

一种自相似网络流量生成器的设计与实现

王晓婷,王忆文,李 平

(电子科技大学 微电子学与固体电子学院,四川 成都 610054)

摘 要: 为了满足网络设备性能测试的需要,流量生成器应能够产生近似真实网络特性的网络流量. 针对现有的流量产生方案的速度限制和精确性不足的问题,设计了一种高效的自相似网络流量生成器. 根据多个独立的 ON/OFF 模型叠加的方式实时产生自相似流量时间序列,并完全利用 FPGA 硬件资源实现了流量生成器的结构设计. 该生成器可实现“Gb/s”的数据速率,且速度可扩展性高. 通过流量统计特性分析,该网络流量生成器生成的流量表现出较准确的自相似性,能够较好地反映真实的网络流量.

关键词: 网络流量生成器;自相似;ON/OFF 模型;FPGA

中图分类号: TP393

文献标识码: A

文章编号: 1000—7180(2016)08—0054—05

Design and Implementation of Self-similar Network Traffic Generator

WANG Xiao-ting, WANG Yi-wen, LI Ping

(School of Microelectronics and Solid-State Electronics, University of Electronic
Science and Technology of China, Chengdu 610054, China)

Abstract: Nowadays it's desirable to generate traffic which can approximate realistic network traffic characteristics for performance evaluation of network equipments. To overcome the speed limit and accuracy insufficiency of existing traffic generation solutions, this paper proposes an efficient self-similar network traffic generator. The traffic generator can generate real-time self-similar time series according to the aggregation of multiple independent ON/OFF models. The main structure of the traffic generator is presented and implemented entirely using FPGA hardware resources. Statistical property analysis of the proposed traffic generator indicates that the generated traffic can achieve accurate self-similarity and better approximate the real network traffic. Moreover, the required data rates can be up to Gbps line rate and scale linearly with the increased port line rate.

Key words: network traffic generator; self-similar; ON/OFF model; FPGA

1 引言

为了保证网络设备的功能和性能,需要在不同的网络流量条件下对其进行测试. 流量生成器,是一种向网络设备施加近似于真实网络流量行为的设备,以对网络设备的性能进行评估的有效测试方法. 大量研究表明,实际的网络业务流量呈现自相似的特性,即在不同的时间尺度下具有突发性. 这种特性对待测网络设备的性能影响很大,因此如何精确地生成近似实际网络业务流的自相似流量,成为流量

生成器研究的一个重要方向.

目前,已有一些用于网络设备测试的流量生成方法,主要包括专用测试仪器^[1]、基于软件的流量生成器^[2]以及基于 FPGA 的流量生成器^[3-4]. 专用测试仪^[1],通常由公司专业开发,能有效地产生具有不同流量分布的高速测试流量,以对网络设备进行测试. 然而,这些测试设备一般非常昂贵,不适用在网络技术的研究开发阶段. 基于软件的流量生成器^[2],可以灵活地配置产生各种流量模式下的网络流量. 但是,这种方法可实现的最大数据速率和流量的统计精度依赖于运行

软件的硬件环境,扩展性差.因此,基于软件的流量生成器不适用于现代高速网络的性能评估.

基于 FPGA 的流量生成器,可以显著改善软件流量生成器的速度限制和流量统计特性不精确的问题. OC-48 流量发生器^[3],结合片上处理器和 FPGA 硬件,可以配置生成满足特定流量模型(自相似过程、马尔可夫调制的过程)的测试流量.然而,OC-48 采用处理器来计算数据包的到达时间间隔,仍然限制了速度的可扩展性.基于 NetFPGA 的流量生成器^[4],通过将记录在给定 PCAP 文件中的信息进行精确重演来产生数据包,并以“Gb/s”线速率传输.这种流量方案依赖于 PCAP 文件中收集的数据信息,随着数据包数量的增加,PCAP 的规模会增大.因此,需要将 PCAP 文件频繁地重新加载到本地存储器,这样会影响传输速度.

为了克服已有流量生成方法的数据速率和精确性两方面的不足,本文提出了一种高效的自相似网络流量生成方案,用于模拟实际的网络流量.

2 自相似流量模型分析

流量模型是分析理解网络流量行为、设计新的网络架构以及评估网络设备性能的基础.不同的网络流量特性对网络设备的性能有显著的影响.因此,选择合适的流量模型以获得准确的性能测试结果是是一个迫切需要解决的问题.

