基于抽样流长与完全抽样阈值的异常流自适应抽样算法

伊 鹏 钱 坤* 黄万伟 王 晶 张 震
(国家数字程控交换系统工程技术研究中心 郑州 450002)

摘 要：高速IP网络的流量测量与异常检测是网络测量领域研究的热点。针对目前网络流量测量算法对小流估计精度偏低，对异常流量筛选能力较差的缺陷，该文提出一种基于业务流已抽样长度与完全抽样阈值 S 的自适应流抽样算法(AFPT)。AFPT算法根据完全抽样阈值 S 筛选对异常流量敏感相关的小流，同时根据业务流已抽样长度自适应调整抽样概率。仿真和实验结果表明，AFPT 算法的估计误差与理论上界相符，具有较强的异常流量筛选能力，能够有效提高异常检测算法的准确率。

关键词：网络测量；自适应流抽样；异常检测
中图分类号： TP393 **文献标识码**： A **文章编号**：1009-5896(2015)07-1606-06
DOI: 10.11999/JEIT141379

Adaptive Flow Sampling Algorithm Based on Sampled Packets and Force Sampling Threshold S Towards Anomaly Detection

Yi Peng Qian Kun Huang Wan-wei Wang Jing Zhang Zhen
(*National Digital Switching System Engineering Technological R&D Center*, *Zhengzhou* 450002, *China*)

Abstract: The network traffic measurement and anomaly detection for high-speed IP network become the hotspot research of network measurement field. Because the current measurement algorithms have large estimation error for the mice flows and poor performance for the sampling anomaly traffic, an Adaptive Flow sampling algorithm based on the sampled Packets and force sampling Threshold S (AFPT) is proposed. According to the force sampling threshold S, the AFPT is able to sample the mice flows which is sensitive to the anomaly traffic, while adaptive adjustment the probability of sampling based on the sampled packets. The simulation and experimental results show that the estimation error of AFPT is consistent with the theoretical upper bound, and provide better performance for the anomaly traffic sampled. The proposed algorithm can effectively improve the accuracy of anomaly detection algorithm.

Key words: Network measurement; Adaptive flow sampling; Anomaly detection

1 引言

网络基础通信设施的大规模部署和网络接入方式的开放性，使得互联网成为一种高度异构与开放的复杂系统[1]。通过网络流量测量技术，可以帮助人们理解掌握网络运行状况，进而优化网络结构和网络应用。网络上的数据报文流经测量节点后，根据系统测量算法必须进行数据压缩[2]或者抽样[3]以减少流量日志，并将测量流量按照一定的数据结构格式存储。测量过程完成后，得到的流量数据可以进一步用于流量分布特征估计、流量计费以及异常检测[4,5]等应用分析。

流量测量时由于大流较小流更容易被抽样检测，因此大流的测量难点主要在于优化数据存储结构或抽样模型，小流由于数量的庞大性和较低的抽样率，与大流的测量估计往往不能兼得[6]。网络研究和管理者仅使用部分原始流量对网络流量的异常检测进行分析，但由于原始流量信息的不完整性使得异常检测算法的分析结果在准确率上不可避免地存在一定偏差[7,8]。针对上述问题，一种理想的流量测量算法首先应当能够为流量分布特征提供较高估计精度信息，同时对异常流量具有准确的筛选抽样能力，或者是具有能够在测量中保留大量与异常流量敏感相关的小流数据的能力。

草图指导的抽样(Sketch Guided Sampling, SGS)算法[9]基本能够解决上述问题，该算法抽样保留每流(per-flow)信息，设置包抽样率为该数据包所属业务流当前流量的单调递减函数。随机共享计数器(Randomized Counter Sharing, RCS)算法[10]是一种较好地满足以上要求的算法，其核心思想是：从

2014-10-29 收到，2015-01-13 改回，2015-05-11 网络优先出版
国家 973 计划项目(2012CB315901, 2013CB329104)资助课题
*通信作者：钱坤 qiank126@126.com

