基于DNS识别恶意域名
1 DNS识别恶意域名的优势
2 DNS安全
2.1 Securing DNS
2.2 Securing Data Provided by DNS
2.3 Securing Users from Attacks Leveraging DNS Disingenuously
2.4 Securing Users from Attacks Leveraging DNS Genuinely
3 恶意域名检测方法
3.1 特征
3.1.1 Internal vs Contextual features
3.1.2 DNS Dataset Dependent vs. Independent features
3.1.3 Mono vs. Multi Domains Features
3.2 检测方法
3.2.1 基于知识的方法（Knowledge Based methods）
3.2.2 基于机器学习的方法（Machine Learning Based methods）
1）监督学习算法
2）半监督学习算法
3）无监督学习算法
3.2.3 混合方法（Hybrid approaches）
4 TLD异常检测
4.1 TLD DDos检测
4.2 TLD Abuse
参考资料

基于DNS识别恶意域名

一直以来，网络被用来攻击不同目标。良性服务和协议被滥用于各种恶意活动，例如传播恶意软件、滥发信息、托管诈骗和钓鱼网页。因此，检测这些恶意活动的来源是非常重要的，可以通过识别URL、域名、IP等手段进行识别。其中DNS数据分析是非常有前途的恶意域名识别方法。本报告简单介绍通过DNS识别恶意域名的方法以及TLD上有关的异常检测相关工作。

1 DNS识别恶意域名的优势

1. DNS数据在网络流量中的比重很小，利于分析。
2. 缓存机制的存在大大减少了网络中DNS的数量，允许研究来自TLD的DNS流量
3. DNS流量包含特别很多与域名有关的信息和特征，这些信息可以用来识别恶意活动
4. 许多信息和特征可以进一步与其他信息关联丰富起来，利用机器学习方法可以做很多识别工作
5. 尽管加密DNS的解决方案已经存在，但是大量DNS流量依旧是未被加密的。

2 DNS安全

2.1 Securing DNS

作为Internet的基础技术，所有DNS组件都遭到了对手的广泛攻击和利用。DNS基础设施已经成为许多拒绝服务尝试的目标。DNS软件多年来一直是被攻击的对象。最早被报道的Linux蠕虫之一，ADM蠕虫，利用了DNS服务器的缓冲区溢出漏洞。为了保护DNS免受这些攻击目前有的解决方案包括：

- DNS服务器副本
- Anti-DDos缓解工具

2.2 Securing Data Provided by DNS

攻击者总是试图破坏合法DNS服务器提供的数据，因为这允许他们将通信重定向到受控制的资源。2008年6月，世界上两个最重要的互联网监管网站，ICANN和IANA被攻击，从而直接导致了一系列安全策略的出现。

2.3 Securing Users from Attacks Leveraging DNS Disingenuously

DNS的第三类安全威胁与该协议的滥用有关。DNS可能是每个计算网络允许使用的为数不多的协议之一。许多恶意软件和僵尸网络都在滥用它，以使受损害的主机和它们的命令和控制服务器之间的通信成为可能。在不久前，攻击者利用的是DNS响应基于UDP，并且发送数量远大于查询数量。攻击者利用这些特性来发起拒绝服务攻击。

