Web通信中可疑域名监控技术的研究

王培新,刘 颖,张思东,陈雨新
(北京交通大学 下一代互联网互联设备国家工程实验室,北京 100044)

摘 要:现有Web服务存在着很多的仿冒、欺骗等安全威胁,而Web通信基本先通过DNS获取IP地址,因此对网络中DNS域名信息的分析有助于加强对可疑非法Web通信的监控。传统的域名分析技术只能进行简单的协议分析,而且耗费资源严重,不能实现安全控制。文中提出了一种针对可疑域名的监控技术,给出了设计方案和具体的编程实现方法,并搭建了校园网环境进行验证,表明该系统有很小的丢包率和及时的安全控制响应,能很好实现对网络域名信息的监控。

关键词:可疑域名监控;域名系统;域名重定向;Web
中图分类号:TP309 **文献标识码**:A **文章编号**:1673-629X(2012)04-0231-04

Research on Suspicious Domain Name Monitoring in Web Communciation

WANG Pei-xin, LIU Ying, ZHANG Si-dong, CHEN Yu-xin
(National Engineering Laboratory for Next Generation Internet Interconnection Devices, Beijing Jiaotong University, Beijing 100044, China)

Abstract: There are a lot of security threats such as counterfeit and cheat on the Web services, many Web services need to firstly get IP address through DNS analyzing, so monitoring the domain name information can help to promote the level of the monitoring of the suspicious or illegal communication. Traditional domain analysis technology can only carry on the simple analysis of protocol, seriously costing resources, and can't achieve safety control. It puts forward a domain name monitoring technology. It gives design, implementation and experiment. Experimental result and the implementations efficiency analysis show that it succeeds in implementing analysis and monitoring on the domain name information of the network.

Key words: suspicious domain name monitoring; DNS; domain name redirection; Web

0 引 言

在2010年众多的网络安全事件中,网络仿冒、网页篡改、恶意代码这3类安全事件占总数的56%[1],这些都是通过Web应用服务而体现的,大量的Web服务是互联网安全威胁传播和非法行为发生的巨大载体。

域名系统DNS(Domain Name System)[2]是网络中提供网络域名和IP地址对应关系的一套映射机制,是工作于UDP之上的应用层协议。客户端通过与服务器之间交互DNS查询报文和应答报文来实现域名到IP地址的查询。由于域名方便人们记忆,现有的Web服务很多都先通过域名解析获取IP地址,可以说DNS是现有互联网中Web通信的先行者[3]。监控DNS能够从通信的第一步就发现非法Web访问过程等,因此很有必要建立一种安全审计系统对网络中的DNS数据进行研究分析,这将有助于提高网络安全性。

因此,文中提出了网络通信中基于DNS的可疑域名的监控技术,能够对目标网络关键出口设备的DNS数据进行统一捕获解析,通过与可疑域名信息数据库进行匹配,并对网络中的可疑或非法行为进行重定向或阻断,便于网络管理者对网络开展监控分析。

1 可疑域名监控技术

1.1 域名分析技术

现有网络中存在着很多能够分析网络中域名信息的手段,比如嗅探技术[4]、分析还原技术[5]。Wireshark等嗅探工具有强大的协议解析功能,但其实质只是一个网络封包分析软件,不具备审计监测功能,并且由于其需要解析出每个协议字段的信息,大大耗费计算机

收稿日期:2011-09-01;**修回日期**:2011-12-04
基金项目:国家自然科学基金(60903150);中央高校基本科研业务费专项资金(2011JBM016);北京市自然科学基金重点资助项目(4091003)
作者简介:王培新(1987-),男,硕士研究生,研究方向为下一代互联网、网络安全;张思东,教授,博士生导师,研究方向为未来网络体系架构、传感器网络、通信与信息系统。

有限的资源,大流量的网络环境下可能会产生丢包甚至死机。分析还原技术能够对网络通信行为进行审计,也可只针对DNS数据进行审计,但只是一种将通信内容展现的技术,不能对发现的非法行为进行及时有效地控制,具有滞后性。

