

基于 IPFIX 的用户网络行为分析系统模型研究

姜巍,秦雅娟,刘颖

(北京交通大学电子信息工程学院,北京 100044)

摘要:网络行为分析是网络安全领域的研究热点。论文以用户使用网络资源产生的流量为依据,对用户的网络行为进行了分类,然后基于正在标准化中与设备不相关的 IP 数据流信息输出(IPFIX)协议,提出了一种用户网络行为分析系统模型,研究了模型中采集点、收集器、分析器的关键技术,并对模型的性能进行了分析。该系统模型具有良好的灵活性和扩展性,并且易于实现,对于网络检测、异常行为发现,以及网络整体规划、网络资源利用等方面都有着重要的意义。

关键词:网络行为分析;网络流量;IP 数据流信息输出;网络安全

中图分类号:TP393.08

文献标识码:A

文章编号:1673-629X(2011)09-0233-04

Research on IPFIX-Based System Model of Users' Network Behaviors Analysis

JIANG Wei, QIN Ya-juan, LIU Ying

(School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Network behaviors analysis is the hot spots of network security. It presents a taxonomy of users' network behaviors that is based on the traffic which is generated by users using network resources. Then based on IPFIX protocol, a system model of users' network behaviors analysis is proposed, the key technologies in this model are studied, and the model's performance is analyzed. The proposed model has good flexibility and scalability, also easy to implement. It is significant to the aspects of network measuring, abnormal behaviors discovering, network planning and network resource utilization.

Key words: network behaviors analysis; network traffic; IPFIX; network security

0 引言

在网络技术飞速发展的今天,网络的复杂性和多样性导致现有的网络安全技术,如防火墙、入侵检测等,已无法满足网络管理人员对于监控网络运行状态、提高网络服务质量等方面越来越高的要求。

与人们在现实社会中的社会行为相对应,网络行为指人们在网络的虚拟社会上,以实现某种特定的目标,采用基于计算机系统的电子网络作为手段和方法而进行的有意识的活动。网络行为分析作为目前网络安全领域的研究热点,在许多方面都有应用。文献[1]综述了网络行为异常检测技术,对现有网络行为异常检测模型进行分类,并做了详细的阐述;文献[2]从用户搜索日志挖掘出用户搜索行为的特点,并以此作为改进搜索引擎性能和服务的重要手段,但文献只

针对用户的网络搜索这一特定行为,应用范围较窄;文献[3]结合网络检测技术和网络数据分析技术,设计了一个面向网络行为特征分析的网络监测系统,实现网络流量实时监控和特定业务行为的分析,但没有对系统的数据采集和输出标准进行定义,系统扩展性不强。IP 数据流信息输出(IPFIX)协议统一了 IP 数据流的统计、输出标准,为基于流量的用户网络行为分析提供了重要依据。文献[4]提出使用 IPFIX 协议对网络行为进行分析,但文中将数据流采集部署在特定的主机监控软件中,只能对运行该软件的用户进行分析,分析范围受限。

文中从网络行为的分类展开讨论,以用户使用网络资源产生的流量为依据,对用户网络行为进行了分类,提出了一种用户网络行为分析系统的模型。该模型利用 IPFIX 协议^[5]采集网络原始流量并转化成数据流信息,从中提取用户行为向量,然后分析用户网络行为。

模型的灵活性高,扩展性强,对于网络检测、异常行为发现,以及网络整体规划、网络资源利用等方面都有着重要的意义。

收稿日期:2011-02-28;修回日期:2011-06-01

基金项目:国家自然科学基金(60833002);中央高校基本科研业务费专项资金(2011JBM016)

作者简介:姜巍(1987-),男,硕士研究生,研究方向为计算机网络安全和下一代互联网;秦雅娟,博士,教授,博士生导师,研究方向为下一代互联网理论和宽带无线通信等。

