恶意域名检测技术综述与展望

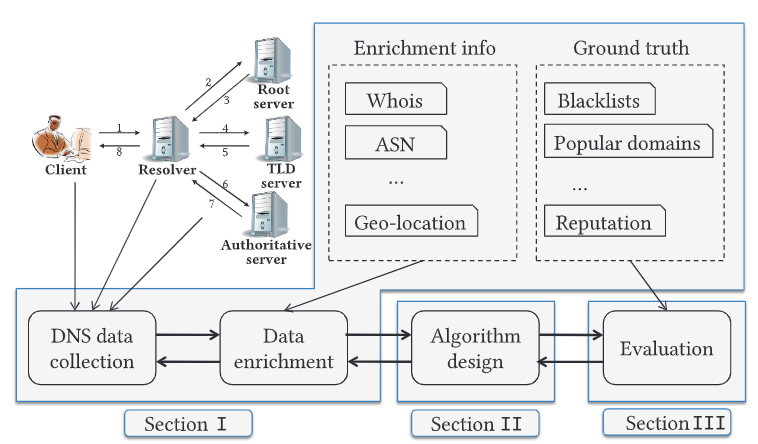
摘要：鉴于恶意域名检测技术的重要性和已有相关研究工作，本文首先介绍目前现有研究的一般方法。其次，从传统的机器学习的检测方法到主流的深度学习检测方法，概述已有的各种恶意域名检测技术。最后，归纳现有恶意域名检测技术存在的不足并对恶意域名检测技术的趋势及未来研究方向进行展望。

# 1引言

域名系统（英语：Domain Name System，缩写：DNS）是互联网的一项服务。它作为将域名和IP地址相互映射的一个分布式数据库，能够使人更方便地访问互联网[1]。作为互联网基础设施，域名系统存在极大的安全隐患，主要表现在：1)协议设计存在较大的脆弱性，缺乏必要的安全性考虑，数据信息真实性和完整性得不到保证；2)系统规模不断扩大导致管理难度急剧增加，人为错误因素使得配置故障普遍存在，系统冗余性大大降低，单点失效问题严重[2]。大量的网络服务依赖于域名服务来展开同时因为DNS的安全隐患，域名系统被滥用于各种恶意活动：散布恶意软件，促进命令和控制服务器之间的（C＆C）通信，发送垃圾邮件，托管诈骗和网络钓鱼网页等。可以看出检测这些恶意活动的起源非常重要，目前可以通过分析网络间流量[3,4]，URL检测[5]等。在此之中，对恶意活动所依托的域名进行检测分析也是一个非常重要的方向。本文综述了恶意域名检测技术的方法及其最新研究进展，结合现在网络中恶意活动的发展趋势指出了检测技术的难点以及未来的研究方向。

本文第2节介绍了恶意域名研究的各个过程。第3节介绍针对恶意域检测的各类研究方案。第4节展望未来的研究方向。第5节为结束语。

# 2 恶意域名研究过程



如图所示针对恶意域名检测的研究大体可以分为3个过程：

1. 相关数据的获取
2. 算法设计
3. 评价方法

# 2.1 相关数据的获取

根据上图客户机发起DNS请求的全过程所示，可以将DNS数据的获取位置分为Host-Resolver与DNS-DNS[6]。

Host-Resolver中获得数据是图中1，8两个位置的DNS数据。BDS[7],Choietal.[8]，BotGAD[9,10],Lee and Lee[11]等都是在这样的位置进行数据获取的。这种方式的好处就是可以获取域名解析器的内部接口，而其提供了可能直接与特定恶意行为相关的用户请求和响应的细节信息。比如，被僵尸网络控制的主机往往会发送相似形式的域名请求信息。此外，任何公司和研究机构都可以直接在自己的解析器中部署检测装置来获取数据，而不需要和其他部门进行合作。此方法的局限之处在于其仅仅获取了某个单独机构的信息，其对于研究各种恶意行为的模式往往不够充分。而像大型ISPs这种机构的解析器拥有较大数量的用户，却由于安全和隐私考虑不对外开放获取途径。

DNS-DNS中获得数据是图中2，7两个位置的DNS数据。Exposure [12,13],Notos[14], Khalil[15]等都是在这样的位置进行数据获取的。最常用的获取DNS-DNS流量数据的位置是：1、授权域名服务器，包括负责TLDs的服务器；2、解析器的外部接口。检测获取的位置越靠近DNS树的根节点，获取到的数据的覆盖范围越大；越靠近根节点能其可获取几乎所有的域名请求，但是无法获取域名的应答报文，反之低级的服务器能获得更多的应答报文，但是缺乏域名请求报文。这种方式相较于Host-Resolver形式最主要的缺点是数据不容易获取，需要通过与服务器管理商合作等形式得到对各服务器进行数据获取的权限。

