

基于分数 Alpha 模型的缓存计算方法

张冰怡¹, 张宏科¹, 边裕兰², 张 辉^{1,3}

(1. 北京交通大学电子信息工程学院, 北京 100044; 2. 上海微创软件有限公司, 上海 200041; 3. 北京航天指挥控制中心, 北京 100720)

摘要: 缓存大小计算是高性能路由器设计中一个必不可少的内容, 常规缓存计算方法是基于 Poisson 通信量模型得到的, 不符合网络通信量的实际特征, 在使用中存在丢包率较高的问题。已提出的分数 Alpha 通信量模型能体现通信量的自相似性和非高斯特征, 用于缓存溢出概率计算, 能得到比其它网络模型更好的结果。基于该模型得到了一个缓存计算方法, 在高速路由器转发引擎的缓存设计应用中得到了满意的结果, 与常规方法相比更体现了实际通信量的变化规律, 计算结果更准确。

关键词: 路由器设计; 缓存计算; Poisson 模型; 分数 Alpha 通信量模型

Method of Buffersize Computation Based on Fractional Alpha Traffic Model

ZHANG Bingyi¹, ZHANG Hongke¹, BIAN Yulan², ZHANG Hui^{1,3}

(1. School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing 100044;

2. Shanghai Wicresoft Co., Ltd., Shanghai 200041; 3. Beijing Aerospace Control Center, Beijing 100720)

【Abstract】 The buffersize is an essential part of a high performance router design. The common method of buffersize computation is based on Poisson traffic model. This method is not in conformity to the actual traffic characteristic and it will cause high rate of packet loss. The proposed fractional Alpha traffic model can denote the characteristic of self-similar and non-Gaussian. The residual distribution function (RDF) based on the fractional Alpha model fit the real traffic better than the RDF based on other models. Based on the new RDF, a method for computing the buffersize is got and used in the high performance router design. The method is simple and the result is satisfying.

【Key words】 Router design; Buffersize computation; Poisson model; Fractional Alpha traffic model

在路由器设计中, 转发引擎和交换引擎中的排队调度都要考虑缓存大小, 缓存大小计算是高性能路由器设计中一个必不可少的内容, 它在QoS的研究中扮演着一个重要的角色, 如设计不合理, 就会明显影响网络性能。例如, 以前按照传统Poisson通信量模型计算的缓存大小, 在服务器上使用后就存在丢包率较高的问题^[1]。原因是实际通信量具有自相似性, 使得通信量会集中在某些时间段内到来, 造成缓存溢出。

为了进一步说明, 分别计算基于 Poisson 模型和 FGN 模型的缓存大小, 对其进行比较。文献[2]基于 FGN 模型提出了一个经典的缓存溢出概率公式, 基于该公式计算不同会话数目的情况下为满足指定丢包率, 缓存大小随 Hurst 参数变化的情况, 分析后发现: 当会话数目小于某一值时, 随着 Hurst 参数的增加缓存大小增加不大, 类似基于 Poisson 模型计算的结果; 而当会话数目超过某一值时, 随着 Hurst 参数的增加缓存大小急剧增加, 明显不同于 Poisson 模型计算结果, 具体参见文献[3]。

大量证据表明, 网络通信量具有自相似性和非高斯特征^[4,5]。Poisson模型已不能准确描述真实网络通信量, 有必要在更准确模型基础上计算网络缓存大小。本文介绍了一个分数Alpha模型, 对比常用的Poisson模型缓存计算方法, 提出一个新的分数Alpha缓存计算方法, 并予以实验验证。

1 分数 Alpha 模型

文献[4]提出并证明了一个分数 Alpha 网络通信量模型, 基于该模型推出了一个缓存溢出概率公式:

$$L_{\alpha,H} = c_1 T^{H \cdot \frac{1}{\alpha} + 1} + c_2 \int_0^T t^{H-1/\alpha} M_s(dt) \quad (1)$$

