计算机研究与发展
Journal of Computer Research and Development
ISSN 1000-1239/CN 11-1777/TP
43(3):417~422, 2006

基于有穷自动机的网络扫描检测算法研究与实现

刘利军　　怀进鹏
(北京航空航天大学计算机学院　北京　100083)
(liulijun@act.buaa.edu.cn)

Research of a Network Scan Detection Algorithm Based on the FSA Model

Liu Lijun and Huai Jinpeng
(*School of Computer Science and Engineering*, *Beihang University*, *Beijing* 100083)

Abstract　Network scan is often the prelude of the network intrusion. Thus precise detection of the network scan plays an important role in the pre-alert of the network intrusion. But the current scan detection technologies are too simple and may be evaded by attackers easily. In this paper, based on the analysis of both the scan and detection technologies, a detection algorithm called SBIPA (FSA-based intrusion pre-alert algorithm) is proposed based on the FSA (finite state automata) model and the key implementation technology is analyzed. The state transfer diagram is used to illustrate the network scan packet series, and three different mechanisms are designed to detect the scan event based on FSA. Experiment reveals that this algorithm not only can detect the single type scan activity more precisely, but also can detect the unobvious scan such as distributed and multi-type mixed scan very well, which can't be detected by other detection technologies. It is believed that it eliminates the limitations of the current scan detection technology and has an important research and practice value.

Key words　network scan; intrusion pre-alert; FSA; detection algorithm

摘　要　网络扫描通常是入侵的前奏,准确的检测网络扫描可以对网络入侵起到重要的预警作用.现有的网络扫描检测机制都过于简单且易于被攻击者逃避.提出了一种基于有穷自动机模型检测网络扫描的入侵预警算法(FSA-based intrusion pre-alert algorithm, SBIPA),用自动机状态迁移图表达扫描报文序列,同时设计了3种不同的机制基于自动机模型对扫描事件进行检测,并讨论了算法实现中的关键技术.实验表明,该算法能在更准确的检测普通扫描的同时,对分布式、多类型混杂扫描等现有技术难以检测的隐蔽扫描也有很好的检测效果,有效弥补了现有同类技术的不足.

关键词　网络扫描;入侵预警;自动机;检测算法

中图法分类号　TP393.08

1　引　　言

在真正发起攻击之前,入侵者会通过网络扫描技术对目标网络进行探测和信息收集,以决定下一步采取的攻击手段和步骤,因此网络扫描通常是入侵事件发起的前奏.因而对网络扫描活动进行准确的检测、报警,对网络入侵具有重要的预警作用.同时,由于在扫描活动中攻击者要收集扫描的反馈信息,通常不能使用IP地址欺骗(IP spoofing)技术来

收稿日期:2004-11-25;修回日期:2005-05-23
基金项目:国家"八六三"高技术研究发展计划基金项目(2001AA110485,2001AA110233,2001AA144150);国家自然科学基金项目(60073006)

隐藏自己,这对准确定位扫描源、防入侵和制定安全策略具有实际意义.

目前,网络入侵检测系统[1](network intrusion detection system,NIDS)所采用的扫描检测技术都过于简单,通常以一个时段内扫描次数阈值作为分析依据(如在 M 时间段内(单位为 s)从相同的源 IP 地址发出的探测超过 N 次[2]),易于被攻击者避开,如分布式扫描就是一种典型的逃避手段,扫描者利用多台主机同时对目标系统发起扫描,使每台主机发起的探测次数限制在报警阈值内,以逃避检测;同时,由于扫描的多样性和动态性,现有简单的扫描检测技术无法适应.

针对上述问题,本文提出了一种基于有穷自动机状态转换分析检测网络扫描的入侵预警算法 SBIPA,利用状态序列刻画扫描特征,利用状态转换树对网络中发生的扫描活动进行全局的分析和检测,不仅提高了对各种单一类型的扫描活动的检测准确度,而且能够对分布式、多类型混杂扫描等现有技术难以检测的隐蔽扫描进行有效检测,提高了网络入侵预警能力.

2 网络扫描技术

目前主要的网络扫描技术[2,3]有如下几类:

(1) TCP connect 扫描

向目标机的某一端口发起连接请求,如果能完成 3 次握手,那么目标机的该端口是开放的,否则就是关闭的.

(2) TCP SYN 扫描

向目标机某端口发送一个 SYN 数据包,如果目标端口处于侦听状态,会返回一个 SYN-ACK 应答,否则会应答一个 RST 报文. 当接收到 SYN-ACK 应答之后,发起者并不完成 3 次握手过程,而是应答一个 RST 报文取消连接请求.

