昆明学院学报　2010, 32(3):79~81　CN 53-1211/G4　ISSN 1674-5639
Journal of Kunming University

网络协议嗅探资源库的建构初探

申浩如,王付艳,邱　莎
(昆明学院 计算机与网络技术系,云南 昆明 650031)

摘要:提出了网络协议嗅探资源库的系统雏形.通过主机内嵌式嗅探软件捕获网络课程教学所需的各种协议数据包、网络安全数据样本,收集用户上传的网络流量数据,配以协议分析报告,分类存储入库,为学习者提供优质、丰富的协议资源.讨论了资源库建设的4个关键问题:嗅探软件选择;实验环境搭建;嗅探主机部署;隐私保护.实践表明,嗅探资源库不仅是一种高效的教辅手段,而且能为网络安全研究提供基础数据.

关键词:网络协议;嗅探软件;资源库;网络安全
中图分类号:TN915.04　**文献标识码**:A　**文章编号**:1674-5639(2010)03-0079-03

Elementary Research on Construction of Network Packet Sniffer Repository

SHEN Hao-ru, WANG Fu-yan, QIU Sha
(Department of Computer and Network Technology, Kunming University, Yunnan Kunming 650031, China)

Abstract: To present an elementary model of network packet sniffer repository through the resources of the repository including network protocol packets and network security data samples which are captured by host-built-in sniffer software, in addition to network traffic traces collected and uploaded by Internet users, all of which are attached to analysis reports, sorted and stored in the database. The purpose is to offer students lots of qualified learning resources about network protocols. Meanwhile, four critical issues relating to building the repository have been also discussed, namely, sniffer software selection, experimental environment set-up, sniffer host deployment, privacy protection. It is proved that the repository is a cost-effective way in computer network teaching but also a useful way for network security research by using the resources.

Key words: network protocol; sniffer software; repository; network security

0　引言

网络协议(Network Protocol)是在网络参考模型规划之下的一组规则集合,用来规定同一层上的对等实体之间所交换信息的格式与含义[1],它是计算机网络技术学习中最重要,也是内涵最丰富的概念之一.由于各种网络协议数据包对终端用户而言几乎是完全透明的,采用协议嗅探技术可以捕获这些数据包,通过分析器自适应地解析包中的字段,理解其工作原理和设计理念,最终推进协议研究与设计.这种方法以实践为手段,原理验证为目的,但又区别于网络协议编程实验,网络协议编程方法主要考虑借助程序编制手段帮助学习者理解复杂的网络原理,好比"用一种新的复杂技术去解释另一种复杂技术[2]",学习者常常在软件编程的"迷局"中消耗大量的时间和精力,教学效果不好,教学质量也难以提高.

国外高校把网络协议分析技术引入计算机网络教学中已有相当长的时间,James F. Kurouse[3]和Jeanna N. Matthews[4]提出了基于协议分析的计算机网络教学法,认为可以"通过观察两个协议实体之间交换的报文序列或钻研协议运行的细节"等方法深化对协议的认识.从2005年开始,CCSC陆续刊发了Felix Fuentes[5],Victor A. Clincy[6]等人关于应用协议分析技术辅助网络教学的论文.国内计算机教师也同时提出了相似方法[7].

网络协议分析也叫网络流量分析,该技术是伴随协议嗅探工具一起发展起来的.网络流量分析就是捕获网络上传输的数据流,对它们进行监视和测量,收集统计量,从中分析得知网络上正在发生的各种事件或行为,由于它在网络规划、业务分析、运行维护和行为检测等方面有十分重要的意义和价值,国际上成立了ITA,CAIDA,WIDE等研究机构,负责监测存储各级各类网络流量[8].2006年Openpacket.org网站开通[9],该站点在Internet上提供各种网络流量的捕获样本,注册用户可以检索并下载这些样本,也可以贡献自己捕获的无安全隐患样本.但Openpacket捕获的数据来源复杂,没有相同的实验环境和统一的捕获标准,主要用于网管人员下载分

收稿日期:2010-03-09
基金项目:昆明学院科学研究资助项目(2008Z010)
作者简介:申浩如(1980—),男(白族),云南昆明人,讲师,硕士,主要从事多媒体通信与信息安全研究;王付艳(1978—),女,云南昆明人,讲师,硕士,主要从事通信理论与编码研究.

