网络信息监听技术的研究*)

陈国龙[1,2] 陈火旺[1] 康仲生[2]

(国防科技大学计算机学院 长沙410073)[1] (福州大学计算机科学与技术系 福州350002)[2]

Research of Network Information Monitor Technology

CHEN Guo-Long[1,2] CHEN Huo-Wang[1] KANG Zhong-Sheng[2]

(Computer School, National University of Defence Technology, Changsha 410073)[1]

(Department of Computer Science and Technology, Fuzhou University, Fuzhou 350002)[2]

Abstract As a sign of information age, Internet offers millions of information and makes the people's work easy. However, when the enterprise gets the information and communicates smoothly, it is at the risk that the interior technologic secret and the commerce secret is betrayed to the rival through the network. So it is necessary to aduit the information of the network. Network information aduit is the first step of network information aduit. This thesis provides a method of the network information monitor, and gives an effective network packet filters technology named BPF aiming at dealing with the large number of data on the network.

Keywords Network information aduit, Network packet filters technology, BPF(BSD PACKET FILTER)

1 引言

作为信息时代的重要标志之一,国际互联网上的万千信息。企业在上网获取信息和沟通便利的同时,也面临着内部技术机密或者商业机密通过网络泄露给外界特别是竞争对手的严重问题。企业网管有必要采用信息监控审计技术在保证企业网络通信正常运作的情况下,通过分析网络信息流,报告可疑的链接和数据,为防止和杜绝企业信息泄漏提供线索和证据。信息监控审计的第一步是信息监听,只有监听到网络信息,才能根据相应规则对网络信息进行审计[1]。

网络信息监控将获取所有的网络流量,包括所有协议端口所有子网主机的所有交互数据,但在实际应用中,其中存在若干用户不需要关心的数据,或者称为垃圾数据,垃圾数据在所有流量中占有极大比重,严重影响了系统工作效率的提高,因此高效的信息过滤机制是信息监听的重要组成部分,它使得用户可以指定特定的子网主机以及特定的协议端口如HTTP、FTP、EMAIL 等进行过滤,只将用户关心的敏感数据向上层提交,从而提高系统工作效率。

2 基于路由器的网络信息监听技术

在不设置监听端口的情况下,网络路由工作原理如图1:

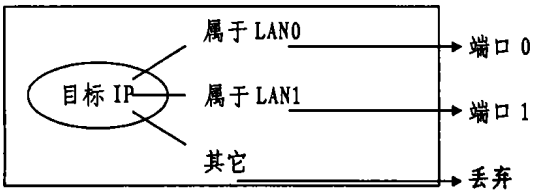

图1

假设设置以太接口2为监听端口,并连接到信息敏感器所在主机,连接图如图2:

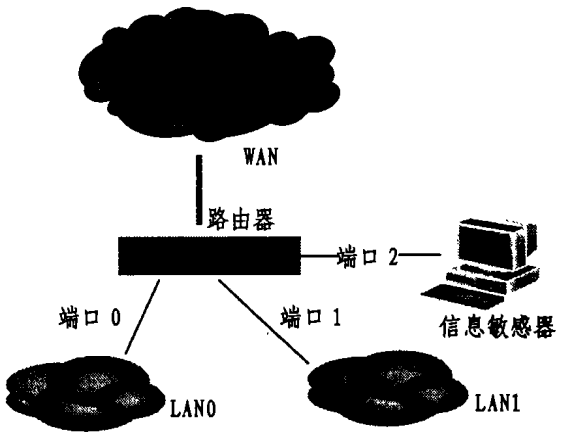

图2

网络路由工作将不同于正常情况,所有的网络信息数据包除按照正常情况转发外,将同时转发到监听端口,从而使得信息敏感器可以监听到所有的网络流量,工作原理示意图如图3:

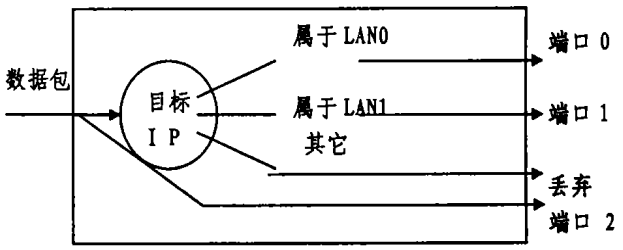

图3

信息敏感器按照单向数据流方式,通过发送 Raw Packet 给控制台实现数据包的采集,敏感器不占用 IP 地址,因此,信息敏感器对任何主机来说都是透明的。通过特定的网络连接方式,任何网络使用者都感觉不到该敏感器的存在,其行为特