(3) TCP ACK 扫描

向某个端口直接发送 ACK 报文,如果返回 RST 报文,则表示该目标端口没有被过滤,否则,端口被过滤掉. 主要用于探测目标网络的防火墙端口过滤规则的设置情况.

(4) 其他

主要是通过对报文设置特殊的、正常通信中不会出现的 TCP 标识位组合进行,如 SYN-FIN 扫描向目标机发送同时设置 SYN 和 FIN 标识位的报文,以穿过不允许单纯的 SYN 数据包通过的防火墙;Null 扫描清空所有标识位;FIN 扫描和 Xmas Tree 扫描(同时设置 FPU 位)也企图通过防火墙的过滤,如果目标端口关闭,则会返回一个 RST 应答,否则没有应答.

3 基于有穷自动机的检测算法

3.1 算法的提出

在当前主要的扫描手段中,设置报文非法标识位组合的扫描活动易于检测,因为其报文特征明显且在正常通信中不可能出现;但对于像 SYN 扫描和 connect 扫描这样的技术,其单个扫描报文在正常通信中也会出现,因而使得检测非常困难. 但它们有一个共同点,即扫描是通过一个特定的报文序列完成的. 根据这个特征,我们基于有穷状态机,通过网络连接的不同状态及状态之间的转换序列来刻画扫描报文序列,提出了一种新的入侵预警算法 SBIPA,并在状态节点中引入数量统计、动作触发等处理机制,从而对现有技术难以检测的分布式扫描等隐蔽扫描进行检测.

首先定义 SBIPA 的状态转换模型如下:

定义 1. SBIPA 的状态转换模型是一个五元组 $M=(Q,\Sigma,\delta,q_0,F)$,其中:

(1) Q 是一个网络连接在扫描序列中所处状态的有穷集合;

(2) Σ 是由数据报文所有可能的标识位组合构成的有穷集合(比如,它的元素有 SYN,SYN|ACK 等);

(3) $\delta:Q\times\Sigma\to Q$ 是状态转换函数;

(4) $q_0\in Q$ 是起始状态,为网络连接上出现第 1 个报文时的状态;

(5) $F\subseteq Q$ 是终止状态集合,是某种扫描报文序列完成时网络连接的状态.

转换函数 δ 指明了 M 处于状态 q_{i-1} 时在报文标识位组合 a 作用下转换到的下一个状态 q_i,即 $\delta(q_{i-1},a)=q_i$,其中 $q_{i-1},q_i\in Q,a\in\Sigma$.

3.2 关键数据结构及算法处理机制

上述模型只定义了一个网络连接的扫描状态转换情况,为了对整个网络的所有连接进行扫描检测,最关键的问题是在算法中如何组织连接信息、如何表达状态转换模型以及如何将连接信息与状态转换模型进行关联处理.

3.2.1 网络连接链表

一个网络连接的信息由一个地址-端口对组成,

即一个四元组〈源 IP 地址,源端口,目标 IP 地址,目标端口〉. 由于网络连接是动态建立、关闭的,在算法处理中需要对连接进行高效的查找定位和新添、删除等操作,因此我们采用链表结构来组织网络连接信息(如图 1 所示). 第 1 层由不同的源地址节点构成主链表,每个源地址节点除了指向主链表中的下一个源地址节点之外还包含一个指针,指向所有与该源地址建立连接的目标地址节点的链表;每个目的地址节点也包含一个指针,指向在对应的源地址和目标地址之间建立的所有连接对应的端口对节点的链表,一个端口对节点即对应一个确定的连接(图 1 中的 *ConnectionInfo*). 这样,多棵树通过一个主链表组织成为一个复杂的多层链表,构成了算法中组织网络连接信息的数据结构.

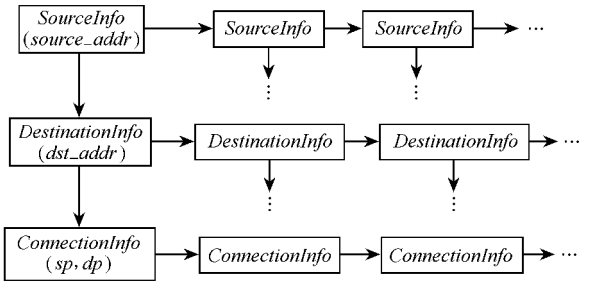

Fig. 1 Network connection list.