其中, C_1, C_2 为常数; α 为 Alpha 随机测度的形状参数; M_s 为 Alpha 平稳随机测度。

基于分数 Alpha 模型的缓存溢出概率如下:

$$\lim_{x \rightarrow \infty} P(Q(x) > x) \geq P_0 \quad (2)$$

$$l(x, t_0)^{-\alpha} C_{\alpha,H} \geq P_0 \geq x^{\alpha(H-1)} K(\alpha, H, C_1, C_2, C) \quad (3)$$

其中, $K, t_0, C_{\alpha,H}$ 均为常数。

选取 Bellcore(现为 Telcordia)实验室的 pOctExt-TL 数据作为真实输入队列流量, 采用分数 Alpha 模型作为理论计算基础, 将理论计算结果和真实数据比较, 然后将基于其它模型的理论计算结果和真实数据比较, 发现基于分数 Alpha 模型的结果更能体现真实数据复杂的变化情况, 真实数据的缓存变化具有两个阶段: 平缓变化阶段和急剧变化阶段。而基于其它模型的结果只有一种变化趋势, 因此, 基于分数 Alpha 模型的缓存计算结果更符合真实数据的整体变化趋势。

2 基于 Poisson 模型的缓存计算方法

文献[6]基于常用的 M/G/1 方法和 Poisson 模型, 针对输

基金项目: 国家自然科学基金资助项目(60473001); 华为高校科技基金资助项目(YJCB2005054RE)

作者简介: 张冰怡(1974-), 男, 博士后, 主研方向: 计算机网络; 张宏科, 教授、博导; 边裕兰, 硕士; 张 辉, 博士生

收稿日期: 2006-07-30 **E-mail:** zby@telecom.njtu.edu.cn

入接口速率高达 5Gbps 的高性能路由器，详细分析了不同情况下转发引擎流水线设计的入口缓存需求，分别得到流水线开启时和关闭时的入口缓存需求。其中流水线开启时的转发引擎入口缓存需求 M' 为

$$\begin{cases} \max(M') = 8(p+q+k-1)l_{\max} & (N \leq T_p) \\ \max(M') = \infty & (N > T_p) \\ E(M') = E(M_p) = 8(p+q+k-1)E(l) & (N \leq T_p) \\ E(M') = E(M_p) + E(M_q) = 8(p+q+k-1+E_{qv})E(l) & (N > T_p) \end{cases} \quad (4)$$

其中， M_p 为转发引擎流水线的缓存需求； M_q 为排队等待的缓存需求； $E(M_q)$ 为 M_q 的平均需求； l_i 为第 i 个 IP 包的包长； N 为转发引擎每秒钟输入的最大 IP 包数； T_p 为转发引擎流水线的吞吐率。

$p\tau$ 表示组播一级查表关键字输入至三态内容地址关联存储器(ternary content addressable memory, TCAM)之后，等待搜索结果所需的等待时间。 $q\tau$ 为组播查表二级关键字输入至静态随机存取存储器(static random access memory, SRAM)之后，等待最终查表结果所需的等待时间。实际设计的转发引擎中， $p\tau$ 为 10 个 FPGA 时钟周期， $q\tau$ 为 4 个 FPGA 时钟周期，该等待时间可视为流水线中的功能段，其段数分别为 2 个和 1 个，用 k 来表示，根据流水线时空图可知，流水线中同时处理的最大 IP 包数为 $p+q+k-1$ 。

在转发引擎的流水线设计中，采用 TCAM 作为转发表存储器件并选择预留表项空间的选择移动法作为 TCAM 中表项维护方案^[7, 8]。得到流水线关闭时的入口缓存需求为

$$\begin{cases} \max(M'') = \frac{wmR}{2f_T} \\ E(M'') = mE(R)E(v)/f_T \end{cases} \quad (5)$$

其中， $E(M'')$ 为流水线关闭时的平均入口缓存； w 为 TCAM 中设置的段数； v 为更新一条表项需要移动相关表项的次数； f_T 为 TCAM 的工作频率； m 为 TCAM 对一条 72bit 宽度表项的写周期数； R 为转发引擎入口的最大接口速率。

