一种检测 NAT 后主机数目的方案

白 雪，钱步仁，梁华庆
（中国石油大学，北京 102249）

摘 要：首先介绍了利用 NAT 技术实现共享接入上网的原理以及对 NAT 后的用户进行识别和精确定位的重要意义，然后在此基础上提出了一种新的数据源追踪技术。通过被动获取的方式来获取 NAT 下不同用户和同一知名网站交互的 HTTP 数据包中的 Cookie ID,如果同一个 IP 地址与该网站有多个 Cookie ID,则 Cookie ID 的数目就是通过该 IP 共享上网的主机个数,并可以通过该 Cookie ID 对用户进行精确识别。该种方案具有更高的准确性和稳定性。

关键词：网络地址转换；共享接入；主机数目

A Scheme for Counting NATted Hosts

BAI Xue, QIAN Bu-ren, LIANG Hua-qing
(China University of Petroleum, Beijing 102249,China)

Abstract: The paper introduces the theory of connecting to the Internet by means of NAT boxes and great importance of detecting NATs. And then presents a new scheme of detecting that registering the hosts' Cookie ID when they located to a famous website. If any of these Internet addresses had more than one Cookie ID of the same website, then the num of the Cookie ID is just the same with the num of hosts under the Internet address. Compared with other techniques, the new scheme has the properties of more exactness and stability.

Key words:NAT; connecting to Internet by means of NAT boxes; num of hosts

1 引言

在 IP 地址日渐匮乏的今天，NAT 技术由于其可以使局域网内的多台主机共享同一个公网 IP 地址上网的特性，逐渐得到了广泛应用。但在互联网安全问题日益严重，对网络数据流量进行安全监控势在必行的今天，当监控设备监控到违法、违规流量时，却无法对 NAT 后众多主机中的违法、违规者进行精确识别和定位。这一问题在对危害国家、社会的危险分子进行监控和跟踪时尤为突出。因此，对 NAT 后主机进行识别和精确定位的数据源追踪技术是保障信息安全、净化网络环境所急需解决的技术问题。

为了解决上述问题，首先要有能够识别 NAT 用户的相关技术，以便在识别出违法、违规数据流后对 NAT 用户进行识别，避免与普通用户混淆；进而再利用相关技术对特定主机进行定位。我们把这两种技术合成为数据源追踪技术。本文针对在 Cookie 有效期内，同一个网站分配给不同用户的 Cookie ID 值不同的原理，提出了一种利用 Cookie ID 来识别 NAT 后主机数目的方案。

2 NAT 技术介绍

随着 Internet 的飞速发展，越来越多的用户加入到使用互联网的行列中。全球 IP 地址资源匮乏的问题也越来越突出。为此，IETF 组织提出了 NAT（Network Address Translation）技术[1]，作为暂时解决 IP 地址耗尽的过渡手段。NAT 设备完成的是网络地址转换的功能，位于 NAT 后的主机拥有自己的私网 IP 地址，并利用 NAT 设备共享一个或几个 IP 上网。当主机需要与位于公网上的设备进行通信的时候，NAT 将对应的私网 IP 地址和端口号映射为自己的公网 IP 地址和端口号。这样位于 NAT 后的主机相对其他公网上设备是透明的，NAT 的网络结构如图 1 所示。

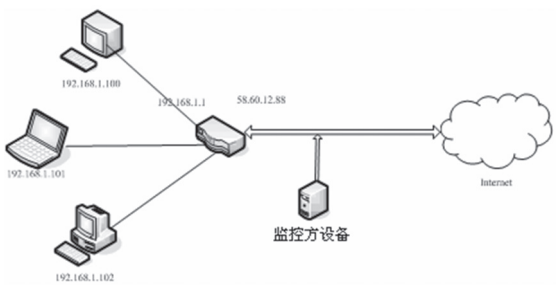

图 1 NAT 网络结构图

图 1 显示出了三台主机利用私有 IP 地址 192.168.1.100~192.168.1.102 通过带有 NAT 功能的路由设备共享一个公网 IP 地址 58.60.12.88 上网的结构图。在该图中，监控方作为中间设备位于 NAT 设备和 Internet 之间，通

