应用交互式网络流模型的高速网络异常行为检测与控制

杨柳静，秦涛，王晨旭
（西安交通大学智能网络与网络安全教育部重点实验室，710049，西安）

摘要：针对网络异常流量的检测与定位问题，提出了一种根据网络流统计量异常变化和不完整网络流来有效识别并定位网络异常流量的方法.该方法建立在交互式网络流模型的基础上，分析了交互式网络流模型下各种网络流的交互特征；为准确实时获取网络异常源，采用中国余数定理，设计了连接度sketch结构中的哈希函数，满足了网络用户信息逆向求解的需要，实现了高速网络中异常网络流特征参数的实时获取；为减缓网络异常行为的扩散速度，提出采用动态软隔离方法实现网络异常行为的控制.真实环境下的实验结果表明，所提方法对于多种类型的网络异常行为具有良好的检测效果，检测的准确率和速率都得到了提高，同时可以准确地定位网络异常源，为有效控制网络异常行为的扩散奠定了基础.

关键词：交互式网络流模型；不完整交互行为；连接度sketch；异常行为检测与控制
中图分类号：TP393 **文献标志码**：A **文章编号**：0253-987X(2012)06-0058-08

Abnormal Behavior Detection and Control in High Speed Networks Based on Bidirectional Flow

YANG Liujing, QIN Tao, WANG Chenxu
(MOE Key Lab for Intelligent and Network Security, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A new method is proposed to effectively identify and locate the abnormal network flows based on the abnormal changes of the flow statistics and the incomplete flows. The method bases on the bidirectional flow model, and analyzes the interactive features of different network flows. A hash function in the structure of the connection degree sketch is designed by using the Chinese remainder theorem, so that the source of the abnormal behaviors can be accurately and timely achieved, and the users' information is obtained from the abnormal flows in the high-speed networks. The dynamic and soft isolation method is used to control the abnormal behaviors and hence to slow down the spread speed of the abnormal behaviors. The experimental results in an actual network show that the proposed method is efficient in improving both the detection accuracy and speed for most kinds of abnormal behaviors. At the same time, the source of the abnormal flow is exactly located, and it is helpful to control the spread of the abnormal behaviors.

Keywords: bidirectional flow; incomplete interactive behavior; degree sketch; abnormal behavior detection and control

Internet是一个开放性的公众服务网络，面临着众多的安全威胁.例如，蠕虫病毒[1-3]和分布式拒绝服

收稿日期：2011-11-22. 作者简介：杨柳静(1987—)，女，硕士生；秦涛(通信作者)，男，讲师. 基金项目：国家自然科学基金资助项目(61103240，91018011，60921003)；中央高校基本科研业务费专项资金资助项目(XJJ2011015).
网络出版时间：2012-03-12 网络出版地址：http://www.cnki.net/kcms/detail/61.1069.T.20120312.1028.003.html

务(DDOS)[4-5]等大规模网络异常攻击往往会在一定时间内造成网络内多数主机被控制，以及隐私信息的大量丢失，而受控主机数量的增多直接威胁着网络的安全运行. 对于这些网络攻击，传播初期的有效检测与控制是降低危害的关键. 网络流量一直是网络测量的重要数据源，早期的方法主要是在宏观上观测单位时间内网络中传输数据包的总数量、总字节以反映网络的特性[3]. 现在越来越多的研究转向网络流量的细粒度分析，为此，CISCO 在高端设备上提供了对 NetFlow 协议的支持，逐渐开展了大量的以此项技术为基础的研究，主要研究方向涵盖网络测量和网络安全两个方面[6-7]. 但是，NetFlow 是一种单向的网络数据包汇聚协议，无法满足网络流量细粒度监控的需要. 针对以上问题，本文提出了交互式网络流模型，并且重点分析了其中存在的不完整网络流的特征，为异常行为检测提供了新的视角.

近几年，为实现网络流量的实时在线监控，高速网络流量的处理方法得到了充分的研究. 其中，sketch 方法是最近几年提出来用于大规模数据处理和计算的一种数据结构，通过多维哈希函数组实现大规模数据近似值的统计. 国内外学者使用 sketch 在网络数据计算和网络监控方面进行了许多研究，包括网络改变的测量[8]、网络流大小分布的估计[9]、网络信息熵的估算和网络异常检测[10]等. 在这些方面的研究中，sketch 通过将大量的数据信息存储在几个哈希表构成的数据结构中，使得其存储空间消耗小并且数据更新时间与数据量的大小成线性关系，为实时网络监控提供了一个新的方向. 但是，以往的一些研究工作是基于传统的网络流模型，并没有使用数据流的交互特性，使用的 sketch 也只能发现网络异常，并不能提供异常的源信息和目的信息. 受到相关工作的启发，本文使用连接度 sketch 数据结构[11-12]存储和测量交互式网络行为特征，可以以很小的空间实现网络交互特征的获取和网络用户行为在相邻时刻变化情况的测量. 并且，采用中国余数定理设计 sketch 中的哈希函数，使该数据结构能够重建不完整网络流产生者的信息，为实现主动防御和异常控制奠定了基础.