结合式(4)、式(5)，得到转发引擎流水线设计入口缓存需求如下

$$\begin{cases} \max(M) = \max \left[8(p+q+k-1)l_{\max}, \frac{wmR}{2f_T} \right] & (N \leq T_p) \\ \max(M) = \infty & (N > T_p) \\ E(M) = \max \left[8(p+q+k-1)E(l), mE(R)E(v)/f_T \right] & (N \leq T_p) \\ E(M) = \max \left[8(p+q+k-1+E_{qv})E(l), mE(R)E(v)/f_T \right] & (N > T_p) \end{cases} \quad (6)$$

结合上述分析，下面以该转发引擎流水线的工程实现为例，计算转发引擎 FPGA 的入口缓存设置值。

首先计算 $E(l)$ 的值，上式中，当转发引擎输入的 IP 包为定长时， $E(l)$ 的值即为该包的长度，当输入 IP 包为变长时，需要分析 IP 包的平均长度。使用统计方法计算实际亚太地区网络 APAN 抽样分组包长均值，统计意义上 IP 包的平均长度约为 508B。考虑到转发引擎流水线设计中内部分组标签所附加的 8B，平均 IP 包字节数应为 516B。

在转发引擎流水线的工程实现中，流水线的系统时钟采用 100 MHz，TCAM 的工作频率 f_T 为 100 MHz，流水线的段数 k 设置为 7，IPv6 单播转发表对应的 w 值为 64，TCAM 对一条 72bit 宽表项写操周期数 m 为 11 个，TCAM 查表结果等待时间对应的 p 值为 2，SRAM 查表结果时间对应的 q 值为 1，转发引擎的最大输入接口速率为 5Gbps，为分析方便取移动表项次数 v 的均值 $E(v)$ 为 4。

该高速 IPv6 路由器设计最大吞吐率 $T_p=125e6$ 。转发引擎的接口速率为 5Gbps，转发引擎每秒中输入的最大 IP 包数 N 为 $5e9/(40 \times 8) = 15.625e6$ 个，即 $N > T_p$ ，根据式(6)可得：转发引擎最大入口缓存需求为： $\max(M) = +\infty$ ，转发引擎入口平均缓存需求为： $E(M) = 41\ 280\text{bit}$ 。

在转发引擎 FPGA 设计中，大容量缓存功能通常利用 FIFO 实现，本项目中，XC2V3000 内部 SelectRam 总量为 1 728Kbit 在数据线宽度为 64b 的情况下，FPGA 内部以 Block Memory 形式实现的 FIFO 的最大深度为 32 768。能够满足上述的转发引擎入口平均缓存需求。因此，转发引擎 FPGA 的入口 FIFO 深度可以设置为 $41280/64 = 645$ 。

3 基于分数 Alpha 模型的缓存计算方法

为了和传统方法比较，针对同一问题，基于分数 Alpha 模型进行重新计算。流水线开启时，入口缓存需求中的流水线的缓存需求 M_p 的平均需求和最大需求仍为式(4)。分析排队等待缓存需求 M_q ，令缓存溢出概率为 ε ，那么根据式(3)可以推出缓存极小值和极大值，这里只考虑极大值的情况，满足式(7)：

$$\begin{aligned} \varepsilon C_2 t^H &= C_{\alpha, H} (M_q - C_1 t^2 + C t) \\ t &= \frac{C(1-H) + \sqrt{(1-H)^2 C^2 - 4HM_q(2-H)C_1}}{2(2-H)C_1} \end{aligned} \quad (7)$$

