文章编号:1001 - 9081(2005)09 - 2175 - 02

基于信息融合的计算机网络信息发现

孙 亮,李 东,张 涛,熊永平,邹百柳
(哈尔滨工业大学 计算机网络与信息安全技术研究中心,黑龙江 哈尔滨 150001)
(sunliang@pact518.hit.edu.cn)

摘 要:引入信息融合技术,利用多个探测工具收集网络信息,并将来自不同类型网络探测工具的信息在不同的层次上进行融合。在数据层融合中,使用模糊逻辑的统计方法识别系统类别、区分网络设备;在决策层融合中,通过系统知识库的支持,获得最可信的网络信息。

关键词:信息发现; 主动探测; 网络监听; 信息融合
中图分类号: TP393.08 **文献标识码**:A

Computer network information discovery based on information fusion

SUN Liang, LI Dong, ZHANG Tao, XIONG Yong-ping, ZOU Bai-liu
(*Research Center of Computer Network and Information Security Technology, Harbin Institute of Technology, Harbin Heilongjiang* 150001, *China*)

Abstract: The available tools for detecting network information can hardly meet the demands of acquiring the completeness and precision of network information for the researchers. The information fusion technology was applied to collect the network information using several detecting tools. The information from different detecting tools was fused in different layers. In data layer, the fuzzy logical statistic method was adopted to identify system type and network device, and in logic layer, the most credible information was obtained with the support of system knowledge database.

Key words: information discovery; active probing; network sniffing; information fusion

0 引言

计算机网络信息发现技术的研究是网络安全领域的一个重要研究方向。目前,网络信息发现采用的探测工具有两类:一类是基于网络监听的方法,捕获网络中的数据包,分析捕获的数据的特征,发现网络信息系统的一些属性;另一类是采用网络扫描的方法,向目的主机发送特征数据包文,并根据收到的响应分析目标系统的属性。

随着网络规模的扩大、网络带宽的增加、网络业务种类的多样化,网络中的各种系统设备变得多种多样,网络管理、网络安全、网络行为等方面的研究者需要更加准确全面地了解网络上各种设备的系统信息、网络结构、网络服务等信息。采用单种或单个网络探测工具提供的信息已经无法满足研究者的需求,需要一种功能更为齐全、性能更为强大的方法来解决这一问题。

网络探测工具作为一种分类器,不同探测工具提供了不同表达方式、不同可信度、不同侧重点和用途的信息。结合信息融合的思想,本文提出一种主要面向于计算机网络安全性分析的、可扩展的、高效的、精确的网络信息发现平台。其可以发现的信息主要有:构成计算机网络的路由器和计算机,网络间的路由信息,以及网络中计算机的信息(IP 地址、操作系统类型、开放的端口与服务、网络服务程序版本等)。

1 相关的研究工作

传统的系统信息扫描工具采用主动发包探测的方法来探测系统信息[1,2,3],例如 Nmap、Nessus 等,ISS 公司也开发了一套针对网络信息系统的包括 Internet 扫描器、数据库扫描器和系统扫描器三个应用程序的安全评价平台。这种方法是抽取已知的系统信息特征,并归纳成规则表达,组成一个"系统特征库",每一个特征对应一个规则,将搜集到的目标系统信息与已有的规则一一匹配。但是这些工具在探测的准确度上存在一定的问题,很容易受到网络过滤设备存在的影响,其产生的大量数据报文也会影响网络的正常运行,使得这一类型的工具只能间断的运行,有很大的时间局限性。

众多的 sniffer 工具常常用来做报文捕获和协议分析等工作[4,5],例如 tcpdump、libpcap 等。Sniffer 工具也可以用来分析系统信息,例如 NeVO 就是一个采用网络监听分析系统信息的工具,它不间断的运行在网络出入口,监听系统内外之间的通讯,通过捕包、协议还原、分析来获取系统的拓扑、服务和端口信息,有效弥补了主动扫描工具间歇工作特点的不足。但是这种工具的局限性在于空间限制,它无法获取系统内部的通讯流量,若检测点置于系统内部,由于多使用交换网络,也是只能捕获部分信息。