1 用户网络行为分类

用户网络行为是一个广义的概念,目前还没有一个比较规范的界定和分类。通常研究问题的侧重点不同,用户网络行为的分类方式也不相同。文献[3]以与网络行为相关的特性参数作为指标,将网络行为划分为流量行为、端到端行为、路由行为和业务行为;文献[6]从网络层用户使用的 IP 地址的角度,把用户网络行为分为单 IP 对单 IP 的访问、单 IP 对多 IP 的访问、多 IP 对单 IP 的访问、多 IP 对多 IP 的访问四类;文献[7]从应用层用户使用网络服务的角度,把用户网络行为分为信息获取、沟通交流、休闲娱乐、电子商务、电子服务五类。

文中从网络安全角度出发,根据用户使用网络资源的习惯模式,以不同行为产生不同的流量作为分类依据,将用户网络行为分为正常流量行为、突发流量行为和非法流量行为三类,能够较全面地概括出用户短期以及长期的行为特征:

(1)正常流量行为。对于不同的行为个体,由于受到个性的影响,其网络行为总是能呈现出明显的稳定性和规律性。这种产生长期、稳定、规律流量的行为,归类为正常流量行为。

(2)突发流量行为。由于行为个体具有学习和认知能力,在这个过程中,个体的行为会出现短期、不规律的特点。经过一段时间之后,这种行为会逐步稳定成为规律行为,称之为突发流量行为。另外,行为个体的误操作,也归类为突发流量行为。

(3)非法流量行为。某些突发流量行为产生的流量除了不规律外,还具有流量大、时间长的特点,这类突发流量行为归类为非法流量行为。

当行为个体的主机感染病毒、蠕虫、木马等,或行为个体本身具有恶意,会出现以影响网络服务质量、破坏网络安全为目的的行为,这类行为在产生初期将其归类为突发流量行为,经过进一步识别后,成为非法流量行为。

2 IPFIX 协议

在对用户网络行为进行分析时,首先需要对网络原始流量进行采集和处理。为提高系统的灵活性和扩展性,文中使用 IPFIX 协议作为原始流量采集和处理的标准。

IPFIX 的全称为 IP Flow Information Export,即 IP 数据流信息输出,它是由 IETF 公布的用于网络数据流信息测量的标准协议。该协议以 Cisco Netflow Version 9 数据输出格式^[8]为基础,为网络设备厂商广泛采用。目前国外的思科公司、国内的锐捷网络公司都在其设备上添加了 IPFIX 功能。

IPFIX 主要包括输出器(Exporter)和收集器(Collector)两类设备,网络流量经过输出器处理后,以 IPFIX 报文的形式传送到收集器,收集器对数据流信息进行存储,以提供给后续的分析。IPFIX 没有对如何分析数据流信息具体说明,可以根据分析需要进行设计。IPFIX 的结构如图 1 所示。

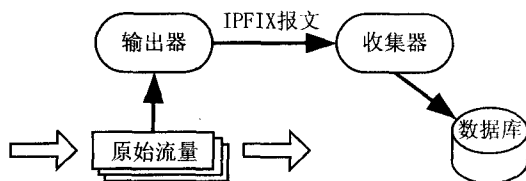


图 1 IPFIX 结构

(1)在输出器中,包含两个处理过程:测量过程(Metering Process)和输出过程(Exporting Process)。测量过程从接收到的网络原始流量中提取符合条件的数据流统计信息;输出过程将统计信息按照定义的模板组成 IPFIX 报文,然后发送给收集器。

(2)在收集器中,包含一个处理过程:收集过程(Collecting Process)。收集过程接收从一个或多个采集器发送的 IPFIX 报文,把报文中的数据流统计信息存储到数据库。

IPFIX 是一种基于模板的格式输出协议,网络管理员可以根据不同的流信息输出要求,定义不同的模板格式。另外,一个输出器可以向多个收集器发送 IPFIX 报文,同时一个收集器可以接收多个输出器的 IPFIX 报文。IPFIX 具有很强的灵活性和扩展性,使得网络管理员很容易地提取和查看存储在网络设备中的重要流量统计信息^[9],非常适用于大型网络的监控系统中。

3 网络行为分析系统模型

基于 IPFIX 的网络行为分析系统模型如图 2 所示。该模型利用 IPFIX 协议,并结合分布采集、集中处理的思想,采用树形结构,能够对网络中,特别是大型网络中用户的网络行为进行分析。

