

校园网钓鱼邮件监控系统的研究与实现

蔡洪民

(广州中医药大学 信息中心, 广东 广州 510006)

摘要:随着计算机网络的发展,垃圾邮件问题和邮件欺骗问题变得越来越严重,给广大网民带来巨大危害。为了保障网络公共安全,减轻垃圾邮件和钓鱼邮件对网络用户的危害,基于深度包检测技术设计实现了一个针对钓鱼邮件的监控系统。通过数据包捕获技术和协议解析技术对电子邮件进行还原,实现对邮件内容的检查,结合关键字匹配技术实现了对垃圾邮件和钓鱼邮件的检测和报警。实验证明,通过协议分析与数据包还原技术,可以对垃圾邮件和敏感邮件进行过滤。

关键词:网络钓鱼;深度包检测技术;LIBNIDS;ICTCLAS;敏感词

中图分类号:TP393.08

文献标识码:A

文章编号:1673-629X(2013)10-0103-04

doi:10.3969/j.issn.1673-629X.2013.10.026

Research and Implementation of a Phishing Email Monitoring System

CAI Hong-min

(Information Center, Guangzhou University of Chinese Medicine, Guangzhou 510006, China)

Abstract: With the development of Internet, the problem of spam and phishing email becomes more and more serious, it has done harm to cyber citizens. In order to strengthen the public network security, reduce the harm of spam and phishing email, design and realize a monitoring system for phishing email based on DPI. Adopt the packet capturing technology and protocol analysis technology to email reverting, succeeding in inspecting the email content, then apply keywords matching technology to detect the phishing email and give a warning for it. The experiments show that it can detect the spam and phishing email by the protocol analysis technology and packet restoration technology.

Key words: phishing; DPI; LIBNIDS; ICTCLAS; sensitive word

0 引言

随着计算机网络的发展,无论是网站浏览、网络聊天、网络视频,还是网络炒股、内容搜索、网络购物,互联网已经在人们的日常生活中扮演了一个越来越重要的角色。因为网络操作系统和网络应用程序存在这样那样的缺陷,互联网的安全问题也越来越严重,给广大网民的日常网络生活带来巨大的危害。据某国内信息安全峰会统计,互联网网络安全事件发生的比例是75%。目前各种各样的木马、蠕虫和病毒程序肆虐互联网,结合社会工程的各种各样的欺骗技术和网络钓鱼横行。由于我国网民数量巨大,对互联网不熟悉的人大有人在,这就给了像网络钓鱼这样的骗术以可乘之机。网络钓鱼,英文名字叫 Phishing,是攻击者利用欺骗性的电子邮件、手机短信、伪造的 Web 站点来进行网络诈骗的一种攻击手段。近年来,数不清的网络

钓鱼事件发生在世界各国,诈骗者通常会将自己伪装成证券公司、银行、在线零售商和信用卡公司等可信的品牌,通过电子邮件、网站链接、手机短信等方式诱骗广大意识薄弱的网民登录从而骗取用户的私人信息,如信用卡号、银行卡账户、身份证号等内容,甚至经济利益遭受损失^[1]。据新民网报道,2011年1月,江苏省发生了百余起中国银行网银受骗案,市民受骗多则高达百万。另据中国反钓鱼网站联盟2012年2月处理报告指出,截止2012年2月,该联盟累计认定并处理钓鱼网站80 076个。

互联网已经成为金融诈骗的主要平台之一。由于广大网民安全意识差,从而使得网络欺骗易于扩散和成功达到目的。邮件一直是人们在网络上通信的重要工具之一,被不法分子利用后已经成为钓鱼网站的重要源头,因此对邮件通信进行监控对于防范网络钓鱼

具有重要意义^[2]。文中设计实现了一个基于邮件还原的钓鱼邮件监控系统,它能够对互联网的收发邮件内容和标题进行敏感词过滤检测,从而发现邮件里面的恶意链接。结果表明它对于公共网络安全是有效的。