图 1 网络连接链表

为了检测由某一源地址发出的扫描,在每个源地址节点中设置存储该地址发起扫描的统计信息的变量,如扫描时间、扫描次数等,当该源地址节点的下属连接中有扫描发生时就更新该节点的存储变量,因而通过检查这些变量就可以检测由单一 IP 地址发起的各种扫描,不论该扫描是通过水平扫描、垂直扫描[2]还是混合方式以及利用不同的扫描技术进行的,都不能逃避算法的检测,因为单一连接上的任何扫描都会引起源地址节点中扫描总次数的增加.但仅通过该机制还不能对分布式扫描等复杂扫描进行检测,这一点将在下面介绍的算法处理机制中解决.

3.2.2 状态转换树

为了利用定义 1 中的自动机模型进行扫描检测,我们用一棵状态转换树表达检测模型的状态转换机制,如图 2 所示.

算法中的状态转换树结构逻辑上对应自动机模型的状态图,树的各节点之间的关系反映了自动机的状态转换函数规定的状态迁移关系.

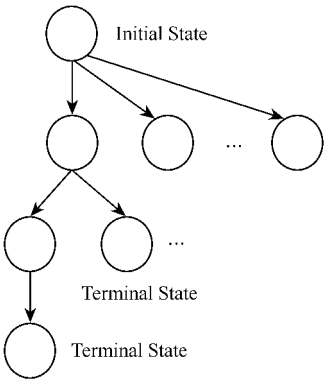

Fig. 2 State transfer tree.

图 2 状态转换树

3.2.3 检测机制

为了对网络内的所有连接进行检测,需要对连接链表和状态转换树中的信息进行关联处理,该机制在算法中设计如下(如图 3 所示):在图 1 所示的 *ConnectionInfo* 节点中包含一个状态节点指针,指向该连接当前所处状态在状态转换树中对应的节点;同时在图 2 所示每个状态节点中包含一个指针链表,链表中的节点分别指向每一个当前处于该状态的 *ConnectionInfo* 节点,这样就在状态节点和连接节点之间提供了一种双向追踪机制. 为了便于对扫描报文进行追踪,在连接节点中还包含了一个指向扫描报文序列的指针.

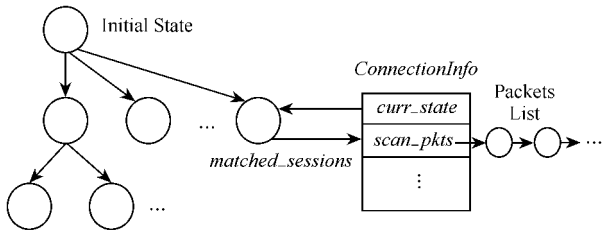

Fig. 3 The match of state tree and connection list.

图 3 状态树与连接表的关联处理

算法采用了如下 3 种不同的扫描检测机制,分别针对不同的扫描方式:

(1) 源地址检测机制

统计连接链表中源地址节点的扫描次数 *scan_cnt* 是否超过极限值 T_1,针对由单一主机发起的扫描(可以是不同类型的扫描方式);

(2) 叶节点检测机制

检查状态树叶节点的匹配次数 *match_cnt* 是否超过极限值 T_2,能够检测第 1 种机制无法检测的由多主机同时发起的单类型的分布式扫描;

(3) 全局检测机制

设置一个扫描次数全局统计变量 $global\_cnt$,所有连接上检测到任何扫描动作发生时都将该变量值递增,通过判断该值是否超过极限值 T_3,可以检测前两种机制无法检测的由多主机同时发起的多类型混杂的扫描.

3.3 检测算法 SBIPA

结合上述对关键数据结构的描述和检测机制的说明,算法 SBIPA 设计如下:

算法 1. 基于有穷自动机的入侵预警算法 SBIPA.

```
void SBIPA (Packet p) /* 以一个新捕获的网络报文作为输入 */
ptrConnNode = FindConnection (p); /* 在连接链表中查找该连接是否已经存在 */
If (ptrConnNode == null) /* 如果连接不存在,创建一个新连接 */
   ptrConnNode = NewConn (p);
   ptrConnNode -> curr_state = q0;
End If
TransConnState(ptrConnNode, p); /* 按照状态转换函数 δ 对连接进行状态迁移 */
If (StateChangedFlag); /* 若连接状态发生变化 */
   AppendToScanPkts(ptrConnNode, p);
   /* 将该数据包追加到连接节点的 scan_pkts 链表 */
End If
Curr_State = ptrConnNode -> curr_state;
If IsLeaf(Curr_State) /* 如果连接状态迁移到状态树的叶节点 */
   Curr_State -> match_cnt ++; /* 增加连接计数 */
   SrcInfoNode = GetSourceNode(ptrConnNode);
   SrcInfoNode -> scan_cnt ++; /* 增加源主机的扫描计数 */
   global_cnt ++; /* 将全局扫描计数器递增 */
End If
CleanOldConnection() /* 清除无效连接 */
If(Curr_State -> match_cnt >= Curr_State -> T2 OR SrcInfoNode -> scan_cnt >= T1 OR global_cnt >= T3)
   alert();
End If
Return
```