式中参数定义同式(1)和式(2)中的参数定义。流水线关闭时的最大入口缓存需求 M'' 同式(5)，主要考虑 $N > T_p$ 的情况，最终转发引擎流水线设计入口缓存需求形式如式(8)：

$$M = \max [M_p + M_q, mRE(v)/f_T] \quad (8)$$

在计算排队等待缓存 M_q 时，把基于分数 Alpha 模型的仿真数据作为真实输入队列流量。包长为 516B，服务速率为 $C=12.5e6$ 包个数/s， $H=0.92$ ， $\alpha=1.16$ ，平均流速 $C_1=7e5$ (包个数/s)， $C_2=100$ 。仿真输入队列见图 1，缓存溢出概率见图 2。

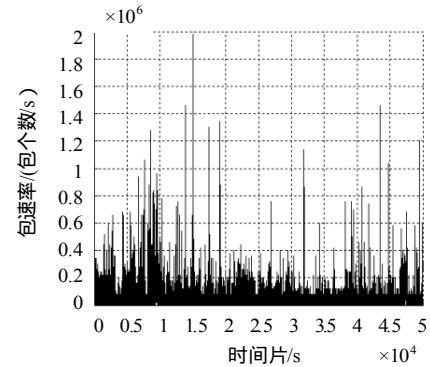


图 1 仿真输入队列

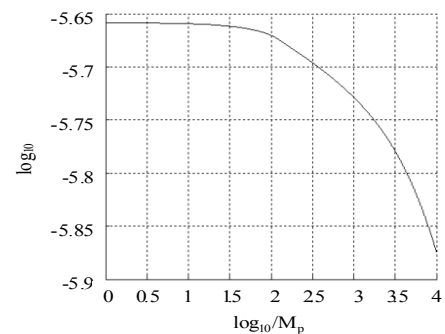


图 2 缓存溢出概率

从图 2 中可以算出 $M_q = 10\ 000\text{bit}$ 可使缓存溢出概率 $\epsilon < 1e-6$, 根据式(8), 整个缓存需求仍然由 M 决定。分析基于分数 Alpha 模型的缓存计算发现, 该方法存在一个关键点, 在关键点之后, 虽然缓存增长很大, 但丢失率变化不大。

4 实验验证

本文通过实验验证了缓存大小计算的正确性^[9]。在高速转发引擎性能测试的环境中, 测试仪的两个 OC-192 端口 A 和 B 分别与 5G POS 接口卡 A 和 5G POS 接口卡 B 相连, 双向同时发送与接收数据。测试过程中, 取包长的范围为 40B ~ 1500B, 包类型为 IPv6 单播包, 得到不同包长业务时的转发引擎的吞吐率和丢包率, 模拟实际包长分布条件下, 不同负荷时的转发引擎丢包率。从测试结果来看, 在单一包长测试条件下, 负荷为 100% 时, 当包长大于等于 109.5B 时的丢包率低于 $1.07e-6$, 吞吐率接近于 1。在端口负荷低于 90% 时, 丢包率低于 $3.0e-4$ 。该转发引擎具有较高的性能, 满速率工作时, 能够支持 110B IPv6 包线速转发, 110B 时的包转发率可以达到 5.7Mpps。基于分数 Alpha 模型的方法和基于传统方法的计算结果相似, 其正确性都在实际实验中得到了证实。这是由于本课题中转发处理能力强大, 与接口速率接近, $N \approx T_p$, 因此, 对缓存要求不高, 使得两种方法结果差别不大。但基于分数 Alpha 模型的方法更符合网络通信量的实际情况, 计算过程也更简单, 更适合接口速率 10Gbps 以上的转发引擎设计。

5 结论

路由器转发引擎缓存设置为 QoS 的实现提供了最基本的保证, 对于转发引擎 FPGA 的选型、流水线设计实现时 FPGA 内部缓存大小的设置, 以及转发引擎的性能估计都具有重要的意义。使用传统分析方法计算排队等待缓存需求的理论模

(上接第 33 页)

首先将 ABM-PIM 中 ActionNode 的名称直接转换为 SAP-PSM 中 WebAction 的名称, 然后依据 3 个条件依次进行转换, 分别对应是否存在参与对象、目标调用对象和导航目标 3 种情况, MsgObjectTrans、InvokeObjectTrans、PageObjectTrans 分别表示为对应的 3 种子转换规则, 需要预先定义。