流量数据的存储优化出发，使用 m 组计数器压缩存储所有数据信息，将不同业务流的数据信息随机存储在 $l(0<l<m)$ 组计数器内。文献[11]提出了一种基于网络流长分布的自适应抽样(Flow Size Adaptive Sampling, FSAS)算法，FSAS 通过估计异常流量的分布特征，得到异常流量的流长阈值，根据阈值对流量进行分段抽样，其自适应抽样的实质是分段的静态流抽样算法。文献[12]提出了一种特征感知的自适应流抽样(Adaptive Flow Sampling, AFS)方法，根据流量五元组特征定义了 2 个特征矩——特征计数和特征熵，在采样前计算特征矩，在抽样时一旦某一特征值出现频度过高立即降低该业务流的抽样率。

本文针对当前流量测量算法对网络流统计特征估计误差偏高、对异常流量抽样能力偏弱的问题，提出一种基于业务流抽样流长与完全抽样阈值 S 的自适应流抽样算法(Adaptive Flow sampling algorithm based on sampled Packets and force sampling Threshold S, AFPT)。第 2 节对 AFPT 算法的抽样模型给出了详细描述，并从理论上对 AFPT 算法的流长度无偏估计、标准差以及存储开销进行了分析证明。第 3 节采用真实链路数据和模拟攻击流量对 AFPT 算法进行仿真测试，实验结果与第 2 节的理论分析吻合。第 4 节总结全文。

2　自适应流抽样算法

2.1 抽样模型

与静态报文抽样算法相比，AFPT 算法的抽样概率函数 $P(s)$ 根据业务流已被抽样的报文分组数量动态调整抽样概率。在最大化降低抽样流量的前提下，AFPT 算法为了获得原始流量的最大信息量，$P(s)$ 随 s 的增加而减小，对于大流采用较小的抽样概率，对于小流采用较大的抽样概率。对于未被抽样的业务流，为保证流间抽样的公平性，采用完全抽样方式以概率 1 进行抽样。对于流长度小于完全抽样阈值 S 的业务流，为保证异常流量的完整性，同样采用完全抽样方式进行抽样。定义 1 给出了 AFPT 算法所使用抽样概率函数 $P(s)$ 的一般定义形式和满足条件。

定义 1　抽样概率函数 $P(s)=1/(p(s+1)-p(s))$，其中抽样函数 $p(s)$ 满足条件：

(1) $p(s)$ 是实数空间内的递增凸函数；

(2) $p(s)$ 满足 $p(0)=0$ 且 $p(1)=\beta$，$\beta>0$，β 可用于调整完全抽样阈值 S；

(3) $p(s+1)<\alpha p(s)+\beta$，$\alpha>1$，$\beta>0$。

2.2 理论分析

定理 1　使用 AFPT 算法抽样时，原始业务流长度 n 的无偏估计是 $\hat{n}(s)=\begin{cases}p(s), & n>S_0\\ s, & n\le S_0\end{cases}$。

证明　对于流长度真实值为 n 的业务流，若计数器计数值 $s=i$，有：

(1)完全抽样阈值 $S|_{\beta=1}=1$ 时，由于 AFPT 算法对流长度为 1 的业务流采用完全抽样，故只需证明 $\hat{n}(s)=p(s)$ 即可。

由概率统计可得：$Q_i(n)=Q_{i-1}(n-1)P(i-1)+Q_i(n-1)[1-P(i)]$，且有 $Q_n(n-1)=0$。

令 $L(n)$ 表示业务流长度的真实值为 n 时，无偏估计 $\hat{n}(s)$ 的期望值，则

$$L(n)=E[p(c)]=\sum_{i=0}^{n}p(i)Q_i(n) \tag{1}$$

由式(1)可得

$$\begin{aligned}L(n)-L(n-1)&=\sum_{i=1}^{n}p(i)[Q_{i-1}(n-1)P(i-1)\\&\quad+Q_i(n-1)(1-P(i))]\\&\quad-\sum_{i=1}^{n-1}p(i)Q_i(n-1)[P(i)+(1-P(i))]\\&=\sum_{i=2}^{n}p(i)Q_{i-1}(n-1)P(i-1)\\&\quad-\sum_{i=1}^{n-1}p(i)Q_i(n-1)P(i)\\&=\sum_{i=1}^{n-1}[p(i+1)-p(i)]Q_i(n)P(i)\end{aligned} \tag{2}$$