1 深度包检测技术和中文分词技术

互联网上所有的通信都需要以某种方式来指定预期接收方的身份,电子邮件系统也是一样。交付电子邮件的主要协议是 RFC2821 描述的 SMTP 协议,接收电子邮件的主要协议是 RFC1081 描述的 POP3 协议。邮件在通信过程中,使用 BASE64 编码进行传输。因此对于邮件内容的检测需要用到 BASE64 的解码。BASE64 编码是一种编码二进制字符串的编码方式。BASE64 的概念起源于特定的 MIME 内容传输编码。BASE64 编码通常用于编码二进制数据并且通过媒介处理正文数据。此编码保证数据在传输过程中是未受损的^[3]。例如,可以通过 LINNIDS 捕获到下面的包含邮件内容的 TCP 数据流:

```
-----=_NextPart_000_0014_01CCC166.0EFEDB30
```

```
Content-Type: text/plain;
```

```
. charset=" gb2312"
```

```
Content-Transfer-Encoding: base64
```

```
tPO9sbXIxOPEw6Osv+y/7NDQtq+joQ==
```

上述截取的 TCP 数据流中 Content-Transfer-Encoding: base64 后面的字符串就是邮件的内容。对其进行 BASE64 解码可以得到原始明文,明文是“大奖等你拿,快快行动!”因此,为了对钓鱼邮件进行内容还原,需要利用 BASE64 解码。

深度包检测技术是一项能够检查每个流经网络的数据包的每一个字节的技术,这意味着数据包头部、应用类型和实际的包内容都可以检查。DPI 技术可以检查 OSI 二层到七层的网络流量,这就是 DPI 中“深度”的意思。例如,对于电子邮件通信来说,IDS 或者 IPS 只能知道邮件的地址,对于邮件内容一无所知;而 DPI 技术检查 OSI 二层到七层就像一个人能够读取邮件并获取里面的内容。DPI 技术目前广泛应用于企业、服务提供商和政府的各类应用程序的开发中。LIBNIDS 是深度包检测技术的开发包之一,它是实现网络入侵检测系统的一个重要组件。它可以和 LINUX 2.0 的 IP 协议栈相媲美。它提供了 IP 包分片、TCP 数据流重组以及 TCP 端口扫描检测等功能。LIBNIDS 最大的优点是可靠性^[4]。一系列的测试表明,它能尽可能地保护 LINUX 主机的行为。目前它可以在 LINUX, BSD, SOLARIS, WIN32 等平台编译。无论攻击者多么有技巧的伪装,LIBNIDS 都能便捷地获取 TCP 的数

据内容。因此,使用 LIBNIDS 开发包,能够获得每一封电子邮件的详细信息。

中国人的母语是汉语,因此大多数国内邮件使用汉字来写。为了避开网络上的关键字过滤系统,很多垃圾邮件的内容一般会经过发送者的技巧来书写。一封经过改造的垃圾邮件如图 1 所示。针对如此复杂的技巧,需要一个工具来对这些钓鱼电子邮件中的特殊字符进行过滤,这样才能提取出所需要的关键词。这就用到了中文分词技术。ICTCLAS 开发包是最好的中文词法分析器之一,它是由中科院计算所研发的^[5]。此工具提供了中文分词、文法标记、新词识别等功能,并且它支持用户字典和多种编码格式。它是最好的中文分词工具并且广泛使用在中文分词领域。通过利用 ICTCLAS 中文分词技术,能够对钓鱼邮件的内容和标题做一个关键词的提取,与定义在数据库中的敏感词进行一一匹配,从而检测有没有钓鱼邮件在网络中传输。



图 1 垃圾邮件示例

2 系统设计与实现

基于上述技术和方法,文中设计实现了一个基于深度包检测技术的钓鱼邮件监控系统。该系统功能模块中包括了数据包捕获模块、协议分析模块、BASE64 解码模块、中文分词模块和敏感词模式匹配等 5 个模块。以下分别以数据包捕获与分析、数据库设计、敏感词过滤等进行阐述。