1　交互式网络流模型及特征参数

1.1　交互式网络流模型的建立

随着网络的快速发展，所提供的服务种类越来越多，用户行为也越来越复杂. 本文首先对常见的网络用户行为进行了分析，如图 1 所示. 作为一个 WEB 服务器，通常情况下只开放一个 80 端口，并且对多个网络用户提供网络访问服务，会和多个用户产生不同的网络流；对于一个文件传输的网络用户行为来说，我们也许只能看到一个简单的网络流，但是这个网络流往往携带大量的数据包；对于 DNS 服务器来说，通常情况下会产生很多网络流，并且这些网络流会有一个固定的端口，例如 53 端口；对于一个传播恶意代码的网络用户，往往会产生众多网络流，每一个网络流都是一次攻击尝试，并且这些网络流往往是单向网络流；对于一个用户的 P2P 行为，往往会产生少数网络流，并且这些网络流之间会传输众多的数据量. 传统的 NetFlow 网络流框架只描述了网络流的整体属性及相关统计信息，没有考虑网络终端之间数据传递的特点，使得不同的交互行为产生的数据在统计层面具有相同的特性，难以进行深层次的网络行为特征刻画.

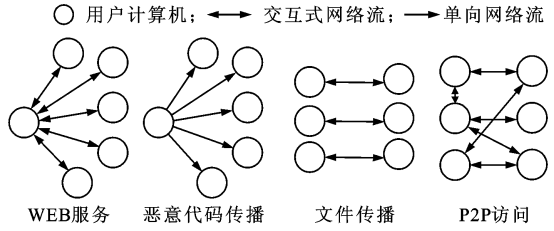


图 1　典型网络用户行为分析

针对上述问题，本文提出交互式网络流模型，如图 2 所示，即单位时间内具有相同 IP 地址的网络数据包集合，包括前向数据包和后向数据包. 当数据包的源 IP、目的 IP 和对应的交互式网络流的源 IP、目的 IP 对应相同时，称为前向数据包；当数据包的源 IP、目的 IP 和对应的交互式网络流的目的 IP、源 IP 对应相同时，称为后向数据包. 利用交互式流框架所提供的属性信息，对网络流量所反映的网络行为进行分类、建模与分析，实现网络用户交互行为的刻画，弥补了 NetFlow 在网络行为刻画方面的不足. 由于交互式网络流定义忽略了端口等信息，在很大程度上缩减了网络流记录，为实时监控和计算奠定了基础.

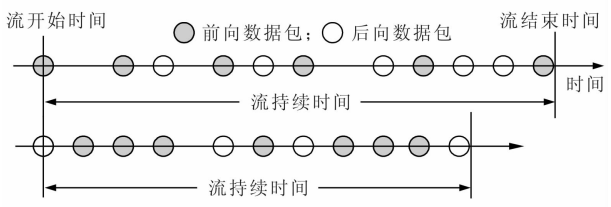


图 2　交互式网络流框架图

1.2 交互式网络流特征测量指标体系

本文所涉及的网络监控点设置于网络的出口路由器处. 假设所监控的网络内部有 m 台被监控主机,它们和外部的 n 台主机进行交互通信. s_{ij} 表示由 S_i 发往 D_j 的数据包个数, d_{ji} 表示由 D_j 发往 S_i 的数据包个数,那么这种通信关系可以用有向图来表示,其中 D_j 和 S_i 表示网络中的节点,s_{ij} 和 d_{ji} 表示两个节点之间的通信关系. 矩阵 $\boldsymbol{M}(t)$ 表示监控网络在 t 时刻的流量特征

$$\boldsymbol{M}(t)=\begin{bmatrix}(s_{11},d_{11}) & (s_{12},d_{21}) & \cdots & (s_{1n},d_{n1})\\ (s_{21},d_{12}) & (s_{22},d_{22}) & \cdots & (s_{2n},d_{n2})\\ \vdots & \vdots & & \vdots\\ (s_{m1},d_{1m}) & (s_{m2},d_{2m}) & \cdots & (s_{nn},d_{nn})\end{bmatrix} \tag{1}$$

根据上述分析,网络的通信行为可以用一系列矩阵 $\boldsymbol{M}(1),\boldsymbol{M}(2),\cdots,\boldsymbol{M}(t)$ 表示,只要分析矩阵随着时间的变化特性就可以分析出网络流量的特征,并据此实现网络流量的动态监控. 基于上述流量矩阵,我们定义了相应的网络流量特征来测量网络流量的动态变化,各参量定义如下.

(1)源主机的出连接度. 定义为在时间窗口 T 内,某源主机所主动连接的不同目的主机的个数,可以表示为

$$X_i=\|\{s_{ij}\,|\,s_{ij}\neq 0,1\leqslant j\leqslant n\}\|\ ,1\leqslant i\leqslant m \tag{2}$$

式中:X_i 表示第 i 个源主机的出连接度.