*)国家自然基金(编号60172017),福建省教育科技基金(编号 JA01008)。陈国龙 博士,副教授,现为国防科技大学计算机学院博士后,主要研究方向为信息安全、计算机网络等。陈火旺 教授,中国工程院院士。康仲生 硕士生,主要研究方向为信息安全。

征就象一个“黑盒子”，从而保证了采集到的信息的完整性和有效性；由于没有IP，现存的任何网络攻击方法都无法对敏感器进行攻击，从而保证了信息采集的安全性。

3 BPF过滤机制

3.1 BPF模型概述

网络信息监听将获得所有的网络流量，包括所有协议端口所有子网主机的所有交互数据，但在实际应用中，其中存在若干用户不需要关心的数据，或者称为垃圾数据，垃圾数据在所有流量中占有极大比重，严重影响了网络信息监控工作效率的提高，因此高效的信息过滤机制是信息监听的重要组成部分，它使得用户可以指定特定的子网主机以及特定的协议端口如HTTP、FTP、EMAIL等进行过滤，只将用户关心的敏感数据向上层提交，从而提高系统工作效率。信息的简单过滤具体有IP地址过滤、数据包类型过滤、TCP端口过滤等几个类别。本文采用BPF信息过滤机制[2]，大大提高了工作效率。

许多版本的UNIX都提供用户级别的网络监听功能，因为监听程序以用户级别进程工作，数据包的拷贝必须跨越内核/用户保护界限，这就需要使用名为包过滤器(packet filter)的内核代理程序，BPF[3](BSD PACKET FILTER)过滤使用了新的基于寄存器的过滤算法，效率比旧的算法提高了20倍，它的缓存机制也对整体效率提高有很大作用。

BPF提供了一种新的流量监控结构，其性能比其它工具有显著提高，主要有如下因素：

(1)BPF使用基于寄存器的过滤机，而CSPF[4]使用基于内存堆栈的过滤机，在现代计算机中，内存操作是系统的瓶颈所在。

(2)BPF使用简单的不共享方式缓存模型，充分利用现代计算机的大地址空间。

3.2 BPF的组成及接口

BPF有两个主要部件，network tap和packet filter。network tap从网络设备驱动程序中搜集数据拷贝并转发到监听程序。过滤器决定是否接受该数据包，该COPY数据包的那些部分。图4表示了BPF和系统其他部分的接口。

