Computer Science and Application 计算机科学与应用, 2021, 11(5), 1349-1356
Published Online May 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.115137

Hans汉斯

基于知识图谱的DNS Query Flood攻击检测研究

夏雪菲，蒋　铜，李天尧，顾海艳

江苏警官学院计算机信息与网络安全系，江苏　南京
Email: ghy7388@126.com

收稿日期：2021年4月20日；录用日期：2021年5月14日；发布日期：2021年5月24日

摘　要

针对具有成本低廉、破坏性大、防御困难特性的DNS Query Flood攻击，本文构建UDP请求的知识图谱。基于攻击者通过发送大量伪造源IP地址的小UDP包冲击DNS服务器实施攻击的原理，本文通过计算客户机对服务器的正常访问频率确定发现DNS Query Flood攻击的流量阈值，基于加州大学洛杉矶分校的DNS Query Flood攻击实验数据集，利用Neo4j可视化分析检验通过阈值判定攻击的准确性。结果表明，阈值检测的方法在混合流量中对攻击流量的检测成功率高达95.04%。

关键词

DDoS攻击，DNS Query Flood攻击，知识图谱，流量检测

Research on DNS Query Flood Attack Detection Based on Knowledge Graph

Xuefei Xia, Tong Jiang, Tianyao Li, Haiyan Gu

Department of Computer Information and Cyber Security, Jiangsu Police Institute, Nanjing Jiangsu
Email: ghy7388@126.com

Received: Apr. 20th, 2021; accepted: May 14th, 2021; published: May 24th, 2021

Abstract

Aiming at the low-cost, destructive, and difficult-to-defense DNS Query Flood attack, this paper constructs a knowledge graph of UDP requests. Based on the principle that the attacker attacks the DNS server by sending a large number of small UDP packets with forged source IP addresses, this

文章引用: 夏雪菲, 蒋铜, 李天尧, 顾海艳. 基于知识图谱的 DNS Query Flood 攻击检测研究[J]. 计算机科学与应用, 2021, 11(5): 1349-1356. DOI: 10.12677/csa.2021.115137

article calculates the normal access frequency of the client to the server to determine the traffic threshold for discovering the DNS Query Flood attack. Based on the DNS query flood attack experiment data set of the University of California, Los Angeles, Neo4j was used to visualize the analysis to test the accuracy of the attack by threshold. The results show that the threshold detection method has a detection success rate of 95.04% for attack traffic in mixed traffic.

Keywords

DDoS Attack, DNS Query Flood Attack, Knowledge Graph, Traffic Detection

1. 引言

网络技术快速发展的同时，各类网络安全问题层出不穷。根据 2021 年 2 月公布的第 47 次《中国互联网络发展状况统计报告》，2020 年，国家信息安全漏洞共享平台收集整理信息安全漏洞 20,721 个，较 2019 年同期增长 28% [1]。在各类漏洞威胁中，DDoS (Distributed Denial of Service，分布式拒绝服务)攻击以容易实施、难以防范、难以追踪等特点成为最难以防范和抵御的网络攻击类型之一，其中 DNS 反射放大攻击(DNS Amplification Attack)又是其中较为常见的方式。DNS 攻击主要是利用 UDP 及 DNS 协议的安全漏洞，以 DNS 服务器作为攻击流量的反射放大器，实现对目标主机的 DDoS 攻击。如果 DNS 服务器被攻击，将会导致大量网站无法被正常访问，对网络安全造成严重威胁。由于计算机根区文件是公开可访问的，因此顶级域的权威服务器很容易被用来作为反射放大器。

2016 年美国 Dyn 公司遭到了一次与以往不同的 DDoS 攻击，此次攻击通过 Mirai 病毒，将攻陷的物联网设备作为“肉鸡”，使得 Dyn 服务器一下子收到了成千上万条请求导致瘫痪。经调查发现，这是通过 DNS 的 TCP 和 UDP 数据包发起的一次攻击，据称此次攻击手段还不是很成熟，但已经对 Dyn 内部造成巨大伤害。DNS 反射放大攻击是一种具有巨大攻击力的 DDoS 攻击方式之一，其危害大、成本低、溯源难，被黑色产业从业者所喜爱。攻击者只需要付出少量的代价，即可对需要攻击的目标产生巨大的流量冲击，对网络带宽资源、连接资源和计算机资源造成巨大的压力。Dyn 公司的 DNS 服务器受到 DNS 反射放大攻击后，导致美国大范围断网，事后的攻击流量分析显示，DNS 反射放大攻击是造成美国大范围断网的 DDoS 攻击的主力之一。