根据定义 1 可得 $p(i+1)-p(i)=1/P(i)$，则有

$$L(n)-L(n-1)=\sum_{i=1}^{n-1}Q_i(n-1)=1 \tag{3}$$

$$L(n)=\sum_{i=1}^{n}[L(i)-L(i-1)]+L(0)=n \tag{4}$$

故

$$E[\hat{n}(s)]=E[p(s)]=L(n)=n \tag{5}$$

(2)完全抽样阈值 $S|_{\beta}=S_0>1$ 时，根据上述推导，得

$$L(n)=\begin{cases}E[p(c)]=\sum_{i=S_0}^{n}p(i)Q_i(n)+S_0, & n>S_0\\ s, & n\le S_0\end{cases}$$

若 $n>S_0$，则

$$\begin{aligned}L(n)-L(n-1)&=\sum_{i=S_0}^{n-1}[p(i+1)-p(i)]Q_i(n)P(i)\\&=\sum_{i=S_0}^{n}Q_i(n-1)=1\end{aligned}$$

$$L(n)=\sum_{i=S_0+1}^{n}[L(i)-L(i-1)]+S_0=n$$

$$E[\hat{n}(s)]=\begin{cases}E[p(s)]=L(n)=n, & n>S_0\\ s, & n\le S_0\end{cases}$$

综合上述推导定理 1 得证。 证毕

定理 2 使用 AFPT 算法抽样时，如果取 $\hat{n}(s)=p(s)$ 作为无偏估计，则标准差的上界是 $\sqrt{\frac{\alpha-1}{2}+\frac{2\beta-\alpha-1}{2n}+\frac{S_0}{n^2}\left[\frac{(\alpha+1)(1-S_0)}{2}-\beta+S_0\right]}$。

证明 对流长度 $n\le S_0$ 的业务流，估计误差始终为 0。

对流长度 $n>S_0$ 的业务流，由 $E[p^2(s)]=\sum_{i=S_0}^{n}p^2(i)Q_i(n)+S_0^2$ 和定义 1 可得：

$$\begin{aligned}
&H(n)-H(n-1)\\
&\quad=\sum_{i=S_0}^{n}(p^2(s))[Q_{i-1}(n-1)P(i-1)\\
&\qquad+Q_i(n-1)(1-P(i))]\\
&\qquad-\sum_{i=S_0}^{n-1}(p^2(s))Q_i(n-1)[P(i)+(1-P(i))]\\
&\quad=\sum_{i=S_0}^{n-1}Q_i(n-1)[p(i+1)+p(i)]\\
&\quad\le\sum_{i=S_0}^{n-1}Q_i(n-1)[(\alpha+1)p(i)+\beta]\\
&\quad=(\alpha+1)(n-1)+\beta
\end{aligned}\tag{6}$$

进一步可得：

$$\left.\begin{aligned}
H(n)&=\sum_{i=S_0+1}^{n}[H(i)-H(i-1)]+H(S_0)\\
&\le\frac{(\alpha+1)n^2+(2\beta-\alpha-1)n}{2}\\
&\quad+S_0\left[(\alpha+1)(1-S_0)/2-\beta+S_0\right]\\
\mathrm{Var}[\hat{n}(c)]&=H(n)-L^2(n)\\
&\le\frac{(\alpha-1)n^2+(2\beta-\alpha-1)n}{2}\\
&\quad+S_0\left[(\alpha+1)(1-S_0)/2-\beta+S_0\right]
\end{aligned}\right\}\tag{7}$$

