

doi:10.3969/j.issn.1001-2400.2013.03.024

一种改进的高速网络分布式流量抽样算法

孙 昱¹, 蒋馥蔚¹, 夏靖波¹,
赵小欢¹, 申 健¹, 李明辉²

(1. 空军工程大学信息与导航学院, 陕西 西安 710077;

2. 空军后勤部, 北京 100720)

摘要: 针对基于掩码匹配的抽样算法中使用某些抽样掩码会造成较大测量误差的问题, 提出了一种增强样本随机性的改进方案. 该方案首先对标识字段进行异或运算, 以得到新的标识值; 然后再进行匹配抽样. 从理论上证明了异或运算能有效提高结果的随机性, 并用实测流量数据进行了统计验证. 最后, 从有效性和稳定性两个方面对改进算法进行了检验. 实验结果表明, 改进算法的各项评价指标均优于文献[10]算法, 任意的抽样掩码均能在改进算法中使用.

关键词: 分布式抽样; 掩码匹配; 标识字段; 随机性; 异或运算

中图分类号: TP393 **文献标识码:** A **文章编号:** 1001-2400(2013)03-0160-06

Improved distributed traffic sampling algorithm for high speed network

SUN Yu¹, JIANG Fuwei¹, XIA Jingbo¹,
ZHAO Xiaohuan¹, SHEN Jian¹, LI Minghui²

(1. School of Information and Navigation, Air Force Engineering Univ., Xi'an 710077, China;

2. Air Force Logistics Department, Beijing 100720, China)

Abstract: The sampling algorithm based on mask matching is a distributed sampling measurement technology which can be used in a high speed network. But there exists a problem that some masks used in the algorithm result in a large measurement error. In order to solve the problem, the paper proposes an improved scheme that aims to enhance the randomness of its sample. The idea of the scheme is to put the identification field into xor operation to get a new identity value before mask matching and sampling. Firstly, the paper demonstrates theoretically that the xor operation can enhance its computed result's randomness effectively. Then it is checked statistically with the realistic traffic data. Finally, the improved algorithm is tested from two aspects of effectiveness and stability. Experimental result shows that the evaluation indexes of the improved algorithm are better than those of the original algorithm. Hence, any sampling mask can be used in the improved algorithm.

Key Words: distributed sampling; mask matching; identification field; randomness; xor operation

网络流量测量和分析是研究网络行为的基础, 其对于网络问题的解决、协议的调试和网络性能评估等方面均有极大的帮助. 但是随着近年来高速网络技术的发展, 传统的全报文、单点的网络流量测量方法不仅难以实施, 而且已经无法全面地反映出网络的性能. 因此, 研究面向高速网络的分布式抽样测量算法势在必行.

1993年, Claffy等^[1]对采用不同触发机制的系统抽样、分层抽样、随机抽样进行了分析对比, 结果表明,

收稿日期: 2012-09-27

网络出版时间: 2013-02-25

基金项目: 陕西省科技计划资助项目(2012JZ8005); 全军军事学研究生课题资助项目(2010XXXX-488)

作者简介: 孙 昱(1989-), 男, 空军工程大学博士研究生, E-mail: 458657673@qq.com.

网络出版地址: <http://www.cnki.net/kcms/detail/61.1076.TN.20130225.1050.201303.193.024.html>

基于计数触发机制的抽样算法要普遍优于基于时间触发机制的抽样算法. 互联网工程任务组通过对这几种传统抽样算法进行性能分析, 在 RFC2330 文档中推荐使用泊松抽样来实现网络流量测量. 为了能更准确地测量网络的性能特征, 自适应抽样算法^[2-4]逐渐发展了起来. 该类抽样方法利用网络流量的相关性预测流量状态, 并通过实时调整抽样策略或参数, 更好地捕捉网络的流量特征. 虽然以上抽样算法得到的样本随机性较好, 但是它们都只是适用于单点测量, 无法满足分布式测量环境对抽样报文的一致性的要求. Cozzani 等^[5]在测量异步传输模式(ATM)网络位元时提出了使用掩码来匹配校验和字段以实现抽样, 但是由于校验和字段在传输过程中会不断发生变化, 因此同样不满足分布式测量的要求. Duffield 等^[6-7]将哈希函数引入到抽样测量中, 该算法将 IP 报文在传输过程中不变且随机性较好的字段映射到一个观察域中, 依据事先选定的抽样域决定是否抽取该报文. 这样虽然能满足对报文一致性的要求, 但是报文样本的随机性好坏不仅取决于选定的报文字段, 还取决于所使用的哈希函数. 目前公认能产生随机性序列较好的哈希函数有 MD5, SHA1, CRC32 等算法, 但是这些算法的时间复杂度很高, 不适合高速的网络测量环境. 潘乔等^[8-9]将哈希函数应用于 IPv6 网络的测量, 所用哈希函数 1 次计算的时间最低也需 68 ns, 并不满足 OC-192 链路对一个报文处理时间小于 32 ns 的要求. 程光等^[10]提出了一种基于掩码匹配的抽样测量算法, 该算法利用掩码来匹配 IP 报文中的选定字段以进行抽样, 由于计算机在进行掩码匹配时只需进行简单的与、或逻辑运算, 可以采用硬件实现, 因此运算速度极快, 可以达到纳秒级, 所以该方法能较好地应用于高速网络的抽样测量^[11-12].

