Journal of Computer Applications ISSN 1001-9081 2016-04-10
计算机应用,2016,36(4):941 -944,984 CODEN JYIIDU http://www.joca.cn

文章编号:1001-9081(2016)04-0941-04 DOI:10.11772/j.issn.1001-9081.2016.04.0941

基于多元属性特征的恶意域名检测

张 洋*,柳厅文,沙泓州,时金桥
(中国科学院 信息工程研究所,北京 100093)
(* 通信作者电子邮箱 zhangyang@iie.ac.cn)

摘 要:域名系统主要提供域名解析功能,完成域名到IP的转换,而恶意域名检测主要用来发现以域名系统为屏障的非法行为,来保障域名服务器的正常运行。总结了恶意域名检测的相关工作,并采用基于机器学习的方法,提出一种基于多元属性特征的恶意域名检测方法。在域名词法特征方面,提取更加细粒度的特征,比如数字字母的转换频率、连续字母的最大长度等;在网络属性特征方面,更加关注名称服务器,比如其个数、分散度等。实验结果表明,该方法的准确率、召回率、*F*1值均达到了99.8%,具有较好的检测效果。

关键词:恶意域名;域名系统;网络钓鱼;随机森林
中图分类号:TP309 **文献标志码**:A

Malicious domain detection based on multiple-dimensional features

ZHANG Yang*, LIU Tingwen, SHA Hongzhou, SHI Jinqiao
(*Institute of Information Engineering, Chinese Academy of Sciences, Beijing* 100093, *China*)

Abstract: Domain Name System (DNS) provides domain name resolution service, i.e., converting domain names to IP addresses. Malicious domain detection is mainly for discovering illegal activities and ensuring the normal operation of the domain name servers. Prior work on malicious domain name detection was summarized, and a new machine learning based malicious domain detection algorithm for exploiting multiple-dimensional features was further proposed. With respect to domain name lexical features, more fine-grained features were extracted, such as the conversion frequency of the numbers and letters and the maximum length of continuous letters. As for the network attribute features, more attentions were paid to the name servers, such as the quantity, and the degree of dispersion. The experimental results show that the accuracy, recall rate, *F*1 value of the proposed method reaches 99.8%, which means a better performance on malicious domain name detection.

Key words: malicious domain; Domain Name System (DNS); phishing; random forests

0 引言

域名系统(Domain Name System, DNS)作为互联网最重要的核心基础设施之一,主要提供域名解析功能,完成域名到IP地址的转换。几乎所有的互联网应用都离不开DNS的域名解析功能。因而,很多网络安全设施都会允许DNS协议类型的数据包通过。正是由于这一原因,DNS受到了攻击者的广泛关注,利用DNS域名进行网络钓鱼、僵尸网络的命令控制(Command and Control, C&C)通信、隐蔽传输敏感信息等威胁行为[1],恶意域名数量在迅速增长[1-2]。因此,恶意域名的检测迫在眉睫。

为了应对日益严重的网络威胁,2013年美国国土安全部发布了爱因斯坦3促进计划(EINSTEIN 3-Accelerated, E3A)和网络安全增强服务计划(Enhanced Cybersecurity Services, ECS),两个计划都提出了将DNS Sinkholing作为应对网络空间威胁的两大手段之一。2014年,在国际安全顶级会议BlackHat上的"APT Attribution and DNS Profiling"报告将DNS域名分析作为应对APT威胁的重要技术手段。

与此同时,OpenDNS、奇虎360等众多安全公司纷纷投入大量的人力物力进行相关研究。例如,OpenDNS提供反钓鱼服务和输入纠正,当用户访问恶意网站时,OpenDNS将封锁这些恶意网站。

综上所述,研究恶意域名检测在对抗网络空间威胁、维护国家稳定和社会经济正常运行等方面具有重大的现实意义和应用价值。

1 相关工作

1.1 恶意域名解析

本节以访问恶意域名"taobao.xxx.com"为例,说明恶意域名的解析过程。当本地没有该域名的缓存时,恶意域名的解析过程如图1所示。

1) 首先请求提供互联网接入的运营商DNS服务器,如果不存在该域名,则请求根域名服务器。

2) 如果根域名服务器中也不存在该域名,则请求".com"服务器。

3) ".com"服务器中存在"xxx.com"域名,则请求"xxx.

