doi:10.3969/j.issn.1000-436x.2014.z1.002

基于 IPFIX 的 DNS 异常行为检测方法

马云龙，姜彩萍，张千里，王继龙
（清华大学 信息化技术中心，北京 100084）

摘 要：提出了一种基于 IPFIX（IP 数据流信息输出）网络流量数据准确检测可疑和异常 DNS、识别 DNS 流量放大攻击行为的算法。该算法已在清华大学校园网实际部署运行，能够有效检测到校园网内部 DNS 的异常行为并发送告警信息，从而及时控制攻击行为，实现异常流量的及时监测和预警。
关键词：异常行为；网络安全；IPFIX；流量分析
中图分类号：TP393 **文献标识码**：A **文章编号**：1000-436X(2014)Z1-0005-05

DNS abnormal behavior detection based on IPFIX

MA Yun-long, JIANG Cai-ping, ZHANG Qian-li, WANG Ji-long
(Information Technology Center, Tsinghua University, Beijing 100084, China)

Abstract: An algorithm based on IPFIX network flow data is proposed. By using proposed algorithm, suspicious and abnormal DNS will be detected accurately, and DNS traffic amplification attack will be distinguished rapidly. This algorithm has been applied in the Tsinghua University campus network. In our practice, DNS abnormal behaviors have been detected and alarm information has been sent to administrators. Thus, abnormal attack behaviors are restrained in time, and the monitoring and warning for abnormal traffic are all realized.
Key words: abnormal behavior; network security; IPFIX; traffic analysis

1 引言

伴随着计算机网络日新月异的发展壮大，针对计算机网络中 DNS 的恶意攻击一刻也没有停止，网络攻击的频率和危害也越来越大。近来多起针对 DNS 的 DDoS 攻击案例被报道[1~4]，这些攻击无一例外都造成了巨大的经济损失。由于 DDoS 攻击利用现有互联网协议漏洞进行，所以无法完全避免针对 DNS 的 DDoS 攻击。因此对 DNS 的异常行为进行实时检测并且对攻击主机进行实时阻断或者防御是当前网络安全领域研究的重要前沿问题。

本文在研究目前国内外针对 DNS 流量攻击的分析和检测方法的基础上，以清华大学校园网络出口路由器的 IPFIX 流数据为基础数据源，筛选出 DNS 流数据作为分析对象，提出了一种有效区别 DNS 异常行为的方法。

2 DNS 流量放大攻击过程

利用 DNS 进行流量放大攻击通常使用大量的僵尸主机同时向 DNS 主机发送海量请求，利用 DNS 流量放大的特点将流量导向目标主机。在一些应用中，请求的数据分组往往可以获得超过请求大小几十倍以上的回复数据流量，由于计算机本身是不能检测收到的数据分组是否合法，最后大量的数据分组会导致受害者主机无法正常工作。DNS 网络是一个分级的分布式系统，文献[5~7]对域名系统进行了详细的描述。这种攻击首先会找到开放的 DNS，这些开放的 DNS 一般安全措施不完善，可以被任意 IP 地址访问和使用，在攻击行为中的角色就是流量放大器，攻击者会控制网络上的僵尸主机发送大量的伪造数据分组到这些开放的 DNS 上；然后，这些开放的 DNS 会开始递归地寻找目标机器，在此过程中，数据分组的大小被放大；最后，开放的

收稿日期：2014-10-15

DNS 将放大后的流量发送给攻击目标。

从以上攻击过程不难看出，利用 DNS 作为流量放大器攻击之所以可行，主要有以下几个原因。1) 互联网中存在海量的开放式递归 DNS，大部分 DNS 都是允许任意请求的递归域名服务（recursive name service）。2) DNS 之间的数据交换使用的是 UDP 无连接协议，而对于 UDP 数据分组的伪造是非常容易的；当然这并不意味着 TCP 数据分组是绝对安全的，有研究表明[8]同样可以伪造出 TCP 的数据分组。3) 数据的接收方没有能力识别出非法数据分组与合法数据分组的区别。分析以上 3 个攻击起因，如果利用网络中的 IPFIX 流数据标记的流量特征，分拣出可能产生攻击的流量数据，进而对这些数据进行异常流量检测并及时提出告警，那么，可以实现对 DNS 异常行为监测和预防，有效地保障 DNS 主机安全，极大提升网络安全预警和预防机制，保障网络安全。