在确定了在何处进行数据获取后，DNS的数据捕获又可以分为主动数据获取和被动数据获取。

主动获取是通过主动发送DNS请求并记录相应的应答来实现的[16-18]。其中请求的域名数据有多种来源，如有名的域名集（Alexa Top Sites等）、黑名单中出现的域名、以及来自授权域名服务器文件的域名等。这种方式主要得到的是域名的DNS记录，比如被解析出来的IP地址，规范化域名，记录的TTL等。

被动获取主要是通过在DNS服务器部署相应的程序来得到包含DNS请求和应答的服务器日志文件[12,14,19]。由于获得的是服务器的数据（数据可能来自于不同地方的机构部门），所以这种被动获取方式得到的数据具有更加丰富的特征和统计信息，可以据此研究更多的恶意行为特征。

# 2.2算法设计

利用相关数据对恶意域检测算法进行设计涉及三个部分：使用了哪些特征，采用什么样的检测方法，产生了什么结论。

特征提取被提取特征的质量对实验方法的成功与否起到了至关重要的作用。尽管有些特征可能具有较高的实验精度，但是其很容易被攻击者通过某些小的修改而失效，所以特征在精度和鲁棒性两方面都需要表现优良。某些特征可以单独从DNS资源记录中获取而不需要额外的数据源，如“域名的平均TTL值”[12,13,20]，从DGA域名中提取特征[14,21,22]。某些特征也可以利用DNS资源记录和外部信息组合起来，如计算某域名对应的IP地址所属的ASN的数量[16,22,23]，其需要知道IP-AS的信息。

恶意域名的检测方法一般可以分为三种：1、基于知识的方法；2、基于机器学习的方法；3、基于混合的方法。各检测方法的具体介绍在第三节。

恶意行为有：垃圾邮件、钓鱼网页、僵尸网络等多种类型，算法的研究结论可以分为不定行为检测即不划分恶意行为的具体类型，而是宏观分析域名的多种行为联系[15,24,25]；特定行为检测即挖掘特定恶意行为的具体特点[26]。

# 2.3评价方法

大多数基于DNS的恶意域名检测方法都利用了机器学习的概念和技术，例如聚类、分类。这些方法中自然会使用机器学习社区开发和使用的评估指标和策略。

一般情况使用混淆矩阵作为评估指标。混淆矩阵是除了ROC曲线和AUC之外的另一个判断分类好坏程度的方法。下面给出二分类的混淆矩阵。

|  |  |  |
| --- | --- | --- |
|  | Predicted as Positive | Predicted as Negative |
| Labeled as Positive | True Positive(TP) | False Negative(FN) |
| Labeled as Negative | False Positive(FP) | True Negative(TN) |

真正例(True Positive, TP)：真实类别为正例，预测类别为正例。

假负例(False Negative, FN)：真实类别为正例，预测类别为负例。

假正例(False Positive, FP)：真实类别为负例，预测类别为正例。

真负例(True Negative, TN)：真实类别为负例，预测类别为负例。

由上述可知：

真正率(True Positive Rate, TPR)：被预测为正的正样本数 / 正样本实际数，即：TPR=TP/(TP+FN)。

假负率(False Negative Rate, FNR)：被预测为负的正样本数/正样本实际数，即：FNR=FN/(TP+FN)。

假正率(False Positive Rate, FPR)：被预测为正的负样本数/负样本实际数，即：FPR=FP/(FP+TN)。

真负率(True Negative Rate, TNR)：被预测为负的负样本数/负样本实际数，即：TNR=TN/(FP+TN)。

# 3 恶意域名研究方法

恶意域名检测方法中，有的方法是基于外部的专家知识，结合统计知识来区别良性域名与恶行域名的，这种只利用专家知识进行检测的方法本文成为基于知识的方法；有的方法是利用机器学习（包含神经网络）中的各种算法来区别良性域名与恶行域名的，这种方法本文成为基于机器学习的方法；有的方法则是混合多种机器学习方法来实现，或者混合使用专家知识与机器学习的方法来实现，本文成为基于混合的方法。