收稿日期:2015-08-31;修回日期:2015-11-02。
基金项目:国家自然科学基金资助项目(61303260);中国科学院战略性先导科技专项(XDA06030200)。
作者简介:张洋(1991—),男,山东临沂人,硕士研究生,主要研究方向:网络与信息安全; 柳厅文(1986—),男,安徽临泉人,助理研究员,博士,CCF会员,主要研究方向:大数据安全分析、知识图谱; 沙泓州(1988—),男,江苏淮安人,博士,主要研究方向:信息安全、数据挖掘; 时金桥(1978—),男,黑龙江哈尔滨人,正研级高级工程师,博士,CCF会员,主要研究方向:信息安全、数据挖掘。

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

com”服务器。

4) “xxx. com”服务器中存在“taobao. xxx. com”域名，则返回其 IP 地址。

5) 用户连接到假冒的网站“taobao. xxx. com”，该网站可能是一个钓鱼网站。

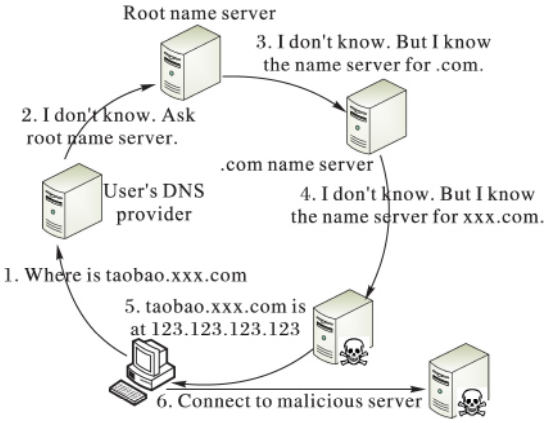

图 1 恶意域名解析过程

通过图 1 可以看到，恶意域名的解析是其他非法活动的重要一环。倘若能够在解析阶段发现可疑的恶意域名，便能将威胁限制在极小的范围之内。

1.2 相关研究

近年来，已经有一些学者、企业、机构等对恶意域名进行了相关研究。Bilge 等[3]把“涉及恶意活动的域名滥用行为”定义为恶意域名。CNCERT/CC 在报告[2]中重点关注涉及网络钓鱼、网页挂马、僵尸网络的恶意域名。目前，针对恶意域名的检测方法可以分为主动分析、被动分析两种。主动分析方法一般包括 DNS 探测、网页内容分析、人工判断。

文献[4]提出一种基于 DNS 探测的检测方法，需要事先对每一个统一资源定位符(Uniform Resoure Locator，URL)进行探测；文献[5]通过对网页内容进行分析来判断是否为恶意域名；人工判断则通过人力来对域名进行分析，进而作出判断。

由此可见，主动分析的方法通常需要较高的分析代价，分析效率相对较低。为此，Weimer[6]在 2005 年提出了基于被动分析的恶意域名检测方法。

目前，学术界对于恶意域名检测方法的研究主要基于被动分析方法，如图 2 所示。

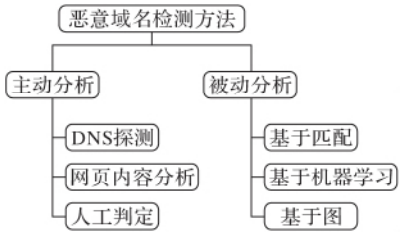

图 2 恶意域名检测方法分类

基于被动分析的恶意域名检测方法可以分为基于匹配、基于机器学习和基于图的方法。本文分析了已有的基于机器学习的检测方法，并将通常选择的特征进行了分类和比较，如表 1 所示。

基于匹配的方法，需要事先维护一个黑名单，通过恶意域名黑名单来匹配恶意域名；基于机器学习的方法是目前研究的热点[5,7]，通过提取特征、建立分类模型来来检测恶意域名；基于图的方法可以挖掘新的恶意域名，但是相关研究较少[13-14]。