3.4 运行实例



图 8 运行实例—添加教职工记录页面

本文开发了支持 ABM 方法的工具原型, 目前基本完成对 J2EE 平台和 ASP.NET 平台开发的支持。基于特定框架 (Struts、JSF) 的代码生成器作为插件实现从 SAP-PSM 到代码的转换。图 3 所示的行为视图和图 4 所示的界面展示视图是一个高校院系综合管理系统 PIM 模型的组成部分, 经过 2 步

型存在问题, 不符合实际通信量的特征, 需要引入更合理的通信量模型计算缓存需求。本文使用分数 Alpha 模型分析了高速 IPv6 路由器的缓存设计并得到了满意的结果。与传统方法比较, 该方法更体现实际通信量的变化规律, 计算结果更准确。比单纯使用真实数据分析具有理论基础和普遍性, 易于调整参数、设置不同场景, 且计算量小, 是一个方便快捷的方法。

参考文献

- 1 Stallings W. 高速网络-TCP/IP 和 ATM 的设计原理[M]. 齐望东, 薛卫娟, 谢希仁, 译. 北京: 电子工业出版社, 1999.
- 2 Norros I. A Storage Model with Self Similar Input[J]. Queueing Systems, 1994, 16(2): 387-396.
- 3 张冰怡. 分数 Alpha 通信量模型的研究与应用[D]. 南京: 南京理工大学, 2004.
- 4 Zhang Bingyi, Sun Yamin. Fractional Alpha Stable Network Traffic Model and Its Application in QoS Routing[J]. Elsevier Journal of Network and Computer Applications, 2006, 29(1): 1-10.
- 5 Zhang Bingyi, Sun Yamin, Bian Yulan, et al. Linear Discriminant Analysis in Network Traffic Modeling[J]. International Journal of Communication Systems, 2006, 19(1): 53-56.
- 6 李玉峰. 基于 IPv6 路由器的高速转发技术研究及实现[Z]. 国家数字交换系统工程技术研究中心. 2004.
- 7 王振兴. NGI 高性能路由器转发处理算法与实现[D]. 南京: 南京理工大学, 2004.
- 8 Gupta P. Algorithms for Routing Lookups and Packet Classification[D]. USA: Stanford University, 2000-12.
- 9 田裕兰. 基于 T 比特路由器高速转发引擎的设计与实现[D]. 武汉: 武汉科技大学, 2004.

转换(PIM 到 PSM、PSM 到代码)后生成的代码在 JSF 框架下运行时的页面如图 8 所示。

4 结束语

本文以抽象代数理论为基础, 介绍了一种基于体系结构映射的模型转换方法, 能够对模型驱动的软件开发提供有力的支持。今后的工作包括进一步完善体系结构模型的形式化描述, 增强其语义表述能力, 并在此基础上加强对模型的验证; 目标平台的多样化以验证该方法的实用性。

参考文献

- 1 Caplat G, Sourrouille J L. Model Mapping Using Formalism Extensions[J]. IEEE Software, 22(2), 2005.
- 2 Gardner T, Griffin C, Koehler J, et al. A Review of OMG MOF 2.0 Query/Views/Transformations Submissions and Recommendations Towards the Final Standard[Z]. <http://www.omg.com/docs/ad/03-08-02.pdf>, 2003.
- 3 Czarnecki K, Helsen S. Classification of Model Transformation Approaches[C]//Proceedings of the 2nd OOPSLA Workshop on Generative Techniques in the Context of the Model Driven Architecture, USA. 2003.
- 4 赵会群, 王国人, 高远. 软件体系结构抽象模型[J]. 计算机学报, 2002, 25(7).
- 5 Kleppe A, Warmer J, Bast W. The Model Driven Architecture: Practice and Promise[M]. Addison-Wesley, 2003.