本文通过构建 UDP 请求的知识图谱，提出运用客户机与服务器的访问频率阈值的方法来检测、发现 DNS Query Flood 攻击。

2. 研究现状

2.1. DNS Query Flood 攻击研究现状

当前 DNS 服务器由于其本身缺陷导致安全事故频发，针对 DNS Query Flood 攻击已有不少研究。严芬等人通过分析一段时间内 DNS 流量变化情况，得到该时间段内域名解析的成功率，计算出信息熵值，从而判断 DNS 服务器是否出现异常，进一步利用滑动窗口算法计算源 IP 地址的信息熵值，与正常情况比较得出结论，并通过实验验证了该检测方法的有效性[2]。NM SAHRI 和 Koji OKAMURA 提出引入一个服务器控制器，负责向先前请求 DNS 服务的每个主机发送身份验证包，通过验证请求客户端网络回复

的“身份验证包”，来确定 DNS 查询是否为合法的查询或攻击包。经实验证实，该方法可以有效阻止来自僵尸网络的所有 DNS 查询[3]。Roberto Alonso 等人提出了一种新的递归抽取 DNS 流量来检测 DNS Query Flood 攻击的方法，建立了一种基于异常的检测机制，给定 DNS 使用的时间窗口，利用 DNS 攻击特征捕获 DNS 社群结构，通过观察 DNS 社群结构的变化来判断是否受到 DNS 攻击[4]。

本文通过对客户机与 DNS 服务器之间的 UDP 数据包特征的研究，分析 DNS Query Flood 攻击流量的判定阈值，以尽早发现可能的攻击行为、减少损失。

2.2. 知识图谱简介

知识图谱是结构化的语义知识库，以符号形式描述物理世界中的概念及其相互关系。其基本组成单位是“实体－关系－实体”三元组，以及实体及其相关属性－值对，实体间通过关系相互联结，构成网状的知识结构[5]。

相对于现有的字符串模糊匹配方式而言，知识图谱一方面能够通过推理实现概念检索从而改变现有的信息检索方式；另一方面能以图形化方式向用户展示经过分类整理的结构化知识，从而更方便人们对知识体系的理解和规律的发现。

实现知识图谱的相关图数据库软件，可将信息数据化、网格化，用可视化的方式呈现大量数据的关系，极大提高数据的可访问性。在图数据库工具中，通过编程可对数据进行多维度的探索，构建数据之间深层次的关系，以提高人工智能的推理能力。知识图谱的方法在金融数据分析中已广泛应用，实现了从简单的量化模型走向更为复杂的价值判断和风险评估的应用，逐步把经验变成可重用、可演化、可验证、可传播的知识模型，实现了数据到智能化决策的提升。

本文将知识图谱方法应用于网络流量分析，以发现 DNS Query Flood 攻击，这是一种新的应用尝试。

3. 基于流量知识图谱的 DNS 攻击检测方法

3.1. DNS Query Flood 攻击基本原理

DNS 服务器进行域名解析时，是通过 UDP 报文进行通信的，而 UDP 协议是一种无连接的服务，在 DNS Query Flood 攻击中，攻击者可发送大量伪造源 IP 地址的小 UDP 包，通过利用大量 UDP 小包冲击 DNS 服务器。

客户端与 DNS 服务器的 UDP 交互流程如下：

① 首先服务器会查找是否有对应缓存；

② 当没有缓存信息，且该域名无法由该 DNS 服务器进行解析时，DNS 服务器会向上层 DNS 服务器递归查询域名信息，直到全球互联网的 13 台根 DNS 服务器，若仍然无法解析，则没有回应的报文。

此交互过程如图 1 所示。

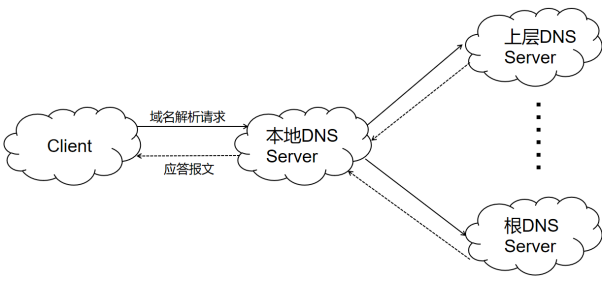

Figure 1. UDP-based domain name resolution process
图 1. 基于 UDP 的域名解析过程

DNS Query Flood 攻击的原理如下：根据图 1 客户端与 DNS 服务器的 UDP 交互流程所述，设有用户机 A、服务器 B、上层服务器 C。当攻击发生时，三者之间的通信过程如下：