3 DNS 流量放大攻击分析和检测研究

在 DNS 流量放大攻击的分析和检测研究中，目前主要从 2 个角度来检测攻击。

一个角度是通过检测数据是否被伪造进行判断。文献[9]介绍了通过硬件功能的拓展，在路由器内部建立表来区别伪造或者合法的数据分组。文献[10]基于攻击者无法干涉的字段 TTL(time-to-live)，提出了一个创造性的攻击检测方法 HCF(hop-count filtering)，避免了对网络拓扑结构的依赖，同时对于伪造分组的检测准确率达到了 90%以上。文献[11]使用 IDPF (interdomain packet filter)的检测方法，避免了对网络拓扑结构的依赖。以上这些检测方法只能检测到存在伪造 IP 分组的攻击，如果攻击者使用合法的 IP 分组攻击，那么这些方法将不会起到作用。当然对于明确检测到的流量放大攻击，这些方法通常能够有效地阻断攻击。

另一个角度就是从受害者的视角出发，检测收到分组的流量特征。文献[12]提出了随机丢弃到达分组的方式来处理这样的攻击，由于合法的发送方倾向于重新发送，而非法的发送者不会有这种倾向。但是这种处理攻击的方式更像是一种拥塞控制的方式，误杀的几率比较高。ICANN SSAC 提出过一个通过给予 IP 认证的方式来处理流量放大攻击的方法，但是这种方法从长期来说会导致更严重的问题，因为一旦攻击者获得了合法的 IP，那么整个网络中的主机将会处于毫无防范的状态。

另外，文献[5]通过建立请求和回复的对应表来检测是否存在流量放大攻击，但是这种方法即使对于一个普通院校的校园网络来说也会形成一张超级巨大的表，会占用非常多的资源进行处理，最后使得处理效率下降到很低。文献[13]使用请求和回复数据的逆差来衡量是否存在攻击，通过参考历史值来平衡数值，同时通过两个缓冲区交替的方法过滤掉攻击。文献[14]提出了黑盒思想，将整个内部网络视为一个大节点，通过分析每天网络中不合法的分组数量，以天为单位来判定有没有攻击，用这种方式成功地在 3 个月内检测到一百多次攻击。不过这种方式不能检测到底是网络中的哪个 DNS 被利用作为流量放大器或者自身直接受到了攻击，同时可以预见，在巨大的网络中，由于大部分 DNS 在大部分时间处于安全的状态，所以攻击者的痕迹可能变得不明显。

综上所述，现在已经有许多方法从不同的角度应对 DNS 流量放大攻击，在一些传统方法中，需要通过提前设定硬性指标来检测异常数据分组，这些硬性指标往往非常依赖于网络环境和结构，不利于重复使用。本文提出了一种通过横向比较和纵向比较清华大学校园网出口网络的 IPFIX 流数据的流量特征来检测 DNS 异常行为的方法，该方法通过从清华大学校园网出口路由器上获取完整的 IPFIX 网络流量数据，包含流数据产生的时间、流数据持续时间、协议类型、数据分组源地址及端口、数据分组目的地址及端口、分组数、字节数等，从中筛选出 DNS 的数据分组（UDP 协议，端口 53），以清华大学校园网为单位，基于筛选后的流数据从横向和纵向进行对比，使用多个指标同时进行评判，从多个角度检测这种攻击，从而提高了检测的准确度，并且避免因为使用一些人为规定的硬性指标如提前预设分组的大小、连接数的范围等来判定 DNS 异常行为而导致算法的局限性。