笔者通过对基于掩码匹配的抽样算法进行分析, 改进文献[10]算法存在的不足, 以提高该算法的有效性和稳定性, 使该算法能更好地应用于高速网络分布式测量环境.

1 问题的提出

文献[10]提出的掩码匹配抽样算法的实质是利用事先选定的掩码与报文中选定的目标字段进行模式匹配, 当匹配结果一致时, 该报文被抽样; 否则, 该报文不被抽样. 由于 IP 报文中的标识字段具备较好的随机性, 因此被文献[10]算法选作为掩码匹配的目标字段. IP 报文中的标识字段一共有 16 位, 所以掩码可以选择 1~16 位不等. 当报文数量在标识字段的各个取值上分布均匀时, 如果掩码只有 1 位, 此时抽样率为 1/2, 而当掩码为 16 位时, 抽样率仅为 1/65 536. 所以设备处理能力一定时, 通过选择不同的掩码长度, 文献[10]算法能较好地适应于各种速率的网络抽样. 如果网络中的所有抽样设备均使用相同的抽样掩码, 那么当一个报文被某台设备抽中时, 它也会被其他设备抽中; 否则, 所有设备都不会抽中它. 故文献[10]算法也能较好地满足分布式测量环境对报文一致性的要求, 既能适应单点测量, 也能适应分布式测量.

虽然文献[10]提出的基于掩码匹配的抽样算法有许多优点, 但是该算法要求在抽样测量的时间粒度内, 总的报文数量在标识字段各个不同的取值上分布均匀. 文献[10]通过对大量的 IP 报文标识字段进行统计, 证明了该字段具有良好的随机性, 但是在用全 0 掩码进行实验时发现, 当全 0 掩码的位数大于 7 时, 实际抽样率将几乎不再变化, 并且严重偏离理论值. 因此, 文献[10]算法建议不要采用全 0 的抽样掩码来进行抽样测量.

为了分析造成这一现象的原因, 笔者分别统计了 10 组流量数据(每组流量数据有 10^6 个数据包)中 IP 报文在标识字段各个取值上的报文数量情况, 发现标识值为 0 的报文数量平均占总数的 3.44%, 标准差为 0.1%, 而标识字段为其他各个值的报文数平均占 0.001 47%, 标准差为 0.000 7%. 因此, 当使用全 0 掩码进行抽样时, 由于标识值为 0 的报文必被抽中, 故实际的抽样率至少为 3.44%. 而当掩码长度增加时, 理论抽样率呈指数级减小, 所以当掩码位数大于某个阈值时, 实际抽样率将逐渐偏离理论抽样率.

虽然不使用全 0 的抽样掩码可以回避这一问题, 但是这样一来, 该抽样算法的可靠性和有效性将会受到影响. 首先, 报文标识字段的随机性是由统计得出的, 因此不能保证在各个测量的时间粒度内, 标识字段为其他值的报文数量不会发生异常. 如果异常情况发生, 就不能选用相应的掩码来进行测量, 但是这事先是并不知道的, 所以算法的可靠性会受到影响. 其次, 由统计信息可知, 标识值为 0 的报文数远大于标识字段为其他值的报文数, 如果使用其他的抽样掩码将这些报文过滤掉, 可能会造成样本无法反映出总体中的一些信息, 导致算法的有效性降低, 测量精度下降. 所以, 文献[10]算法给出的建议并不可取.