(2)源主机的入连接度. 定义为在时间窗口 T 内,访问某源主机的不同目的主机的个数,可以表示为

$$Y_i=\|\{d_{ji}\,|\,d_{ji}\neq 0,1\leqslant j\leqslant n\}\|\ ,1\leqslant i\leqslant m \tag{3}$$

式中:Y_i 表示第 i 个源主机的入连接度.

(3)目的主机的出连接度. 定义为在时间窗口 T 内,某目的主机所主动连接的不同源主机的个数,可以表示为

$$Z_j=\|\{d_{ji}\,|\,d_{ji}\neq 0,1\leqslant i\leqslant m\}\|\ ,1\leqslant j\leqslant n \tag{4}$$

式中:Z_j 表示第 j 个目的主机的出连接度.

(4)目的主机的入连接度. 定义为在时间窗口 T 内,访问某目的主机的不同源主机的个数,可以表示为

$$W_j=\|\{s_{ij}\,|\,s_{ij}\neq 0,1\leqslant i\leqslant m\}\|\ ,1\leqslant j\leqslant n \tag{5}$$

式中:W_j 表示第 j 个目的主机的入连接度.

上述4个特征属性从不同的角度描述了网络用户的通信行为,源主机的出入连接度主要描述了源主机的通信行为,目的主机的出入连接度主要描述了目的主机的通信行为,这种内外相结合的行为刻画方式可以有效地实现网络行为的全面监控,不但可以监控发生在监控网络内部的网络异常行为,还可以实现外部网络对监控网络渗透行为的检测.

2 流量参数获取与异常检测及控制

2.1 连接度 sketch 的设计及其在参数获取中的应用

近些年来,sketch 方法得到了充分的研究,并广泛地应用于高速网络流量监控中. 在已有的工作中,连接度 sketch 被用来检测具有大连接度的网络主机[13-14]. 本文主要利用连接度 sketch 实现高速网络中用户连接度的实时统计,并通过设计 sketch 中的哈希函数,实现网络异常主机详细信息的获取和分析. 连接度 sketch 可以表示为 $\boldsymbol{B}=(\boldsymbol{B}_1,\cdots,\boldsymbol{B}_i,\cdots,\boldsymbol{B}_H)$,其中 $\boldsymbol{B}_i(1\leqslant i\leqslant H)$ 是 $v\times m_i$ 大小的矩阵,H 是数据矩阵元的个数. 每个 $\boldsymbol{B}_i$ 都关联一个哈希函数 $h_i:\{0,1,\cdots,n-1\}\rightarrow\{0,1,\cdots,m_i-1\}$, 其中 n 表示数据包源空间的大小. 所有的 $\boldsymbol{B}_i(1\leqslant i\leqslant H)$ 共享一个哈希函数 $f:\{0,1,\cdots,E-1\}\rightarrow\{0,1,\cdots,v-1\}$,其中 E 是源地址与目的地址相互关联的地址空间大小. 连接度 sketch 的更新过程如图3所示,在每一次测量开始的时刻,$\boldsymbol{B}_k(1\leqslant k\leqslant H)$ 中的每一位都先置0,当一个数据包 $p_i=(s_i,d_i)$ 到达时,$\boldsymbol{B}_k(1\leqslant k\leqslant H)$ 通过设置每一位的行 $f(s_i\parallel d_i)$ 和列 $h_k(s_i)$ 实现其值的更新,更新规则为

$$\boldsymbol{B}_k[f(s_i\parallel d_i)][h_k(s_i)]=1,\quad 1\leqslant k\leqslant H \tag{6}$$

本文采用源 IP 和目的 IP 作为数据包标识,记为 $s<\text{sip},\text{dip}>$. 哈希函数 h_i 具体为

$$h_i(x)=(a_ix+b_i)\bmod m_i,\quad 1\leqslant i\leqslant H \tag{7}$$

式中:m_i、$m_j(i\neq j)$ 为互质正整数;a_i 和 b_i 均为正整数. 在对 sketch 结构 $\boldsymbol{B}$ 进行分析时,需要通过 $\boldsymbol{B}_i(1\leqslant i\leqslant H)$ 中的异常位置来确定异常源,即通过异常位置反向得到流标识 $s<\text{sip},\text{dip}>$. 问题可以描述为求解映射到哈希表 $\boldsymbol{B}_1,\cdots,\boldsymbol{B}_H$ 中位置分别为 $c_1,\cdots,c_H$ 的原始网络流量特征信息,即求解下述问