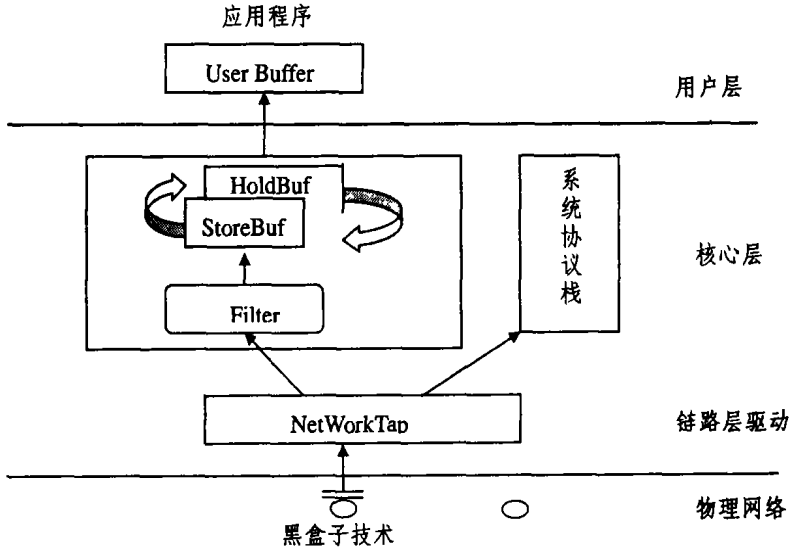

图4

3.3 BPF过滤过程描述

BPF过滤器的过滤功能是通过虚拟机执行过滤程序来实现的。过滤机主要由累加器、索引寄存器、数据存储器和隐含的程序计数器几部分组成。过滤程序实际上是一组过滤规

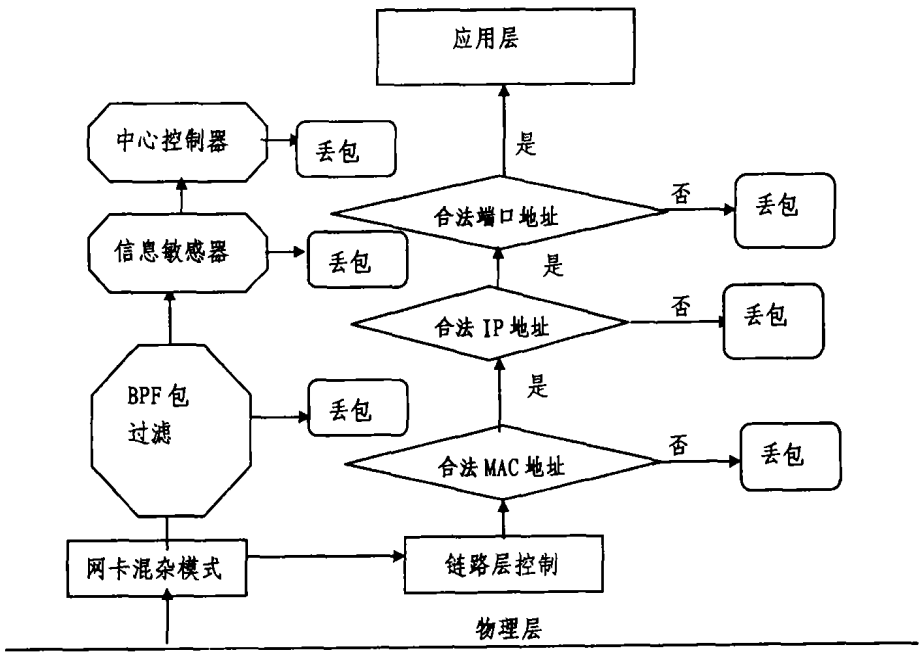

图5

(下转第115页)

1999
2 Berners-Lee T, Hendler J, Lassila O, 魏丰译. 语义网. http://www.xml.org.cn
3 Nigam K, Mccallum A, Thrun S, Mitchell T. Text Classification from Labeled and Unlabeled Documents uring EM. Machine Learning, 1999
4 Yang Yiming, Pedersen J O. A Comparative Study on Feature Selection in Text Categorization. In: Proc. of ICML-97, 14th Intl. Conf. on Machine Learning, 1997
5 Joachims T. A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization. In: Proc. of ICML-97, 14th Intl. Conf. on Machine Learning, 1997
6 Luke S. Ontology-Based Knowledge Discovery on the World-Wide Web http://www.cs.umd.edu/~seanl/, 1996
7 Iwazume M, Shirakami K, Hatadani K, Takeda H, Nishida T. IICA: An Ontology-based Internet Navigation System. JL41-96 Workshop on Internet-based Information Systems, Portland, OR, 1996
8 Rocchio J. Relevance Feedback in Information Retrieval, in The SMART Retrieval System: Experiments in Automatic Document Processing, Chapter 14, Prentice-Hall Inc. 1971. 313~323
9 Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm. Journel of the Royal Ststistical Society, Series B, 1977, 39(1): 1~38
10 朱明, 王军, 王俊普. Web 网页识别中的特征选择问题研究. 计算机工程, 2000, 26(8)
11 卢增祥, 李衍达. 交互支持向量机学习算法及其应用. 清华大学学报(自然科学版), 1999, 39(7)
12 邹涛, 孙赛. 文档自动分类技术及其实现. 计算机系统应用, 1999(4)

(上接第92页)

则。过滤规则由用户定义,以决定是否接受数据包和需要接收多少数据。每一条规则执行一组操作,具体操作可以分为指令装载、指令存储、执行算术指令、执行跳转指令、执行返回指令等几个类别。

结合图5,过滤过程描述如下:当一个数据包到达网络接口时,链路层驱动程序将其提交到系统协议栈;如果 BPF 正在此接口监听,驱动程序将首先调用 BPF,BPF 将数据包发送给过滤器,过滤器对数据包进行过滤,并将数据提交给过滤器关联的上层应用程序;然后链路层驱动将重新取得控制权,将数据包提交给上层的系统协议栈处理。