1) 用户机 A 对服务器 B 发动 DNS Query Flood 攻击；

2) 服务器 B 向其上层服务器 C 进行递归查询。

利用知识图谱概念对 DNS Query Flood 攻击进行图形描述，如图 2 所示。

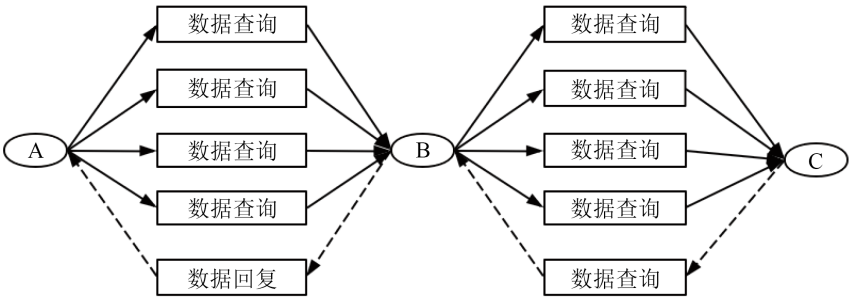

Figure 2. Knowledge graph of DNS Query Flood Attack
图 2. DNS Query Flood 攻击的知识图谱

3.2. 流量检测及确定阈值方法

DNS Query Flood 攻击难以应对的主要原因在于这是一种攻击流量和正常流量混合在一起的攻击方式。在对于流量进行检测时，如何区分正常流量和攻击流量成为一个难点。

通过图 2 可直观地发现：在 DNS Query Flood 攻击过程中，用户机 A 与服务器 B 之间、服务器 B 与服务器 C 之间的大部分网络流量都是用户机 A 对服务器 B、服务器 B 对服务器 C 的单方面数据查询和关联建立。

因此，本文研究提出通过构建历史正常流量知识图谱，对知识图谱中的正常流量进行分类，抽出每一条流量的时间戳，以此计算出正常的源地址对 DNS 服务器的访问频率，找出其中的最大值作为阈值。确定阈值之后，在实际网络管理过程中，即可以源地址为基础，对流量进行分类，计算出源地址对 DNS 服务器的访问频率，并与阈值进行比较，小于阈值的则为正常流量，大于阈值的即作为攻击流量的可疑对象，可进一步跟踪处理。

4. DNS Query Flood 检测实验流程

4.1. 实验数据的选取及预处理

本文分析的 DDoS 攻击流量数据集，来自加州大学洛杉矶分校计算机科学系的网络研究实验室和高级系统研究实验室的数据(https://lasr.cs.ucla.edu/ddos/traces/)。该数据通过在测试机上运行 TFN 攻击工具并附加在另一台测试机而生成，同时也存在一些合法流量。实验中，分别抽取了两份攻击流量和两份正常流量作为本文研究的对象，将它们记作 attacktrace1，attacktrace2，trace1 和 trace2。因为 DNS Query Flood 攻击中，攻击者可发送大量伪造源 IP 地址的小 UDP 包，通过利用大量 UDP 小包冲击 DNS 服务器，所以本文主要研究其中 UDP 数据包，能更好地对 DNS Query Flood 攻击流量进行检测、防御。

数据预处理时，首先将获取到的格式为 txt 文件转化为 csv 格式。

然后根据官网给出的关于 UDP 封包的列名信息进行字段处理，生成的数据如表 1 所示，表中各段含义如下：发送数据包的时间 Packet_TIME；掩盖数据包源的 IP 地址 IP_from；掩盖数据包目标的 IP 地址 IP_to；原始源端口 PORT_from；原始目标端口 PORT_to；数据包的长度 LENGTH。

Table 1. Preprocessed data table
表 1. 预处理后的数据表

序号	Packet_TIME	IP_from	IP_to	PORT_from	PORT_to	LENGTH
1	0.015674	1.1.139.167	1.1.236.8	9997	3	1001
2	0.01574	1.1.139.92	1.1.236.8	9996	4	1001
3	0.015824	1.1.139.210	1.1.236.8	9995	5	1001
4	0.015945	1.1.139.71	1.1.236.8	9994	6	1001
5	0.016028	1.1.139.142	1.1.236.8	9993	7	1001
6	0.01612	1.1.139.131	1.1.236.8	9992	8	1001
7	0.0162	1.1.139.84	1.1.236.8	9991	9	1001
8	0.016283	1.1.139.80	1.1.236.8	9990	10	1001
……						