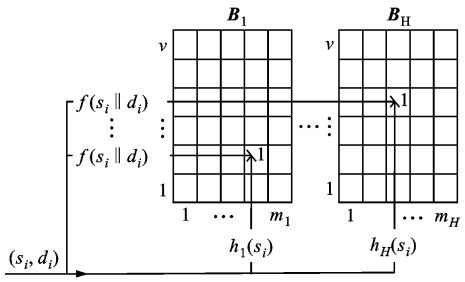


图 3　连接度 sketch 的更新过程

题[15]

$$\left.\begin{aligned} a_1 x + b_1 &\equiv c_1 (\mathrm{mod} m_1) \\ &\vdots \\ a_H x + b_H &\equiv c_H (\mathrm{mod} m_H) \end{aligned}\right\} \tag{8}$$

利用孙子定理和费马小定理可以求解出所有满足条件的候选集

$$x \equiv \sum_{i=1}^{H} M_i M_i^{-1}\, a^{\varphi(m_i)-1}(c_i - b_i) \mathrm{mod} \hat{m} \tag{9}$$

式中：$\hat{m} = m_1 \cdots m_H$；$M_i = \hat{m}/m_i$，m_i^{-1} 可以由式 $M_i M_i^{-1} \equiv 1 (\mathrm{mod} m_i)$ 求得；φ 是欧拉函数，其求解过程如下. 设正整数 n 的质因数分解式为

$$n = p_1^{k_1} \cdots p_L^{k_L} \tag{10}$$

式中：$p_1, \cdots, p_L$ 是不同的素数，则欧拉函数为

$$\varphi(n) = (p_1 - 1) p_1^{k_1 - 1} \cdots (p_L - 1) p_L^{k_L - 1} = n \prod_{p|n} (1 - 1/p) \tag{11}$$

令 $\hat{x} = \sum_{i=1}^{H} M_i M_i^{-1}\, a^{\varphi(m_i)-1}(c_i - b_i) (\mathrm{mod} \hat{m})$，即 $\hat{x}$ 为小于 $\hat{m}$ 的唯一潜在异常键值，那么所有潜在的异常键值如式(12)所示，x 对应着所求取的网络异常行为的持有者

$$x = k\hat{m} + \hat{x}, \quad 0 \leqslant k < N/\hat{m} \tag{12}$$

2.2　网络异常行为发现方法

网络异常行为通常指网络流量的行为偏离了正常行为的情形，会对网络的正常使用造成不良影响. 对网络流量异常进行检测和分析，是网络监控分析的一个重要目的. 造成网络流量异常的原因大概有以下几类：①与网络安全有关，例如 DDoS 攻击和 DoS 攻击等拒绝服务攻击、黑客入侵、端口扫描、蠕虫爆发；②网络配置错误等故障或者性能方面带来的问题，例如服务器由于猛增的计算请求等突发事件造成的服务失效；③与网络的使用有关，例如 P2P 的大量应用严重占据网络带宽，影响正常用户的网络使用. 网络异常行为通常体现为相邻时刻统计量发生突变，或者短时间内产生大量的不完整网络流.

根据网络的平稳运行特征，网络用户行为统计特征的变化是平稳的，即在短时间内，用户的行为统计特征不会出现明显的变化. 但是，当网络发生异常行为时，网络用户的统计行为会在短时间内发生大的变化. 基于此，我们通过分析对比用户在相邻时刻的变化特征来实现网络异常行为的检测.

计算机网络主要应用于信息交互和信息共享，其数据传输是双向交互的过程，在交互过程中，会产生前向和后向两个方向的数据包，如果只存在从交互的一方发送到另一方的数据包，那么这个交互行为是不完整的. 不完整网络流定义为只有前向数据包或者只有后向数据包的网络流. 不完整网络流的产生原因有很多种. 首先，由于网络原因用户所发送的请求没有得到响应，就会产生一个不完整网络流，这种情况很少出现，并且不完整网络流也较小. 其次，网络异常行为是产生不完整网络流的重要原因，并且产生的不完整网络流都比较大. 表 1 给出了当前主要异常行为与不完整网络流特征的对应情况. 据此，我们可以通过统计分析不完整网络流的特征来实现网络异常行为的检测. 文中以被监控用户为参照，当交互过程只有被监控用户发向对方的数据包时，称为缺乏后向数据包的不完整交互行为，所有不完整网络流均由被监控用户产生. 当交互过程只

表 1　常见网络攻击产生不完整网络流的状况

攻击类型	行为描述	不完整网络流特征
SynFlood(DDOS)	多个源地址向目标发送大量的 SYN 数据包	目标主机产生大量不完整网络流
UDPFlood(DDOS)	多个源地址向目标发送大量的 UDP 数据包	目标主机产生大量不完整网络流
Flash Crowd	多个源地址向单个目的地址发送大量数据	目标主机产生大量不完整网络流
Port Scan	对于单个或多个目的主机的端口扫描	源主机产生大量不完整网络流
Network Scan	对于某个网段内主机的端口扫描	源主机产生大量不完整网络流
Point to Multi-Point	单源地址向多个目的地址发送大量数据	源主机产生大量不完整网络流
Worms	大规模的对于漏洞主机扫描	源主机产生大量不完整网络流

有外界向被监控用户发送的数据包时，称为缺乏前向数据包的不完整交互行为，所有不完整网络流均由被监控用户接收. 通过这种内外相结合的监控方法，能够全面实现网络异常行为的监控.