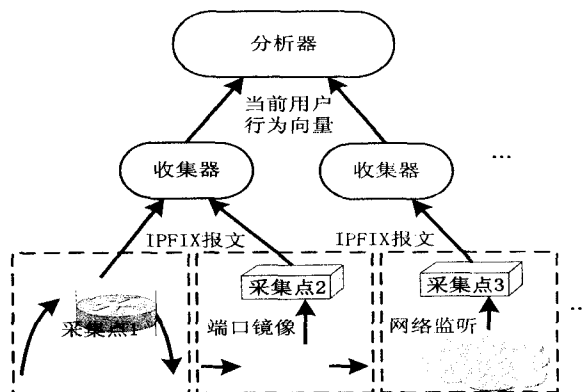


图 2 用户网络行为分析系统模型结构

3.1 采集点

采集点的处理过程如图 3 所示,每个采集点负责一个采集区域的网络原始流量采集,并从中提取出所需的流信息,组成 IPFIX 报文发送给收集器。多个采集点的 IPFIX 报文可以发给同一个收集器。

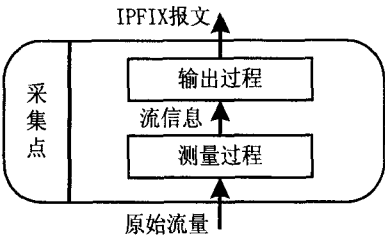


图 3 采集点处理过程

可以直接使用提供 IPFIX 功能的网络设备(如路由器)作为采集点,对流经该设备的流量进行处理;也可以独立部署采集点,并通过交换机的端口镜像功能或是使用网络监听的方式获得网络原始流量的拷贝;甚至可以读取存储数据包的文件,对历史流量进行处理。

测量过程对原始流量中的数据包按照“流”进行分类。IPFIX 中对每股流用源 IP 地址、目的 IP 地址、TCP/UDP 源端口、TCP/UDP 目的端口、三层协议类型、服务类型、输入逻辑接口七个域来表示。不同数据包,如果所有的七个关键域都匹配,将被视为同一股流。每股流的开始时间、持续时间,流中报文数、字节数等信息都被记录下来。当该股流结束或到期后,输出过程将这些流信息根据如图 4 所示的模板组成 IPFIX 报文^[10],发送给收集器。

Set ID = 2		Length = 56 octets	
Template ID = 256		Field Count = 12	
0	ipVersion = 60	Field Length = 1	
0	protocolIdentifier = 4	Field Length = 1	
0	sourceIPv4Address = 8	Field Length = 4	
0	destinationIPv4Address = 12	Field Length = 4	
0	sourceIPv6Address = 27	Field Length = 16	
0	destinationIPv6Address = 28	Field Length = 16	
0	sourceTransportPort = 7	Field Length = 2	
0	destinationTransportPort = 11	Field Length = 2	
0	flowStartSysUpTime = 22	Field Length = 4	
0	flowEndSysUpTime = 21	Field Length = 4	
0	octetDeltaCount = 1	Field Length = 4	
0	packetDeltaCount = 2	Field Length = 4	

图 4 用于网络行为分析的 IPFIX 报文模板

该模板是根据用户网络行为分析的需要所定义,模板中的各字段域在文献[11]中进行了详细说明。采集点在发送 IPFIX 报文前,需要先将模板发送给收集器,使收集器正确处理 IPFIX 报文中的流信息。

除了模板信息和流信息外,IPFIX 报文中还包含了报文输出时间、序列号和采集区域 ID,收集器可以以此来区分每个 IPFIX 报文的到达时间和发送它的采集点。

3.2 收集器

本模型的收集器在收集过程之上,增加了行为向量生成过程,如图 5 所示。收集器接收来自各个采集点的 IPFIX 报文,按照模板提取出报文中的流信息,利用流信息生成用户行为向量,然后将用户行为向量交给分析器。