4.2. 核心代码

本实验中采用 Python 语言中的 py2neo 模块和 Neo4j 可视化数据库进行构造知识图谱。代码整体可分为三个部分，核心部分就是根据正常流量数据计算其阈值。

1) 通过 py2neo 模块连接 Neo4j 数据库，并导入数据。

2) 创建函数 create_rel，此函数主要功能是以导进来的数据为基础，构建知识图谱。首先以数据中的 IP_from (原 IP)和 IP_to (目的 IP)作为图谱中的两个节点，并把 Packet_TIME (数据的时间戳)作为 IP_from 的一个属性，visit 作为连接两个节点的联系，依次构建 UDP 数据包知识图谱。

3) 以第二步构建好的知识图谱为基础，对 IP_from 和其时间属性进行统计，并计算 IP_from 对于 DNS 服务器的访问频率。具体流程为：

① 通过 NodeMatcher 查询 Neo4j 中相同的 IP_from 地址，并统计其个数，并将其时间属性放进已经定义好的列表 attr。

② 对 IP_from 的访问频率进行计算，在创建好的 attr 列表中寻找最大时间和最小时间并求其差值，结合前面统计的 IP_from 个数计算出 IP_from 对服务器的访问频率。

③ 将这部分代码放进一个 for 循环中，计算每一个 IP_from 的访问频率，并将每一个频率放进列表 Threshod 中，找出 Threshod 中的最大值，作为本实验的阈值。

4.3. 阈值的计算

实验中随机抽取攻击流量 5000 条和正常流量 722 条，以及 2000 条混合流量(其中 1000 条正常流量，1000 条攻击流量)。

第一步：构建知识图谱。运用创建函数 create_rel 构建出 DNS 访问图谱，构建的图谱如图 3 所示。

第二步：计算检测阈值。查找出知识图谱中每个 IP_from 的个数，并将同一 IP_from 的时间属性放到列表 diff 中，计算最大时间差，结合 IP_from 的个数来计算相应 IP 每分钟访问的次数，之后将不同 IP 每分钟访问次数加入到列表 Threshod 中，找出列表 Threshod 中的最大值作为检测阈值为 28.75721。

第三步：检验阈值的有效性。将 5000 条攻击流量数据包，按照第一步工作步骤构建攻击流量的知识图谱，之后计算攻击流量每个 IP 每分钟的访问次数，利用计算出的攻击流量和第二步得出的阈值进行比较，高于阈值为攻击流量，低于阈值的为正常流量。经过实验检验，发现利用阈值 28.75721 可成功检测出绝大部分攻击流量，正确率高达 95.04%。

Figure 3. DNS access map visually constructed by Neo4j
图 3. Neo4j 可视化构建的 DNS 访问图谱

4.4. 攻击流量的检测

为了进一步检验本文方法的可靠性，在数据集中另外选取一部分攻击流量和正常流量，将它们混合在一起构成新的数据集合来进行测试。此混合数据集共计 2000 条数据，包括 1000 条正常流量和 1000 条攻击流量，如表 2 所示。

Table 2. Example of mixed traffic data
表 2. 混合流量数据示例

序号	IP_from	IP_to	数量	类型
1	1.1.139.83	1.1.236.54	315	攻击
2	1.1.139.104	1.1.236.8	215	攻击
3	1.1.139.12	1.1.236.27	225	攻击
4	1.1.139.157	1.1.236.56	246	攻击
5	1.1.139.25	1.1.12.7	196	正常
6	1.1.139.177	1.1.12.20	345	正常
7	1.1.139.202	1.1.12.66	228	正常
8	1.1.139.66	1.1.12.96	231	正常
……				

利用 python 编写的程序，2000 条混合流量计算出的各 IP 地址的访问频率检测值如表 3 所示。

Table 3. System resulting data of standard experiment
表 3. 标准试验系统结果数据

序号	IP_from	IP_to	访问频率
1	1.1.139.83	1.1.236.54	39.80007
2	1.1.139.104	1.1.236.8	29.66331
3	1.1.139.12	1.1.236.27	36.41097
4	1.1.139.157	1.1.236.56	31.05044
5	1.1.139.25	1.1.12.7	24.73938
6	1.1.139.177	1.1.12.20	28.27390
7	1.1.139.202	1.1.12.66	24.06290
8	1.1.139.66	1.1.12.96	24.37952
……			