表 1 相关论文的特征选取

类型	方法所来自的文献序号						
	[5]	[7]	[8]	[9]	[10]	[11]	[12]
URL	√	√	—	—	√	—	—
whois	—	√	—	—	—	√	—
TTL	√	—	√	√	—	√	√
A 记录	√	—	√	√	—	√	√
AS	—	—	—	√	—	√	√
生存时间	—	—	—	—	√	√	—

注 “√”表示文献方法使用了对应的类型特征
“—”表示文献方法未使用对应的类型特征。
AS 为自治系统(Autonomous system)。

而基于机器学习的方法无需人工过多干预，大大降低了分析成本，并且该方法无需维护黑名单，可以检测黑名单以外的恶意域名。

恶意域名通常具有域名伪装、页面内容伪装、生命周期短、域名频繁跳变、A 记录频繁跳变的特点[4,14]，即“伪装”和“跳变”。本文总结了不同种类恶意域名的特点，如表 2 所示。

表 2 恶意域名特点

特点	恶意域名类型		
	网络钓鱼	网页挂马	僵尸网络
域名伪装	√	√	—
生命周期短	√	√	√
页面内容伪装	√	√	—
域名频繁跳变	√	√	√
A 记录域名频繁跳变	—	—	√

注 “√”表示文献方法使用了对应的类型特征
“—”表示文献方法未使用对应的类型特征。

2 基于多元属性特征的恶意域名检测

针对域名的“伪装”特点和“跳变”特点，本文针对性从多个角度地提取一些特征进行恶意域名检测。如“taobao. xxx. com”，可以从域名词法角度提取属性特征，也可以通过其网络属性来进一步丰富属性特征，如生存时间值(Time-To-Live，TTL)。

本文采用基于机器学习的方法，提出一种基于多元属性特征的恶意域名检测方法，如图 3 所示。

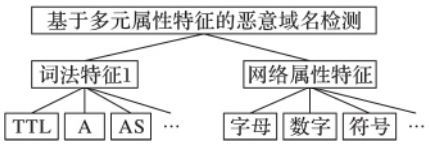

图 3 基于多元属性特征的检测方法

在域名词法特征方面，提取更加细粒度的特征，比如数字字母的转换频率、连续字母的最大长度等；在网络属性特征方面，更加关注名称服务器(Name Server，NS)，比如 NS 的个数、分散度、TTL 值。

2.1 特征选取

表 3 给出了本文使用的 11 个词法特征，并给出了恶意域名在每个特征上的特点。表 4 给出了本文使用的 9 个网络属性特征，并给出了恶意域名在每个特征上的特点。

对 2869 个已标注域名的每个特征进行了统计分析，部分

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

分析结果如图 4 ~ 5 所示。可以发现词法特征和网络属性特征对于检测恶意域名有较好的区分度。

表 3　本文使用的词法特征

编号	词法特征	恶意域名可能的特征
1	域名长度	域名长度较长
2	是否存在 IP 地址	存在 IP 地址
3	分隔符个数	分隔符较多
4	特殊字符个数	特殊字符较多
5	数字个数	数字较多
6	数字占总长度的比例	数字占有较大比例
7	数字、字母转换频率	数字、字母转换频率较大
8	大写字母个数	含有大写字母
9	域名分隔符间的最大长度	域名分隔符间的长度较大
10	连续数字的最大长度	连续数字较长
11	连续字母的最大长度	连续字母较长

表 4　本文使用的网络属性特征

编号	网络属性特征	恶意域名可能的特征
1	TTL 平均值	TTL 平均值较小
2	A 记录个数	多个 A 记录
3	所属网段个数	所属于多个网段
4	AS 个数	AS 个数较多
5	NS 个数	NS 个数较多
6	NS 分散度	NS 分散度较大
7	NS 对应 TTL 平均值	NS 对应的 TTL 平均值较小
8	注册时间	注册时间较晚
9	所属国家	国家分布不均匀