2.1 自相似过程

随着网络应用的多样化,实际的网络流量呈现出自相似特性,而这种特性无法由传统的泊松过程或马尔可夫调制的过程(Markov-modulated)模型^[5-6]进行建模.传统的流量模型中,流量的突发性在较长的时间尺度下变得很平滑,这可能会导致网络设备的性能测试结果与实际有很大偏差.相反,自相似过程在很宽的时间尺度下都具有突发性,能够较好地描述实际网络的流量行为.因此,精确地生成自相似过程以近似真实的网络流量至关重要.具体来说,一个自相似过程^[5]定义如下:

给定一个时间序列 X ,表示在每个单位时间间隔 t 内到达的数据包数量.假设 $X = (X_t; t = 0, 1, 2, 3, \dots)$ 是一个平稳随机过程,其统计特性(如均值、方差等)不会随着时间的推移而改变.其中, X 的均值为 $\mu = E[X_t]$,方差为 σ^2 ,自相关函数为 $r(k)$.自相关函数 $r(k)$ 定义为

$$r(k) = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{E[(X_t - \mu)^2]}, k = 0, 1, 2, \dots \quad (1)$$

通过将时间序列 X 划分为若干组互不重叠的大小为 m 的序列,产生新的 M -聚集时间序列 $X^{(m)} = (X_k^m; k = 1, 2, 3, \dots)$,其中

$$X_k^m = \frac{1}{m} \sum_{t=(k-1)m+1}^{km} X_t, k = 1, 2, 3, \dots \quad (2)$$

若时间序列 X 为自相似过程,则其自相关函数 $r(k)$ 满足式(3)中的条件.而且,对于所有的 $m(m = 1, 2, \dots)$, $X^{(m)}$ 的自相关函数 $r^m(k)$ 满足式(4):

$$r(k) \sim k^{-\beta} \text{ as } k \rightarrow \infty, 0 < \beta < 1 \quad (3)$$

$$r^m(k) = r(k) \quad (4)$$

很显然,自相似过程 X 的自相关函数 $r(k)$ 按照双曲线形式衰减,而传统的泊松流量模型等是以指数形式衰减,双曲线形式的衰减速度比指数衰减要慢.

2.2 基于 ON/OFF 源汇聚的自相似流量模型

ON/OFF 模型能够从物理意义上较好地解释自相似流量的产生,已经有研究证明无穷多个独立的具有重尾分布的 ON/OFF 过程的叠加收敛于自相似过程分形布朗运动(fractional Brownian motion, FBM)^[5].自相似流量可以通过大量的独立的 ON/OFF 数据源进行汇聚而有效产生^[7],如图 1 所示.每个 ON/OFF 源由严格交替的 ON 和 OFF 周期构成.在 ON 周期中,数据源以固定的速率发送数据包;而在 OFF 周期中,数据源保持空闲,不发送数据包.ON 周期和 OFF 周期的时间长度独立地符合重尾分布.假设随机变量 X 服从重尾分布,则累积分布函数 $F(x)$ 满足:

$$1 - F(x) = P[X > x] \sim x^{-\alpha}, x \rightarrow \infty, 0 < \alpha < 2 \quad (5)$$

也就是说,不考虑随机变量 X 的较小值的分布特性,如果大部分取值的分布形状趋近于双曲线,则随机变量 X 为重尾分布.

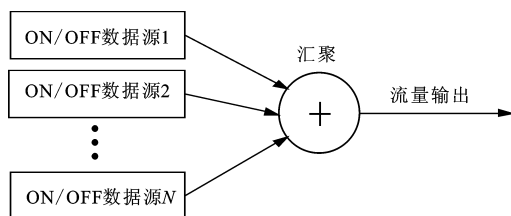


图1 基于 ON/OFF 源汇聚的自相似流量模型

3 自相似网络流量生成器的设计

3.1 流量生成器的整体结构

通过 MATLAB 建模发现,将 20 个独立分布的 ON/OFF 数据源进行叠加,即可很好地近似具有自相似特性的流量行为.因此,对于自相似流量的生

成,基于 ON/OFF 数据源汇聚的思想进行硬件设计.图 2 所示为自相似网络流量生成器的整体架构,主要包括六大功能模块:20 个独立的数据源、仲裁器、请求 FIFO(先入先出队列)、流量控制器、数据包封装器及数据包 FIFO.在每个时钟周期,每个数据源按照一定的概率产生 1 位的数据包生成请求,每个请求能够使能产生一个完整的数据包.基于重尾分布的 ON/OFF 数据源的汇聚过程能够有效地近似自相似流量,其实现方法将在 3.2 节中进行详细介绍.