造成这一问题的根本原因是文献[10]算法测量精度的好坏在很大程度上依赖于 IP 报文标识字段的随机性. 而报文数量并不总是在标识字段的所有取值上均匀分布, 例如, 标识值为 0 的报文数就远高于标识字段为其他各个值的报文数. 因此, 若要提高算法的可靠性和有效性, 还需进一步提高标识字段的随机性.

2 算法的改进

为了衡量标识字段的随机性好坏, 首先定义信息熵的概念.

定义 1 记一个 m 位的二进制序列 $a_1 a_2 \cdots a_m$ 为 s , s 共有 $n = 2^m$ 种取值可能, 若每种取值的概率分别为 p_1, p_2, \cdots, p_n , 则 s 的熵为 $H(s) = - \sum_{i=1}^n p_i \lg p_i$.

标识字段的熵越大, 说明该字段越随机. 当标识字段取各个值的概率相等时, 它的熵达到最大值. 由于网络的突发性和不可预测性, 在某个时间粒度内, 流量数据可能会出现具有某个特定标识值的 IP 报文数量过多的情况, 从而导致 IP 报文标识字段的随机性变差. 因此, 在进行掩码匹配前, 希望能将标识值为该特定值的报文通过某种方式比较均匀地映射到其他标识值上. 考虑将标识字段与一个 16 位的随机二进制序列作异或运算, 所得结果将均匀地分散到标识字段的各个取值上. 若标识字段以这种方式得到的结果的随机性更好, 那么将其作为掩码匹配的目标字段时, 算法的测量精度应该提高. 将异或运算后的结果称为新标识字段, 下面证明新标识字段的熵高于标识字段.

定理 1 一个 m 位的二进制序列 $s: a_1 a_2 \cdots a_m$ 分为两个子序列 $s_1: a_1 a_2 \cdots a_r$ 和 $s_2: a_{r+1} a_{r+2} \cdots a_m$, 其中, $1 \leq r \leq m-1$, 且 r 为整数. 若子序列 s_1 和 s_2 相互独立, 则 $H(s) = H(s_1) + H(s_2)$.

证明 s_1 共有 $n_1 = 2^r$ 种取值可能, 设其概率分别为 $p_1, p_2, \cdots, p_{n_1}$, 则 $\sum_{i=1}^{n_1} p_i = 1$. s_2 共有 $n_2 = 2^{m-r}$ 种取值可能, 设其概率分别为 $q_1, q_2, \cdots, q_{n_2}$, 则 $\sum_{j=1}^{n_2} q_j = 1$.

根据定义 1, 则有

$$\begin{aligned} H(s) &= - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} p_i q_j \lg(p_i q_j) = - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} p_i q_j (\lg p_i + \lg q_j) = \\ &= - \sum_{i=1}^{n_1} \left[\left(\sum_{j=1}^{n_2} q_j \right) p_i \lg p_i + p_i \sum_{j=1}^{n_2} (q_j \lg q_j) \right] = - \sum_{i=1}^{n_1} p_i \lg p_i + \left(\sum_{i=1}^{n_1} p_i \right) \left(- \sum_{j=1}^{n_2} q_j \lg q_j \right) = \\ &= - \sum_{i=1}^{n_1} p_i \lg p_i + \left(- \sum_{j=1}^{n_2} p_j \lg p_j \right) = H(s_1) + H(s_2) \quad \text{证毕.} \end{aligned}$$

推论 一个二进制序列 $s: a_1 a_2 \cdots a_m$ 中的各个二进制位 a_1, a_2, \cdots, a_m 之间相互独立, 则 $H(s) = \sum_{i=1}^m H(a_i)$.

证明 记 $s_{i,j}$ 为 s 中的第 i 位至第 j 位构成的子序列, 其中, $1 \leq i \leq j \leq m$.

由定理 1 可知, $H(s) = H(s_{1,1}) + H(s_{2,m})$,

同理, $H(s_{2,m}) = H(s_{2,2}) + H(s_{3,m}), \cdots, H(s_{m-1,m}) = H(s_{m-1,m-1}) + H(s_{m,m})$,

所以, $H(s) = H(s_{1,1}) + H(s_{2,2}) + \cdots + H(s_{m-1,m-1}) + H(s_{m,m}) = \sum_{i=1}^m H(a_i)$. 证毕.