2.3 网络异常行为控制方法

利用网络统计量异常变化和不完整网络流所检测到的网络异常源属于网络中的高危用户，它们可能被传播性病毒感染，并在尝试探测新的网络脆弱主机，以进一步扩大病毒的传播范围. 为此，我们提出采用动态隔离的病毒传播控制方法，该方法以“未证明无罪之前即是有罪”的假设为前提，采用动态软隔离方法实现网络异常主机行为的控制. 该隔离方法可以较好地实现传染病的控制，例如几年前成功地应用于SARS病毒在中国传播的控制. 另外，Zou等[16]也应用该方法控制网络蠕虫传播. 理论分析证明，该方法在不影响网络用户正常行为的条件下，可以较好地控制病毒的传播速度，降低由于网络病毒攻击所造成的损失.

动态软隔离方法描述如下：动态地阻断所检测到的网络异常用户的网络连接行为，阻断方法采用动态隔离方法，即在所检测的网络异常源还没有得到安全管理人员确认的情况下，首先阻断所检测到的网络异常源的网络连接，在阻断1 min后，释放对该用户连接行为的阻断，在没有得到安全管理员的确认以前，反复重复此隔离措施. 为了不影响用户的正常网络连接需求，对网络连接请求的阻断措施要尽量地详细，例如，仅阻断相应通信的目的端口或者相应通信的目的地址的网络连接行为. 我们所提出的这种动态软隔离方法具有两方面的好处：首先，不会对用户的正常网络行为产生较大的影响，却可以有效地控制网络攻击传播的速度；其次，对检测方法的误报率有一定的容忍性.

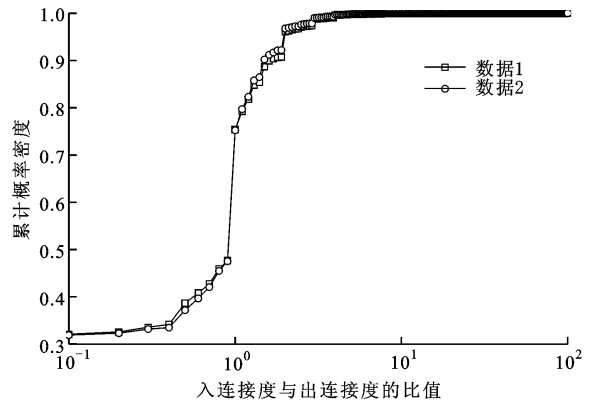
3 实验分析结果

3.1 网络流量样本收集

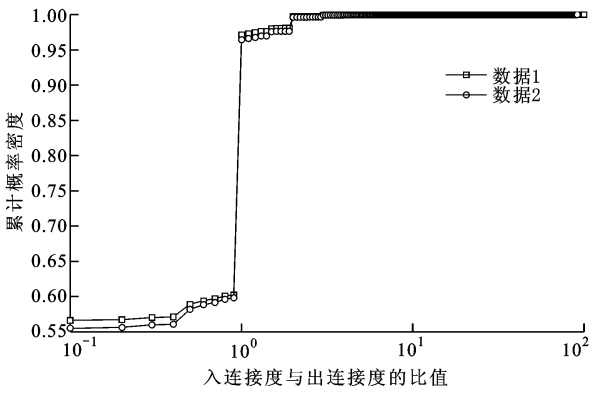
在CERNET（中国教育科研网）西北网络中心设置一个网络流量监控平台，数据采集节点位于教育网西北网络主干网的网络出口路由器处，它是西北科研网络接入教育网的唯一路由器，收集的流量包括入口流量和出口流量，并且收集的流量数据包含用户的全部交互信息. 文中所用的数据是西安交通大学学生区网络出口流量，样本收集于2008年12月1日23:00至12月2日14:50，包含受监控网络用户10 081名，所有用户都具有固定的公网IP地址. 将数据分为两段，数据1以12月1日23:00为起始时刻，数据段较小，网络用户活跃程度低；数据2以12月2日10:00为起始时刻，数据段相对较大，网络用户较活跃.

3.2 网络连接度分析

结合所获取的网络流量特征，采用连接度对称度描述网络主机入连接度和出连接度之间的关系，连接度对称度定义为某台主机的入连接度和出连接度的比值，主机采用IP地址标识. 分析结果如图4所示，图4a是监控源主机的分析结果，图4b是监控目的主机的分析结果. 由累积概率密度函数可以看出，在我们所监控的网络中，大部分主机的连接度对称度近似等于1，即具有相同的入连接度和出连接度，这是比较正常的网络交互行为. 但是，也有很多网络主机的入连接度和出连接度具有较大的差别.