信息融合又称为多传感器融合,是将来自于多传感器或多源的信息和数据进行综合处理,从而对目标系统得出更为准确、全面、可信的结论。它最早应用于军事领域,经过多年发展目前已广泛应用于目标检测、图像分析、机器人和智能仪器系统等方面。按照数据抽象的层次,信息融合可以分为数据层融合、特征层融合和决策层融合[6]。

为了提高信息的准确性和完备性,国外研究者已经开始把这种基于多源数据的信息获取、分析、分类和决策的信息融合技术应用在网络安全分析方面。马萨诸塞州立大学研究了一种决策树算法用来分析 UNIX 系统日志,使用多种信号检测技术来确定是否有入侵事件发生,Nong Ye 为了提高检测

收稿日期:2005 - 03 - 12;**修订日期**:2005 - 05 - 17 **基金项目**:国防十五预研安全性分析项目(41315.7.1)
作者简介:孙亮(1980 -),男,辽宁丹东人,硕士研究生,主要研究方向:网络安全评估; 李东(1967 -),男,广西南宁人,教授,主要研究方向:计算机系统结构、并行处理、网络安全; 张涛(1977 -),男,山东德州人,博士研究生,主要研究方向:网络安全评估.

的准确性,使用决策级的融合来获取不同入侵检测工具的结果权值[7]。Alfonso Valdes 在 EMERALD 的基础上提出一种入侵检测检测工具关联决策,分为三个层次:事件聚集、传感器耦合,和警报融合,并且实现了一个原型系统[8]。Tim Bass 提出集成多种入侵检测工具实现分布式网络入侵检测模型,他从数据、信息、知识的三个层次来综合分析可能发生的网络入侵事件。Oleg Sheyner 在开发攻击图的自动生成系统时,需要首先建立了网络信息模型,其中信息采集和处理部分使用了多种探测工具来获取系统设备信息和系统连接信息[9]。

2 系统设计

信息系统的结构具有一定的稳定性,同时其局部存在着动态变化。要完成对信息系统全面的信息发现,必须消除单种探测工具的信息获取能力在时间上或空间上的局限性。本文提出一种集成多个和多种网络探测工具的可扩展信息发现平台,结构模型如图 1 所示。n 个探测节点 $S_1, S_2, \cdots, S_n$ 分别代表 n 个分布在网络上的不同类型信息探测节点,探测节点收集到原始信息 $Y_1, Y_2, \cdots, Y_n$ 之后,在局部节点分别处理,并把获取的信息 $u_1, u_2, \cdots, u_n$ 送到融合中心,融合中心根据检测方法做出判决。融合中心采用数据级融合和决策级融合两种方法。同时系统设计考虑到信息系统规模大小的变化,可以通过增加或减少探测点来提供适当的覆盖范围。

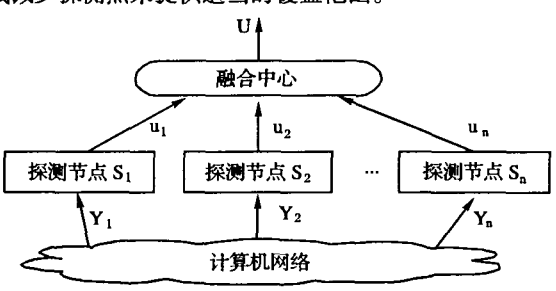

图 1 并行分布式结构

2.1 探测节点类型与融合层次

数据层的融合方法直接分析处理来自于探测节点的数据,这种融合的主要优点是能够保持尽可能多的现场数据,提供其他融合层次所不能提供的细微信息,数据通讯量较大。在决策层的融合方法中,每个探测节点首先完成分析处理以获得独立的分析结果,然后再对来自每个探测节点的分析结果进行融合,这种层次的融合具有很高的灵活性,通讯量小,对探测节点的要求比较高。