可疑域名监控只涉及到网络中的DNS数据包,并不需要分析所有的数据报文的所有字段的信息;另外,监控系统需要对网络中的可疑行为按照安全规则采取相应的策略,涉及到了识别可疑行为和安全控制,这些都是嗅探技术和分析还原技术不能满足的。因此,需要一种专门的域名监控系统,能够只对网络中DNS域名信息进行解析,并能实施安全控制策略。

1.2 可疑域名监控技术

1.2.1 系统整体设计

文中所阐述的可疑域名监控技术主要分为域名解析模块、域名匹配模块、安全控制模块。其中域名解析模块将网络中捕获的DNS数据包还原出包含的域名信息,是系统的数据来源;域名匹配模块可以判断域名是否是可疑的,是引擎工具;安全控制模块提供了控制网络可疑非法行为的产生,是一个对非法可疑行为的响应机制。系统结构如图1所示:

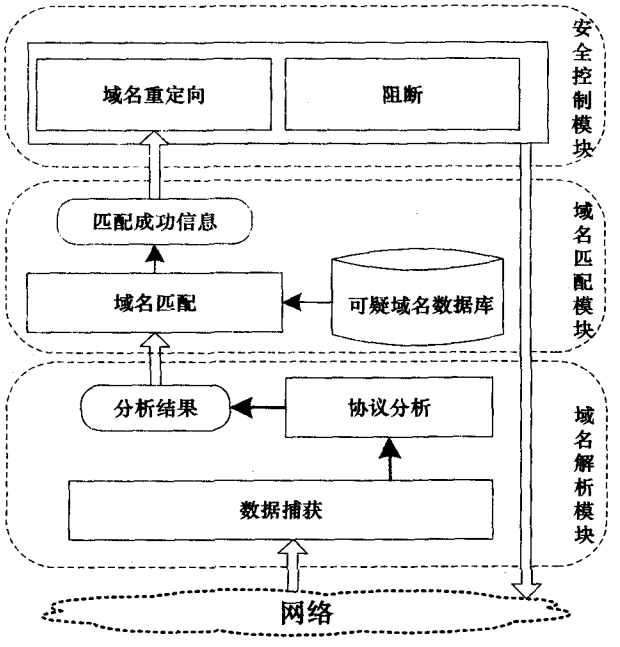

图1 系统结构示意图

可疑域名监控系统架构在目标网络的网关等流量汇聚关口,捕获并解析网络中的DNS数据报文,然后将还原出DNS域名信息采用字符匹配算法与可疑域名数据库中的内容进行匹配,如果匹配成功,则代表网络中出现可疑域名的访问情况,可以根据网络管理员设定的安全级别调用安全控制模块(重定向和阻断)作用于目标网络的关键出口设备,从而实现网络的监管控制。

1.2.2 域名解析模块

域名解析是为了获取在目标网络中传输交互的域名信息,包括数据的捕获和解析。监控系统需要对部分用户的非法访问进行定向或阻断,因此这种实时性需求要求域名解析模块要有很低的丢包率和高效的解析能力。

因为网络中的域名信息只通过DNS报文交互,所以只需要捕获网络中的DNS报文进行分析,也就是包含53端口的报文,这大大减少了捕获和解析数据包的数量。另外,采用多线程并行工作思想[6],数据捕获和协议分析可以在两个并行独立的线程中实现。这样既不延误DNS数据包的捕获,也不影响DNS信息的实时解析。

捕获线程采用基于NAPI机制[7]的Libpcap不断地抓取网络中的源端口或目的为53的数据包并保存在报文缓存队列中。解析线程从缓存队列头部逐个读取数据包,解析出域名和其他信息存储在result结构中,这些信息包括域名、服务器IP、源地址、目的地址、源端口、目的端口等。对每个数据包的协议分析,从物理层开始依次向上逐层分析直至应用层。其中,关键是对传输层UDP和应用层DNS协议的解析,并且需要存储关键信息以便后续进行域名匹配和安全控制。