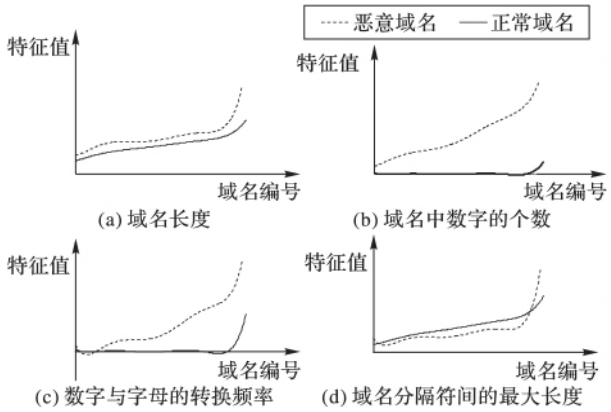

图 4　部分词法特征统计情况

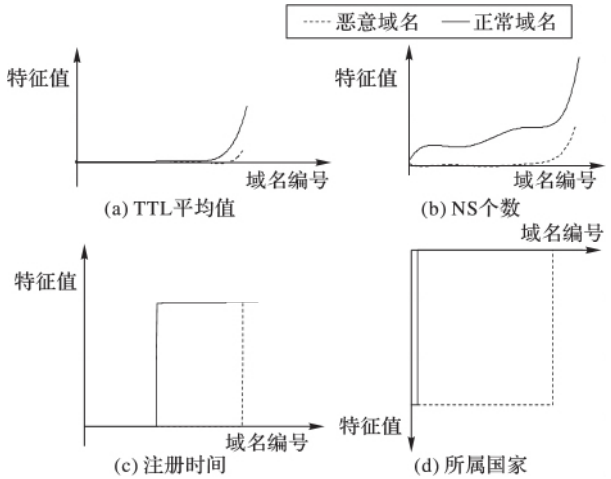

图 5　部分网络属性特征统计情况

从图 4 可看出: 恶意域名通常是数字与字母混合的长串，这与作者的认知基本吻合。

从图 5 可看出: 恶意域名具有 TTL 平均值较小、NS 个数较多、注册时间较晚、所属国家比较分散的特点。

2.2　词法特征转换

词法特征提取算法是将域名字符串转换为词法特征向量:

$$V_L = F_L(D)$$

其中: V_L 为词法特征向量; D 为域名字符串; F_L 为词法特征转换函数，F_L 计算表 3 中的词法特征。

2.3　网络属性特征转换

网络属性特征提取算法是根据域名的网络属性特征，将其映射为网络属性特征向量:

$$V_N = F_N(D)$$

其中: V_N 为网络属性特征向量; D 为域名字符串; F_N 为网络属性特征转换函数，F_N 计算表 4 中的网络属性特征。

2.4　分类算法

本文使用随机森林分类器对恶意域名和正常域名进行分类。随机森林分类器具有诸多优势: 能够处理高维数据，能够处理多种格式的数据，能够处理缺失的特征，能够平衡误差，不会出现过拟合的现象等。鉴于此，随机森林分类器非常适合对域名进行分类。

随机森林分类器作为基于机器学习的一种集成学习方法，结合了多个决策树的分类效果，最终通过“投票”方式选出得票最多的类别作为最终的分类。

每一个树通过下列算法构建:

1) 用 N 来表示训练用例(样本) 的个数，M 表示特征数目。

2) 输入特征数目 m，用于确定决策树上一个节点的决策结果; 其中 m 应远小于 M。

3) 从 N 个训练用例(样本) 中以有放回抽样的方式，取样 N 次，形成一个训练集，并用未抽到的用例(样本) 作预测，评估其误差。

4) 对于每一个节点，随机选择 m 个特征，决策树上每个节点的决定都是基于这些特征确定的。根据这 m 个特征，计算其最佳的分裂方式。

5) 每棵树都会完整生长而不会被剪枝。

3　实验与分析

3.1　数据来源与指标

从网络安全联盟、PhishTank 等网站获取了大量的已知恶意域名，由于恶意域名的持续时间非常多，选择了当前能够提取上述所有特征的 500 个恶意域名作为黑名单。使用 Alexa 排名比较靠前的域名作为正常域名，因为 Alexa 根据域名三个月累积的访问信息为排名依据，因而使用这些数据作为白名单是合理的。

使用 1 162 个来自 Alexa 排名的域名作为正常域名来构建白名单。对于每个域名的网络属性特征，使用 dig、nslookup、whois、bgp. he. net 来得到。