# 3.1基于知识的方法

基于知识的方法往往依赖大量的专家知识与经验，这一般来自于在对恶意域名的探索过程。有很多文献是采用这样的方法。如Sato[27]观察到属于一个恶意软件家族的恶意域名往往会被多次查询。Krishnan[28]和Guerid[29]观察到在DNS查询方面，网络中存在的机器人往往表现出类似的模式，而DNS基础设施无法解决这些问题。Yadav等[26]通过查看同一组IP地址的所有域中的unigram和bigram特征分布，查找算法生成的域名的固有模式来检测DNS流量中的域名。Mowbray等[39]在域名查询服务中通过使用不寻常的字符串长度分布来检测恶意域名。

这些方法也都有局限性，专家知识也可能有意或无意间对数据产生一定的偏见。如Grill[30]在观察到DGA恶意软件会生成大量DNS的解析以找到能够与正常域名进行通信的DNS解析，通常感染了这类恶意软件的主机其DNS解析数量大于后续通信的数量。然而现代浏览器会预测用户的互联网行为，并提前解析一些域名即使他们从来没有被查询，这就会导致这种方法会产生误报，因为这种行为对于专家知识是未知的。

# 3.2基于机器学习的方法

大多数恶意域名检测的方法采用的都是以数据驱动为核心的机器学习算法[31]，一般我们可以分为监督学习，无监督学习，神经网络三类。

监督学习中需要完整的训练集，与数据样本相对应的每个特征向量都必须与代表该样本所属类别的标签相关联，即意味着训练集中的每个域名都必须明确标记为恶意或良性。依靠这种方法的研究人员仅需要从原始数据中提取特征，并在标记的数据集上训练分类器，将训练有素的分类器应用于新数据。如DomainProfiler [16]使用随机森林算法用于发现被滥用的域名，其提取了一共55个特征（考虑了相关的IP地址和域名）。Antonakakis et al. [14]也使用了随机森林方法。

无监督学习又称聚类技术[32]即只使用数据的内部属性自动将域名划分为聚类。理论上恶意域名与正常域名具有完全不同的行为特征，可以使用聚类算法将数据样本划分为两类，然后由研究人员决定哪个集群属于恶意[9,10]，

上述监督学习与无监督学习基本都采用的是手工提取特征，有两个缺点：手工提取的特征容易规避，手工提取特征耗时。而深度学习方法则为恶意域名检测技术提供了新的思路。在使用神经网络的文献中[33]提出了一种基于word-hashing技术的深度学习网络对域名进行分类，其不仅避免了手工提取特征还发现了传统统计方法无法发现的特征。Woodbrideg[34]利用LSTM网络实现对DGA域名的实时预测，而无需上下文信息或手动创建的特征。Tran[35]提出了一种改进的成本敏感的LSTM算法来应对DGA域名据多类不平衡的问题，相对原始敏感LSTM算法，具有较高的准确率。Yin[36]提出了一种基于生成对抗网络的僵尸网络检测增强框架，该网络通过生成器连续生成“假”样本，并扩展标记的数量，以帮助原始模型进行僵尸网络检测和分类。陈立皇等[47]也提出了一种基于注意力机制的深度学习模型，不同的是他们提出一种域名的多字符随机性提取方法，提升了识别低随机DGA域名的有效性。

# 3.3基于混合的方法

虽然现在可以利用单个检测算法对域名进行分类，但大多数工程中使用的方法都是混合的，并采用几种不同类型的算法来产生结果，如使用多种机器学习算法的组合[14,20,21,37]

# 4 未来展望

恶意域名检测领域中大部分的数据都是有标注的，目前使用机器学习中的各类算法可以有效的检测出恶意域名，但是随着恶意域名生成技术的发展，从简单的利用域名生成算法生成伪随机子符串的域名发展到可以躲避神经网络检测的更加智能化的域名。

域名检测面临的一个挑战便是新域名知识验证的困难。大多数的方法只能很好的应用在其论文自身的测试集上，并没有真正做到预测新型恶意域名，利用GAN生成对抗网络来预测未来可能出现的DGA变体域名[38]是一个很好的研究方向。

因为面临法律、隐私等相关的阻碍，恶意域名检测领域中缺少公开可用的数据集，需要有研究人员来构建这样一个基准测试数据集来进一步推动恶意域名检测领域的发展。

基于现有的检测模型，如何设计一个高效的检测模型是一个难点，因为伪域名越来越智能化，可以逃避一般的神经网络模型的检测。

# 5 结束语

以互联网为依托的经济贸易圈日益增大，网络信息安全成为了近年来关注的热点。随着5G的快速发展，网络空间会存在着越来越多的网络攻击，恶意域名只是冰山一角，我们需要加快网络空间安全建设，完善恶意域名检测系统，做到对恶意域名的实时检测。