2.1 数据包捕获与协议分析设计

深度包检测技术广泛应用在网络程序开发中。通过 DPI 技术,能够解码每个数据包的所有层次协议和数据负载,如图 2 所示。在这一点上,使用网络入侵检测开发包 LIBNIDS 来捕获网络数据包,然后对捕获数据进行分层协议解码获得 TCP 数据流,接着提取里面的 SMTP 和 POP3 协议数据,最后按照电子邮件格式进行邮件报头与数据还原邮件^[6]。

利用 BASE64 编码解码邮件数据。一旦获得邮件主题和邮件内容,就利用 ICTCLAS 中文分词工具进行

中文分词,然后针对中文分词的结果使用敏感词的模式匹配技术来判断是否存在网络钓鱼相关信息。

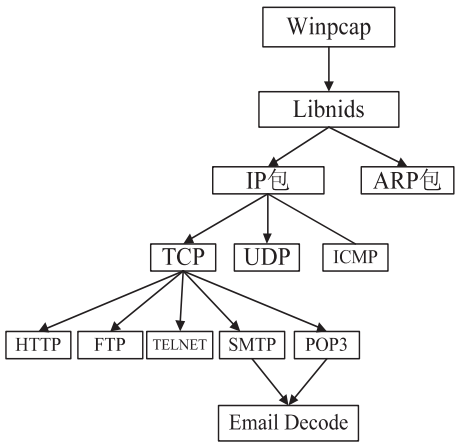


图 2 系统捕包与协议分析模块

2.2 数据库设计

该系统数据库涉及三个表,EmailInfo 表存储系统检测到的每一封钓鱼邮件的详细信息;Keyword 表存储管理员定义的与网络钓鱼相关的敏感词,如“中奖”、“恭喜”、“大奖”等;Log 表存储整个监控的日志,如正常运行多长时间等。其中 Email 表的设计如表 1 所示。

表 1 EmailInfo 表结构设计

字段	ID	From	To	BaddyInfo	Link	Subject	Content
注释	钓鱼邮件 ID	发信人	收信人	邮件中诈骗人的详细 信息	邮件中包含的 钓鱼网 站链接	钓鱼邮件主题	钓鱼邮件内容

2.3 系统实现采用的技术

该系统使用 VisualC++6.0 平台开发,因为系统中所存储数据量相对较小,数据库使用小巧的 MySQL。使用 LIBNIDS 开发包进行数据捕获与协议分析的工作,通过解析 SMTP 协议和 POP3 协议来获得邮件详细信息。因为电子邮件网络传输过程中是使用 BASE64 编码的,因此需要使用 BASE64 解码邮件信息。邮件还原后,利用 ICTCLAS 中文分词工具进行中文分词。系统采用的主要技术如下:

- (1)多线程技术。作为一个广泛的程序和运行模型,多线程技术允许一个进程中多个线程共存。这些线程共享进程资源但是单独运行。系统采用多线程技术进行数据包捕获与协议分析。一个线程捕获并分析数据包,另一个线程执行敏感词模式匹配。
- (2)深度包检测技术。系统使用 LIBNIDS 开发包捕包后进行协议解码整个数据包,包括应用层数据。它首先解码数据帧到数据包,然后解码数据包到 TCP 数据流,接着从 TCP 数据流中解码得到 HTTP,SMTP,POP3,FTP,TELNET 等应用层数据。根据系统需要,只提取 SMTP 协议和 POP3 协议通信的数据。最后,它解

码 SMTP 协议和 POP3 协议来获得钓鱼电子邮件的完整数据。

(3)中文分词技术。系统采用著名的中文分词工具 ICTCLAS 对邮件还原后的数据进行中文分词。中文分词结果用来匹配敏感词。一旦系统检测到邮件主题或邮件内容中存在敏感词,它会报警并且保存这封邮件的详细信息到数据库中。