使用准确率 P、召回率 R、$F1$ 值对结果进行评价:

$$P = \frac{TP}{TP + FP}$$

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$$R = \frac{TP}{TP + FN}$$

$$F1 = \frac{2PR}{P + R}$$

如果分类器能够正确地检测出恶意域名，则 $TP = 1$；如果分类器能够正确地检测出正常域名，则 $TN = 1$；如果分类器不能正确地检测出恶意域名，则 $FP = 1$；如果分类器不能正确地检测出正常域名，则 $FN = 1$。

3.2 实验结果与分析

使用随机森林分类器并采用十折交叉验证对 1 662 个域名进行分类验证。

每个词法特征和网络属性特征对检测效果的影响曲线如图 6～7 所示。

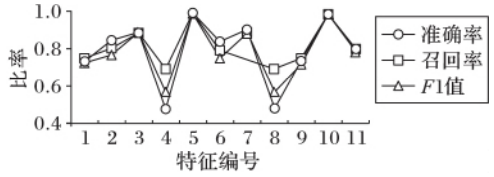

图 6 词法特征对检测效果的影响曲线

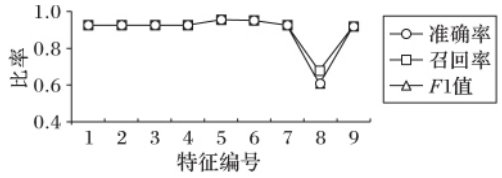

图 7 网络属性特征对检测效果的影响曲线

从实验结果可看出：

1）网络属性特征对实验效果的影响整体好于词法特征，具体来说，网络属性特征 1～7 和特征 9 均使效果保持在较好的水平，而网络属性特征 8（注册时间）相对于其他网络属性特征，对于效果的影响较小。

2）词法特征 5（数字的个数）和词法特征 10（连续数字的最大长度）对于实验结果的影响较好。可知恶意域名经常会包含多个数字，这与本文的认知是比较吻合的，因为正常域名通常使用有意义的字母组合。

3）网络属性特征 5（NS 个数）、6（NS 分散度）、7（NS 对应 TTL 平均值）在很大程度上利于检测效果，所以，可以推断出恶意域名时常会涉及到 NS 服务器。

从表 5 可看出，已使用词法特征 5、10 和网络属性特征 5、6、7 进行恶意域名检测 $F1$ 值就能够达到 0.993。对于其他的词法特征和网络属性特征，虽然单个特征的检测效果有效，但结合在一起对于检测效果的提升也起了一定的作用。

表 5 实验结果

特征组合	准确率	召回率	$F1$ 值
词法特征 5、10 和网络属性特征 5、6、7	0.993	0.993	0.993
所有 20 个特征	0.998	0.998	0.998

综上所述，从域名词法特征、网络属性特征来表达域名的“伪装”和“跳变”特点，使用 20 个特征来构建随机森林分类模型，进而完成恶意域名的检测，具有较好的检测效果。

4 结语

本文研究了恶意域名的检测问题，提出一种基于多元属性特征的恶意域名检测方法。该方法考虑到了恶意域名表现出的“伪装”和“跳变”特点，从域名词法特征、网络属性特征中选取了具有较好检测效果的 20 个特征来构建随机森林分类模型，进而完成恶意域名的检测。实验结果表明，该方法的准确率、召回率、$F1$ 值均达到了 99.8%，具有较好的检测效果。

参考文献：

[1] 王永杰，刘京菊. 基于 DNS 协议的隐蔽通道原理及性能分析[J]. 计算机工程，2014，40(7)：102－105.（WANG Y J，LIU J J. DNS-based covert channel principle and performance analysis [J]. Computer Engineering，2014，40(7)：102－105.）

[2] CNCERT/CC. 2014 中国互联网网络安全报告[EB/OL]. [2015-08-15]. http://www.cert.org.cn/publish/main/upload/File/2014%20secirity%20situation%20report.pdf.（CNCERT/CC. The 2014 Internet security report in China [EB/OL]. [2015-08-15]. http://www.cert.org.cn/publish/main/upload/File/2014%20secirity%20situation%20report.pdf.）

[3] BILGE L，KIRDA E，KRUEGEL C，et al. EXPOSURE：finding malicious domains using passive DNS analysis [EB/OL]. [2015-07-06]. http://seclab.ccs.neu.edu/static/publications/ndss2011dns.pdf.