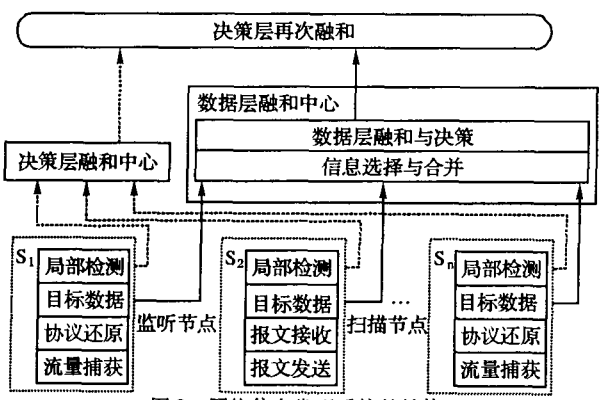

图 2 网络信息发现系统的结构

网络信息探测节点包括两种类型:基于网络监听的探测节点和基于网络扫描的探测节点。对于基于网络监听的探测节点,采集到的数据量巨大,分析结果具有较强的可信度,此类探测节点主要采用决策层融合的方法,同时记录少量的特殊类型报文(具有特定标志的 TCP 报文和 ICMP 报文)用作数据层决策。对于基于网络扫描的探测节点,采集到的数据量相对较少,自身的判断能力较弱,此类探测节点主要采用数据层融合的方法,同时也将其判断结果输入到决策层融合中心。网络信息发现系统的结构如图 2 所示。

2.2 融合方法

2.2.1 *数据层融合方法*

数据层融合的目的主要是为了精确识别操作系统的类型。不同操作系统中 TCP/IP 协议栈的实现上存在着差异,因此可以利用特征报文来识别不同的系统类别。其中特征报文类型主要包括:FIN 探测,BOGUS 标记探测器,TCP 初始化序列号取样,不分片标志,TCP 初始化窗口,ACK 值,ICMP 错误信息终结,ICMP 消息引用,ICMP 错误消息回应完整性,服务类型,分段控制,TCP 选项,SYN 洪水限度等。

由于网络过滤设备和拓扑结构的不同,探测节点获取的特征报文不可避免地存在精度问题,而且特征报文受系统配置和安装软件的影响也存在一定的模糊度。系统类型对于特征报文的匹配程度可以分为:测试满足、测试可能满足、测试可能不满足、测试不满足,如图 3 所示。本文提出的高精度辨识方法,通过增加特征报文的种类,使用基于模糊逻辑的统计方法来判断系统类别。

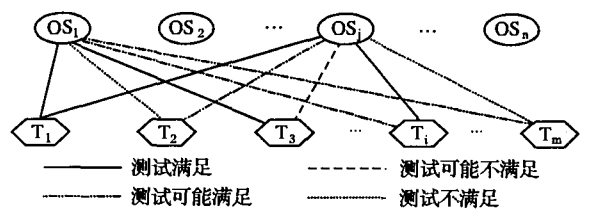

图 3 特征报文与系统对应关系

设 T_i 为使用一种特征报文的方法样例,共有 m 种测试方法样例,所有的探测方法样例可以表示为测试方法集 $(T_1, T_2, \cdots, T_i, \cdots, T_m)$,$OS_j$ 为一种操作系统类型,共有 n 种操作系统 $(OS_1, OS_2, \cdots, OS_j, \cdots, OS_n)$;对应于测试方法样例 T_i 的返回结果,目标系统隶属于操作系统 OS_j 的程度可以用 $\mu_{(i,j)}$ 表示,则可以用 $(\mu_{(i,1)}, \mu_{(i,2)}, \cdots, \mu_{(i,j)}, \cdots, \mu_{(i,n)})$ 表示此测试样例隶属于所有操作系统的程度;目标系统类别与特征报文样例对应关系如下表 1 所示,根据已知的先验知识 $\mu_{(i,j)}$ 的取值可以根据表 2 获得。