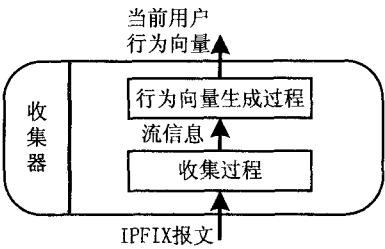


图 5 收集器处理过程

假设一个 IP 地址对应一个用户,可以使用行为向量 {Source、Destination、StartTime、Duration、Protocol、Traffic} 来描述某个用户的行为特征,即行为源端(Source)、行为目的端(Destination)、行为发生的时间(StartTime)、行为持续的时间(Duration)、使用的协议(Protocol)、行为产生的流量(Traffic)。

行为向量可以通过 IPFIX 报文中的各个字段域的信息来确定:

(1)行为源端。TCP 或 UDP 流的行为源端用“源 IP 地址+源端口”表示,ICMP 流或其他协议流使用“源 IP 地址”表示。对于 IPv4 流,源 IP 地址为 sourceIPv4Address 字段域的值,对于 IPv6 流,源 IP 地址为 sourceIPv6Address 字段域的值;源端口为 sourceTransportPort 字段域的值。

(2)行为目的端。TCP 或 UDP 流的行为目的端用“目的 IP 地址+目的端口”表示,ICMP 流或其他协议流使用“目的 IP 地址”表示。对于 IPv4 流,目的 IP 地址为 destinationIPv4Address 字段域的值,对于 IPv6 流,源 IP 地址为 destinationIPv6Address 字段域的值;源端口为 destinationTransportPort 字段域的值。

(3)网络行为发生的时间。flowStartSysUpTime 字段域的值是 1970 年 1 月 1 日 00:00:00(UTC)到数据流开始时的时间秒数。通过计算得到数据流开始时间,再根据人的活动特征,将其离散化为几个时间段:清晨(6:00~8:00)、上午(8:00~12:00)、中午(12:00~14:00)、下午(14:00~18:00)、晚上(18:00~24:00)、深夜(24:00~6:00)。

(4)网络行为持续的时间,即 flowEndSysUpTime

字段域的值减去 flowStartSysUpTime 字段域的值。并将其离散化为极短(0min ~ 1min)、短(1min ~ 5min)、中(5min ~ 30min)、长(30min ~ 120min)、极长(120min ~)五个区间。

(5) 网络行为使用的协议,即 protocolIdentifier 字段域的值。分为网络层协议 ICMP,传输层协议 TCP、UDP,或是其他协议。

(6) 网络行为产生的流量。可以基于不同的粒度,如字节、包、流,分析不同粒度的速率。octetDeltaCount 字段域的值是数据流的字节数,packetDeltaCount 字段域的值是数据流的包数。

3.3 分析器

分析器的处理过程如图 6 所示,分析器将收集器发来的当前用户行为向量与正常流量行为库进行对比,判断当前行为是否正常,如果正常则更新正常流量行为库,否则更新突发流量行为库。

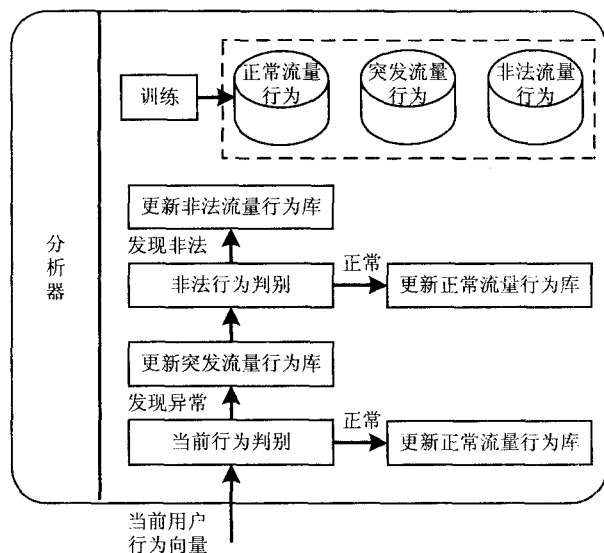


图 6 分析器处理过程

使用计算量很小的非参数 CUSUM (Cumulative Sum) 算法^[12]进行非法流量行为的判别。如果满足非法流量条件的行为则更新非法流量行为库,否则更新正常流量行为库,完成用户行为的分类。

