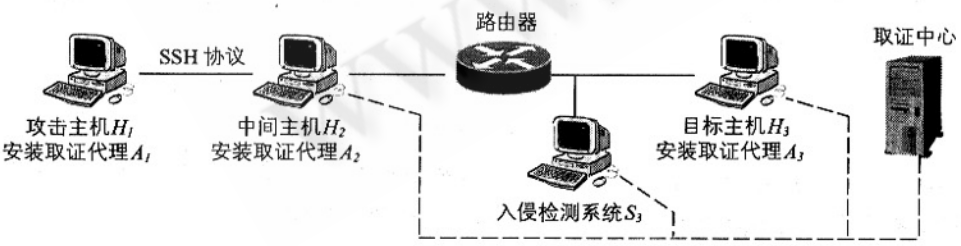

图 3　主机 H_1 控制主机 H_2 攻击主机 H_3 示意图

4.2 内部网络攻击源定位

据统计，网络犯罪大部分来自于网络内部。对于常见的难以伪造 IP 地址类的网络攻击，根据入侵检测 Snort 系统提供的报警信息、网络数据流信息和内部网络内路由器 IP 地址配置表，系统可以对发生在内部网络的网络攻击实现攻击路径识别和攻击源定位。

对于控制中间主机而进行的隐匿攻击，如图 3 所示，系统根据数据流协议类型、数据流 IP 地址间的相关性特征等确定攻击路径和攻击源。该方法能对控制安装有取证代理的主机作为中间主机发动的攻击，实现攻击路径识别和攻击源定位。

该方法具体步骤描述如下：

(1) 取证中心根据数据库中收集存贮的入侵检测系统 S_3 提供报警信息、被攻击主机 $H_{victim}=H_3$ 网络数据流信息以及路由器 IP 地址配置表确定发出攻击命令的主机 $H_{attack}=H_2$。根据网络攻击过程取证分析确定网络攻击时间范围$[t_1, t_2]$。

(2) 取证分析对数据库存贮的 H_{attack} 网络数据流信息进行分析，确定在攻击时间范围$[t_1, t_2]$内是否满足：一是否有主机 H_i 通过 Telnet 协议或 SSH 协议同 H_{attack} 远程交互；二是否有数据流流入主机 H_{attack}，立刻有数据流流出 H_{attack}，表现为多对相邻数据报源 IP 地址和目的 IP 相互连接，形成连接链。即设 n(n > 设定阈值)对 Telnet 协议或 SSH 协议相邻数据报记录 p_i =(源 IP 地址 s_i，目的 IP 地址 d_i)，p_{i+1} =(源 IP 地址 s_{i+1}，目的 IP 地址 d_{i+1})，如果满足 $d_i=s_{i+1}$，$s_i \neq d_{i+1}$，s_i 和 d_{i+1} 是主机 H_{attack} 或 H_{victim} 的 IP 地址，则 $H_{remote}=H_i$ 是远程控制主机(如图 3 中 H_1)。如果满足以上条件则转到步骤 3，否则转到步骤 4；

(3) $H_{victim}=H_{attack}$，$H_{attack}=H_{remote}$ 转到步骤 2；

(4) 如果 H_{remote} 未安装有取证代理，则该主机是可以反向追溯到的最终的攻击源；如果 H_{remote} 安装有取证代理，则该主机就是真正的攻击源[8]。根据步骤 2 中确定的 Hattack 和内部网络内路由器 IP 地址配置表可以进一步建立攻击路径[5]。

5 小结

蜜罐技术已经成为安全专家们所青睐的对付黑客的有效工具之一。它的最大优势在于发现新型的攻击工具。值得一提的是入侵者们也在开发各种工具来对抗现有的侦听技术，如将数据包分片再进行重组等，入侵检测系统要想检查出其攻击特征将变的极其困难，所以必须对各种新的攻击方法保持高度的关注。另外管理和分析 Honeynet 要耗费管理员大量的精力，管理员需要经常对可疑的网络事件进行深度分析，这需要很长的时间和熟练的分析能力，因此提高系统的智能分析能力也是非常重要的。

另外蜜罐技术面临着一定的法律问题。蜜罐技术的最初目标是为起诉攻击者提供收集证据的手段，但是一些国家的法律规定蜜罐收集的证据不能作为起诉的证据，因此必须明确各种与蜜罐有关的法律事物，才能更好地应用蜜罐进行更有效的工作。

参考文献

1 吴际、黄传河、王丽娜等，基于数据挖掘的入侵检测系统研究[J]，计算机工程与应用，2003，40(4)：166－168。

2 薛静锋、曹元大，基于数据挖掘的入侵检测[J]，计算机工程，2003，29(9)：17－18。

3 温智宇、唐红、吴渝，数据挖掘技术在入侵检测系统中的应用[J]，计算机工程与应用，2003，40(17)：153－156。

4 熊华等，网络安全－取证与蜜罐[M]，北京 人民邮电出版社，2003。

5 段海新等译 Computer Forensics[M]，北京：人民邮电出版社，2003。

6 张世永，网络信息安全[M]，北京 科学出版社，2003。

7 韩东海等，入侵检测系统实例剖析[M]，北京：清华大学出版社，2002。

8 戴江山，主动型网络取证模型及其关键问题研究[D]．南京，博士论文，2005。