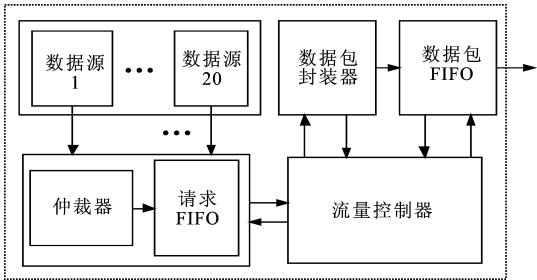


图 2 流量生成器的整体架构

每个时钟周期,当有任意数量的数据包生成请求产生时,由请求仲裁器进行仲裁管理,并将这些请求依次压入宽度为 1 位、深度为 1 024 的共享请求 FIFO 队列中.流量控制器,根据请求 FIFO 和数据包封装器的状态进行整个数据包生成过程的控制:当数据包封装器为空闲状态且请求 FIFO 为非空时,从请求 FIFO 中取出一个生成请求;然后,通知数据包封装器按照特定的协议类型和包格式,填充实际的有效载荷,进而封装产生一个完整的数据包,并将其发送到数据包 FIFO 队列中等待传输.

3.2 ON/OFF 数据源的实现

ON/OFF 数据源中,ON 和 OFF 周期长度的重尾分布采用最简单的 Pareto 分布来表示. Pareto 分布的累积分布函数为

$$F(x) = \begin{cases} 0, & x \leq k \\ 1 - (\frac{k}{x})^\alpha, & x > k, \alpha > 0 \end{cases} \quad (6)$$

式中,参数 k ($k > 0$) 为随机变量 x 的最小值, α ($0 < \alpha < 2$) 决定 x 的均值和方差.

由于 Pareto 分布很难用硬件电路实现,设计两个线性分段函数拟合器来分别产生服从 Pareto 分布的 ON 和 OFF 周期长度.如图 3 所示,每个拟合器采用一个 14 位的随机数 rand 和 19 个具有相同位数的边界值($B_1, B_2, \dots, B_{19}$)实现.第 x 个边界值定义为 ON 或 OFF 周期长度为 2^x 的累积分布概

率.采用 2^x 对 ON 或 OFF 周期长度进行分段拟合,原因在于 Pareto 分布的重尾特性导致极大值出现的概率不可忽略,而利用 2^x 能够尽可能多地涉及较大值,以提高拟合精度.

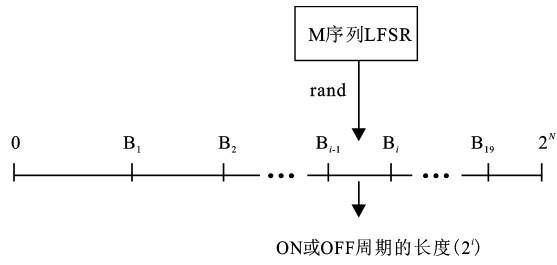


图 3 Pareto 分布分段拟合器

随机数 rand 通过一个 32 位 M 序列的线性反馈移位寄存器(LFSR)产生. M 序列的 LFSR,能够不重复地遍历除全零的所有二进制数值,32 位的 LFSR 具有 2^{32} 种数值. rand 设置为 32 位 LFSR 中的低 14 位,每个时钟周期随 LFSR 的改变而变化. 19 个边界值将整个范围 $[0, 2^{14}]$ 划分为 20 个区域,通过比较随机数 rand 和 19 个边界值,确定对应的周期长度:若 $B_{i-1} < \text{rand} \leq B_i$,当前 ON 周期或 OFF 周期的长度为 2^i .

符合 Pareto 分布的 ON/OFF 数据源的硬件结构如图 4 所示,主要包括 ON 周期拟合器、OFF 周期拟合器、周期状态控制器、多路选择器、比较器以及计数器.两个拟合器分别产生 ON 周期和 OFF 周期的长度 2^i .多路选择器,根据周期状态控制器所指示的当前周期信息(ON 或 OFF 状态),从两个拟合器中选择一个周期长度.计数器用于与周期长度 2^i 进行比较:在 ON 周期内,若计数器的值未达到 ON 周期的长度值,则生成器每个时钟周期产生一个数据包产生请求;当计数器的值到达 ON 周期长度 2^i 时,由 ON 周期切换到 OFF 周期,反之亦然.