分析器在对当前用户行为进行分析处理前,需要一段训练时间,收集适当时间(通常以一周为周期)内的正常用户行为向量,用于建立一个正常流量行为库的基准库。文献[13]使用数据挖掘的方法,对用户行为向量集进行挖掘,寻找出向量集中项与项之间的关联规则,并利用关联规则的支持度,判别正常流量行为;文献[14]使用加权近期最多使用算法,按时间对用户行为赋予不同的权值,越接近当前时间的行为,其权值越高,然后对行为进行加权统计,统计值最高的即为正常流量行为。

具体的训练算法这里不再细述。

4 模型性能分析

文中提出的用户网络行为分析系统模型易于实现,扩展性强。通常情况下,在网络规模较小的应用中,使用一个分析器、一个收集器和若干个采集点就可以完成分析功能。采集点使用分布式部署,可以以模块形式集成在路由器中,也可以单独实现。收集器和分析器作为一个设备上的两个模块,便于系统的维护管理。

当网络规模扩大,除了增加相应的采集点数量外,可以将收集器分离出来单独作为一个设备,并部署多个收集器,每个收集器处理若干个采集点发送的 IPFIX 报文,将生成的行为向量发送给分析器进行分析。这种树形结构提高了整个系统处理效率,适于大型网络应用。

采集点的流信息输出格式基于自定义的 IPFIX 模板,具有很高的灵活性。文中的模板是按照目前生成网络行为向量的需求定义的,在未来系统的应用中,可以根据分析需求的变动修改模板。例如针对 IPv6 的 ICMP 流,当需要分析流的具体类型时,可以在模板中增加 icmpTypeIPv6 字段域。

模型以流量为依据对用户网络行为进行分类,能够实时地对网络运行状态进行监控,发现异常行为。同时,网络管理员通过统计分析正常流量行为库中存储的数据,可以对网络进行合理规划,提高网络资源利用率。

另外,本系统模型基于网络数据流信息测量的 IPFIX 协议,使用标准的、一致的流输出架构,具有广阔的应用前景。

5 结束语

随着 IPFIX 协议标准化进程的加快,基于 IPFIX 的用户网络行为分析技术将得到广泛应用。文中在 IPFIX 结构的基础上,提出了一种网络行为分析系统的模型。该模型在未来网络规模不断扩大、新业务不断增加的网络行为分析中,具有很好的应用前景。在今后的工作中,需要对突发流量行为判别和非法流量行为判别方面进行进一步的研究,提高系统对网络行为分类的准确率。

参考文献:

- [1] Shu Y L, Andy J. Network Anomaly Detection System: The State of Art of Network Behaviour Analysis[C]//International Conference on Convergence and Hybrid Information Technology. Korea: [s. n.], 2008.
- [2] 岑荣伟, 刘奕群, 张 敏, 等. 基于日志挖掘的搜索引擎

(下转第 241 页)