[4] 洪博，耿光刚，王利明，等. 一种基于 DNS 主动检测钓鱼攻击的系统[J]. 计算机应用研究，2013，30(12)：3771－3774.（HONG B，GENG G G，WANG L M，et al. System to discover phishing attacks actively based on DNS [J]. Application Research of Computers，2013，30(12)：3771－3774.）

[5] ZHANG Y，HONG J I，CRANOR L F. Cantina：a content-based approach to detecting phishing Web sites [C]// Proceedings of the 2007 16th International Conference on World Wide Web. New York：ACM，2007：639－648.

[6] WEIMER F. Passive DNS replication [EB/OL]. [2015-07-06]. http://www.first.org/conference/2005/papers/florian-weimer-paper-1.pdf.

[7] PAN Y，DING X. Anomaly based Web phishing page detection [C]// Proceedings of the 22nd Annual Computer Security Applications Conference. Washington，DC：IEEE Computer Society，2006：381－392.

[8] HOLZ T，GORECKI C，RIECK K，et al. Measuring and detecting fast-flux service networks [EB/OL]. [2015-07-12]. http://user.informatik.uni-goettingen.de/~krieck/docs/2008-ndss.pdf.

[9] ZHOU C V，LECKIE C，KARUNASEKERA S，et al. A self-healing，self-protecting collaborative intrusion detection architecture to trace-back fast-flux phishing domains [C]// Proceedings of the 2008 IEEE Network Operations and Management Symposium Workshops. Piscataway，NJ：IEEE，2008：321－327.

[10] BASNET R，MUKKAMALA S，SUNG A H. Detection of phishing attacks：a machine learning approach [M]// PRASAD B. Soft Computing Applications in Industry. Berlin：Springer，2008，226：373－383.

[11] PASSERINI E，PALEARI R，MARTIGNONI L，et al. FluXOR：detecting and monitoring fast-flux service networks [M]// ZAMBONI D. Detection of Intrusions and Malware，and Vulnerability Assessment，LNCS 5137. Berlin：Springer，2008：186－206.

（下转第 984 页）

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表2 不同方法恢复隐藏文件的测试结果

操作	文献[7]方法	文献[8]方法	本文方法
文件内容操作	不可恢复	可恢复	可恢复
重命名操作	可恢复	可恢复	可恢复
文件整体操作	可恢复	可部分恢复	可恢复
格式化	不可恢复	不可恢复	可恢复

从表2可知:文献[7]方法的鲁棒性最差,而本文提出的文件隐藏方法的鲁棒性最好。按照本文方法实现文件隐藏时,所有对存储器的操作都不会影响隐藏文件的恢复,这是因为隐藏文件存储在专门开辟的隐藏区,而系统操作的存储区为普通区,普通区与隐藏区是相互独立的,所以对普通区的各种操作不会影响隐藏文件的恢复。综上分析,FHCD具有较好的鲁棒性。

4 结语

针对传统文件隐藏方法对操作系统不透明的缺点,本文分析了Nand Flash存储系统的FTL管理机制,讨论了USB移动存储系统的Bulk-Only传输协议和SCSI命令规范。然后,在此基础上,提出了一种新的文件隐藏方法,并通过搭建设备进行了具体实现。该方法采用了软硬件结合技术,利用了USB设备主控芯片对主机透明的优势,通过建立双文件系统,克服了隐藏文件容易受文件操作影响的缺点,使得被隐藏的文件具有更好的隐蔽性。与其他文件隐藏方法相比,具有更高的鲁棒性和隐藏强度,在隐藏文件的同时也隐藏了其存在性,使攻击者很难感知隐藏文件的存在,如何对隐藏区的数据进行加密保护,以抵抗存储器芯片攻击,是下一步需要研究的重要内容。

参考文献:

[1] FRINCKE D. Balancing cooperation and risk in intrusion detection [J]. ACM Transactions on Information and System Security, 2000, 3(1): 1-29.

[2] LIN X J, LI Q B, WANG W, et al. Information hiding based on CAVLC in H.264/AVC standard [C]// Proceedings of the 4th International Conference on Multimedia Information Networking and Security. Piscataway, NJ: IEEE, 2012: 900-904.