(a)源主机连接度对称度分布图



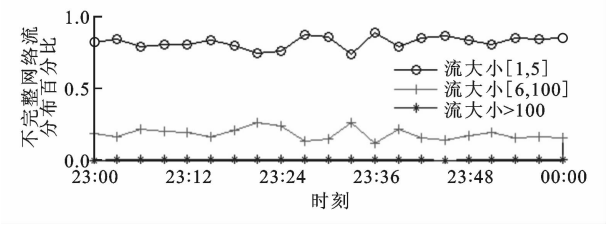
(b)目的主机连接度对称度分布图

图4 网络主机连接度对称度分析图

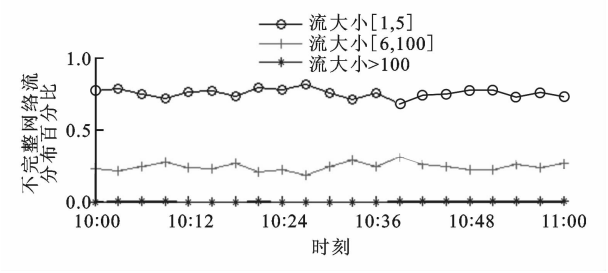
3.3 不完整网络流及其特征分析

我们分析了协议类型为TCP的不完整网络流的大小分布特征，其中网络流大小采用其所携带的数据包个数定义，如图5所示，我们分析了两段数据

中的前 1 h 不完整网络流的分布. 可以看出，80%以上的不完整网络流少于 5 个数据包，实际有 70%的不完整网络流只有 1 个数据包. 这可能是用户偶尔的错误点击或者网速缓慢导致 TCP 没有进行 3 次握手便终止而产生的，是正常现象. 同时，也存在少量的不完整网络流，其所携带的数据包超过了 100 个. 这部分网络流极可能是由网络中最常见的扫描探测攻击造成的，而这类攻击正是网络攻击者寻找攻击目标最常用的方法. 正常网络中每个用户所拥有的不完整网络流数量应该大致相等并且数量较少，这样才能满足不完整网络流是随机事件产物的假设，因而这些较大的不完整网络流应该是来源于网络异常行为.



(a)2008-12-01

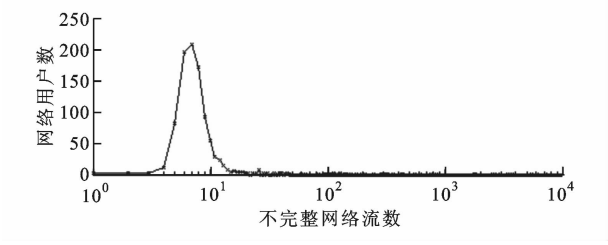


(b)2008-12-02

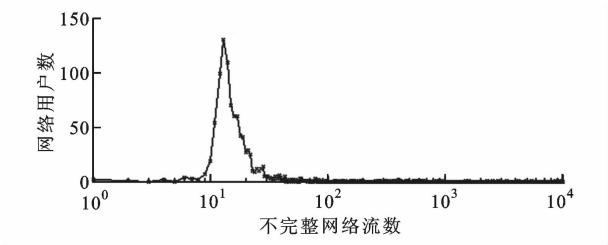
图 5　不完整网络流分布图

3.4　不完整网络流持有者分析

不完整网络流的持有者是指产生了不完整网络流的用户，用户以 IP 地址为标识. 我们忽略了动态 IP 配置情况，这符合西安交通大学生活区的 IP 地址分配情况. 如图 6 所示，我们分析了两段数据 2 h 内协议类型为 TCP 的不完整网络流的持有者状况，可以看出，不完整网络流数为 10 左右的用户很多，并且大致呈正态分布. 大多数不完整网络流的持有者产生较小的不完整网络流，而正态分布的重尾部分则表明较大的不完整网络流主要是由少数的几个特定的 IP 地址产生的. 我们分析了这种不完整网络流在不同用户之间的分布情况，发现在较短时间内(3～4 h)，这种分布具有一定的稳定性，部分特定的网络用户会持续产生不完整网络流. 结合上述分析结果，可以推断这些不完整网络流是因为部分网络用户的异常行为产生的，而不是网络随机性的产物.



(a)2008-12-01



(b)2008-12-02

图 6　不完整网络流持有者分布图

4　sketch 方法性能分析

可以通过两种方式获取网络异常数据源，第 1 种是根据相邻时刻的主机连接度剧变来确定异常时刻，即发生异常的时刻获取异常数据源；第 2 种是测量获取的不完整网络流的特征. 在获取网络异常数据源之后，采用 sketch 方法对网络异常的产生者进行定位分析，为网络流量控制奠定基础.