AFPT 算法的标准差上界为

$$\frac{\sqrt{\mathrm{Var}[\hat{n}(c)]}}{n}\le\sqrt{\frac{\alpha-1}{2}+\frac{2\beta-\alpha-1}{2n}+\frac{S_0}{n^2}\left[\frac{(\alpha+1)(1-S_0)}{2}-\beta+S_0\right]}\tag{8}$$

证毕

表 1 是 SGS, RCS 和 AFPT 算法的抽样概率函数与理论标准差。图 1(a)~图 1(c)分别是 3 种算法的标准差对比。SGS 算法和 RCS 算法的理论误差相近，误差均随流长度的增加而减小，AFPT 算法的理论误差随流长度的增加略有增大，但误差最大值仍然小于 0.071。

表 1 理论误差

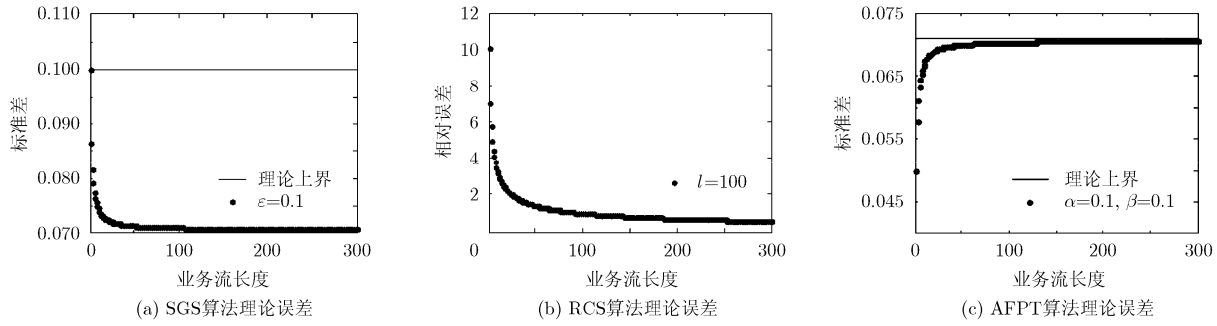
算法	抽样概率函数	标准差
SGS	$1/(1+\varepsilon^2 i)$	$\varepsilon\sqrt{(n+1)/2n}\le\varepsilon$
RCS	–	$\sqrt{\frac{(l-1)}{n}+\frac{l^2(1-1/m)k}{n^2m}}$
AFPT	$1/(\beta\alpha^n)$	$\le\sqrt{\frac{\alpha-1}{2}+\frac{2\beta-\alpha-1}{2n}}$

说明：RCS 算法是数据流压缩算法，无抽样函数。

2.3 存储开销

抽样算法的主要存储开销集中在流长度计数器的使用上。为尽可能保留原始流量信息，抽样算法通常采用每流抽样模型，为每个业务流创建流长度计数器；在计数器的设计上，由于业务流长度的差异跨度非常大，抽样算法为保证所有业务流的统计需求，必须根据最大业务流长度值确定计数器位宽。

定理 3 对真实大小为 n 的业务流，AFPT 算法计数器位宽的数学期望上界是 $p^{-1}(n)$，其中 $p^{-1}(n)$ 是 $p(c)$ 的反函数。



(a) SGS算法理论误差 (b) RCS算法理论误差 (c) AFPT算法理论误差

图 1 3 种算法的理论误差

证明　根据定义 1 中条件(1)，$p(c)$ 是凸函数且满足 $p(x) \geq p(y) + (x-y)p_r^{'}(y)$，$\forall x, y > 0$。

其中，$p_r^{'}(\bullet)$ 是 $p_r(\bullet)$ 的右导数。令 $x = c$，$y = E[c]$ 可得

$$p(c) \geq p(E[c]) + (c - E[c])f_r^{'}(E[c])$$

$$E[p(c)] \geq E[p(E[c]) + (c - E[c])p_r^{'}(E[c])]$$

$$E[p(c)] = n \geq f(E[c])$$

由 $p(c)$ 是递增函数可得 $E[c(n)] \leq p^{-1}(n)$。　　证毕

SGS 算法需要的计数器位宽是 $\lceil lb(n+1) \rceil$（“$\lceil\bullet\rceil$”表示上取整），RCS 算法所需的计数器位数与测量流量的报文总数及计数器组的数量有关，但一般取 9 bit 即可满足流长度在 6000 以内的业务流，AFPT 算法的计数器位宽是 $\lceil lb(p^{-1}(n)) \rceil$。