定理 2 两个相互独立的 1 位序列 s_1 和 s_2 , 其熵分别为 $H(s_1)$ 和 $H(s_2)$, 若 $s = s_1 \oplus s_2$, 则 $H(s) \geq \max(H(s_1), H(s_2))$.

证明 首先证明 $H(s) \geq H(s_1)$. 设 s_1 取 0 的概率为 q_1 ($0 < q_1 < 1$), s_2 取 0 的概率为 q_2 ($0 < q_2 < 1$). 则 s 取 0 的概率为 $q_1 q_2 + (1 - q_1)(1 - q_2)$. 若 1 位随机序列 w 取 0 的概率为 p , 则其熵 $H(w) = -p \lg p - (1 - p) \lg(1 - p) = f(p)$, 如图 1 所示.

根据图 1, 要证明 $H(s) \geq H(s_1)$, 即要证明 $(q_1 q_2 + (1 - q_1)(1 - q_2) - 0.5)^2 \leq (q_1 - 0.5)^2$. 化简得 $(1 -$

$2q_1)^2(1 - q_2) \geq 0$, 由于 $0 < q_2 < 1$, 所以上式恒成立. 同理可证, $H(s) \geq H(s_2)$. 因此, $H(s) \geq \max(H(s_1), H(s_2))$. 证毕.

定理 3 两个 m 位的二进制序列 s_1 和 s_2 相互独立, 且 s_1 中的各位相互独立, s_2 中的各位也相互独立, 若 $s = s_1 \oplus s_2$, 则 $H(s) \geq \max(H(s_1), H(s_2))$.

证明 $\alpha_i, \beta_i, \gamma_i$ 分别表示 s_1, s_2, s 中的第 i 位, 由推论知,

$$H(s_1) = \sum_{i=1}^m H(\alpha_i), H(s_2) = \sum_{i=1}^m H(\beta_i), H(s) = \sum_{i=1}^m H(\gamma_i).$$

又由定理 2, $\forall i (1 \leq i \leq m)$, 有 $H(\gamma_i) \geq \max(H(\alpha_i), H(\beta_i))$, 因此,

$$\sum_{i=1}^m H(\gamma_i) \geq \max\left(\sum_{i=1}^m H(\alpha_i), \sum_{i=1}^m H(\beta_i)\right), \quad \text{即 } H(s) \geq$$

$\max(H(s_1), H(s_2))$. 证毕.

定理 3 从理论上证明了两个随机的二进制序列异或运算的结果更加随机. 由于文献[10]算法能运用于分布式的抽样测量环境中, 所以与标识字段进行异或运算的随机二进制序列只能从 IP 报文的其他字段中找. 文献[10]经过统计分析后指出, 除标识字段外, IP 报文目的 IP 字段的后 16 位随机性也比较好的. 又由于这两个字段相互独立, 且各位之间的相关性较弱, 因此, 可以利用它与标识字段进行异或运算. 分别统计 10 组流量数据中标识字段与新标识字段的熵, 所得结果如图 2 所示.

标识字段的熵的均值为 15.61, 标准差为 0.02, 而新标识字段的熵的均值为 15.87, 标准差为 0.007. 统计结果表明, 标识字段与目的 IP 字段后 16 位进行异或运算后, 熵值增加了. 虽然从结果上看熵值提升的幅度不是很大, 这是因为所统计的流量数据中除了标识值为 0 的 IP 报文数量很多外, 具有其他标识值的 IP 报文的数量的分布是比较均匀的, 如果在较多的标识值上报文的数量异常, 那么异或运算以后熵值的增加是显著的. 由于新标识字段的熵值更高, 所以改进算法将使用新标识字段作为掩码匹配的目标字段, 而不像文献[10]算法直接使用未经任何优化处理的标识字段.

3 实验验证

文中验证所用的流量数据来自于 MAWI^[13]工作组 2009 年不同时间点在互联网的一条骨干链路上通过全流量采集得到的数据, 共计 10 组, 每组数据共有 10^6 个 IP 报文.

3.1 稳定性分析

采用异或运算不仅使新得到的标识字段的随机性提高了, 但更为重要的是, 标识值为特定值的报文被比较均匀地分散到了其他的标识值上, 这样就不用再回避使用某些抽样掩码.