表 3 检测结果在 Neo4j 中的图表现形式，如图 4 所示。

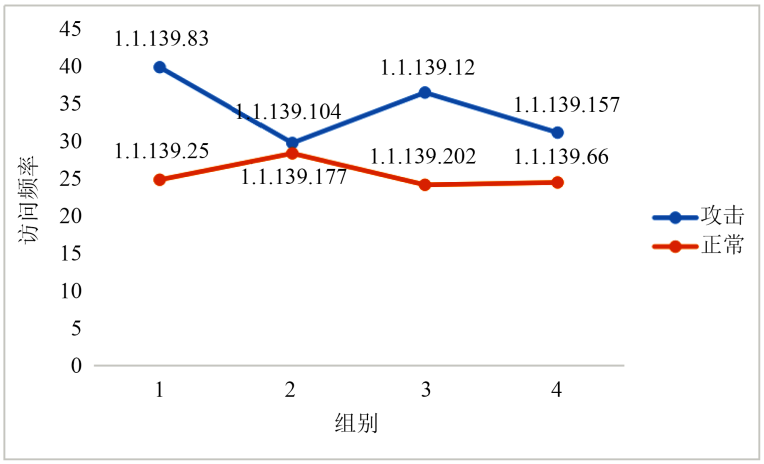

Figure 4. Access frequency knowledge graph
图 4. 访问频率知识图谱

从图 4 可以看出攻击数据和正常数据的访问频率分别处于以阈值 28.75721 为分界线的不同区间，这表明利用本文方法可成功区分出 DNS Query Flood 攻击的攻击流量和正常流量。

5. 结束语

DNS Query Flood 攻击会造成服务器资源被攻击者占用，导致 DNS 服务器无法对其他正常用户提供服务。由于 DNS 具有逐级向上查询的特性，当 DNS Query Flood 攻击出现时，域名服务器与上级域名服务器之间的通信次数急剧增加，最终可能会使上级域名服务器资源也被耗尽，导致更大范围的网络崩溃，这是一种危害极大的攻击手段[6]。为此，本文提出了通过构建网络流量知识图谱，运用正常的源地址对 DNS 服务器的访问频率阈值来检测 DNS Query Flood 攻击的方法，并在加州大学洛杉矶分校分布网络研究实验室采集的 DNS Query Flood 攻击流量数据集上通过 Neo4j 进行可视化分析。实验结果表明该方法确定的阈值对于攻击流量的检测成功率高达 95.04%，验证了该检测方法的有效性和可靠性。

当然对于一些大型的 DNS 服务器而言，由于每天会有各种新增的访问，仅仅运用历史数据可能会影响阈值的准确性，可以在本研究的基础上结合机器学习的方法，结合历史访问数据实时进行阈值的计算和调整，以增加算法的自适应能力[7]，可进一步提高其可靠性。

基金项目

江苏省高等学校大学生实践创新训练计划基于典型案例的网站安全问题研究项目，项目编号：202010329035Y。

参考文献

[1] 中共中央网络安全和信息化委员会办公室, 中华人民共和国国家互联网信息办公室, 中国互联网络信息中心. 中国互联网络发展状况统计报告[R], 2020.

[2] 严芬, 丁超, 殷新春. 基于信息熵的 DNS 拒绝服务攻击的检测研究[J]. 计算机科学, 2015, 42(3): 140-143.

[3] Sahri, N.M., Okamura, K. and Auth, C. (2016) Protecting DNS Application from Spoofing Attacks. *IJCSNS International Journal of Computer Science and Network Security*, **16**, 125-134. https://doi.org/10.1145/2935663.2935666

[4] Alonso, R., Monroy, R. and Trejo, L.A. (2016) Mining IP to Domain Name Interactions to Detect DNS Flood Attacks

on Recursive DNS Servers. *Sensors*, **16**, 1311. https://doi.org/10.3390/s16081311

[5] 刘峤, 李杨, 杨段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.

[6] 王文蔚, 肖军弼, 程鹏, 张悦. 基于 SDN 的 DDoS 攻击防御系统[J]. 计算机与现代化, 2021(2): 117-118.

[7] 陈佳. 基于知识图谱的 DDoS 攻击源检测研究[J]. 信息安全研究, 2020(1): 91-96.