2.4 Securing Users from Attacks Leveraging DNS Genuinely

为了运行恶意活动，攻击者需要在远程服务器中托管的各种服务。在早期，恶意软件的常见做法是硬编码服务器的IP地址，以接收命令或窃取数据。这种做法很快就被放弃了，因为捕获一个恶意软件样本就可以提取所有ip，这足以关闭整个僵尸网络。很明显，这些服务器需要能够跨IP空间移动。这正是创建DNS的目的。此外，为了不被列入黑名单，域名也应该跨域名空间移动。用于实现这种敏捷行为的主要技术有两种：**Domain-Flux and IP-Flux (or Fast-Flux)**。前者指的是让几个FQDNs与一个IP地址相关联的策略。使用一个域名生成算法(DGA)，一个恶意软件能够动态地生成新的域名，通常作为日期和时间的函数。除非对DGA进行逆向工程，否则这种技术很难阻止僵尸网络使用的域名，因为这些域名的生命周期很短。后者的特征是与特定域名相关联的IP地址的不断变化。在这种情况下，恶意软件会建立一个快速流量服务网络（Fast Flux Service Network，FFSN），由分配给一个给定域名的数百甚至数千个IP地址组成。当查询这样的域时，它被解析为这些频繁更改的ip，从而保护了恶意服务的真实位置。通常，大量的IP地址池并不是内容请求的最终目的地，它们只是中途停留，也就是说，到达最终目的地可能要经过几次停留。双通量网络（Double-flux networks）是一种更复杂的技术，提供了额外的冗余层。具体来说，无论是DNS的A记录集，还是恶意域名的权威NS记录，都会以轮询的方式不断地改变，并广告到快速通量服务网络中。显然，这些技术也可以组合使用，提供FQDNs和IP地址之间的多对多关系。　　尽管这些技术符合DNS协议的规范，恶意软件还是以各种方式滥用它们，以提高其服务器的移动性，从而提高其弹性。好消息是，这些技术在DNS数据中留下了痕迹。这样的跟踪为研究人员提供了重要的线索，以便开发检测方法，利用DNS流量观察提供的独特视角来考虑域名-ip映射的变化。

3 恶意域名检测方法

本节主要介绍一些通过DNS检测恶意域名的方法，主要分为两个方面：

- 特征的选取
- 检测的方法

3.1 特征

特征提取（又称特征工程）是一项具有挑战性的任务，它对检测方法的质量(准确性和鲁棒性)有很大影响。精心设计的特性对方法的成功有很大的帮助，相反，糟糕的特性甚至可能毁掉好的检测算法。另一方面，即使一个特征具有很好的预测能力，因此检测精度较高，但如果它很容易被攻击者伪造，那么依赖它的检测方法的鲁棒性就会较低。因此，在选择特征时，成功的检测方法必须考虑到准确性和鲁棒性的平衡。

很少有方法简单地解析来自DNS流量的资源记录，并在它们出现时使用来自特定字段的值。相反，在使用这些原始值进行检测之前，可以对它们进行多种处理(平均值、标准差、最大值、最小值、速率、离群值等)。此外，可以使用DNS环境之外的外部数据来丰富初始数据集。有些方法在检测方法中使用DNS数据之前，需要将其转换为不同的数据结构，例如图。

具体可以考虑三个维度来区分特征:

- Internal vs Contextual features
- DNS dataset Dependent vs Independent features
- Mono vs Multi domains features

3.1.1 Internal vs Contextual features

Internal和contextual feature的区别就像主动特征和被动特征。

internal features：这些特征可以从DNS资源记录中单独提取，不需要外部附带的数据源，然而这些特征可能在被放入检测方法之前就被转换。例如“domain average TTL value”。另外从域名中提取的特征也属于internal features，这在DGA检测中很流行。此外，基于图的关联性特征也通常由这类特征建立

Contextual features：contextual feature是结合DNS和外部资源建立起来的，例如计算“一个域的ip所述的ASNs数量”，这就需要IP-AS映射信息。

一些上下文特性需要查询由攻击者控制的资源。例如，有的工作使用**domain web presence**作为特征之一，即每次当一个新域名出现在他们的列表中，他们就会检查该域名是否有可用的网页。另一种特殊类型的上下文特性使用DNS数据本身来充实内容。例如，检查域是否有一个相关联的MX记录。因此，使用这种类型的特性可能会警告攻击者该域正在受到监视。

尽管internal feature的使用有许多好处，主要是简单性，但是它们捕获信息的能力是有限的，而漏掉的信息又是区分恶意和良性域名的重要因素。例如，**给定域的注册时间**通常是一个非常重要的特性，但它不能仅从DNS数据获得。有时攻击者在恶意活动开始前几个月就大量注册了域名，检测这些注册模式使研究人员能够主动检测恶意域名。但是，国家代码顶级域名(ccTLD)通常无法获得这些信息，因为ccTLD注册中心很少提供对它们的区域文件的访问。同样，由于可访问性有限、隐私问题、成本过高等原因，其他一些有用的信息也很难获得。

3.1.2 DNS Dataset Dependent vs. Independent features

区分**受特定DNS数据集影响的特征**（DDD）和**独立于DNS数据集的特征**（DDI）是很重要的。

单一依赖DDD特征的方法的性能受到所选数据集的高度影响。因此，为了评估这些方法的质量，执行跨数据集验证是非常重要的，使用来自不同地方、不同时期、不同规模的数据集等。相反，依赖于DDI特性的方法更加稳定，并且可以在不同的环境中平等地运行。