表 1 特征报文样例与系统对应关系

	OS_1	OS_2	…	OS_n
T_1	$\mu_{(1,1)}$	$\mu_{(1,2)}$	…	$\mu_{(1,n)}$
T_2	$\mu_{(2,1)}$	$\mu_{(2,2)}$	…	$\mu_{(2,n)}$
…	…	…	…	…
T_m	$\mu_{(m,1)}$	$\mu_{(m,2)}$	…	$\mu_{(m,n)}$

表 2 特征报文隶属度赋值

$\mu_{(i,j)}$	描述
0	T_i 测试结果不满足 OS_j 的特征
0.3 ~ 0.4	T_i 测试结果可能不满足 OS_j 的特征
0.7 ~ 0.8	T_i 测试结果可能满足 OS_j 的特征
1	T_i 测试结果满足 OS_j 的特征

对于一次目标探测,按照对应的响应和表 1 赋值给 $\mu_{(i,j)}$,目标系统的类型可以根据如下公式计算获得:

$$OS = OS_j\left[\mathrm{MAX}\left(\sum_{i=1}^{m}\mu_{(i,1)}, \sum_{i=1}^{m}\mu_{(i,2)}, \cdots, \sum_{i=1}^{m}\mu_{(i,j)}, \cdots, \sum_{i=1}^{m}\mu_{(i,n)}\right)\right]$$

(下转第 2195 页)

[2] COMBA JLD, DIETRICH CA, PAGOT CA. Computation on GPUs: From a Programmable Pipeline to an Efficient Stream Processor[J]. Revista Informática Teóricae Aplicada, 2003, X(2): 41 - 70.
[3] 吴恩华,柳有权. 基于图形处理器(GPU)的通用计算[J]. 计算机辅助设计与图形学学报,2004,16(5):601 - 612.
[4] HOPF M, ERTL T. Hardware Accelerated Wavelet Transformations [A]. Proceedings EG/IEEE TCVG Symposium on Visualization VisSym '00[C], 2000. 93 - 103.
[5] 吴仲乐,王遵亮,罗立民. 基于 GPU 的快速 Level Set 图像分割[J]. 中国图象图形学报,2004,9(6):679 - 683.
[6] KRüGER J, WESTERMANN R. Linear Algebra Operators for GPU implementation of Numerical Algorithms[J]. ACM Transactions on Graphics, 2003, 22(3): 908 - 916.
[7] NVIDIA Corporation . NVIDIA OpenGL Extension Specifications [EB/OL] http://developer. nvidia. com /object/nvidia _ opengl _ specs. html, 2004-8-10/2004-11-10.
[8] (日)谷萩隆嗣. 快速算法与并行信号处理[M]. 薛培鼎,徐国弼,译. 北京:科学出版社,2003. 24 - 25.
[9] LAN B, PAT H. Data parallel Computation on Graphics hardware [EB/OL]. http://graphics. stanford. edu /projects/brookgpu/, 2004-8-10/2004-11-10.

(上接第 2176 页)

此种方法可以有效的提高探测的精度,缺点是增加了特征报文的数量。一种有效的改进途径是结合先验知识和实验结果,动态地删除对 $\sum_{\mu(i,j)}(j=1,2,3,\cdots,n)$ 影响小的测试方法和增加影响比较大的测试方法,并且根据大量的统计数据结果提高或者减少 $\mu_{(i,j)}$ 的取值。同传统的严格匹配方法,此方法有了更强的抗干扰能力。