图 2 对比了 SGS 算法与 AFPT 算法所需计数器位宽，对于流长度在 100 以内的小流，SGS 算法和 AFPT 算法的计数器位宽差距较小。随着业务流长度的增加，SGS 算法的计数器位宽成对数比例增加，AFPT 算法的计数器位宽趋近于 10 且此时已能够满足流长度在 10^7 以内的所有业务流。即使在提高完全抽样阈值 $S|_{\beta=0.61} = 50$ 时，AFPT 算法的计数器位宽仍然保持趋近于 10，存储开销未出现大幅增加。与 RCS 算法相比，在小流与中大流计数器位宽的使用上，AFPT 算法优于 RCS 算法，随着流长度的增加，AFPT 算法所需的计数器位宽趋近于 10 略大于 RCS 算法。

3　仿真与结果分析

3.1　真实网络流量下的仿真实验

本文使用 CAIDA[13] 在 2012 年 3 月 10 日采集的互联网实际网络 40 Gbps 链路的流量数据。仿真过程中，SGS 算法参数设置为 $\varepsilon = 0.1$，计数器位宽为 14；AFPT 算法抽样概率 $P(s) = 1/\beta\alpha^s$，$\alpha = 1.01$，$\beta = 1$；RCS 算法的计数器组数目 $m = 10^6$，随机映射计数器数目 $l = 100$，计数器位宽 $b = 8$。

AFPT 算法抽样仿真结果与标准差如图 3 所示。对图 3(b)中的标准差数据点进行统计，结果如表 2 所示，AFPT 的标准差稳定在 0.071 以内，97.67% 的数据点误差范围在 0.071 以内，个别小流的标准差稍大于 0.07 但原始流量的整体误差仍然小于 0.5。随着业务流长度的增加，标准差完全落在区间[0,0.071]以内，这与定理 2 的标准差理论上界保持一致。

表 2　3 种算法标准差的区间比例

算法	误差区间	比例(%)
AFPT	<0.071	97.67
	[0.071,0.1)	2.18
	[0.1,0.5)	0.15
SGS	<0.100	82.64
	[0.1,1]	17.36
RCS	<0.300	81.28
	[0.3,0.5)	15.64
	[0.5,1)	2.50

图 4 是 SGS 算法的流长度估计值与标准差，估计误差显著高于 AFPT 算法。尽管大流的估计误差较低，但小流的估计误差是 AFPT 的数十倍以上。由于 SGS 使用哈希引入了误判误差，SGS 的标准差在 0.1 以上的比例超过了 17.36%，且小流的估计精度差于 AFPT，标准差上界较 AFPT 高出约 30%。

RCS 算法的流长度估计值与标准差如图 5 所示。与 SGS 相比，RCS 对小流的标准差较大，当业务流长度达到 100 左右时标准差基本稳定在 1 以内，且快速降低趋近于 0。RCS 约 81.28%的标准差在 0.3 以内，稍差于 SGS。而 AFPT 仿真结果的标准差仅在 0.01 以内的比例就高达 99.85%，且 AFPT 算法的标准差基本落在 0.071 以内，误差上界较 RCS 降低约 76.3%。