4 检测 DNS 流量放大攻击的算法

通过分析 DNS 流量放大攻击的过程，攻击流量具有几个很明显的特征：1) 在受到攻击时，作为流量放大器的 DNS 的请求数和回复数会发生非常明显的倾斜；2) 由于攻击者大多采用分布式攻击，因此受到攻击时虽然每个连接的数据量可能不大，

但是被用作流量放大器的 DNS 在受攻击时段的连接数必然非常大；3) 由于攻击的目标是消耗带宽、内存，这是不可避免的，此时该 DNS 的数据流量会出现非常大的增长。

但是，衡量正常或者攻击状态下的流量或者连接数的多少没有一个明确的标准。对于不同的网络环境和拓扑结构，不同位置的 DNS 流量特征数值是不一样的。同时，由于日期、时段的不同，在相同的网络环境下也会不同。从清华大学校园网出口网络采集的 DNS 流量数据看，2013 年 11 月 1 日全天只有 3 GB 左右，但 2013 年 12 月 10 日数据有 7 GB 左右。因此设定固定的指标会导致结果出现很大的偏差。

本文提出的方法基于一个重要的合理假设，即在清华大学校园网内的近 100 台 DNS 服务器基本不会存在同时受到 DDoS 攻击的情况，因为同时攻击这 100 台服务器所产生的放大流量通常会使得整个校园网络出现极大的拥塞，从而使得这种攻击变得没有意义。所以首先对校园网内所有的 DNS 进行横向比较，用连接数作为指标来检测存在异常的 DNS，根据出口路由器的 IPFIX 流数据每 5 min 产生的数据文件进行分析数据的筛选，考虑到发生网络异常行为时持续时间很少超过 24 h，所以在筛选数据时去除了持续时长超过 24 h 的 IPFIX 流。统计所有 DNS 的连接数 $total\_flows=\sum_{i=1}^{n}(host\_flows)$，求出平均值 $\mathrm{Mean}(host\_flows)=\mathrm{average}(host_1,host_2,\cdots,host_n)$。然后，采用每个 DNS 对于平均值的方差贡献作为评判标准，即 $Hf=\left(\frac{host\_flows-Mean(host\_flows)}{Mean(host\_flows)}\right)^2$ 当某个 DNS 的 Hf 大于 100 时标记这个 DNS 的行为可疑，为了保障检测的广度，增加流量值 Hb(host_bytes) 和数据分组值 Hp(host_packets)也作为评判指标。

$$Hb=\left(\frac{host\_bytes-Mean(host\_bytes)}{Mean(host\_bytes)}\right)^2$$

$$Hp=\left(\frac{host\_packets-Mean(host\_packets)}{Mean(host\_packets)}\right)^2$$

当流量值 Hb 或者数据分组值 Hp 大于 100 时进一步标记相应的 DNS 行为可疑。

通过连接数 Hf、流量值 Hb 和数据分组值 Hp 等指标筛选出标记为行为可疑的 DNS 主机后，为了进一步判断出该 DNS 是否行为异常以及自身受到攻击还是作为流量放大器攻击其他人。所以进行纵向比较，进一步统计这些标记为可疑的 DNS 请求分组数和回复分组数的失衡情况，通过分析发现，即使正常的 DNS 主机，由于回复和请求之间存在延迟，所以也存在一定的不平衡，对于行为异常的 DNS，往往请求和回复分组数之间存在非常明显的失衡情况，采用 exceed 评价失衡的状况。

$$exceed=\left(\frac{in\_packets-out\_packets}{\min(in\_packets,out\_packets)}\right)^2$$

正常情况下，exceed 值是比较稳定的，这是由于请求和回复地址基本会保持一致，偏离的程度大多由当前网络状况决定。当该 DNS 被用作流量放大器攻击时，因为攻击目标往往集中在少数地址上，所以这个倾斜会非常的明显。因此当 exceed 值大于 10 000 时，标记该 DNS 行为异常，判断异常 DNS 主机是自身受到攻击或者被用作流量放大器，只需判断该主机入出的分组数倾斜方向，如出方向数据分组远超入方向分组时标记该主机是流量放大器，入方向数据分组远超出方向数据分组则为受害者。