由于所统计的流量数据中标识值为 0 的报文比较多, 因此文献[10]算法建议不选择全为 0 的掩码来进行匹配抽样. 但是经过改进之后, 报文的标识字段会首先与其目的 IP 字段的后 16 位进行异或运算, 这样一来, 原来标识值为 0 的报文经异或运算后, 会随着目的 IP 字段后 16 位的随机性比较均匀地分布到其他标识值上, 此时, 即使使用全 0 的抽样掩码也能得到理论上的抽样结果.

定义抽样率 p 为抽中报文个数与报文总个数之比, 为了证明改进以后的算法稳定性更优, 分别使用 1~16 位的全 0 掩码对 10 组流量数据进行掩码匹配抽样, 所得实际抽样率如图 3 所示. 图 3 中每个点的纵坐标都是 10 组流量数据统计结果的均值, 正如所分析的那样, 当全 0 掩码的位数达到一定长度时, 文献[10]算法的实际抽样率开始偏离理论值, 而改进后的算法得到的抽样率却和理论值吻合得很好. 这不仅说明了新标识字段具有

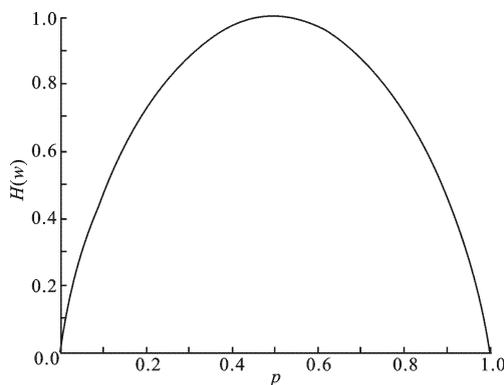


图 1 随机序列 w 的熵值

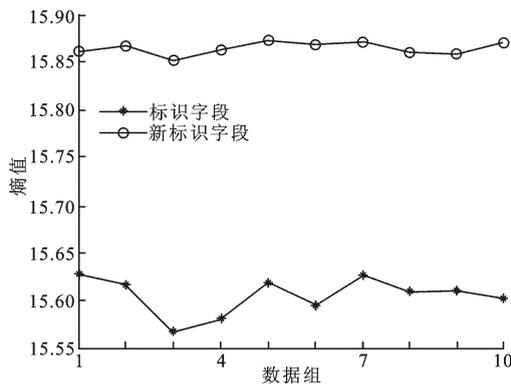


图 2 熵值比较

良好的随机性,同时也证明了与目的 IP 字段的后 16 位进行异或运算确实能将具有特定标识值的报文均匀分散到各个新的标识值上. 所以改进后的算法不需要再依据经验来回避某些特定的抽样掩码, 算法的稳定性得到了提高. 当设备的处理能力一定时,掩码位数越长,越能适应更高速率网络的抽样. 而改进算法能较好地克服文献 [10]算法中某些抽样掩码位数过长时,实际抽样率与理论值有较大偏差的不足,算法应用的可扩展性也得到了增强.

3.2 有效性分析

算法的有效性可以通过以下两点进行说明:(1)抽样算法所得样本的分布能较好地代表总体的分布.(2)改进算法所得样本的分布比文献[10]算法获得的样本的分布更接近于总体分布. 由于报文总体的长度分布是网络测量中最为关注的分布之一,所以可通过比较报文样本的长度分布和报文总体的长度分布来对算法的有效性进行检验.

因为所用流量数据中 IP 报文的长度集中在 32 B~ 2 048 B,故将其划分为 32 B~ 63 B,64 B~ 127 B, 128 B~ 255 B,256 B~ 511 B,512 B~ 1 023 B,1 024 B~ 2 048 B 6 个区间. 假设总体中报文长度落在这 6 个区间的概率为 $p_i (1 \leq i \leq 6)$, 样本报文长度落在这 6 个区间的概率为 $q_i (1 \leq i \leq 6)$, 样本容量为 n , 那么可使用卡方检验量^[14] $\chi^2 = \sum_{i=1}^6 (n/p_i)(q_i - p_i)^2$, 来观察样本分布是否符合总体分布. 文献[10]已通过卡方分布检验证明了这一点是成立的,因此,文中将直接验证改进算法所得样本的分布是否比文献[10]算法获得的样本的分布更接近于总体分布.

