NAT识别相关技术概述

NAT(Network Address Translation)[1],是一种对IP数据报某些字段进行重写的技术，该技术是由IETF组织制定的一个标准，是暂时解决公网Ipv4地址资源匮乏问题的过渡技术。

NAT可以分为静态NAT(Basic NAT)以及网络地址端口转换NAPT。其中静态NAT只是将内网中的IP地址按照某种策略映射至某些公网IP上，主要用于公网IP略少于内网IP的情况，其目的主要在于隐藏内网主机，实际应用并不广泛。NAPT网络地址端口转换则是将内网（IP/端口）映射至外网（IP/端口）上，在这个方式下NAT服务器会在其内部维护一张内网IP/端口和外网IP/端口之间的映射关系表，如表1所示。NAPT是目前使用最多广泛的NAT方式，这也是本文研究的重点，如无特别说明本文所指的NAT都是NAPT。

在实际场景中，NAPT又分完全圆锥型NAT,受限圆锥型NAT,端口受限圆锥型NAT以及对称性NAT 四种工作模式。这四种NAPT的特点如表2所示。

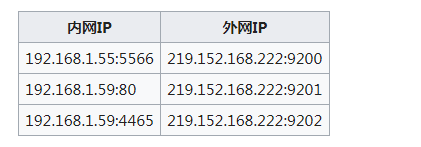
表1 NAPT映射关系表示意

表2 NAPT四种工作模式

|  |  |
| --- | --- |
| 工作模式类型 | 特点 |
| 完全圆锥型NAT | * 一旦一个内部地址（iAddr:iport）映射到外部地址(eAddr:eport),所有发自iAddr:iport的包都经由eAddr:eport向外发送。 * 任意外部主机都能通过访问eAddr:port来访问iAddr:port。 |
| 受限圆锥型NAT | * 内部地址（iaddr:iport）映射到外部地址(eAddr:eport),所有发自iAddr:iport的包都经由eAddr:eport向外发送; * 外部主机hostAddr能够通过eAddr:eport访问iAddr:iport当且仅当内部地址先向该外部主机hostAddr发送过数据包。 |
| 端口受限圆锥型NAT | * 内部地址（iaddr:iport）映射到外部地址(eAddr:eport),所有发自iAddr:iport的包都经由eAddr:eport向外发送; * 只有iaddr:iport向特定外部主机(hostAddr:hostport)发送过数据，（hostAddr:hostport）才能通过eaddr:eport访问iaddr:iport. |
| 对称型NAT | * 每一个来自相同内部IP与端口，到一个特定目的地地址和端口的请求，都映射到一个独特的外部IP地址和端口。 |

其中对称型NAT和圆锥型NAT的区别主要在于：圆锥型NAT下来自内网的iaddr:iport网络地址端口对会映射到固定的eaddr:eport上;而对称型NAT下内网的iaddr:iport会依照目的主机和目的端口的不同而映射到不同的eaddr:eport上。

虽然通过NAT技术可以缓解当前Ipv4地址不足的困境，但是NAT的引入也给整个网络带来新的问题。首先，NAT破坏IP协议中对等通信的原则，NAT使得公网主机无法主动访问内网主机;其次NAT会使IP会话的保持时效变短，会影响某些网络应用的运行。再者NAT将网络环境变的复杂，使得依赖IP地址进行主机跟踪的机制失效，同时影响网络运营商进行正常的网络计费、加大网络管理员管理网络的困难。最后NAT会修改IP数据报的某些字段，这使得某些传统的安全协议受到一定的影响，例如IPSec协议。

为了解决上述问题，人们提出NAT穿透技术。NAT穿透技术主要基于UDP进行打浻，如RFC3489中定义的STUN协议以及其补充协议TURN，同时也有基于构造特殊的ICMP进行NAT穿越的技术[10]。

为了净化网络环境、保障网络安全、有效地管理与监控网络，人们又开始寻找行之有效的NAT流量识别方法。

NAT流量识别方法主要分为主动和被动检测方法。

主动检测NAT的技术不多。STUN协议的NAT发现过程可以用来检测当前主机是否位于NAT之后以及所处的NAT类型。然而该方法需要拥有NAT后某台主机的控制权。还有使用端口扫描进行NAT检测的方法，这主要是因为某些NAT服务器会开放一些控制端口（例如SNMP相关端口）以方便用户监控和控制他们的服务器，通过端口扫描技术可以很容易的检测出这些端口。但只要NAT设备关闭和修改相应的端口即可逃避检测，而且主动端口扫描很容易被IDS中途拦截。

被动探测技术通过被动监听网络中的数据流量，监测数据包的首部或内容信息来进行NAT流量的判断。根据各种识别的不同特点，大致可以把现有的NAT被动检测方法分为两大类：基于IP数据报特定字段的识别方法以及基于IP数据报流量特征的识别方法。