3.4 BPF 过滤模型

网络监听应用可能只关心网络流量数据的一个子集,因此包的过滤将大大提高系统的性能,为了减少内存操作,以尽量避开瓶颈,包过滤应该在包的原始位置 (in-place,即 DMA 操作存放数据的内存位置)进行过滤,而不是首先拷贝数据。

包过滤函数一般可用布尔函数表示,如函数返回为 ture,则将包拷贝到上层,反之忽略该包。通常有两种方法可以实现该函数:布尔表达树方式和可控制流图方式(CFG, Control Flow graph),如图6,7用两种方式表示了接受 IP 包和 ARP 包的函数。

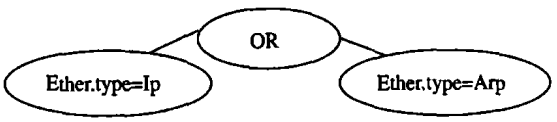

图6

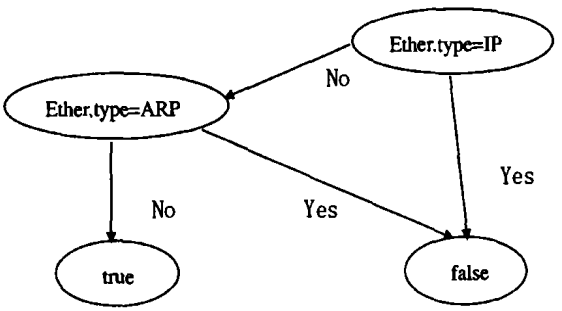

图7

在树型模型中,每一节点代表一个布尔关系如 AND、OR,每个叶子代表一个谓词断语,如 type=IP,边表示布尔操作和操作数的关系。

在 CFG 模型中,每个节点代表一个谓词断语,而边代表控制转换,如果谓词断语为 true,则转向右边,反之转向左边。每个 CFG 图有两个终结节点,表示返回 true 或者 false。

以上两个模型的计算功能大致相同,用一个模型可以表示的函数用另一个模型也同样可以表示,但是在实现的时候却大不相同,树型图的操作往往基于操作数堆栈操作,而 CFG 模型可以基于寄存器操作。

树型模型基于操作数堆栈运算,指令将常数或包数据压入堆栈,在堆栈顶层执行布尔运算或位运算,通过连续运算,在堆栈清空时返回结果。堆栈需要内存模拟,其中的 pop 和 push 操作需要指针的加减运算将数据从内存读取和写入内存。另外树型结构可能进行重复操作,例如某个值可能在多个叶子中分别从内存读取并计算,因此这种方式效率较低。

BPF 采用 CFG 过滤模型而非过滤树模型,因为过滤树模型在解释包时可能进行重复计算,而 CFG 模型解释包时数据包的解释状态和路径是可记忆的,不需要重复计算。

例如采用 CFG 模型接受来自主机 FOO 的所有数据包,包括 IP、ARP 和 RARP 包,在所有包中都需要查询地址,但不同包的格式是不一样的,采用树型模型时,为检查数据包是 ARP 还是 RARP,在表达树模型中,需要7次比较操作和6次布尔操作,而采用 CFG 模型的最长路径需要5次比较操作,平均只需要3次。

结束语 因为危害网络安全的有害信息能顺利地通过防火墙和入侵检测系统,它们不在像几年前那样明目张胆地从某一个地址发出,而往往是包装成合法的报文,或者加载到合法的报文中间,通过合法的用户发布出去。针对这类有害信息,只有通过特定的网络信息审计系统才能将其检查出来。而网络的审计的第一步是网络信息的监听,本文提出采用基于路由器的网络信息监听技术。然而,其面临的一个问题是网络信息监听主机工作负载过大,原因是监听的信息数据中包括大量不需要关心的数据,或者称为垃圾数据。因此论文提出采取高效的 BPF 信息过滤机制对监听的信息数据进行过滤,从而大大提高工作效率。

参 考 文 献

1 监听与隐藏—网络侦听揭密与数据保护技术. 求是科技. 北京:人民邮电出版社, 2000. 12~234
2 Braden R T. A pseudo-machine for packet monitoring and statistics. In: Proc. of SIGCOMM '88 (Stanford, CA, Aug. 1988), ACM
3 Steven McCanney and Van Jacobson. The BSD Packet Filter: A New Architecture for User-level Packet Capture. Lawrence Berkeley Laboratory (December 19, 1992)
4 张兴虎. 黑客攻防技术内幕. 北京:清华大学出版社, 2002. 45~67