析或导入其它系统进行研究,对初学者而言,直接"读懂"那些原始数据显然是不现实的. 网络协议分析是网络流量分析的基础,前者侧重教育价值的挖掘,后者则强调工程实践意义.

1 嗅探资源库的建构

1.1 资源库建构的目的

建构网络协议嗅探资源库的目的是:1)为师生提供主流的、优质的网络协议捕获资源,以及正确规范的协议分析报告. 建立这样一个低成本的实验研究体系,使高校本科计算机网络基础课程的教学从宏观过渡到微观、从定性走向定量;2)为工程技术人员提供一个开放的网络协议和网络安全技术交流平台,对大量网络流量数据样本的分析能使他们更加深入地了解网络运行规律、网络应用程序的运行规律、网络用户行为以及对异常流量的监控.

1.2 嗅探资源库的组成和结构

嗅探资源库面向计算机网络专业教学,其组成成分包含主流的网络协议、部分处于研发阶段的新型协议以及网络安全数据样本,其建构过程应当遵循计算机网络本身的发展规律. 入库协议主要按照 TCP/IP 模型分类,尽管在体系结构的完备程度和协议标准化程度上 TCP/IP 不及 OSI 模型,但 TCP/IP 模型拥有较成功的商业运作和广泛的用户基础,事实上,绝大多数网络协议分析工具都是按照 TCP/IP 的层次结构解析所捕获的数据包的,但 TCP/IP 模型并不通用,不适合用来描述 TCP/IP 之外的任何其他协议栈. 另外,入库协议应配有由专家撰写的协议分析报告或网络安全分析报告,报告参考格式如下:1)协议概述,包括协议演化历史、RFC 编号、功能描述、封装层次关系、优缺点评述等;2)协议工作原理描述;3)协议字段格式描述;4)其它相关内容,如协议存在的安全威胁和脆弱性分析等.

资源库的另一个部分是来自其他用户捕获的网络流量数据. 一份网络流量数据不仅仅针对某种协议进行捕获,它没有明确的逻辑边界,捕获持续时间越长所能反映出的网络运行规律就越准确,故需要更大的空间去存储这些数据,而且要确保数据不存在泄密的危险.

图 1 描述了嗅探资源库的系统结构. 采用 B/S 架构,Server 端由资源数据库和 Web 服务器构成,

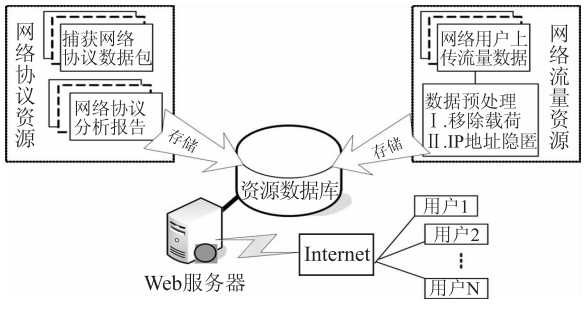

图1 网络协议嗅探资源库的系统结构

Web 服务器提供协议查询、显示等服务. Browser 端为普通 Web 浏览器. 系统采用 AJAX 技术来缩短服务器对用户查询请求的响应时间.

2 嗅探资源库建构的几个关键问题

2.1 嗅探软件的选择和实验环境的搭建

嗅探资源库采用开源的主机内嵌软件 Wireshark 作为协议捕获分析工具,在主机与网络的通信接口位置嵌入 Wireshark(实际是由 WinPcap 执行包捕获功能),通过监测对通信模块的调用来截获往返通信的全部内容[10]. 几乎所有主机内嵌软件在捕获网络流量时都利用了网卡的混杂模式(Promiscuous Mode),无论这些数据包是否是发送给该网卡物理地址,其都能够接收网段共享介质上流经它的所有报文. 由此可见,直接捕获实际网络流量,会使我们真正关心的协议报文"淹没"在大量背景流量中,给初学者造成不小的干扰. 因此,选择在网络工程实验室搭建统一的实验环境,服务器 OS 采用 Windows Server 2003,客户机 OS 采用 Windows XP Professional,还可选用数据流量生成软件 PCATTCP 生成定制参数的 TCP 或 UDP 流量,做到 1 个协议配置 1 个网络环境,形成 1 个捕获文件,撰写 1 份分析报告,这样大大减少了捕获文件中所包含的协议数量和文件本身的容量,学习者也就能把精力集中在关键报文的研究上.