3.4 实现关键技术

(1) 状态转换序列规则描述语言

为了在算法运行时建立状态树,需要对不同扫描的状态转换序列进行描述. 该描述不仅要表达状态之间的转换关系,而且要描述实现状态转换时需要的触发条件. 为此,我们设计了一个脚本语言 StateSDL (state sequence description language),StateSDL 的文法用巴克斯范式(BNF)描述如下:

〈状态描述〉::=〈状态名〉{〈关键字-值对〉}|〈状态名〉:〈状态名〉{〈关键字-值对〉}

〈状态名〉::=任意字符串

〈关键字-值对〉::=〈关键字〉:〈值〉|〈关键字〉:〈值〉;〈关键字-值对〉

〈关键字〉::=direction|server_ignore|protocol|tcp_flag|flag_mask|threshold|action|msg

〈值〉::=字符串或数字

状态名惟一标识了一个状态,为避免状态树解析时出现二义性,在同一个状态序列描述文件中不应该出现两个同名状态;当在一条规则中顺序出现两个状态名时,第 2 个表示第 1 个状态的父状态,一个状态至多只能有一个父状态. 状态名后的关键字-值对描述了从父状态转移到该状态的触发条件,如报文的传输方向、协议类型等.

(2) 无效连接的清除

网络连接是动态创建和关闭的,为保证检测的准确性,对无效连接的及时清除是十分关键的. 连接的清除可以有两种方案,第 1 种是定时清除,第 2 种是在符合报警条件准备报警时清除. 定时清除的优点是可以根据报警的频度灵活的调整清除的间隔时间,从而可以最大可能的减小连接清除带来的负载,但缺点是报警时间和连接清除的时间不一致,因此进行报警时报警条件的统计数量中可能有一定数量的无效连接,从而造成误报;因此在实现中我们采用了第 2 种方式,即在出现报警条件时进行连接清除. 在实际应用中,由于报警频度并不会太高,因此这种清除方式带来的额外负载很低.

4 实验分析

实验运行在 100Mbps 的共享局域网环境下,1 台主机运行实现 SPIBA 算法的 PreIDS 系统,5 台主机运行测试软件,对网络内的主机发起各种扫描. 测试工具采用漏洞扫描器 Nessus 和自行开发的发包软件 PacketSender.

4.1 普通扫描测试

使用 Nessus 从一台主机对另一台主机的 1～256 号端口分别发起 TCP connect 扫描、SYN 扫描、FIN 扫描、Xmas Tree 扫描、Null 扫描,检测结果如表 1 所示:

Table 1 Common Scan Test Result
表 1 普通扫描测试结果

No.	Scan Type	Scan Port	Alert Times
1	TCP connect	1～256	68
2	SYN	1～256	90
3	FIN	1～256	153
4	Xmas Tree	1～256	333
5	Null	1～256	51

对不同类型的常规扫描,算法都能够及时检测到并报警. 对不同类型扫描的报警次数不同是因为不同类型的扫描报警极限值不同.

4.2 隐蔽扫描测试

利用 PacketSender 同时从多主机对另一主机发起相同类型的扫描,然后发起不同类型混杂的扫描,两种情况下扫描端口的数量都很少,以逃避通常的检测技术.

(1) 多主机发起同类型分布式扫描

测试结果如表 2 所示.

SYN 扫描的状态树叶节点可以很准确的累计扫描动作,从而进行报警;同时由于扫描总次数超过了全局阈值,全局检测机制也做出了准确报警.

(2) 多主机发起多类型分布式扫描

测试结果如表 3 所示.

叶节点检测机制报警的 4 次为 1 次 FIN 扫描报警,3 次 Xmas Tree 扫描报警,这是因为这两种扫描的特征明显,报警阈值较低;全局检测机制累计了不同扫描的总次数,从而能够检测到这些非常隐蔽的扫描.