2.2.2 决策层融合方法

决策层融合的数据为主机上开放服务服务与端口的对应关系结论、服务程序版本结论和操作系统类型结论。网络服务与端口对应关系的寻找方法有两种:第一种是查找默认端口/服务对照表,将端口直接对应某个网络服务,网络中服务不开放在默认端口的情况很多,这种直接对应的方法可信度较差;第二种是通过应用层特征匹配的方法找出对应关系,例如在基于监听的方法中,数据包应用层数据的前 3 个字节为"GET"时,可以标识出服务器的端口上开放着 HTTP 服务。显然这两种方法得到的结论有着不同的可信度。同样,不同扫描方法、利用不同特征进行匹配得到的操作系统类型、网络服务类型、服务程序版本都具有不同的可信度。

探测节点首先完成方法可信度的判断,决策层融合中心然后再根据节点的可信度完成最终的可信度判断,找出最佳的结论。可信度用一个 0 ~ 1 的小数表示。设某探测节点利用可信度为 mt_i 的方法得到某个结论 R_i,该探测节点得到该类结论的可信度为 nt_i,则决策层融合中认为该探测节点得到的结论 R_i 可信度为 $mt_i * nt_i$。决策中心计算针对同一问题所有结论的可信度,将可信度最大的作为最终结论。

系统中的每个探测节点上维护一张检测方法与得到结论之间可信度对照表 MT,决策层融合中心中维护一张结论类型与探测节点之间的可信度对照表 NT。其表结构如表 3、表 4 所示。

表 3　方法与结论可信度的对应关系表

结论可信度	方法
M_1	mt_1
M_2	mt_2
…	…
M_n	mt_n

表 4　结论类型与探测节点可信度对应关系

结论类型	探测节点	可信度
R_1	1	nt_1
R_1	2	nt_2
…	…	…
R_i	j	nt_m

2.3 系统知识库

参照 RFC 和已有的各种网络系统信息,本文提出并建立了一个较详细的系统知识库,其内容包括:

1) 默认网络端口与服务类型对照表后门程序开放端口等。

2) 特征关键字表,存放应用协议自身的特征关键字和服务程序特征关键字。

3) 系统协议栈指纹特征表,对应于表 1 的一个协议栈特征库。

4) 探测节点中每种方法与结论可信度的对照表。

5) 结论类型与探测节点可信度的对照表。

3 结语

计算机网络信息发现是一个具有广泛应用背景的课题。随着网络规模的扩大和网络应用技术的发展,针对网络安全、网络管理方面研究的需求,本文提出了基于信息融合的计算机网络信息发现方法。这种方法可以有效的解决单个或单种探测工具在空间和时间上的限制,充分发挥了各种探测工具的优势,具有较强的实用性。

参考文献:

[1] ISS Website[EB/OL]. http://www. iss. com/, 2003.
[2] NMAP Website[EB/OL]. http://www. insecure. org/nmap/index. html, 2003.
[3] ARKIN O. ICMP Usage in Scanning Version 3.0[EB/OL]. http://www. sys-security. com/ archive/papers/ICMP_Scanning_v3. 0. pdf, 2002.
[4] libpcap-0. 7. 2. tar. gz. http://www. tcpdump. org/release/, 2002.
[5] Passive Vulnerability Scanning[EB/OL]. http://www. tenablesecurity. com/nevo. html, 2004.
[6] 何友, 王国宏. 多传感器信息融合及应用[M]. 北京:电子工业出版社, 2000.
[7] YE N, GIORDANO J, FELDMAN J, *et al.* Information fusion techniques for network intrusion detection[A]. Information Technology Conference[C], 1998. 117 - 120.
[8] VALDES A, SKINNER K. An Approach to Sensor Correlation[EB/OL]. http://www. raid-symposium. org/raid2000/Materials/Abstracts/41/corr_approach. pdf, 2000.
[9] SHEYNER O. Scenario Graphs and Attack Graphs[D]. Carnegie Mellon University, 2004.
[10] STEVENS WR. TCP/IP Illustrated, Vol. 1[M]. Addison-Wesley, 1994.
[11] VIVO M, CARRASCO E. A review of port scanning techniques[J]. ACM SIGCOMM Computer Communication Review, 1999, 29(2): 42 - 48.