综上所述，本文提出的检测 DNS 流量放大算法，通过连接数 Hf、流量数 Hf、分组数 Hp 可检测出行为可疑的 DNS，通过 exceed 值可进一步检测出某个 DNS 是否行为异常，将检测到的可疑和异常 DNS 汇总到统一报警平台，即时向管理员发送报警邮件、报警短信，从而做出及时响应处理，遏制攻击行为，检测算法流程如图 1 所示。

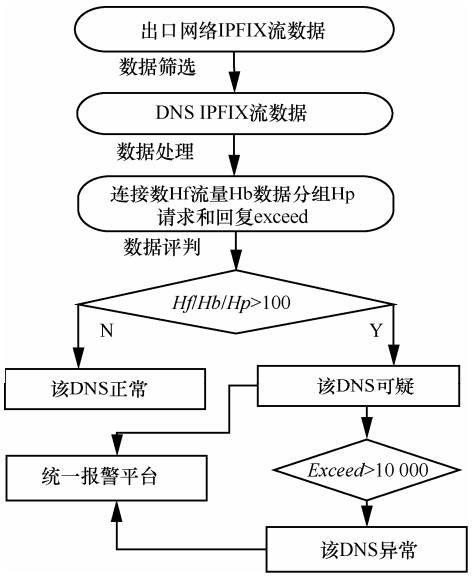


图 1　检测 DNS 流量放大攻击算法流程

5　算法有效性验证

为了验证算法的有效性，在清华大学校园网部署了算法程序。从清华大学校园网出口路由器获得的 IPFIX 原始数据每天大约 300 GB，经过对原始数据进行筛选，仅留下 DNS 通信的记录后，每天的数据大约有 6 GB，使用一台普通 PC 服务器筛选一天的原始数据时间约 100 min 左右，通常以 5 min 为粒度，筛选程序处理 5 min 的 IPFIX 数据所需时间为 20 s 左右。因此，可以认为当前分析处理的是前一个 5 min 所发生的流量，从攻击持续时间的特点来说可以算作实时检测攻击的发生。

本文的检测分析程序部署在清华大学校园网后，根据报警平台的自动报警邮件发现，在 2014 年 3 月某天的 14:00，166.111.77.2 这个清华大学校园网内部的某台 DNS 出现了异常，在这个异常发生时刻，DNS 的连接数 Hf 和流量 Hb，数据分组 Hp 如图 2~图 4 所示。

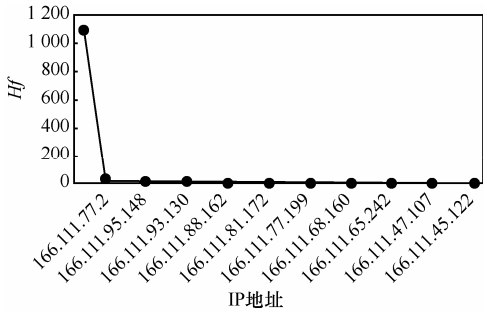


图 2　报警时刻 *Hf*

从图 3 中发现，在报警时刻标记了一个异常 DNS 主机，而在图 3 和图 4 标记了 2 个异常 DNS 主机，其中一个就是 166.111.77.2。此时 166.111.77.2 的连接数 Hf 为 7 703 520，该主机平均每秒的连接数为 25 679，该 5 min 内全体 DNS 的平均连接数为 223 357，平均每秒连接数为 744，因此 166.111.77.2 被标记为行为可疑主机。同时，它的流量 Hb 和