为了衡量样本分布对总体分布的接近程度,首先定义绝对接近度和相对接近度的概念. 定义样本的绝对接近度 $d = \sum_{i=1}^6 (q_i - p_i)^2$, 该指标描述了样本分布对总体分布的接近程度, d 值越小,说明样本分布越接近于总体分布;当 $d=0$ 时,说明样本分布与总体分布完全一致. 定义样本的相对接近度 $r = d_2/d_1$, 其中, d_1 为文献 [10] 算法所得样本的绝对接近度, d_2 为改进算法所得样本的绝对接近度. 当 $r > 1$ 时,说明文献 [10] 算法所得样本更接近于真实的分布;当 $r < 1$ 时,说明改进算法所得样本更接近于真实的分布;当 $r = 1$ 时,说明二者的接近程度相当. 随机产生 3 组 1~16 位的非 0 抽样掩码(用十进制表示): 1, 3, 5, 12, 8, 57, 122, 202, 96, 978, 1 960, 1 957, 6 489, 2 317, 25 952, 60 020; 1, 3, 6, 14, 20, 2, 108, 140, 224, 988, 1 521, 1 606, 5 369, 2 804, 26 223, 42 356; 1, 2, 7, 10, 3, 17, 70, 112, 174, 599, 458, 3 920, 3 976, 8 289, 4 649, 27 640. 然后,利用这 3 组掩码对 10 组流量数据进行掩码匹配抽样,分别计算样本的相对接近度 r , 所得结果如图 4 所示.

图 4 中每个点的纵坐标都是 10 组流量数据统计结果的均值. 从图 4 中可以看出,当掩码位数比较少时,相对接近度 r 明显小于 1,这是因为改进后的算法不用像文献[10]算法那样通过选择合适的掩码过滤掉具有特定标识值的报文,从而使得样本能更好地代表总体. 当掩码位数比较多时,由于抽样率呈指数减少,导致抽中的报文很少. 例如,当掩码位数为 16 位时,从 10^6 个报文中理论上只能抽中 15.36 个报文,所以此时样本分布的好坏在很大程度上取决于样本的随机性. 在所得的这 10 组流量数据中,改进算法用作匹配的新标识字段的熵只略高于文献[10]算法所使用的标识字段,故相对接近度比较接近 1. 可以看出,图中所有点的纵坐标均小于 1,这说明改进算法得到样本的分布比原抽样算法更接近于总体分布,算法的有效性得到了证明.

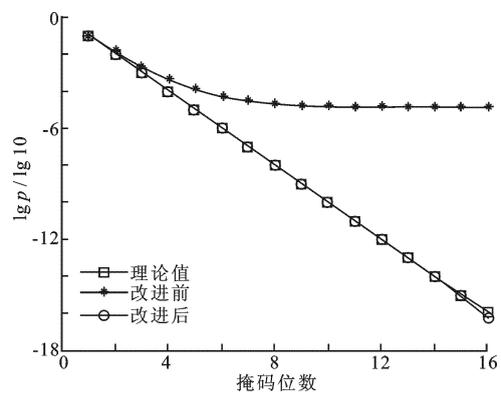


图 3 抽样率比较

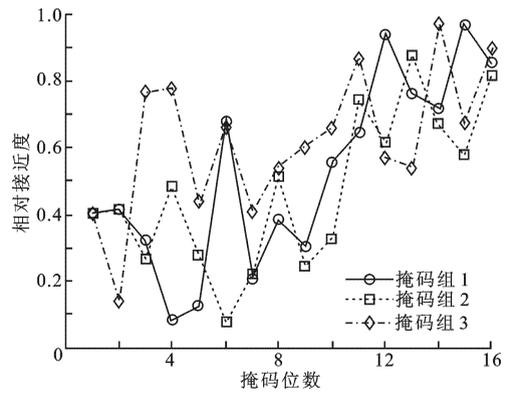


图 4 样本相对接近度

4 结束语

面向高速网络的分布式抽样测量技术是目前研究的一个重点问题,而基于掩码匹配的抽样测量技术既能适应高速网络,同时又能满足分布式测量环境的要求.文中通过分析文献[10]算法中的不足之处,提出了将标识字段与目的 IP 字段后 16 位进行异或运算的改进方案,不仅从理论上证明了该方案的可行性,也使用实际的流量数据检验了改进后算法的有效性和稳定性.文中对文献[10]算法进行改进,实现简单,可扩展性较强,可以应用于类似的抽样算法中.