2.4 系统工作流程

整个钓鱼邮件监控系统流程如图 3 所示。首先,使用 LIBNIDS 开发包来捕获数据包并进行协议分析。为了满足系统的需求,仅仅在 TCP 数据流中提取与电子邮件相关的 SMTP 和 POP3 的协议数据。其次,使用 BASE64 编码解码捕获的邮件数据信息,从而得到一封电子邮件的各个部分的信息,例如邮件主题、邮件内容等等。此时,已经可以看到邮件的明文内容了,但因为钓鱼邮件发送者往往会人为在邮件中添加一些特殊字符或别的手段,妄图逃过关键字过滤系统,这样还是无法实现自动化地检测钓鱼邮件^[7]。第三步,调用 ICTCLAS 中文分词工具对邮件主题和邮件内容进行中文分词,这样就可以过滤掉邮件内容中那些无意义的填充或混淆字符数字。最后,使用模式匹配技术在邮件内容中来匹配关键字。如果一个或多个敏感词匹配了,就认为此邮件带有网络钓鱼信息,就将其视为一封钓鱼邮件,同时将钓鱼邮件内容中的诸如“请访问网站……”等详细的网络链接信息和钓鱼邮件中诈骗人的联系方式等信息保存在数据库中。因此,把此系统部署在主干网络上,就可以监控整个网络的邮件通信信息,从而可以对流经整个网络的钓鱼邮件详细进行统计分析,从而给出详细的报告和警报^[8]。

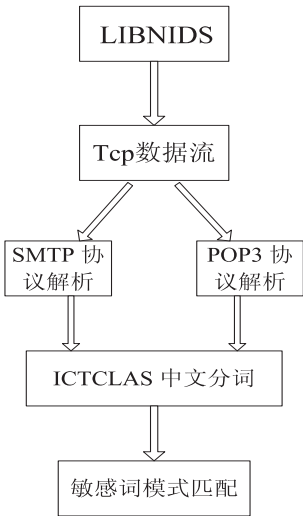


图 3 系统流程图

3 结果分析与讨论

该钓鱼邮件监控系统在某校园网进行测试。为了

便于测试,选择一些子网来执行^[9]。为了监控整个子网的数据流,配置三层交换机,把三层交换机的主干接口镜像到另一个接口。然后,部署一台主机通过双绞线连接到交换机的镜像接口上面。在这台主机上运行钓鱼邮件监控系统后,就可以对网络中的邮件通信进行实时监控。经过数天的运行观测与数据分析,它能够捕获整个子网的数据流并且分析解码 SMTP 和 POP3 协议,然后告诉管理员网络中通信的钓鱼邮件情况。在实验中,截获了 70 多封邮件,检测到有 2 封是钓鱼邮件。图 4 显示了系统运行情况。总之,此监控系统在一定程度上保障了公共网络安全。

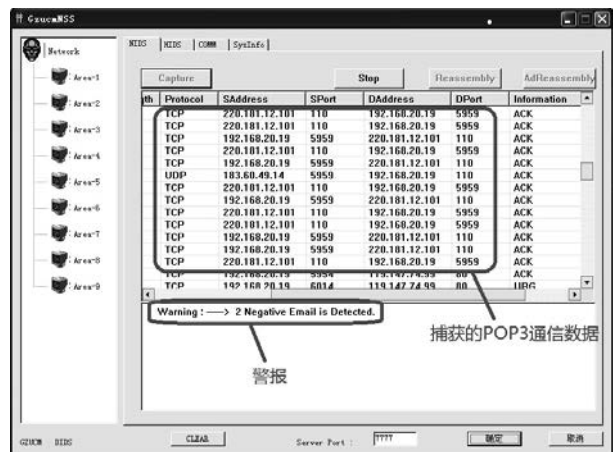


图 4 该系统在某校园网进行运行测试情况

4 结束语

由于互联网的发展和普通网民网络安全意识的亟待提高,在今后相当长的一段时间内,网络欺诈与网络钓鱼等欺骗技术会一直在互联网上发生,并且受到相关监控^[10]。

文中设计实现了一个钓鱼邮件监控系统来保障公共网络安全。实验表明系统能检测到钓鱼邮件里面的敏感词并且有一个良好的警报效果。而如何检测邮件内容是图片且图片中存在钓鱼网站相关信息是下一步的工作。