# 参考文献

1. Mockapetris P, Dunlap K J. Development of the domain name system[C]//Symposium proceedings on Communications architectures and protocols. 1988: 123-133.
2. 王垚, 胡铭曾, 李斌, 等. 域名系统安全研究综述[J]. 通信学报, 2007, 28(9): 91-103.
3. Seungwon Shin, Zhaoyan Xu, and Guofei Gu. 2012. EFFORT: Efficient and effective bot malware detection. In Pro-ceedings of the IEEE INFOCOM. 2846–2850.
4. Ting-Fang Yen and Michael K. Reiter. 2008. Traffic aggregation for malware detection. In Proceedings of the Interna-tional Conference on Detection of Intrusions and Malware, and Vulnerability Assessment. 207–227.
5. Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. 2011. Learning to detect malicious URLs. ACM Trans. Intell. Syst. Technol. 2, 3 (2011), 30:1–30:24.
6. Zhauniarovich Y, Khalil I, Yu T, et al. A survey on malicious domains detection through DNS data analysis[J]. ACM Computing Surveys (CSUR), 2018, 51(4): 1-36.
7. Iria Prieto, Eduardo Magaña, Daniel Morató, and Mikel Izal. 2011. Botnet detection based on DNS records and active probing. In Proceedings of the International Conference on Security and Cryptography. 307–316.
8. Hyunsang Choi, Hanwoo Lee, Heejo Lee, and Hyogon Kim. 2007. Botnet detection by monitoring group activities in DNS traffic. In Proceedings of the IEEE International Conference on Computer and Information Technology. 715–720.
9. Hyunsang Choi and Heejo Lee. 2012. Identifying botnets by capturing group activities in DNS traffic. Comput. Netw. 56, 1 (2012), 20–33.
10. Hyunsang Choi, Heejo Lee, and Hyogon Kim. 2009. BotGAD: Detecting botnets by capturing group activities in net-work traffic. In Proceedings of the International ICST Conference on Communication System Software and Middleware. 2:1–2:8.
11. J. Lee, J. Kwon, H. J. Shin, and H. Lee. 2010. Tracking multiple C&C botnets by analyzing DNS traffic. In Proceedings of the IEEE Workshop on Secure Network Protocols. 67–72.
12. Leyla Bilge, Engin Kirda, Christopher Kruegel, and Marco Balduzzi. 2011. EXPOSURE: Finding malicious domains using passive DNS analysis. In Proceedings of the Network and Distributed System Security Symposium.
13. Leyla Bilge, Sevil Sen, Davide Balzarotti, Engin Kirda, and Christopher Kruegel. 2014. Exposure: A passive DNS analysis service to detect and report malicious domains. ACM Trans. Info. Syst. Secur. 16, 4 (apr 2014), 14:1–14:28.
14. Manos Antonakakis, Roberto Perdisci, David Dagon, Wenke Lee, and Nick Feamster. 2010. Building a dynamic reputation system for DNS. In Proceedings of the USENIX Security Symposium. 273–290.
15. Issa M. Khalil, Ting Yu, and Bei Guan. 2016. Discovering malicious domains through passive DNS data graph analysis. In Proceedings of the ACM Symposium on Information, Computer and Communications Security. 663–674.
16. Daiki Chiba, Takeshi Yagi, Mitsuaki Akiyama, Toshiki Shibahara, Takeshi Yada, Tatsuya Mori, and Shigeki Goto. 2016. DomainProfiler: Discovering domain names abused in future. Proceedings of the Annual IEEE/IFIP International Conference on Dependable Systems and Networks, 491–502.
17. Thorsten Holz, Christian Gorecki, Konrad Rieck, and Felix C. Freiling. 2008. Measuring and detecting fast-flux ser-vice networks. In Proceedings of the Network and Distributed System Security Symposium.
18. Maria Konte, Nick Feamster, and Jaeyeon Jung. 2009. Dynamics of online scam hosting infrastructure. In Proceedings of the International Conference on Passive and Active Network Measurement. 219–228.
19. Manos Antonakakis, Roberto Perdisci, Wenke Lee, Nikolaos Vasiloglou, II, and David Dagon. 2011. Detecting mal-ware domains at the upper DNS hierarchy. In Proceedings of the USENIX Security Symposium. 27–27.