过分光、镜像等技术获取NAT设备发出的数据流，从而对NAT后的主机数量进行分析。

由于NAT的特性，在经过NAT的数据包中很难观察出NAT后的主机信息，对NAT后的主机数量的统计变得非常困难。因此，若想正确地统计出NAT后的主机数量，必须要获悉每台机器独有的一些特性。

3 现有检测技术

3.1 IPID技术[2]

IPID指的是IP报文首部的标识(identification)域，长度为16比特，用来惟一标识一个IP报文，Windows操作系统将IPID作为一个计数器，不论数据包属于哪个连接，同一主机每发出一个数据包，IPID值递增1[3]。通过分析一个指定IP地址发出的数据包的IPID值有多少个连续的轨迹，可以确定这个地址后的主机数。

但是这种算法的缺点在于：

(1) 只能针对操作系统是Windows的主机；

(2) 因为IPID轨迹的获取过分依赖于是否能够连续地获得目标IP发出的数据包。当NAT设备后的主机使用多线程下载工具，或主机之间有大量数据交互时，都会有IPID值的变化，使得IPID值从NAT外看失去规律性，从而影响IPID检测方法的准确性，导致误判或多检。

(3) 另外，IPID属于底层协议字段，现在很多的NAT设备都会将其后不同主机的IPID值进行修改，使修改后的数值看似同一台主机连续发出的IPID值，使得IPID检测方法失效。

3.2 其他技术

还有一些基于行为特征的检测方案，如利用IP地址的并发端口数是否多于设定阈值的方法等[4]，但这些基于行为特征的检测方案的准确性无法保障。尤其在P2P类应用广泛使用的今天，更使其准确性大打折扣，经常会出现误判的情况。另外，这种方案只能统计是否存在共享，而不能统计精确的数量。

针对现有技术存在的各种问题，本文提出了一种基于Cookie ID的共享主机检测方案，该方案较好地解决了现有技术的各种缺陷，是一种非常有效的共享主机检测方法。

4 Cookie ID技术

4.1 原理介绍

Cookie是大部分网站为了辨别用户身份而存储在用户本地终端上的数据。当用户浏览某网站时，Web服务器会发送给用户一个包含日期时间和用户ID信息的Cookie[5]。用户的浏览器在获得页面的同时会将这个Cookie保存在用户硬盘上的某个文件夹下。当用户再次访问该网站时，会带出该Cookie，网站根据Cookie得到用户的相关信息，就可以做出相应的动作，如用户不必每次输入ID、密码就可直接登录等。

在HTTP协议中，网络服务器会给初次访问该网站的用户通过HTTP 200 OK回应包中的Set-Cookie字段分配Cookie，用户获得Cookie之后在每次发送给该网站的HTTP请求包中都会有该Cookie的信息。Cookie的格式如下：

Set-Cookie: NAME=VALUE; Expires=DATE; Path=PATH; Domain=DOMAIN_NAME;SECURE

其中，只有NAME=VALUE为必选项，是网站分配给用户的惟一ID值，Expires属性用以确定Cookie的有效期。在Cookie有效期内，同一网站为不同用户分配的ID值不同。当指定IP地址的数据包通过监控设备时，服务器可以对其进行分析，监听HTTP请求报文，读取其中的网站地址和相应的Cookie ID并记录下来。如果访问相同网站的Cookie ID值不只一个，就可以确定这个IP地址有多台主机共享，并且，有几个Cookie ID就说明有几台主机。

4.2 具体实现

由上面的叙述可以看出，利用Cookie ID进行主机数目判断的实现过程其实就是捕获数据，然后对数据进行统计分析，最后上报的过程。因此本方案分为监控模块、统计模块和上报模块三部分。