1.2.3 域名匹配模块

域名匹配是为了检测域名解析模块所提供的域名信息是否属于网络管理者认定的可疑域名以及安全控制需要的一些信息。能否快速确定一个域名的属性决定了系统对可疑行为的响应速度,进而影响了系统功能实现和性能的提高。

可疑域名等信息存储在一个基于MySQL黑名单机制的一个字符列表,由管理员进行配置。总共有三个字段,分别用来记录网络管理者输入的各种可疑域名、解析应答数据包时提取的域名IP、传递给重定向模块用来将当前的访问转移到指定页面的IP地址。当系统启动或者有新的域名条目添加时,系统会将可疑域名读入内存,形成可疑域名链表,方便进行域名匹配。

确定一个域名是否属于可疑域名,就是判断数据包域名信息与数据库中的可疑域名是否匹配,即字符串的模式匹配,因此模式匹配算法的效率是系统性能的关键因素。提高模式匹配效率的办法就是利用模式串与文本串之间的关系,增大匹配窗口的有效偏移距离,减少匹配次数。域名一般长度不超过20字节,而且数量众多,可能会有上千条,因此对于模式算法的匹配窗口偏移距离不需要很多,字符串众多且长度都很相近。这些条件很适合一种高效算法—MWM(Modified Wu Manber)[8,9]的应用场景,该算法分为预处理

阶段和查找阶段,具有很小的时间复杂度。

在拥有了可疑域名数据库和优良的模式匹配算法后,系统就能高效地处理域名匹配的问题。可疑域名首先初始化可疑域名表,然后将域名解析模块提交的数据进行域名检测,检查是否匹配。若不匹配,结束处理返回。若匹配则提交给安全控制模块实现访问控制,这将在下文介绍。

1.2.4　安全控制模块

安全控制方法根据网络安全控制需求,对可疑域名匹配的相关数据连接进行安全控制,包括重定向和阻断策略。

重定向模块主要针对 DNS 请求报文,能够将域名匹配成功的连接重定向至指定的其他页面。目标网络的主机产生的可疑 DNS 请求数据包通过汇聚网关被系统截获后,重定向模块将会根据 DNS 请求数据包,伪造一个 DNS 应答数据包发给目标主机,当目标网络主机收到伪造的 DNS 应答后将会访问重置的页面,而由 DNS 服务器发送的真正的 DNS 应答报文将会因为前一个 UDP 的到达而被系统丢弃[10,11]。伪造的 DNS 数据包的每一层构造如表 1 所示:

表 1　伪造报文包各字段填充内容

层次	字段名称	长度	填充信息
网络层	源 IP 地址	32 位	查询报文中的目的 IP 地址填充
	目的 IP 地址	32 位	查询报文中的源 IP 地址填充
传输层	UDP 源端口	16 位	知名端口 53 填充
	UDP 目的端口	16 位	查询报文中的源端口填充
应用层	标识字段	16 位	查询报文中的标识字段填充
	问题区部分	与查询包一致	所有字段都用查询问题部分来填充
	回答区资源记录部分	不定	使用压缩方式的指针和偏移量填充
	资源数据字段	32 位	所要重定向到的域名的 IP 地址填充

重定向模块只是阻止了主机获取域名对应的 IP 地址,直接采用 IP 地址访问是能够通行的。因此需要阻断模块,当通过 DNS 应答包产生的域名匹配成功时,同样可以实现过滤 IP 地址的访问。因为监控系统是旁路部署,不能实现对数据包的阻断,需要通过与汇聚网关的防火墙配合才能实现。监控系统与网关之间可以构建 C/S 模型的规则分发服务端和客户端。当域名匹配成功后,客户端发送阻断规则给服务端,服务端收到阻断规则后,调用 IPtables[12] 执行防火墙规则,禁止网络主机继续访问该域名的网络服务。当然,当网络管理员从数据库中删除域名阻断规则时,系统会调用客户端发送放行规则给网关,取消阻断规则。