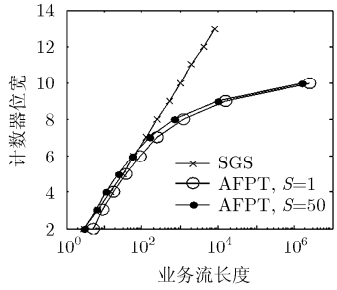
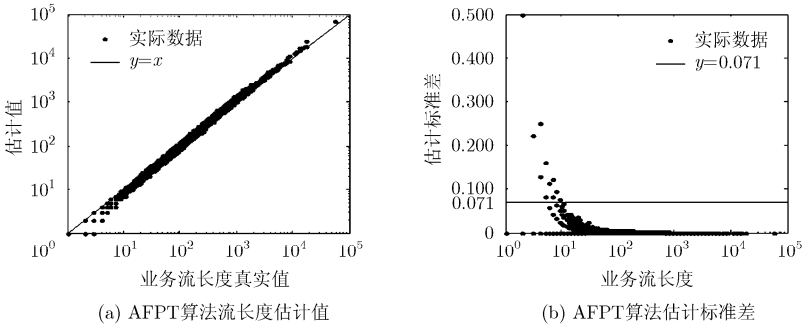
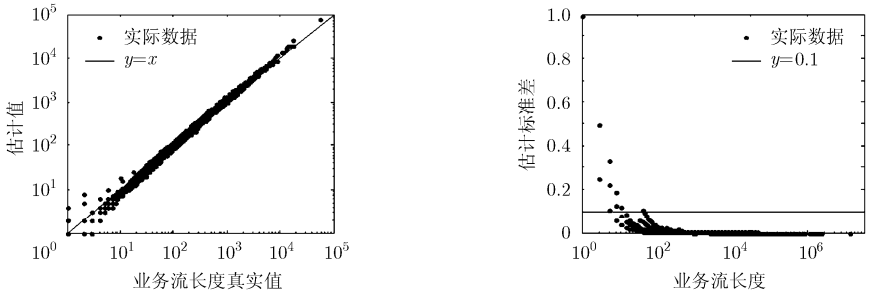


图 2 报文抽样与 AFPT 算法需要的计数器位数



(a) AFPT算法流长度估计值　　(b) AFPT算法估计标准差

图 3　AFPT 算法仿真结果



(a) SGS算法流长度估计值 (b) SGS算法估计标准差

图 4 SGS 算法仿真结果

3.2 异常流量下的仿真实验

在 DoS、端口扫描或者蠕虫攻击等网络攻击下会使网络链路产生大量的小流，抽样测量时如果能够对这些小流尽可能地抽样就可以获得更多的异常流量数据，为异常检测算法的准确率提供更加可靠的保证。由于不同类型攻击流的流长并没有相关的明确定义，综合考虑存储资源的消耗，本文选择设置完全抽样阈值 $S=50$ 以满足大多数攻击流都可被抽样，实际阈值的设置视需求而定。

AFPT 的抽样概率函数 $P(s)$ 值取决于 α 与 β，对于不同的取值组合 $P(s)$ 可能出现大于 1 的情况。对 $P(s)$ 取值大于 1 的概率点，均以概率 1 进行完全抽样。若取 $\beta\in[0,1]$ 间任意值，可调整完全抽样阈值 S，将需要深入研究的业务流全部保留，仅对流长度超过阈值 S 的业务流进行抽样测量。图 6 是 $\alpha=1.01$ 时，完全抽样阈值 S ($S=\lfloor\log_\alpha(1/\beta)\rfloor$)随 β 的变化曲线。

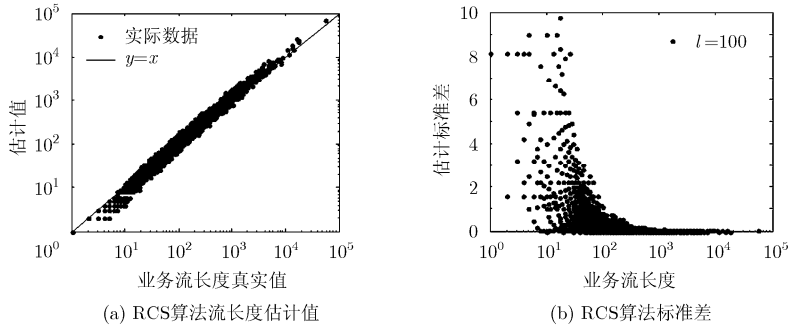
图 7(a)是 SGS, RCS 和 AFPT 这 3 种算法对合成流量的测量仿真结果，SGS 和 RCS 算法的参数设置与 3.1 节的仿真实验一致，AFPT 算法取完全抽样阈值 $S|_{\beta=0.61}=50$。异常流量的合成采用文献[14]提出的构造方法，基于实验中使用的正常情况下网络数据流量和 DARPA 入侵检测流量[15]，模拟异常流量的攻击强度为每秒 30000 包，其中异常流量比例约 20.8%，攻击流量在 100 s, 200 s, 300 s 时逐渐达到 60%。