另一项技术也可以大幅减少捕获的协议数量,即 Wireshark 提供的 2 种包过滤器. 第 1 种是捕获过滤器(Capture Filters),在现场捕获时只捕获"感兴趣"的包,丢弃其他"不相关"的包;第 2 种是显示过滤器(Display Filters),允许显示流量中符合条件的包,同时也保留那些未被显示的包,类似于视图的功能.

2.2 嗅探主机的部署

如何在网络中正确部署嗅探主机是网络协议捕获成功实现的保障. 这个问题在共享式以太网上可忽略不计. 以太网的设计初衷就是让所有接入网段的计算机共享传输介质,接入同一网段的机器能够侦听通过该网段的任何数据,嗅探主机可放置在网段的任何位置. 然而,对于交换式以太网或路由网络,需要谨慎部署嗅探主机,毕竟交换机和路由器属于流量隔离设备. 那种在每一个点对点连接上部署嗅探主机的做法既繁琐又会影响网络性能. 现有如下两种部署方式,综合考虑性能、经费等问题,我们选择交换机端口镜像来部署嗅探主机.

1)端口镜像(Port Mirroring):某些交换机具备将选定端口(如图 2 中的端口 A 和 B)或 VLAN 中传输的流量复制到一个指定端口(如图 2 中端口 C),端口 C 为流量监测端口,嗅探主机只需接入到该端口就可以捕获到其它端口的网络流量. 但有许多交换机不支持这种端口数据多路转发操作,或对其功能加以限制—只允许数据流定向从低速端口镜像到高速端口. 另一问题是某些种类的数据包错误(如帧结构错误)会被交换机丢弃而非转发,正常情

况下,当然希望交换机丢弃这些产生错误的数据包,以提高网络的实际吞吐量,但在网络安全研究的前提下却希望在镜像中捕获到那些出问题的包. 因此,必须搞清楚实验交换机执行镜像功能的粒度.

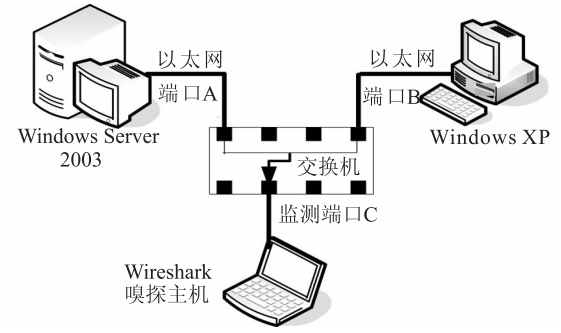

图2　使用交换机端口镜像部署嗅探主机

2)TAP 设备:分路器 TAP 是一种用于直接复制网络链路上的电信号或光信号的设备. 图 3 给出了使用 TAP 部署嗅探主机的示意图,图中 TAP 采用半双工监测端口与嗅探主机相连. 与端口镜像技术相比,TAP 采用容错设计,任何情况下都不会产生丢包,也不会干扰网络的正常运行,可以获得数据包的准确时间戳,但 TAP 1 次只能复制 1 条链路的流量,常用于骨干链路,IDS 或要求精确度量时间的 VoIP 测量领域.