2　实验验证

2.1　实验环境

实验拓扑环境如图 2 所示,包括以下网络设备:目标网络客户端(系统为 Linux 2.6)、可疑域名监控设备(系统为 Linux 2.6,2G)、交换机和出口路由器。其中交换机所带的客户端形成监控系统的目标网络,里面的客户端通过出口路由器访问 Internet,可疑域名监控设备连接在交换机上捕获解析目标网络的 DNS 数据包。

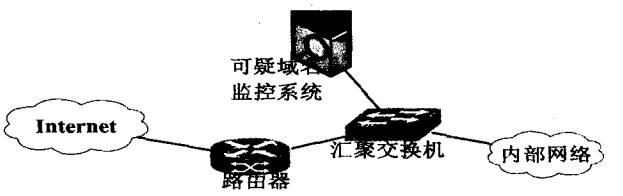

图 2　测试环境拓扑

基于以上环境进行以下实验:

1. 启动出口路由器阻断服务端程序,开始监听;

2. 启动可疑域名监控系统,开始捕获分析 DNS 数据包,在可疑域名数据库中写入重定向规则:1,bjtu,NULL,119.75.217.56,其中 119.75.217.65 是百度地址;

3. 在客户端主机上通过浏览器访问 www.bjtu.edu.cn;

4. 在客户端主机上通过浏览器访问 www.bjtu.edu.cn 的地址 202.112.154.27;

5. 在可疑域名数据库中删除 bjtu 相关规则,重复 3、4。

2.2　实验结果

当数据库中有 bjtu 条目时,用域名访问 www.bjtu.edu.cn 被重定向到百度页面;用 202.112.154.27 访问时页面显示不了。取消数据库中的 bjtu 条目时,3 和 4 步骤均能正常进行。可以看出,该技术能够对网络中的可疑域名进行有效地监控,并实现有效的安全控制措施。

为了测试系统的性能,主要考察系统在高强度工作环境下的丢包率。域名监控系统的数据捕获端口接收镜像数据为 890Mb/秒,在数据库中输入不同数量的可疑域名,丢包率与域名数量的对应关系如图 3 所示:

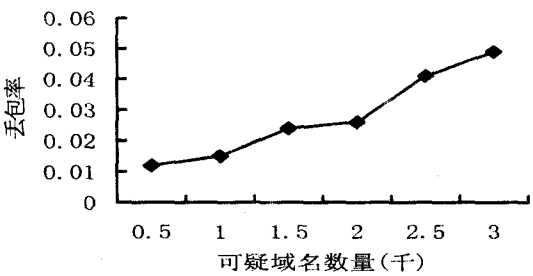

图 3　丢包率与可疑域名数量关系

由图 3 可知,当系统的重定向规则中设置 3000 条规则,丢包率不足 5%,在这样的大数量域名、大流量下的丢包率是完全可以接受的,因此系统能够提供对

目标网络及时有效的监控。丢包率之所以低主要有3方面原因：

1. 本系统采用NAPI机制来实现基于Libpcap的捕包机制；

2. 测试环境流量虽然有近千兆，但是DNS数据量并不是很大，程序设计为丢弃除DNS数据包的所有数据包；

3. 用于测试的系统硬件配置相对较高，可以最大程度的降低丢包率。

3 结束语

针对Web通信的安全问题，文中提出了一套涉及分析、匹配以及控制的域名监控方案。设计的域名解析能够高效地捕获解析网络中的DNS报文获取域名信息，设计的匹配方案可以快速判断域名是否属于可疑或者非法域名，设计的安全控制模块能够重定向和阻断通信连接，从而在Web通信的第一步就防止了网络非法行为的产生，提供了网络的安全性。

参考文献：

[1] 国家互联网应急中心. 2011年中国互联网网络安全报告[EB/OL]. [2011-04-22]. http://www.cert.org.cn/articles/docs/common/2011042225342.shtml.

[2] Mockapetris P. RFC1034-Domain names-concepts and facilities[S]. [s.l.]: Network Working Group, 1987.

[3] 郑海涛. 基于网络信息内容的DNS检测系统的设计与实现[D]. 北京：北京交通大学，2009.