SGS 的异常流量抽样比例接近于 0.7，最差约 0.4。由于 RCS 对所有数据报文采用了数据压缩方式存储，对异常流量的抽样原则仅与压缩率有关，比例值稳定在 0.8 左右，即使在过渡压缩数据信息时的抽样比例降低至约 0.6，仍优于 SGS。AFPT 通过设置完全抽样阈值，抽样比例显著优于 SGS 和 RCS，比例值稳定在 1.00 左右，在攻击流量较为集中时也能够保证抽样比例不小于 0.75，基本实现对异常流量的逐包抽样统计。

AFPT 与 AFS, FSAS 算法的比较结果如图 7(b)所示，其中，FSAS 的参数设置为：大中小流阈值分别是 1000, 500, 50，抽样概率分别是 0.01, 0.10, 0.20。由于 FSAS 需要在抽样的同时估计业务流长度，极其耗费计算资源，因此在攻击强度增大时，对小流的估计偏差也相应增大，导致其抽样性能下降非常明显。AFS 的抽样效果略差于 AFPT，这是由于 AFS 的特征矩是根据流五元组特征信息定义的，而抽样攻击流量中的 Probe 与 U2R 等攻击流仅仅依靠这些特征信息是不够的。

4 结束语

高速 IP 网络的流量测量在抽样与异常检测之间，一直存在由于抽样导致的异常检测偏差。本文提出一种可设置完全抽样阈值的抽样算法，在对网



(a) RCS算法流长度估计值 (b) RCS算法标准差

图 5 RCS 算法仿真结果

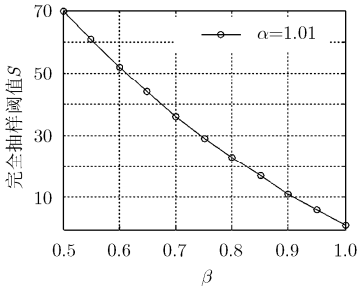
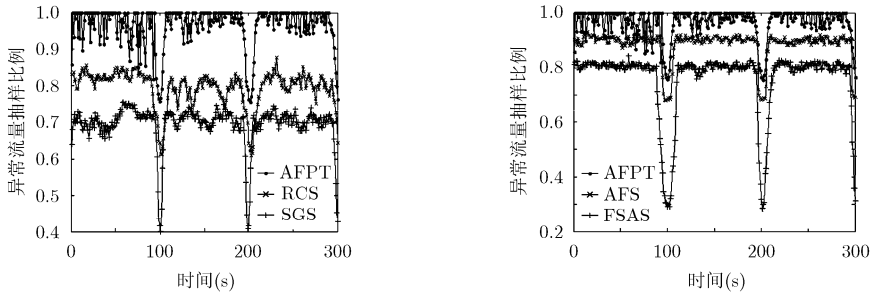


图 6 完全抽样阈值 S 的变化曲线



(a) AFPT, RCS, SGS算法仿真结果　　(b) AFPT, AFS, FSAS算法仿真结果

图 7 异常流量抽样比例

络业务流抽样估计原始流量分布特征的同时，将与异常流量敏感相关的小流完全抽样，使得异常检测处理的流量数据中包含了至少 75%以上的异常流量，有助于提高检测算法的准确率。该算法将抽样流量的估计误差上界降低了 30%以上，在存储开销上降低了对计数器位宽的需求，与其他每流测量算法相比所需的存储开销基本持平并未明显增加。通过实验仿真的进一步验证，该算法完全能够适用于 40 Gbps 速率的骨干链路，对大型骨干网络的管理规划具有重要意义。

参 考 文 献

[1] Zhou Ai-ping, Cheng Guang, and Guo Xiao-jun. High-speed network traffic measurement method[J]. *Journal of Software*, 2014, 25(1): 135−153.