4.1　sketch 在检测异常时刻具有明显连接度变化的主机中的性能分析

我们分析了前面所设计的连接度 sketch 方法通过第 1 种方式检测异常及定位异常源时的性能，参考文献[8]中的连接度参数为 $H=2$，$\hat{H}=1$，$\boldsymbol{B}_1$ 矩阵大小为 256×216，$\boldsymbol{B}_2$ 矩阵大小为 $256\times(216+1)$，验证矩阵 $\hat{\boldsymbol{B}}_1$ 大小为 $256\times(216-1)$. 阈值指的是连接度变化量的范围，为了估计所选择的阈值对检测结果的影响，我们根据不同的阈值做了相应的实验. 实验分析结果如表 2 所示，在测量分析中，我们采用漏报率(FNR)来衡量本文所提方法的性能. 可以看出，随着阈值的增加，检测准确率逐渐升高，检测漏报率逐渐下降. 由此说明，在合适的阈值下，本文所设计的方法可以准确地获取网络异常变化剧烈的主机，具有较好的实用性.

表2 异常时刻定位到异常源的情况

阈值	异常数量	检测到异常数量	漏报率/%
>100	301	291	3.43
>200	98	96	2.08
>300	57	56	1.78
>400	17	17	0

4.2 sketch在不完整网络流持有者定位中的性能分析

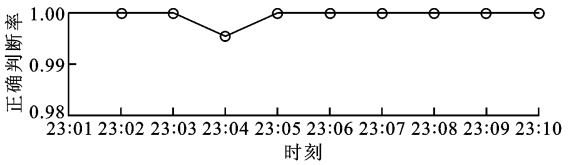
实时监控西安交通大学数个宿舍楼的出口流量,包含网络用户8 108个,具有独立的公网IP地址.监控分析结果如表3所示,不完整网络流数阈值表示要监控的不完整网络流产生者产生的不完整网络流数目下限,主要持有者数表示符合这一阈值特征的用户个数.可以看出,使用sketch不仅在更新过程中有常数时间消耗的性能,在小批量异常集的逆运算过程中也非常快,因而在测量实际网络中的异常用户集时,sketch的逆运算能够迅速完成,内存消耗小且不随数据量的变化而增大.

表3 实时监控网络中不完整网络流产生者定位结果

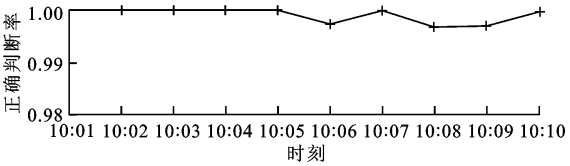
监控时段	流数阈值	主要持有者数	运算耗时/ms
10:00~11:00	>200	116	4 391
12:00~14:00	>500	57	546
15:00~16:00	>800	46	197
19:00~20:00	>1 000	37	172

4.3 交互式网络流模型结合可逆sketch的性能分析

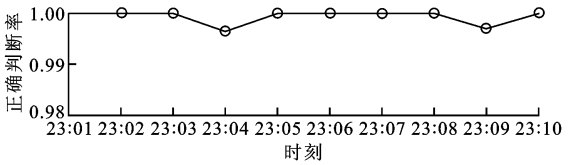
为了验证sketch在不完整网络流统计方面的正确性,我们利用离线数据进行了测量分析.图7给出了可逆sketch对于其存储的交互式网络流模型中不完整网络流持有者标识的测量结果,可以看出,sketch测量的准确率很高,在所测的每段数据的前10 min内,只在个别时间间隔内有漏报现象出现.图7的测试结果分别是缺乏后向数据包和前向数据包的不完整网络流,图7a为被监控的用户产生不完整网络流,图7b为被监控用户接收到的不完整网络流.根据表1的描述,在图7a中这些测量得到的流标识集合中,那些持有大量不完整网络流的被监控用户极有可能产生了表1中后4种异常行为,而图7b中持有大量不完整流的被监控用户可能遭受了前3种攻击.同时,可逆sketch方法还原了图7中全部异常的IP信息,这表示我们不仅能够得到被控用户的信息,而且能够找到攻击被控用户或者被被控用户攻击的IP地址.



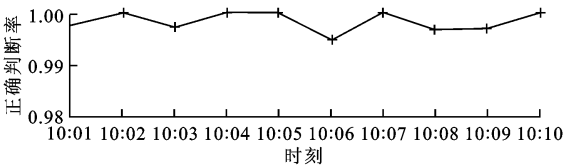
(a)缺少后向数据包(数据1)



(b)缺少后向数据包(数据2)



(c)缺少前向数据包(数据1)



(d)缺少前向数据包(数据2)

图7 不完整网络流产生者定位准确度

5 结 论

本文提出了交互式网络流模型,并用于描述网络终端数据之间的交互特征.在此基础上,对真实样本数据进行了分析,发现存在众多的不完整网络流.通过对不完整网络流特征的分析,发现了不完整网络流与网络异常行为的关系.用可逆sketch数据结构存储交互式网络流模型描述其交互特征,能够准确检测出部分网络异常并且对异常进行准确的定位,为有效控制网络异常行为的扩散奠定了基础.真实网络环境下的实验结果表明,所提方法能够准确地检测及定位异常.在下一步工作中,将主要分析这些不完整网络流产生的原因,即对异常进行分类,为网络异常检测提供更具参考价值的数据源.