Table 2 Multi-Host Single Type Distributed Scan Test Result
表 2 多主机单类型分布式扫描测试结果

No.	Scan Host	Scan Type	Scan Port	Alert
1	Host1	SYN	1～3	
2	Host2	SYN	4～6	The leaf node mechanism alerts 2 times. The global mechanism alerts 1 time.
3	Host3	SYN	7～9	
4	Host4	SYN	10～12	
5	Host5	SYN	13～15	

Table 3 Multi-Host Multi-Type Distributed Scan Test Result
表 3 多主机多类型分布式扫描测试结果

No.	Scan Host	Scan Type	Scan Port	Alert
1	Host1	TCP connect	1～3	
2	Host2	SYN	4～6	The leaf node mechanism alerts 4 times. The global mechanism alerts 1 time.
3	Host3	FIN	7～9	
4	Host4	Xmas Tree	10～12	
5	Host5	Null	13～15	

5 相关工作比较

网络扫描的检测对于入侵预警有非常重要的作用,但与问题本身的重要性相比,在过去的研究中得到的关注却相当少[2].

在 NSM[4],GrIDS[5],Snort[6] 等著名的入侵检测系统中,检测扫描的依据是在 X 秒时间内,从任一源 IP 地址向超过 Y 个其他主机发起的连接企图.与本文中的检测算法相比,这种机制有两个缺点:一是对所有类型的扫描只能采用统一的检测条件(即固定的 X,Y),不能根据不同扫描的特点单独规定报警阈值;二是只能检测从相同扫描源发起的扫描,不能检测分布式扫描等隐蔽扫描.

在 Emerald[7] 系统中,检测扫描使用了异常检测的方法,该系统将远程通信主机视为一个主体,利用统计技术对主体的通信流量建立统计轮廓,并不断将主体当前的通信与其统计轮廓进行比较,如果

偏差过大则认为是异常的.文献[2]也提出了一种基于扫描报文异常评估和异常事件关联、分组的扫描检测机制.这类异常检测方法的主要缺陷是统计轮廓的建立需要相当长时间和大量正常通信的采样,同时误报率较高.

6 结束语

为了准确的检测网络扫描活动,本文提出了一种基于有穷自动机状态转换分析检测网络扫描的入侵预警算法SBIPA,并对系统实现中的部分关键问题进行了讨论.实验证明,基于该算法实现的检测系统在检测普通扫描和隐蔽式扫描方面均比现有的同类技术有明显优越性,能够对网络入侵起到更好的预警作用.

参 考 文 献

1 J. Allen, A. Christie, A. Fithen, *et al*. State of the practice of intrusion detection technologies. Software Engineering Institute, Carnegie Mellon University, Tech. Rep.: CMU/SEI-99-TR-028, 2000

2 S. Staniford, J. A. Hoagland, J. M. McAlerney. Practical automated detection of stealthy portscans. The 7th ACM Conf. Computer and Communications Security, Athens, Greece, 2000

3 Fyodor. The art of port scanning. http://www.insecure.org/nmap/nmap_doc.html, 2004

4 L. Heberlein, G. Dias, K. Levitt, *et al*. A network security monitor. IEEE Symposium on Research in Security and Privacy, Oakland, CA, 1990

5 Steven Cheung, Rick Crawford, Mark Dilger, *et al*. The design of GrIDS: A graph-based intrusion detection system. U. C. Davis Computer Science Department, Tech. Rep.: CSE-99-2, 1999

6 Martin Roesch. snort. http://www.snort.org/, 2004

7 P. Porras, A. Valdes. Live traffic analysis of TCP/IP gateways. 1998 Internet Society Symposium on Network and Distributed System Security, San Diego, 1998

Liu Lijun, born in 1977. Ph. D. candidate. His main research interests is network security.

刘利军,1977 年生,博士研究生,主要研究方向为网络安全.

Huai Jinpeng, born in 1962. Professor and Ph. D. supervisor. His main research interests include computer new technology and network security.

怀进鹏,1962 年生,教授,博士生导师,主要研究方向为计算机软件新技术、网络安全.

Research Background

Network intrusion became more and more serious and brought heavy damages to the network-based business development during the past ten years. Network scan is often the prelude of the network intrusion. Thus precise detection of the network scan can play an important role in the pre-alert of the network intrusion. However, the scan detection technologies nowadays are too simple and may be evaded by attackers easily. Under this background, we propose a detection algorithm called SBIPA based on the FSA(finite state automata) model in order to detect the network scan more precisely. In this algorithm, we use the state transfer diagram to illustrate the network scan packet series, and design three different mechanisms to detect the scan event based on the FSA. Experiment result reveals that this algorithm not only can detect the single type scan activity more precisely, but also can detect the unobvious scan such as distributed and multi-type mixed scan very well, which can't be detected by other detection technologies. This work eliminates the limitations of scan detection technologies nowadays and can play a better role in the pre-alert of the network intrusion. Our work is supported by the 863 High Technology Research and Development Program of China and the National Science Foundation.