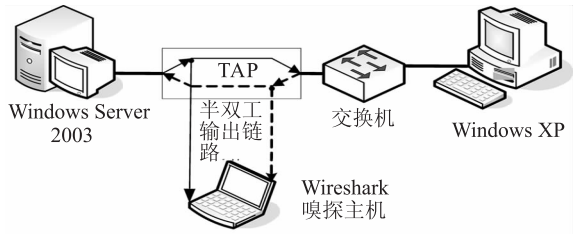

图3　使用TAP设备部署嗅探主机

2.3 去隐私处理

资源库中的数据样本面向 Internet 公众开放,在无偿提供这些资源的时候,应意识到资源中可能携带着用户隐私信息和控制信息,例如网络 IP 地址和应用层数据载荷. 在实验网络中捕获的协议报文,不可能产生隐私信息泄露的危险,但对其他人员上传的捕获流量,需严格审核,必要时要进行去隐私处理才能对外发布. 去隐私处理涉及两个主要问题:1)移除含有用户隐私信息的 TCP 或 UDP 包的有效载荷(Payload);2)尽管 IP 地址本身有层次结构,亦或有一些特殊 IP 地址(广播和多播地址、私有地址等)的存在,但 IP 地址匿名化(Anonymization)是必须的. 可以把流量中原始 IP 地址 Hash(哈希)为一个新的 IP 地址,也可以把几个地址的相同前缀(Prefix)哈希为另一个地址前缀,以便模拟流量中的路由信息,不过该方法容易遭到逆向工程破解[11-12],有关 IP 地址保留前缀的匿名化算法请参阅文献[13~14]. IP 地址隐匿还存在其他问题,上层协议报文中也可能含有 IP 地址信息,如 ICMP 和 DNS 数据包载荷中就含有 IP 地址,IP OPTION 字段中也含 IP 地址. 目前大多采用改进后的 tcpdpriv 工具移除隐私信息[8].

3　结论

嗅探资源库的建设是一项长期的工作,其设计初衷并不是要囊括所有已知的网络协议,事实上也不太可能做到,我们的目的是要捕获并分析诸如 HTTP,SMTP,TCP 一类经典协议或对未来有重大影响的新型协议,把这些资源分发出去供学习者分析、验证和测试,同时,培养学习者从不同的流量中辨识正常流量与异常流量的网络安全技能. 下一步的工作重心是嗅探资源库的设计与实现,包括网络协议的遴选、资源库的具体实现方法以及面向新协议的网络嗅探工具二次开发等问题.

[参考文献]

[1]TANENBAUM A S. 计算机网络[M]. 潘爱民,译. 第 4 版. 北京:清华大学出版社,2004.

[2]陈鸣,常强林,岳振军. 计算机网络实验教程[M]. 北京:机械工业出版社,2007.

[3]KUROUSE J F,ROSE K W. Computer Networking:A Top-Down Approach Featuring the Internet[M]. 3th ed. Beijing:Higher Education Press,2005.

[4]MATTHEWS J N. Hands-on Approach to Teaching Computer Networking Using Packet Traces[C] // Special Interest Group on Information Technology Education(SIGITE). New York:ACM,2005:167-173.

[5]FUENTES F,KAR D C. Ethereal vs. Tcpdump:a comparative study on packet sniffing tools for educational purpose[J]. Journal of Computing Sciences in Colleges,2005,21(4):169-176.

[6]CLINCY V A, ABU-HALAWEH N. A Taxonomy of free Network Sniffers for teaching and research[J]. Journal of Computing Sciences in Colleges,2005,21(1):64-75.

[7]黄俊,韩玲莉,陈光平,等. 基于协议数据包分析的计算机网络课程教学方法探索[J]. 实验室研究与探索,2006,25(6):3-5.

[8]CHO K,MITSUYA K,KATO A. Traffic Data Repository at the WIDE Project[C] // Proceedings of the annual conference on USENIX Annual Technical Conference. California:USENIX,2000:51.

[9]BEJTLICH R. OpenPacket. org: the challenge of a free,public packet capture repository[C] // Proceedings of the 1st ACM workshop on Network data anonymization. New York:ACM,2008:1-2.

[10]刘芳,窦伊男,陈陆颖,等. 网络流量监测与控制[M]. 北京:北京邮电大学出版社,2009.

[11]YIONEN T. Thoughts on how to mount an attack on tcpdpriv's "-a50" option[EB/OL]. [2009-06-21]. http://www.ita.ee.lbl.gov/html/contrib/attack50/attack50.html.

[12]陆正福,李敏,何英,等. 覆盖多播点失效检测的分布式算法的改进[J]. 昆明学院学报,2009,31(6):67-68.

[13]FAN J L, XU J,AMMAR M H,et al. Prefix-preserving IP address anonymization:measurement-based security evaluation and a new cryptography-based scheme[C]// Proceedings of the 10th IEEE International Conference on Network Protocols. New York:IEEE Computer Society,2004,46(2):253-272.

[14]史冰,吴连国,丁伟. IP 地址前缀保留匿名化算法的改进[J]. 微电子学与计算机,2007,24(10):167-170.