20. R. Perdisci, I. Corona, D. Dagon, and Wenke Lee. 2009. Detecting malicious flux service networks through pas-sive analysis of recursive DNS traces. In Proceedings of the Annual Computer Security Applications Conference. 311– 320.
21. Manos Antonakakis, Roberto Perdisci, Yacin Nadji, Nikolaos Vasiloglou, Saeed Abu-Nimeh, Wenke Lee, and David Dagon. 2012. From throw-away traffic to bots: Detecting the rise of DGA-based malware. In Proceedings of the USENIX Security Symposium. 24–24.
22. Hyunsang Choi and Heejo Lee. 2012. Identifying botnets by capturing group activities in DNS traffic. Comput. Netw. 56, 1 (2012), 20–33.
23. Xin Hu, M. Knysz, and K. G. Shin. 2011. Measurement and analysis of global IP-usage patterns of fast-flux botnets. In Proceedings of the IEEE INFOCOM. 2633–2641.
24. J. Lee and H. Lee. 2014. GMAD: Graph-based malware activity detection by DNS traffic analysis. Comput. Commun. 49 (2014), 33–47.
25. Pratyusa Manadhata, Sandeep Yadav, Prasad Rao, and William Horne. 2014. Detecting malicious domains via graph inference. In Proceedings of the European Symposium on Research in Computer Security. 1–18.
26. Sandeep Yadav, Ashwath Kumar Krishna Reddy, A. L. Narasimha Reddy, and Supranamaya Ranjan. 2010. Detecting algorithmically generated malicious domain names. In Proceedings of the ACM SIGCOMM Conference on Internet Measurement. 48–61.
27. Kazumichi Sato, Keisuke Ishibashi, Tsuyoshi Toyono, and Nobuhisa Miyake. 2010. Extending black domain name list by using co-occurrence relation between DNS queries. In Proceedings of the 3rd USENIX Conference on Large-scale Exploits and Emergent Threats: Botnets, Spyware, Worms, and More.8 – 8 .
28. Srinivas Krishnan, Teryl Taylor, Fabian Monrose, and John McHugh. 2013. Crossing the threshold: Detecting net-work malfeasance via sequential hypothesis testing. In Proceedings of the Annual IEEE/IFIP International Conference on Dependable Systems and Networks. 1–12.
29. H. Guerid, K. Mittig, and A. Serhrouchni. 2013. Privacy-preserving domain-flux botnet detection in a large scale network. In Proceedings of the International Conference on Communication Systems and Networks.1 – 9 .
30. M. Grill, I. Nikolaev, V. Valeros, and M. Rehak. 2015. Detecting DGA malware using netflow. In Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management. 1304–1309.
31. Matija Stevanovic, Jens Myrup Pedersen, Alessandro D’Alconzo, Stefan Ruehrup, and Andreas Berger. 2015. On the ground truth problem of malicious DNS traffic analysis. Comput. Secur. 55 (Nov. 2015), 142–158.
32. A. K. Jain, M. N. Murty, and P. J. Flynn. 1999. Data clustering: A review. Comput. Surveys 31, 3 (Sept. 1999), 264– 323.
33. 赵科军, 葛连升, 秦丰林, 等. 基于 word hashing 的 DGA 僵尸网络深度检测模型[J]. 东南大学学报: 自然科学版, 2017, 47(A01): 30-33.
34. Woodbridge J, Anderson H S, Ahuja A, et al. Predicting domain generation algorithms with long short-term memory networks[J]. arXiv preprint arXiv:1611.00791, 2016.
35. Tran D, Mac H, Tong V, et al. A LSTM based framework for handling multiclass imbalance in DGA botnet detection[J]. Neurocomputing, 2018, 275: 2401-2413.
36. Yin C, Zhu Y, Liu S, et al. An enhancing framework for botnet detection using generative adversarial networks[C]//2018 International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 2018: 228-234.
37. R. Perdisci, I. Corona, and G. Giacinto. 2012. Early detection of malicious flux networks via large-scale passive DNS traffic analysis. IEEE Trans. Depend. Secure Comput. 9, 5 (2012), 714–726.
38. 王媛媛, 吴春江, 刘启和, 等. 恶意域名检测研究与应用综述[J]. 计算机应用与软件, 2019, 36(09): 310-316.
39. Mowbray M, Hagen J. Finding domain-generation algorithms by looking at length distribution[C]//2014 IEEE international symposium on software reliability engineering workshops. IEEE, 2014: 395-400.
40. 陈立皇, 程华, 房一泉. 基于注意力机制的 DGA 域名检测算法[J]. 华东理工大学学报 (自然科学版), 2019, 45(3): 478-485.