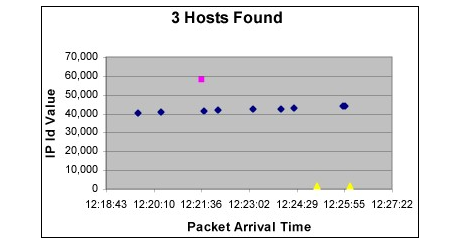
Bellovin 提出通过观察IP 报文中的ID字段来发现NAT设备后面的主机数量.[2]这种方法的实现是基于当主机每发一个IP数据包则将IP ID增加一定数值，且每个主机IP ID的初值不同。如果在NAT后面存在多个主机则这些IP ID的值将会发散开来。如图1所示，可以发现三条不同的直线，因此有三个主机。但是这种方式不适用于BSD系统，因为BSD系统的IP ID是随机的。

图1 IP ID检测示意图

Beverly 提出使用TCP/IP报文里面的TTL ,DF,窗口大小以及SYN初始值来区分NAT和主机的方法。[3] 其实就是基于OS的协议栈指纹，来统计有多少种不同指纹进而判断主机数。

白雪提出基于统计一定时间段内访问同一web网站Cookie ID数量的方式来判断NAT以及NAT后面的主机数量[4]。还有提出统计一段时间内IM数据包内的QQ号码数量来判断是不是NAT以及计算NAT的规模[11]。Zhu[12]提出来使用User-agent 字段是web服务器用于确定用户操作系统、浏览器及版本的字段 ,因此通过统计User-agent可以在一定程度上发现NAT设备以及NAT后主机的数目。

Kohno提出使用TCP的时间戳字段来当作OS的指纹[5]，以此来识别NAT流量以及NAT规模。但是该方法需要操作系统默认启动TCP时间戳字段方可。

Sophon Mongkolluskamee [6] 提出通过观察IPID,TCP 序列号,TCP 源端口的方式来计算Nat后的主机数目。他将这三者关联起来，可以识别Windows,linux,freeBSD,mac os ,openbsd系统以及对应的数量。他通过nat流量合成和办公室网关抓包的方法收集数据集，最后把该方法用在一个公开流量数据集中计算2001-2010每年的nat比例和nat后主机的平均规模。

以上几种方法都是基于特定报文字段，与传统单主机的流量相比，经过NAT后的主机流量应该是具有以下特征：更多byte,更多连接;更多DNS查询;更多协议类型等等。因此有学者基于IP数据报的流量特征来识别NAT流量。

Martin Grill [8]提出基于HTTP 访问记录使用SVM分类器进行被动检测NAT流量的方法。他统计一定时间段内的日志反映的8个属性：不同IP地址数;不同User-Agent数目;不同的OS数目;持续的流数目;上传字节数;下载字节数;http请求数。他生成数据集的方法，先收集明确是非nat主机的访问记录;然后将不同端的访问记录组合在一起，人工构建nat数据;这是因为若干个主机将经nat后，就相当于把这8个向量整合在一起。

Rui[9]提出一种使用SVM进行检测NAT流量的方法。他生成数据集的方式：1台主机不进行NAT，2～5台主机nat。使用的特征向量有：出包数量;入包数量;UDP报文数量;TCP报文数量;DNS查询数量;FIN数量;RST数量;SYN数量。为了过滤掉一些在线率不高的主机,通过某种方式把休眠的主机过滤了。

高骥翔[7]提出使用SVM,C4.5,朴素贝叶斯方法,K-means聚类,多层感知机的方法进行NAT流量的判断。他提取16个特征作为特征向量输入到这些机器学习的模型中进行学习等。同时，他还探究p2p流量对分类器的影响。

参考文献

[1] P. Srisuresh and K. Egevang. Traditional IP network address translator (Traditional NAT). RFC 3022, Internet Engineering Task Force, January 2001, 2001.

[2] S. M. Bellovin, "A technique for counting NATted hosts" in , New York, USA:, 2002.

[3] R. Beverly, "A robust classifier for passive TCP/IP fingerprinting", Passive and Active Measurement Workshop:5th international workshop, pp. 158-167, 2004.

[4] 白雪.一种检测NAT后主机数目的方案.[J].计算机安全.2009

[5] T. Kohno, A. Broido, K. C. Claffy, "Remote physical device fingerprinting", IEEE Transactions on Dependable and Secure Computing, vol. 2, pp. 2, 2005.

[6]Sophon Mongkolluksmaee,Kensuke Fukuda,Panita Pongpaibool.”Counting Nated Hosts by Observing TCP/IP Field Behaviors.

[7] 高骥翔.基于网络流量特征的NAT识别方法.电子科技大学硕士学位论文.2012

[8] Tom´aˇs Kom´arek, Martin Grill, Tom´aˇs Pevn´y, Passive NAT Detection Using HTTP Access Logs, Information Forensics and Security (WIFS).2014

[9] L. Rui, Z. Hongliang, X. Yang, X. Yang, and W. Cong, “Remote nat detect algorithm based on support vector machine,” in Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on. IEEE, 2009, pp. 1–4

[10]Andreas Muller,Nathan Evans,Christian Grothoff.“Autonomous NAT Traversal” in Peer-to-Peer Computing (P2P), 2010 IEEE Tenth International Conference on.2010

[11]梁峰.代理服务器及NAT网关检测技术的研究生.山东大学硕士学位论文.2009

[12]Zhu Hongliang,Li Rui et al.An Integrative Model and System for detecting NATed Hosts.ICIECS 2009,Wuhan,China.Dec 2009.