[2] Peter Lieven and BjörnScheuermann. High-speed per-flow traffic measurement with probabilistic multiplicity counting [C]. Proceedings of the INFOCOM 2010, San Diego, CA, USA, 2010: 1−9.

[3] Cheng Guang and Tang Yong-ning. Estimation algorithms of the flow number from sampled packets on approximate approaches[J]. *Journal of Software*, 2013, 24(2): 255−265.

[4] Lee Y J, Yeh Y R, and Wang Y C F. Anomaly detection via online oversampling principal component analysis[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2013, 25(7): 1460−1470.

[5] Pham D S, Venkatesh S, Lazarescu M, *et al*.. Anomaly detection in large-scale data stream networks[J]. *Data Mining and Knowledge Discovery*, 2014, 28(1): 145−189.

[6] Cai Yuan-jun, Wu Bin, Zhang Xin-wei, *et al*.. Flow identification and characteristics mining from internet traffic with hadoop[C]. Proceedings of the Computer Information and Telecommunication Systems (CITS), Jeju Island, Korea, 2014: 1−5.

[7] Brauckhoff D, Tellenbach B, Wagner A, *et al*.. Impact of packet sampling on anomaly detection metrics[C]. Proceedings. of the 6th ACM Sigcomm conference on Internet measurement, Rio de Janeiro, Brazil, 2006: 159−164.

[8] Mai Jian-ning, Chuah C N, Sridharan A, *et al*.. Is sampled data sufficient for anomaly detection?[C]. Proceedings of the 6th ACM Sigcomm Conference on Internet Measurement, Rio de Janeiro, Brazil, 2006: 165−176.

[9] Kumar A and Xu J. Sketch guided sampling using on-line estimates of flow size for adaptive data collection[C]. Proceedings of IEEE INFOCOM 2006, Barcelona, Spain, 2006: 1−11.

[10] Li Tao and Chen Shi-gang. Per-flow traffic measurement through randomized counter sharing[J]. *IEEE ACM Transactions on Networking*, 2012, 13(5): 325−336.

[11] 王苏南. 高速复杂网络环境下异常流量检测技术研究[D]. [博士论文], 信息工程大学, 2012:38-49.
Wang Su-nan. Research on anomaly detection technology in high-speed complex network environment[D]. [Ph.D. dissertation], The PLA Information Engineering University, 2012: 38−49.

[12] 郭通. 基于自适应流抽样测量的网络异常检测技术研究[D]. [博士论文], 信息工程大学, 2013: 38−49.
Guo Tong. Research on network anomaly detection technology based on adaptive flow sampling measurement[D]. [Ph.D. dissertation], The PLA Information Engineering University, 2013: 38−49.

[13] CAIDA. Cooperative as-sociation for internet data analysis [OL]. http://www.caida.org/data, 2012.

[14] Lakhina A, Crovella M, and Diot C. Mining anomalies using traffic feature distributions[C]. Proceedings of the 5th ACM Sigcomm Conference on Internet Measurement, Philadelphia, PA, USA, 2005: 217−228.

[15] MIT Lincoln Laboratory. DARPA Intrusion Detection Evaluation[OL]. http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html, 1999.

伊　鹏：　男，1977 年生，副教授，博士，研究方向为宽带信息网络.

钱　坤：　男，1990 年生，硕士生，研究方向为宽带信息网络、网络测量、异常流检测.

黄万伟：　男，1979 年生，讲师，博士，研究方向为宽带信息网络.

王　晶：　女，1980 年生，讲师，博士生，研究方向为可重构网络与网络测量技术.

张　震：　男，1985 年生，讲师，博士，研究方向为网络测量技术.