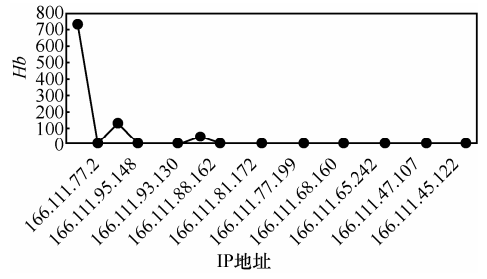


图 3　报警时刻 *Hb*

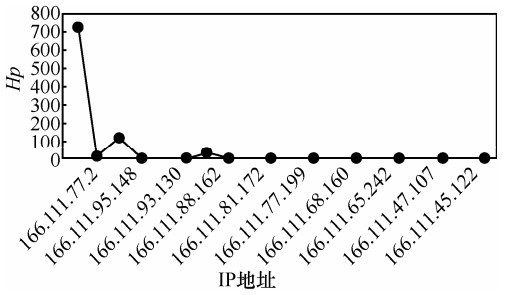


图 4　报警时刻 *Hp*

数据分组数 Hp 也有一个非常明显的峰值，根据这 3 种流量特征的判定，可以标记这个 DNS 行为可疑。对于图 3 和图 4 中出现的 2 个 DNS 主机，由于它们的连接数表现不突出，所以未标记该主机行为可疑，通过核对原始数据，发现这 2 个主机此时行为正常。

为了进一步确认，发现此时 166.111.77.2 的 exceed 值达到了 12 553，即此时该 DNS 入出方向的数据分组倾斜超过了 100 倍。此时基本可以确定这个 DNS 行为异常，因此报警程序向管理员邮箱发送了报警邮件。接下来观察 166.111.77.2 连续的 exceed 变化，如图 5 所示。

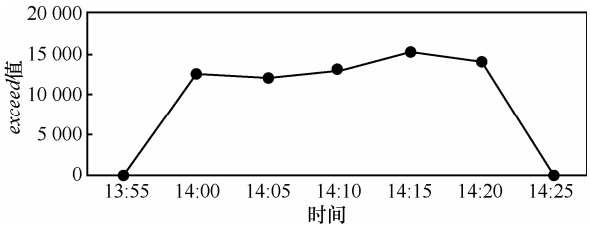


图 5　*exceed* 数据变化

从图 5 中可以看出在 13:55 该 DNS 主机行为还属正常，在 14:00 突然出现跃升，而且持续到 14:25 突然回落，通过核对 IPFIX 原始数据，发现该主机确属异常，而且通过入出分组倾斜方向的比对发现出方向数据分组超过入方向数据分组 100 倍，该 DNS 主机的确被用作流量放大器，向校外的某几个目标地址集中发送流量。综上所述，这种检测方法能够实时有效地检测出 DNS 异常行为。

在随后的运行实验中发现，当校内出现热点页面时，某个 DNS 地址的连接数和流量也会出现较大的增长，但是通常比较缓慢，而且入出数据分组的倾斜值通常没有出现很大的偏离，所以这种情况通常不会出现在报警信息中。

另外，对积累的报警信息经过统计发现，在 2014 年 6 月，清华校园网内几乎每天都存在行为可疑的 DNS 主机，行为可疑持续时段通常不超过

60 min，而且大多集中在一些固定的 DNS 主机上，应该是这些 DNS 主机一直缺乏细致的安全策略实施。从攻击者使用的 IP 地址分析来看，通常来自于 2 个方面，一方面是中国教育科研计算机网络中其他高校的 IP 地址，一方面是国外 IP 地址，其主要原因是这些主机通常采用静态 IP 地址，感染木马程序后更容易被控制。从受害者的 IP 地址来看，有一些集中在公网的热点应用上，例如某游戏公司的网站。从攻击流量上来看，某次被用作流量放大器的某台 DNS 主机在 5 min 内共计发送了 10.9 GB 的数据，检测来看其中攻击流量集中在 UPD 数据上，有 9.2 GB，即每秒持续 30 MB 的 UDP 攻击流量。

6　结束语

开放式 DNS 服务器被利用作为流量放大器的问题越来越严重，本文提出的检测算法，利用校园网出口路由器的 IPFIX 网络流量数据，能够准确地检测到校内的可疑和异常行为 DNS，并能进一步识别出 DNS 流量放大攻击行为，同时向管理员发送告警信息。本算法的特点是使用连接数、流量和数据分组等流量特征的相对值来衡量 DNS 的异常行为，所以在不同规模或不同流量特征的网络中均可以不需调整使用。本算法已在清华大学校园网实际部署运行，经实际运行验证表明本算法能够及时有效地检测到校园网内部 DNS 的异常行为，并向管理员发送告警信息，从而及时控制攻击行为，实现异常流量的及时监测和预警。基于 IPFIX 流的大数据研究工作可以此为基础进一步进行。

参考文献：

[1] MARIOS A, GEORGIOS K, PANAGIOTIS K, *et al*. DNS amplification attack revisited[J]. Computers & Security, 2013,39:475-485.