参考文献:

- [1] 吕述望,王昭顺,李 剑,等. 针对电子银行的网络钓鱼攻击及其防范策略[J]. 信息安全,2011(7):1-3.
- [2] 朱 红,刘宝成,张 开. 规避网络钓鱼给证券行业带来的安全风险[J]. 信息安全与技术,2011(5):67-69.
- [3] 丁岳伟. 基于 SMTP 协议电子邮件的还原[J]. 小型微型计算机系统,2002,23(3):290-293.
- [4] 吴 勋,刘嘉勇. 基于网络数据包的邮件还原技术研究[J]. 通信技术,2011,44(4):124-126.
- [5] 郑 魁,疏学明,袁宏永. 网络舆情热点信息自动发现方法[J]. 计算机工程,2010,36(3):4-6.
- [6] 吴志强,马春波,敖发良. 基于 Winpcap 的邮件还原系统的实现[J]. 微型机与应用,2011,30(2):58-61.
- [7] 何高辉,邹福泰,谭大礼,等. 基于 SVM 主动学习算法的网络钓鱼检测系统[J]. 计算机工程,2010,37(19):126-128.
- [8] Liu Wenyin. An antiphishing strategy based on visual similarity assessment[J]. Internet Computing,2006,10(2):58-65.
- [9] Fu A Y. Detecting phishing web pages with visual similarity assessment based on earth mover's distance[J]. IEEE Trans on Dependable and Secure Computing,2006,3(4):301-311.
- [10] Raffetseder T. Building anti-phishing browser plug-ins: an experience report [C]//Proceedings of the 3rd International Workshop on Software Engineering for Secure Systems. Austria:[s. n.],2007:1-7.

(上接第 102 页)

参考文献:

- [1] 杨启亮,邢建春,王 平. 安全关键系统及其软件方法[J]. 计算机应用与软件,2011,28(2):129-138.
- [2] 贾旭杰. 安全关键系统可靠性与安全性的研究与分析[M]. 北京:中国科学技术出版社,2011.
- [3] SAE International. Architecture analysis and design references language(AADL)[S]. AS5506,2004.
- [4] SAE-AS5506/1. Architecture analysis and design language annex volume 1[S]. 2006.
- [5] 杨志斌,皮 磊,胡 凯,等. 复杂嵌入式实时系统体系结构设计与分析语言: AADL[J]. 软件学报,2010,21(5):899-915.
- [6] ErrorModelAnnex-phf-JuneMtg2009[EB/OL]. 2009-04. http://www.aadl.info/aadl/documents/.

- [7] 史定华,王松瑞. 故障树分析技术方法和理论[M]. 北京:北京师范大学出版社,1993.
- [8] Li Yue,Zhu Yian, Ma Chunyan, et al. A method for constructing fault trees from AADL models[C]//Proc of the 8th International Conference on Autonomic and Trusted Computing. Berlin:Springer-Verlag,2011:245-258.
- [9] 周建军. 基于有向图和故障树的导弹故障诊断系统研究[D]. 北京:航天工业总公司四部,2001.
- [10] 李堂经,王新阁,杨 哲. 动态故障树的综合分析方法[J]. 装备制造技术,2009(8):22-23.
- [11] 朱正福,李长福,何恩山,等. 基于马尔可夫链的动态故障树分析方法[J]. 兵工学报,2008,29(9):1104-1107.
- [12] 高顺川. 动态故障树分析方法及其实现[D]. 长沙:国防科技大学,2005.

校园网钓鱼邮件监控系统的研究与实现

作者：[蔡洪民, CAI Hong-min](#)

作者单位：[广州中医药大学 信息中心, 广东 广州, 510006](#)

刊名：[计算机技术与发展](#)

英文刊名：

Computer Technology and Development

ISTIC

年, 卷(期):

2013(10)

本文链接：http://d.wanfangdata.com.cn/Periodical_wjtz201310026.aspx