[4] Ansari S, Rajeev S G, Chandrashekar H S. Packet sniffing: a brief introduction[J]. IEEE Potentials, 2003, 21(5): 17-19.

[5] 万国根，秦志光，刘锦德. 网络内容安全分析及审计技术研究[J]. 计算机应用研究，2004，21(1)：117-118.

[6] Stevens W R. UNIX网络编程第一卷：套接口API[M]. 杨继张，译. 第3版. 北京：清华大学出版社，2006.

[7] Liu Bin, Li Zhitang, Li Yao. High Speed Network Packet Capture Based on Linux[J]. Application Research of Computers, 2006, 23(5): 225-227.

[8] 陶善旗，李 俊，郭伟群，等. 入侵检测系统中模式匹配算法的研究与改进[J]. 计算机技术与发展，2010，20(2)：168-170.

[9] 孙晓妍，武东英，祝跃飞，等. Wu_Manber多模式匹配算法的研究与改进[J]. 计算机工程，2008，34(8)：85-89.

[10] Green I. DNS spoofing by the man in the middle[EB/OL]. 2005. http://www.sans.org/rr/whitepapers/dns/1567.php.

[11] 刘 扬，刘 杨，胡仕成，等. 基于ARP与DNS欺骗的重定向技术的研究[J]. 计算机工程与设计，2007，28(23)：5604-5609.

[12] Andreasson O. Iptables-tutorial[EB/OL]. 2003. http://www.frozentux.net/documents/iptables-tutorial/.

(上接第230页)

小、占用存储空间小、所需带宽要求低的非对称加密算法，目前主要应用于快速加密解密、身份鉴别、数字签名认证、密钥信息交换、移动传输通信、智能卡安全等安全需求高的领域，所以椭圆曲线密码具有十分广阔的应用价值和理论研究意义。同时随着椭圆曲线密码理论的不断突破和新型技术的持续更新，椭圆曲线密码算法的实现速度也将会大幅提升，因而其安全性强度也将必然更高，所以椭圆曲线密码将是更适合于当今社会电子商务/政务和智能卡等需要安全和高效密码系统的加密算法，对于椭圆曲线密码的研究将会有更加广泛的应用前景和实际应用价值。

参考文献：

[1] Koblitz N. Elliptic curve cryptosystem[J]. Mathematics of Computation, 1987, 48(177): 315-322.

[2] Miller V. Uses of elliptic curves in cryptography[C]//Advances in Cryptology-CRYPTO'85 Proceedings. [s.l.]: [s.n.], 1986: 417-426.

[3] 孟春岩，范 辉，余雪丽. 椭圆曲线用于加密的安全性讨论[J]. 微型机与应用，2001(6)：59-60.

[4] 白国强，马润年，肖国镇. 化离散对数问题为特殊的椭圆曲线离散对数问题[J]. 西安电子科技大学学报，2001，28(2)：254-257.

[5] 徐秋亮，李大兴. 椭圆曲线密码体制[J]. 计算机研究与发展，1999，36(11)：1281-1288.

[6] Xu Guangwu. Short vectors, the GLV method and discrete logarithms[J]. Journal of Lanzhou University (Natural Sciences), 2009, 45(1): 73-77.

[7] 杨君辉，戴宗铎，杨栋毅，等. 一种椭圆曲线签名方案与基于身份的签名协议[J]. 软件学报，2000，11(10)：1303-1306.

[8] 黄建华，马大朋. 椭圆曲线密码体制理论与安全性分析[J]. 网络安全技术与应用，2008(7)：91-93.

[9] 张晓丰，樊启华，程红斌. 密码算法研究[J]. 计算机技术与发展，2006，16(2)：179-180.

[10] 张海波，王小非，夏学知，等. 一个改进的离散对数问题攻击算法[J]. 计算机应用，2007，27(4)：843-845.

[11] 陈智华. 基于DNA计算自组装的Diffie-Hellman算法破译[J]. 计算机学报，2008，31(12)：2116-2122.

[12] 孟春岩. 椭圆曲线加密算法密钥长度讨论[J]. 电力学报，2007，22(4)：479-481.