DDD：例如，在观察期间，“观察到分配给一个域的IP地址的数量”是一个DDD特性，因为它的值取决于特定的数据集。类似地，Khalil等人[3]在域名之间建立关联时所使用的“一对域名共享的可观察到的常见ASNs的数量”特征也依赖于数据集，因为使用这种关联建立的图取决于数据集的收集地点和方式。

DDI：另一方面，“流行搜索引擎中特定域名的命中数”是一个DNS数据集独立的特性，因为它不依赖于所选DNS数据集的内容。类似地，一个域名的“n-gram”分布是DNS数据集独立的，因为它不依赖于所选择的数据集。

3.1.3 Mono vs. Multi Domains Features

Mono domain features：对每个单独的域名都提取这个特征，例如“托管一个给定域的国家数量”，使用这类特性的优点之一是这些方法依赖于它们可以在完全不同的数据集上被训练和操作。

Multi domain features：在一对域上计算的域关联特征被用于许多基于图和聚类方发。依赖于多域特性的方法通常需要更大的数据集才能正常工作。实际上，两个任意域之间的关联可能是间接的，因此为了建立这样的关联，也应该考虑中间域，以使方法正常工作。

3.2 检测方法

3.2.1 基于知识的方法（Knowledge Based methods）

为了检测涉及恶意活动的域名，基于知识的方法依赖于专家的洞察。这样的洞察可以通过测量研究来获得，它探索与恶意域活动相关的异常。例如，属于一个恶意软件家族的恶意域名往往会同时被查询。因此，通过测量已知恶意域和未知域之间的共同发生程度，并将结果与某个阈值进行比较，就有可能检测到新的恶意域。在DNS查询方面，网络中的机器人群体往往表现出类似的模式，而DNS基础设施无法解决这些问题。

不幸的是，这种方法有其局限性。专家可能会有意或无意地产生偏见。此外，专家通常不擅长分析高维数据，因为人类很难掌握从数据中提取的特征之间的所有关联和依赖关系。

3.2.2 基于机器学习的方法（Machine Learning Based methods）

基于机器学习的方法可以分为三个子类：

- 监督学习算法（Supervised learning）
- 半监督学习算法（Semi-supervised learning）
- 无监督学习算法（Unsupervised learning）

1）监督学习算法

监督学习算法。这些算法要求得到完整的训练集，即一个样本数据对应的每个特征向量必须与一个代表该样本所属类别的标签相关联。也就是说训练集中的每个域名必须明确地标记为恶意或良性。然而，考虑到实验训练期间观察到的典型域的数量，几乎不可能对所有域进行正确的标记。因此，通常在监督学习方法中，训练数据集会被裁剪，只包含那些标记为高置信度的数据。监督机器学习方法在这一领域非常受欢迎，因为它们简单，自动选择最相关的特征和有效性。事实上，依赖于这种方法的研究人员只需要从原始数据中提取特征，然后在标记数据集上训练分类器。将经过训练的分类器应用于新数据是直接的。例如，DomainProfiler[2]使用55个特征提取考虑相关IP地址和域名。应用随机森林算法发现滥用域。Antonakakis等人的[4]也使用了随机森林。然而，在本工作中，特征提取从被动DNS数据的权威名称服务器。

监督学习方法有几个缺点。首先，他们需要一个带标签的数据集来训练。由于DNS和黑名单数据的变化无常，很难获得完整和完全正确的数据集。手工标注是耗时的，并且不会产生大量的训练数据集。同样，使用来自不同白名单和黑名单的信息进行自动标记也容易导致不正确的数据纳入。其次，监督学习方法更容易对特定数据集进行过拟合。如果标记的数据集是有偏差的，这可能会无意中导致分类器学习到不正确的特征变量分布。此外，在真正的DNS数据中，只有一部分域可以用标签分配。在实际应用中，绝大多数样本没有进行标记，因此无法参与分类器学习的过程，使得训练数据集不一致。