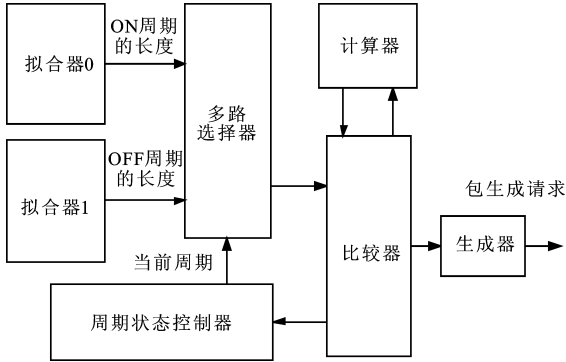


图 4 ON/OFF 数据源的实现框图

4 性能评估

4.1 自相似流量特性分析

本节对流量生成器产生的自相似序列进行统计特性分析. Hurst 参数 H 是评估自相似过程精确度的最重要的特征参数^[8]. 通常 Hurst 参数 $H(1/2 < H < 1)$ 的值越大, 自相似性越高. 在自相似流量模型中, Hurst 参数 H 定义为

$$H = (3 - \alpha)/2, \alpha = \min(\alpha_{\text{on}}, \alpha_{\text{off}}) \quad (7)$$

式中, α_{on} 和 α_{off} 分别对应 ON 周期和 OFF 周期的 Pareto 分布的形状参数.

本文设计了一个包捕获器来捕捉流量生成器产生的数据包到达时间间隔, 并采用 Hurst 参数估计器分析流量行为是否符合自相似模型. 一个好的估计器能够产生准确的 H 估计值. 目前已有几种常用的 Hurst 参数估计方法, 如小波分析估计法、R/S 分析法、方差时间法等, 其中偏差最小的是小波分析估计法^[8]. 因此, 在流量性能的评估实验中, 采用基于小波的 H 估计法来获得时间序列的 Hurst 参数估计值. 对于给定 $H = 0.6, 0.7, 0.8$ 和 0.9 , 通过对具有足够数量的 10 组时间间隔序列的 Hurst 估计值求平均值, 得到实验结果 H_{ave}^e . 通过式(8)计算 Hurst 参数 H 的精确值与估计平均值之间的相对误差 ΔH :

$$\Delta H = \frac{H_{\text{ave}}^e - H}{H} \times 100\% \quad (8)$$

式中, H 为精确值; H_{ave}^e 为估计平均值.

表 1 所示为流量生成器产生的自相似序列的 Hurst 参数 H 值的结果. 其中, H_{max}^e 和 H_{min}^e 分别为估计 H 的最大值和最小值. 对于理想值 $H = 0.7, 0.8$ 和 0.9 , 相对误差 ΔH 的绝对值都小于 5%. $H = 0.6$ 时, Hurst 参数 H 的相对误差稍大于 5%. 总体来看, 流量生成器产生的时间序列表现出较精确的自相似性, 能够很好地模拟理想的自相似流量模型.

表 1 自相似时间序列的估计 H 值与误差结果
($H = 0.6, 0.7, 0.8$ 和 0.9)

H	$\alpha_{\text{on}}/\alpha_{\text{off}}$	H_{max}^e	H_{min}^e	H_{ave}^e	$\Delta H/\%$
0.6	1.8	0.695	0.545	0.650	+8.356
0.7	1.6	0.782	0.668	0.730	+4.302
0.8	1.4	0.865	0.737	0.785	-1.880
0.9	1.2	0.944	0.821	0.880	-2.163

4.2 硬件资源开销

自相似网络流量生成器基于 Xilinx Virtex-5 XC5VFX70T FPGA 实现, 表 2 所示为流量生成器

在 FPGA 上的硬件资源消耗情况. 可以看出, 流量生成器占用了 5 310 个 6 输入查找表(LUT), 其中 LUT 最大的消耗来源于 20 个 ON/OFF 数据源逻辑、仲裁器以及数据包封装器; 而少部分 LUT 作为 RAM, 用于请求 FIFO 和数据包 FIFO. 此外, 流量生成器结构占用 1 995 个触发器(Flip Flop), 主要用在 ON/OFF 数据源中的 M 序列 LFSR. 本文提出的自相似流量生成方案具有较好的可实现性.