[3] Al-FRAJAT A K, JALAB H A, KASIRUM Z M, et al. Hiding data in video file: an overview [J]. Journal of Applied Sciences, 2010, 10(15): 1644-1649.

[4] ZHANG H, WU C, NIU X, et al. A disk disguising and hiding method [C]// Proceedings of the 3rd International Conference on Convergence and Hybrid Information Technology. Piscataway, NJ: IEEE, 2008: 517-520.

[5] RAY R, SANYAL J, DAS D, et al. A new challenge of hiding any encrypted secret message inside any text/ASCII file or in MS word file: RJDA algorithm [C]// Proceedings of the 2012 International Conference on Communication Systems and Network Technologies. Piscataway, NJ: IEEE, 2012: 889-893.

[6] 姜子峰,曾光裕,王炜,等. 基于磁盘冗余空间的数据隐藏[J]. 计算机应用研究, 2014, 31(3): 839-842. (JIANG Z F, ZENG G Y, WANG W, et al. Data hiding based on disk slack space [J]. Application Research of Computers, 2014, 31(3): 839-842.)

[7] NIU X P, LI Q B, WANG W, et al. G bytes data hiding method based on cluster chain structure [J]. Wuhan University Journal of Natural Sciences, 2013, 18(5): 443-448.

[8] 刘玉,刘洋,饶炤骅,等. 基于FAT32磁盘文件系统结构的文件隐藏方法: CN1434450 [P]. 2003-08-06. (LIU Y, LIU Y, RAO S H, et al. File hidden based on FAT32 disk file system structure: CN1434450 [P]. 2003-08-06.)

[9] 蔡风华. 基于FAT32文件系统的文件隐藏研究与实现[D]. 武汉:华中科技大学, 2007: 1-52. (CAI F H. File Hidden Research and Implementation based on the FAT32 File System [D]. Wuhan: Huazhong University of Science and Technology, 2007: 1-52.)

[10] 袁杰,江祖敏. 基于FAT32的文件隐藏方法及在Linux上的实现[J]. 电子设计工程, 2012, 20(13): 15-18. (YUAN J, JIANG Z M. Implementation of file hidden based on FAT32 in Linux system [J]. Electronic Design Engineering, 2012, 20(13): 15-18.)

[11] 郭松辉,王玉龙,牛小鹏,等. 基于闪存冗余块的文件隐藏技术[J]. 计算机应用研究, 2014, 32(8): 900-904. (GUO S H, WANG Y L, NIU X P, et al. File hiding based on flash redundant blocks [J]. Application Research of Computers, 2014, 32(8): 900-904.)

[12] BURTON B J. How to hide file in slack file space with Slacker.exe [EB/OL]. [2013-04-18]. http://cfile27.uf.tistory.com/attach/214FDB4555D96EB418C14A.

Background

WANG Kang, born in 1990, M. S. candidate. His research interests include network information security.

LI Qingbao, born in 1967, Ph. D., professor. His research interests include computer system architecture, information security and trusted computing.

(上接第944页)

[12] PERDISCI R, CORONA I, DAGON D, et al. Detecting malicious flux service networks through passive analysis of recursive DNS traces [C]// Proceedings of the 2009 Annual Computer Security Applications Conference. Washington, DC: IEEE Computer Society, 2009: 311-320.

[13] CHAU D H, NACHENBERG C, WILHELM J, et al. Polonium: Tera-scale graph mining for malware detection [EB/OL]. [2015-07-12]. http://epubs.siam.org/doi/pdf/10.1137/1.9781611972818.12.

[14] MANADHATA P, YADAV S, RAO P, et al. Detecting malicious domains via graph inference [C]// Proceedings of the 2014 Workshop on Artificial Intelligent and Security Workshop. New York: ACM, 2014: 59-60.

Background

This work is supported by the National Natural Science Foundation of China (61303260), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA06030200).

ZHANG Yang, born in 1991, M. S. candidate. His research interests include network and information security.

LIU Tingwen, born in 1986, Ph. D., research associate. His research interests include big data security, knowledge graph.

SHA Hongzhou, born in 1988, Ph. D. His research interests include information security, data mining.

SHI Jinqiao, born in 1978, Ph. D., Senior Engineer. His research interests include information security, data mining.

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net