(1) 监控模块

该模块负责监控NAT设备的上行流量或者下行流量。所谓上行流量是指客户端发到Internet上的数据流，下行流量是指外端网络发回给客户端的数据流。监控设备过滤出被监控流量中的HTTP类型数据流。如果监控的是上行流量，提取出其中含有Cookie字段的数据包；如果监控的是下行流量，提取出其中含有Set-Cookie关键字段的数据包。

然后，提取这些数据包的源IP地址、目的IP地址、Cookie名称及Cookie ID值。将这些信息提交给统计分析模块。

(2) 统计模块

该模块主要负责统计来自监控模块的数据并保存在数据库中，因此，需要事先在数据库中建立如图2所示的存储表，用以保存Cookie的统计信息。

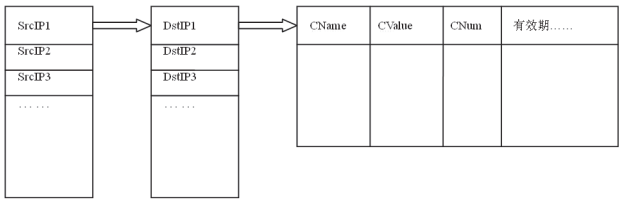

图 2 存储表格式

其中，SrcIP 为被监控的 NAT 的公网 IP 地址，DstIP 为被访问的网站的地址，最后一栏分别存储 Cookie 的名称、数值、个数及生存期等信息。

接收到来自监控流量模块的数据后，统计模块根据源 IP 地址和目的 IP 地址找到对应的 Cookie ID 记录表，然后在该记录中查找该 Cookie ID 值，如果找到，不做任何处理；如果没有找到，就添加该 Cookie ID 值的信息。总之，一定要保证所有的 Cookie ID 信息都有记录。

(3) 上报模块

上报模块每到一个上报周期，就查找同一个源 IP 下每一个网站的 Cookie ID 记录表，将 Cookie 的有效期与系统当前的时间做比较，如果 Cookie ID 已经过期，就将其删除。

然后统计每一个源 IP 对应的所有门户网站的 Cookie ID 中 CNum 的最大值。如果该最大值为 1，则说明该 IP 下面没有 NAT 环境，否则，该最大值即为 NAT 环境下的主机数目。

4.3 与其他检测技术比较

本方案以 Cookie ID 值为依据，对 NAT 环境进行识别并精确统计 NAT 后面的主机数量。不会发生误判或多判的情况，完全符合业界的检测标准。另外，克服了 IPID 检测方案准确度受内网的流量及用户使用协议的影响的弊端，消除了当 NAT 设备修改底层协议字段后 IPID 检测方案无效的重大缺陷，同时相对于行为特征检测方案，还拥有较高的检测准确率，因此，是一种非常实用的检测方案。

5 结束语

本文提出了一种利用 HTTP 协议中的 Cookie ID 进行数据源追踪的方案，通过本方案可以准确地检测出 Internet 上的 NAT 环境以及其后的主机数量，并能精确定位到 NAT 后的某台主机。本方案通常应用于网络监控领域，用来对违法、违规主机进行定位，以便后续其他技术的实施。整个网络监控流程包括流量识别、目标追踪、目标控制三步骤，本方案是其中目标追踪这一环，在整个网络监控技术中位于非常重要的地位，具有较大的实现意义和价值。

参考文献：

[1] K. Egevang, Cray Communications, P. Francis. Rfc1631:The ip network address translator (nat). Technical report, IETF, 1994.

[2]Steven M. Bellovin .“A Technique for Counting NATted Hosts”. ACM New York, NY, USA ,2002.

[3]Peter Phaal .“Detecting NAT Devices using sFlow”,2004.

[4]antirez. dumbscan. Technical report, Bugtrack, 1998.

[5]Internet Engineering Task Force (IETF) . Rfc2965:HTTP State Management Mechanism ,2000.

作者简介：白 雪（1984-），女，硕士研究生，主要研究方向：检测技术与自动化装置。

收稿日期：2008-10-09