参考文献：

[1] MOORE D, SHANNON C. Code-red: a case study on the spread and victims of an internet worm[C]// Proceedings of the 2002 ACM SICGOMM Internet Measurement Workshop. New York, NY, USA: ACM, 2002:273-284.

[2] KIENZLE M, ELDER M. Recent worms: a survey and trends[C]// Proceedings of the ACM CCS Workshop on Rapid Malicious Code. New York, NY, USA: ACM, 2003:1-10.

[3] STANIFORD S, PAXSON V, WEAVER N, et al. How to own the internet in your spare time [C]// Proceedings of the 11th USENIX Security Symposium. Berkeley, CA, USA: USENIX Association, 2002:149-167.

[4] DOULIGERIS C, MITROKOTSA A. DDoS attacks and defense mechanism: classification and state of the art [J]. Computer Networks, 2004, 44(4):643-666.

[5] PENG Tao, LECKIE C, RAMAMOHANARAO K. Survey of network based defense mechanisms: countering the DoS and DDoS problems [J]. ACM Computing Survey, 2007, 39(1): 1-42.

[6] 肖志新，杨岳湘，杨霖. 一个基于 NetFlow 的异常流量检测与防护系统[J]. 微电子学与计算机，2006，23(5):209-213.
XIAO Zhixin, YANG Yuexiang, YANG Lin. An anomaly traffic detection and network defending system based on NetFlow [J]. Microelectronics and Computer, 2006, 23(5):209-213.

[7] KIM M S, KONG H J, HONG S C, et al. A flow-based method for abnormal network traffic detection [C]// Proceedings of the Network Operations and Management Symposium. Piscataway, NJ, USA: IEEE, 2004:599-612.

[8] KRISHNAMURTHY B, SEN S, ZHANG Yin, et al. Sketch-based change detection: methods, evaluation, and applications[C]// Proceedings of the ACM SIGCOMM Internet Measurement Conference. New York, NY, USA: ACM, 2003:234-247.

[9] GIBBONS P B, MATIAS Y. Synopsis structures for massive data sets[C]// Proceedings of the 10th Annual ACMSIAM Symposium on Discrete Algorithms. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1999: 909-910.

[10] MUTHUKRISHNAN S. Data streams: algorithms and applications [M]. Boston, MA, USA: Now Publishers Inc., 2003.

[11] 冯文峰，黄永峰，李星. 可逆概要数据结构[J]. 清华大学学报，2008，48(10):1625-1628.
FENG Wenfeng, HUANG Yongfeng, LI Xing. Reversible sketch data structure [J]. Journal of Tsinghua University, 2008, 48(10):1625-1628.

[12] SCHWELLER R, LI Zhichun, CHEN Yan, et al. Reversible sketches: enabling monitoring and analysis over high-speed data streams [J]. Transactions on Networking, 2007, 15(5):1059-1072.

[13] ZHAO Qi, KUMAR A, XU Jun. Joint data streaming and sampling techniques for detection of super sources and destinations[C]// Proceedings of ACM SIGCOMM Internet Measurement Conference. New York, NY, USA: ACM, 2005:77-90.

[14] GUAN Xiaohong, WANG Pinghui, QIN Tao. A new data streaming method for locating hosts with large connection Degree [C]// Proceedings of IEEE Global Communications Conference. Piscataway, NJ, USA: IEEE, 2009:6421-6426.

[15] STALLINGS W. Cryptography and network security: principles and practice [M]. 4th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[16] ZOU C C, GONG Weibo, TOWSLEY D. Worm propagation modeling and analysis under dynamic quarantine defense. [C]// Proceedings of ACM Workshop on Rapid Malcode. New York, NY, USA: ACM, 2003: 51-60.

[本刊相关文献链接]

丁要军，蔡皖东. 采用两阶段策略模型(KTSVM)的 P2P 流量识别方法. 2012,46(2):45-50.

侯重远，江汉红，芮万智，等. 工业网络流量异常检测的概率主成分分析法. 2012,46(2):70-75.

任浩，王劲林，尤佳莉. 采用节点优先级的对等网络流媒体请求量分配算法. 2011,45(12):10-15.

褚伟波，蔡忠闽，管晓宏，等. 采用子网流量组合优化的网络流量分割方法. 2011,45(12):22-27.

陈刚，蔡远利，穆静，等. 海量信息异常检测问题的异常概率排序算法. 2011,45(4):36-40.

崔筱宁，赵保华，李青，等. 传感器数据中事件样本与错误样本的系统化区分框架. 2010,44(10):30-35.

颜若愚，郑庆华. 使用交叉熵检测和分类网络异常流量. 2010,44(6):10-15.

颜若愚，郑庆华，牛国林. 自适应滤波实时网络流量异常检测方法. 2009,43(12):1-5.

（编辑　武红江）