表 2 基于 Xilinx Virtex-5 XC5VFX70T FPGA 的硬件资源开销

资源	占用	可用	利用率/%
Flip Flops	1 995	44 800	4
6-input LUTs	5 310	44 800	11
Block RAMs	0	296	0
I/O Buffers	120	640	18

已有的 OC-48 发生器^[3]和基于 NetFPGA^[4]的流量产生器, 采用软件来生成流量的时间间隔序列, 因此速度是一大瓶颈. 此外, 在相应的文献中缺乏自相似序列的统计特性结果, 其流量的准确性和效率无法判断. 相反, 本文提出的自相似网络流量生成器, 完全通过 FPGA 硬件资源有效产生流量序列, 具有良好的速度可扩展性, 可以实现的最大数据速率能够随着硬件运行速度的增大而增加. 而且, 自相似时间序列的 Hurst 参数平均误差小, 自相似性准确度高, 能够很好地近似自相似流量. 因此, 提出的自相似网络流量生成器更有优势.

5 结束语

由于真实的网络流量具有自相似特性, 这种特性对网络性能测试的结果影响很大. 本文首先分析了自相似流量模型, 然后应用 FPGA 硬件设计并实现了一种基于 ON/OFF 模型汇聚过程的自相似网络流量生成器, 可以产生高速率的测试流量. 流量性能测试结果显示, 该流量生成器产生的流量时间序列表现的自相似统计特性参数十分接近理想值, 能够较好地模拟具有典型自相似特性的真实网络流量, 因此适用于网络设备的性能测试.

参考文献:

- [1] Ixia. IxNetwork [EB/OL]. [2015-10-20]. <http://www.ixiacom.com/products/ixnetwork>.
- [2] Avallone S, Guadagno S, Emma D, et al. D-ITG distributed Internet traffic generator [C] // International Conference on Quantitative Evaluation of Systems, Enschede, Netherlands: IEEE, 2004: 316-317.

- [3] Abdoa A, Awad H, Paredes S, et al. OC-48 configurable IP traffic generator with DWDM capability [C] // Proc Canadian Conference on Electrical and Computer Engineering. Ottawa, Canada: IEEE, 2006: 1842-1845.
- [4] Covington G A, Gibb G, Lockwood J W, et al. A packet generator on the NetFPGA platform [C] // Proc IEEE Field-Programmable Custom Computing Machines. Napa, California: IEEE, 2009: 235-238.
- [5] Leland W, Taqqu M, Willinger W, et al. On the self-similar nature of ethernet traffic (extended version) [J]. IEEE/ACM Transactions on Networking, 1994, 2(1): 1-15.
- [6] Willinger W, Taqqu M, Sherman R, et al. Self-similarity through high-variability: statistical analysis of ethernet LAN traffic at the source level [J]. IEEE ACM Transactions on Networking, 1997, 5(1): 71-86.
- [7] Crovella M E, Bestavros A. Self-similarity in World Wide Web traffic: evidence and possible causes [J]. IEEE/ACM Transactions on Networking, 1997, 5(6): 835-846.
- [8] Jeong H D. Modelling of self-similar teletraffic for simulation [D]. Christchurch: University of Canterbury, 2002.

作者简介:

王晓婷 女, (1987-), 博士研究生. 研究方向为高性能交换机结构与算法设计、SOC 实现.

E-mail: wangxiaoting12409@163.com.

王忆文 男, (1966-), 博士, 教授. 研究方向为高性能交换机技术、可重构 SOC 架构、指纹识别.

李平 男, (1957-), 教授, 博士生导师. 研究方向为 FPGA 设计与测试、SOC 设计与验证.

(上接第 53 页)

- [5] Cash D, Hofheinz D, Kiltz E, et al. Bonsai trees, or how to delegate a lattice basis [C] // Proc of Eurocrypt 2010 LNCS6110. Berlin: Springer, 2010: 523-552.
- [6] Micciancio D, Regev O. Worst-case to average-case reductions based on gaussian measures [J]. SIAM Journal on Computing, 2007, 37(1): 267-302.
- [7] Micciancio D, Vadhan S P. Statistical zero-knowledge proofs with efficient provers: Lattice Problems and more [C] // Proceedings of the Advances in Cryptology-CryPTP'03. Berlin: Springer-Verlag, 2003: 282-298.
- [8] Ajtai M. Generating hard instances of the short basis problem [C]. ICALP 1999. Berlin: Springer Verlag, 1999: 1-9.
- [9] 谢璇, 喻建平, 王廷, 等. 基于格的变色龙签名方案 [J]. 计算机科学, 2013, 40(2): 117-119.

作者简介:

彭春燕 女, (1980-), 博士研究生, 副教授. 研究方向为无线网络与安全研究. E-mail: 7456911@qq.com.

杜秀娟 女, (1970-), 教授, 博士生导师. 研究方向为水下通信协议、无线网络安全.