- [2] Chow C, Mokbel M F. Enabling privacy continuous queries for revealed user locations[C]//Proc of the Int Symposium on Advances in Spatial and Temporal Databases (SSTD). Boston: Springer, 2007.
- [3] Mokbel M F, Chow C Y, Aref W G. The new casper: query processing for location services without compromising privacy[C]//Proc of the 32nd International Conference on Very Large Data Bases (VLDB). New York: ACM, 2006: 763-774.
- [4] Gruteser M, Grunwal D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//Proc of the Int Conference on Mobile Systems, Applications, and Services (MobiSys). New York: ACM, 2003: 163-168.
- [5] Dewri R, Ray I, Ray I, et al. Query m-Invariance: Preventing Query Disclosures in Continuous Location-Based Services[C]//11th International Conference on Mobile Data Management (MDM). Kansas City, Missouri, USA: [s. n.], 2010: 95-104.
- [6] Liu F, Hua K A, Cai Y. Query l-Diversity in Location-Based Services[C]//Proceedings of the 10th International Conference on Mobile Data Management: Systems, Services and Middleware. [s. l.]: [s. n.], 2009: 436-442.
- [7] Machanavajjhala A, Gehrke J, Kifer D, et al. Diversity: Privacy Beyond k - Anonymity[C]//Proceedings of the 22nd International Conference on Data Engineering. [s. l.]: [s. n.], 2006.
- [8] Wang Yiming, Wang Lingyu, Fung B C M. Preserving Privacy for Location-Based Services with Continuous Queries[C]//IEEE International Conference on. [s. l.]: [s. n.], 2009.
- [9] Pan Xiao, Hao Xing, Meng Xiaofeng. Privacy Preserving towards Continuous Query in Location-based Services[J]. Journal of Computer Research and Development, 2010, 47(1): 268-281.
- [10] Xu T, Cai Ying. Location Anonymity in Continuous Location-based Services[C]//Proc. of Int Symposium on Advances in Geographic Information Systems (GIS). New York: ACM, 2007.
- [11] Amoli A S, Kharrazi M, Jalili R. 2PLoc: Preserving Privacy in Location-Based Services[C]//IEEE International Conference on Privacy, Security, Risk and Trust (PASSAT10). [s. l.]: [s. n.], 2010.
- [12] Beresford A R, Stajano F. Location privacy in pervasive computing[J]. IEEE Pervasive Computing, 2003(1): 46-55.
- [13] Duckham M, Kulik L. A formal model of obfuscation and negotiation for location privacy[C]//In: Pervasive 2005. Berlin: Springer, 2005: 152-170.
- [14] Mokbel M F, Chow Chi-Yin, Aref W G. The New Casper: Query Processing for Location Services Without Compromising Privacy[C]//Proceedings of the International Conference on Very Large Data Bases. [s. l.]: [s. n.], 2006: 763-774.

(上接第 236 页)

- 用户行为分析[J]. 中文信息学报, 2010, 24(3): 49-54.
- [3] Zeng Bin, Zhang Dafang, Li Wenwei, et al. Design and Implementation of a Network Behavior Analysis-Oriented IP Network Measurement System[C]//the 9th International Conference for Young Computer Scientists. China: [s. n.], 2008.
- [4] 马延鹏, 苏金树, 王勇军. 一种基于 IPFIX 协议的网络行为分析方法[J]. 福建电脑, 2008(11): 150-151.
- [5] Leinen S. Evaluation of Candidate Protocols for IP Flow Information Export (IPFIX); IETF RFC 3955 [S/OL]. 2004-10. <http://www.ietf.org/rfc/rfc3955.txt>.
- [6] 马力, 焦李成, 董富强. 一种 Internet 的网络用户行为分析方法的研究[J]. 微电子学与计算机, 2005, 22(7): 124-126.
- [7] 中国城市网民行为与互联网市场演进研究报告[EB/OL]. 2003. <http://jiuban.chinalabs.com/cache/doc/03/05/15/88.shtml>.
- [8] Cisco Netflow Version 9[EB/OL]. 2004. http://www.cisco.com/en/US/products/ps6645/products_ios_protocol_option_home.html.
- [9] Farzaneh F, Mohammad H Y. Design and Implementation of a Monitoring System Based on IPFIX Protocol[C]//The Third Advanced International Conference on Telecommunications. Mauritius: [s. n.], 2007.
- [10] Claise B. Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information; IETF RFC 5101 [S/OL]. 2008-01. <http://www.ietf.org/rfc/rfc5101.txt>.
- [11] Quittek J, Bryant S, Claise B, et al. Information Model for IP Flow Information Export; IETF RFC 5102 [S/OL]. 2008-01. <http://www.ietf.org/rfc/rfc5102.txt>.
- [12] Leu F Y, Yang W J. Intrusion Detection with CUSUM for TCP-based DDoS[C]//The First IFIP Workshop on Trusted and Autonomic Ubiquitous and Embedded Systems. [s. l.]: [s. n.], 2005: 1255-1264.
- [13] 缪红保, 李卫. 基于数据挖掘的用户安全行为分析[J]. 计算机应用研究, 2005, 22(2): 105-107.
- [14] 孙知信, 王汝传, 王绍棣, 等. 基于用户行为的入侵检测系统的设计与实现[J]. 计算机工程与设计, 2004, 25(10): 1633-1635.