2）半监督学习算法

半监督学习算法从有标签和没有标签的数据中学习。未标记数据帮助机器学习算法修改从标记数据集获得的假设。然而，这种算法的采用往往是相当具有挑战性的，需要研究者付出更多的努力。基于图的推理方法是这类方法中最流行的方法之一。例如，Manadhata等人[5]将belief propagation算法应用到从企业HTTP代理日志中提取的主机域图中检测到恶意域。

3）无监督学习算法

引入无监督学习方法不仅是为了消除对标记数据集的依赖。无监督学习方法，又称**聚类技术**，只使用数据的内部属性自动将域划分为聚类。理论上，通过对恶意域和良性域表现出完全不同行为的特征的仔细选择，可以使聚类算法将提供的样本划分为两个聚类。然后，研究人员决定哪些集群包含恶意和良性域名。然而，一些方法并没有遵循这条路径，而是更进一步。它们将与不同恶意行为相关的多个维度的域分组，然后通过将已识别的组相互关联来选择恶意域的集群。　　尽管这些方法在独立于标记数据方面有明显的优势，但它们在文献中并不常见。此外，考虑到标签数据集通常存在于这一领域（尽管既不完整也不完全正确），研究人员更喜欢探索更容易使用的监督和半监督方法。

3.2.3 混合方法（Hybrid approaches）

尽管可以根据所提供的分类对单一检测算法进行分类，但现实中大多数现有的方法是混合的，并使用几种不同类型的算法来产生结果。可以是机器学习技术的组合。例如，在Notos系统[6]中就使用了这种方法。它在第一阶段训练了5个元分类器，使用监督学习技术来评估一个域与预定义的一组域（Popular、Common、Akamai、CDN、动态DNS）的紧密程度。然后将计算出的贴近度分数作为第二阶段监督学习算法的特征。Oprea等人[7]将半监督方法（belief propagation）与监督学习算法（linear regression）相结合。机器学习和基于知识的方法的混合也被用于该领域。例如，Segugio系统[8,9]结合了基于图的预滤波和监督机器学习。

4 TLD异常检测

本节主要简单介绍TLD上的一些异常检测相关工作。

4.1 TLD DDos检测

DDos攻击严重影响网络的正常运行，尤其是DNS服务的运行。由于DDoS攻击的破坏性影响，实现能够及时检测并减轻它们的有效对策（特别是针对权威DNS服务器的对策）非常重要。针对DNS服务器的最流行的DDoS攻击之一是DNS放大。放大攻击的主要目标是耗尽受害者的带宽，耗尽它的CPU或内存。

通常，放大效应通过两种方式来实现:**消息的数量**或**消息的大小**。DNS放大采用后一种方法;因此，攻击者欺骗目标的IP并发送一个小的DNS请求，目的是使该IP成为更大响应的接收方。

Trejo L A[10]等人设计了DNS-ADVP，进行TLD服务器上的DDos异常检测，并以可视化的方式展示。

Wang Z[11]等人采用协方差分析的方法在中国顶级域名服务器上检测中国5.19事件的DNS流量异常。对不同时间片长度的协方差变化进行归一化、扩展和平均，增强检测的鲁棒性。基于协方差变化异常的聚类分析，检测特征异常。提出了一种初始聚类选择技术来降低算法的复杂度，并对其性能进行了分析。定义了瞬态异常和时间跨度异常，并给出了一种有效的实时逼近算法。对5.19事件的流量检测结果表明，该方法能够准确检测出网络异常。

4.2 TLD Abuse

顶级域名(TLD)注册，允许注册商出售高数量的域名给专业垃圾邮件制造者和恶意软件操作员，本质上帮助和唆使了在互联网上的滥用行为。一些注册商和经销商明知卖高数量的域名，这些行动者的利润，和许多注册没有做足够的阻止或限制这无尽的供应域名。下图列出了截止至2020年9月5日的TLD滥用top 10。

The 10 Most Abused Top Level Domains

As of 05 September 2020 the TLDs with the worst reputations for spam operations are:

1	**.email**	**Badness Index: 4.52**	Domains seen: 10,104 Bad domains: 5,318 (**52.6%**)
2	**.fit**	**Badness Index: 4.10**	Domains seen: 9,185 Bad domains: 4,479 (**48.8%**)
3	**.run**	**Badness Index: 3.25**	Domains seen: 2,186 Bad domains: 1,024 (**46.8%**)
4	**.gq**	**Badness Index: 2.89**	Domains seen: 14,281 Bad domains: 4,857 (**34.0%**)
5	**.tk**	**Badness Index: 2.80**	Domains seen: 45,772 Bad domains: 13,468 (**29.4%**)
6	**.ml**	**Badness Index: 2.76**	Domains seen: 19,457 Bad domains: 6,159 (**31.7%**)
7	**.cn**	**Badness Index: 2.50**	Domains seen: 385,017 Bad domains: 84,929 (**22.1%**)
8	**.work**	**Badness Index: 2.34**	Domains seen: 53,040 Bad domains: 13,117 (**24.7%**)
9	**.cf**	**Badness Index: 2.33**	Domains seen: 22,056 Bad domains: 5,915 (**26.8%**)
10	**.loan**	**Badness Index: 2.06**	Domains seen: 248 Bad domains: 109 (**44.0%**)

报告[12]分析了2016年gTLD滥用的统计情况。检查全球DNS中的恶意行为，并比较新旧通用顶级域名的滥用率。报告结合了来自许多来源的数据集，包括区域文件、域WHOIS信息、活动测量获得的数据，以及代表恶意软件、钓鱼和垃圾邮件的11个有信誉的黑名单。报告发现新的通用顶级域名已经影响了旧通用顶级域名的垃圾邮件数量，在新的通用顶级域名中，滥用域名并没有增加恶意注册总数，相反，旧通用顶级域名中的恶意注册数量有所下降。在2016年最后一个季度，旧的g顶级域名的垃圾域名率为56.9 / 10000，而新的g顶级域名的垃圾域名率为526.6 / 10000，几乎高出一个数量级。此外，报告还分析了所收集的安全指标与新通用顶级域名的结构特性和滥用之间的关系，在通用顶级域名的水平上。结果表明，滥用计数主要与更严格的注册策略相关。我们的发现表明，一些新的通用顶级域名已经成为恶意行为者日益增长的目标。对垃圾邮件黑名单的分析显示，2016年第四季度，大约有三分之一的新注册通用顶级域名没有发生过一次事故，而Spamhaus将15个新通用顶级域名中至少10%的注册域名列入了黑名单。

参考资料

[1] 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.

[2] 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 (2016), 491–502.

[3] 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

[4] Manos Antonakakis, Roberto Perdisci, Wenke Lee, Nikolaos Vasiloglou, II, and David Dagon. 2011. Detecting Malware Domains at the Upper DNS Hierarchy. In Proceedings of the USENIX Security Symposium. 27–27.

[5] 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.

[6] 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.

[7] A. Oprea, Z. Li, T. F. Yen, S. H. Chin, and S. Alrwais. 2015. Detection of Early-Stage Enterprise Infection by Mining Large-Scale Log Data. In Proceedings of the Annual IEEE/IFIP International Conference on Dependable Systems and Networks. 45–56.

[8] B. Rahbarinia, R. Perdisci, and M. Antonakakis. 2015. Segugio: Efficient Behavior-Based Tracking of Malware-Control Domains in Large ISP Networks. In Proceedings of the Annual IEEE/IFIP International Conference on Dependable Systems and Networks. 403–414.

[9] Babak Rahbarinia, Roberto Perdisci, and Manos Antonakakis. 2016. Efficient and Accurate Behavior-Based Tracking of Malware-Control Domains in Large ISP Networks. ACM Transactions on Privacy and Security 19, 2 (Aug. 2016), 4:1–4:31.

[10] Trejo L A, Ferman V, Medina-Pérez M A, et al. DNS-ADVP: A Machine Learning Anomaly Detection and Visual Platform to Protect Top-Level Domain Name Servers Against DDoS Attacks[J]. IEEE Access, 2019, 7: 116358-116369.

[11] Wang Z, Tseng S S. Anomaly detection of domain name system (DNS) query traffic at top level domain servers[J]. Scientific Research and Essays, 2011, 6(18): 3858-3872.

[12] Korczy'ski M, Wullink M, Tajalizadehkhoob S, et al. Statistical Analysis of DNS Abuse in gTLDs Final Report[R]. Technical Report. https://www. icann. org/en/system/files/files/sadag-final-09aug17-en. pdf, 2017.