参考文献:

- [1] Claffy K C, Polyzos G C, Braun H W. Application of Sampling Methodologies to Network Traffic Characterization [J]. SIGCOMM Computer Communication Review, 1993, 23(4): 194-203.
- [2] 陈松, 王珊, 周明天. 基于实时分析的网络测量抽样统计模型 [J]. 电子学报, 2010, 38(5): 1177-1180.
Chen Song, Wang Shan, Zhou Mingtian. Network Data Measurement and Statistics Model Based on Real-time Analysis [J]. Acta Electronica Sinica, 2010, 38(5): 1177-1180.
- [3] 潘乔, 罗辛, 王高丽, 等. 基于 FARIMA 模型的流量抽样测量方法 [J]. 计算机工程, 2010, 36(15): 7-11.
Pan Qiao, Luo Xin, Wang Gaoli, et al. Traffic Sampling Measurement Method Based on FARIMA Model [J]. Computer Engineering, 2010, 36(15): 7-11.
- [4] 陈庶樵, 张果, 朱柯. 一种基于包速率自适应的报文抽样算法 [J]. 计算机应用研究, 2010, 27(7): 2727-2729.
Chen Shuqiao, Zhang Guo, Zhu Ke. Algorithm Based on Packet Rate Adaptive for Packet Sampling [J]. Application Research of Computers, 2010, 27(7): 2727-2729.
- [5] Cozzani I, Giordano S. A Passive Test and Measurement System: Traffic Sampling for QoS Evaluation [C]// Proceedings of GLOBECOM 1998. Sydney: IEEE, 1998: 1236-1241.
- [6] Duffield N, Grossglauser M. Trajectory Sampling for Direct Traffic Observation [J]. IEEE/ACM Trans on Networking, 2001, 9(3): 280-292.
- [7] Duffield N, Lund C, Thorup M. Learn More, Sample Less: Control of Volume and Variance in Network Measurement [J]. IEEE Trans on Information Theory, 2005, 5(51): 1756-1775.
- [8] 潘乔, 裴昌幸. 用于高速 IPv6 网络流量抽样测量的算法 [J]. 西安电子科技大学学报, 2007, 34(3): 377-381.
Pan Qiao, Pei Changxing. Method for Traffic Sampling in High-speed IPv6 Networks [J]. Journal of Xidian University, 2007, 34(3): 377-381.
- [9] 潘乔, 裴昌幸. 基于信息熵理论的高速 IPv6 网络流量抽样测量方法 [J]. 吉林大学学报(工学版), 2009, 39(5): 1338-1341.
Pan Qiao, Pei Changxing. Information Entropy Theoretic Approach to Traffic Sampling Measurement in High-speed IPv6 Networks [J]. Journal of Jilin University(Engineering and Technology Edition), 2009, 39(5): 1338-1341.
- [10] 程光, 龚俭, 丁伟. 基于统计分析的高速网络分布式抽样测量模型 [J]. 计算机学报, 2003, 26(10): 1266-1273.
Cheng Guang, Gong Jian, Ding Wei. Distributed Sampling Measurement Model in a High Speed Network Based on Statistical Analysis [J]. Chinese Journal of Computers, 2003, 26(10): 1266-1273.
- [11] 王海涛, 吴连才, 武媛媛. 基于掩码匹配的 IP 网络抽样测量系统的设计和实现 [J]. 数据通信, 2011(2): 18-21.
Wang Haitao, Wu Liancai, Wu Yuanyuan. Design and Implementation of a Flow Sampling Measurement System Based on Mask Matching on IP Network [J]. Data Communications, 2011(2): 18-21.
- [12] 李志敏, 赵治国, 朱安新, 等. 校园网流量抽样测量系统设计与实现 [J]. 网络安全技术与应用, 2011(11): 70-72.
Li Zhimin, Zhao Zhiguo, Zhu Anxin, et al. Design and Implementation of a Flow Sampling Measurement System on Campus Network [J]. Network Security Technology and Application, 2011(11): 70-72.
- [13] MAWI Working Group Traffic Archive. Traffic Trace Info [EB/OL]. [2012-08-20]. <http://mawi.wide.ad.jp/mawi>.
- [14] 盛骤, 谢式千, 潘承毅. 概率论与数理统计 [M]. 北京: 高等教育出版社, 2008: 199-200.

(编辑: 齐淑娟)