[2] Arbor Networks. 2012 infrastructure security report[EB/OL]. http://tinyurl.com/ag6tht4.

[3] CHEUNG S. Denial of service against the domain name system[J]. IEEE Security and Privacy, 2006, 4(1):40-45.

[4] 孙红杰,方滨兴,张宏莉. 基于链路特征的 DDoS 攻击检测方法[J]. 通信学报, 2007,28(2):88-93.
SUN H J, FANG B X, ZHANG H L. DDoS attacks detection based on link character[J]. Journal on Communications, 2007, 28(2): 88-93.

[5] GEORGIOS K, TASSOS M, DIMITRIS G, *et al*. A fair solution to DNS amplification attacks[A]. Proceedings - 2nd International Annual Workshop on Digital Forensics and Incident Analysis[C]. 2007.38-47.

[6] MOCKAPETRIS P. Domain names-concepts and facilities[EB/OL]. http://www.cse.ohio-state.edu/cgi-bin/rfc/rfc1034.html.

[7] MOCKAPETRIS P. Domain names-implementation and specification[EB/OL]. http://www.cse.ohio-state.edu/cgi-bin/rfc/rfc1035.html.

[8] CHANG Y H, YOON K B, PARK D W. A study on the IP spoofing attack through proxy server and defence thereof[A]. 2013 International Conference on Information Science and Applications[C]. 2013.

[9] FERGUSON P. Network ingress filtering: defeating denial of service attacks which employ IP source address spoofing[EB/OL]. http://www.cse.ohio-state.edu/cgi-bin/rfc/rfc2827.html.

[10] GHOSH A, WONG L, CRESCENZO G D, *et al*. Infilter: predictive ingress filtering to detect spoofed IP traffic[A]. Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops[C]. 2005.

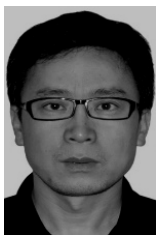
[11] HAINING W, CHENG J, KANG G S. Defense against spoofed IP traffic using hop-count filtering[J]. IEEE/ACM Transactions on Networking, 2007, 15(1):40-53.

[12] ZHEN H, XIN Y, JAIDEEP C. Controlling IP spoofing throughinter-domain packet filters[J]. IEEE Transactions on Dependable and Secure Computing, 2008, 5(1):22-36.

[13] CHANGHUA S, BIN L, LEI S. Efficient and low-cost hardware defense against DNS amplification attacks[A]. 2008 IEEE Global Telecommunications Conference, GLOBECOM[C]. 2008.2062-2066.

[14] DESHPANDE T, KATSAROS P, BASAGIANNIS S, *et al*. Formal analysis of the DNS Bandwidth amplification attack and its countermeasures using probabilistic model checking[A]. Proceedings of IEEE International Symposium on High Assurance Systems Engineering[C]. 2011.360-367.

作者简介：



马云龙（1972-），男，黑龙江嫩江人，清华大学工程师，主要研究方向为宽带认证、计费、Oracle 数据库、大规模邮件系统等。

姜彩萍（1968-），女，甘肃白银人，清华大学高级工程师，主要研究方向为互联网网络管理、网络安全、网络监测等。

张千里（1975-），男，内蒙古包头人，清华大学副研究员，主要研究方向为 IPv6 网络真实源地址、互联网网络安全、网络监测等。

王继龙（1973-），男，黑龙江大兴安岭人，清华大学研究员，主要研究方向为大规模互联网的规划、建设、运行和研究等。