

基于仿真的并发流媒体汇聚特性分析

赵 昕, 陈庶樵, 张校辉

(国家数字交换系统工程技术研究中心, 河南 郑州 450002)

摘 要: 通过仿真的方式, 分析不同汇聚程度下的流媒体业务时延特性和汇聚流量自相似特征变化情况. 实验结果显示: 流媒体流量科学汇聚能够明显改善节点处理时延, 同时随着汇聚规模的增大汇聚流量的自相似性能够显著弱化. 这些结果对于当前正在开展的下一代广电网(NGB)等网络的建设有重要的参考意义.

关键词: 流媒体; 汇聚特征; 并发; 自相似; 时延

中图分类号: TP39

文献标识码: A

文章编号: 1000-7180(2011)01-0123-04

Analysis of the Aggregation Characteristics of Concurrent Streaming Media Traffic

ZHAO Xin, CHEN Shu-qiao, ZHANG Xiao-hui

(National Digital Switching System R&D Center, Zhengzhou 450002, China)

Abstract: This paper analyzed the self-similarity degree and waiting time under varying aggregating level. The experiment reveals that waiting time can be improved efficiently with scientific aggregating and the self-similarity is degraded significantly with the increase of concurrent level. The results of this paper are fairly useful in the design of the Next Generation Broadcasting network (NGB).

Key words: streaming media; aggregating characteristic; concurrent; self-similarity; delay

1 引言

近年来, 互联网出现了越来越多的基于流媒体技术的应用. 即将开始建设的下一代广播电视网(Next Generation Broadcasting Network, NGB)的一个根本出发点就是构建以IP技术为基础的支持大规模高质量流媒体业务的新型网络. NGB以有线电视网数字化整体转换和移动多媒体广播的成果为基础, 以“高性能宽带信息网(3TNet)”自主创新的核心技术为支撑, 旨在开发适合我国国情“三网融合”的, 有线无线相结合、全程全网的中国下一代广播电视网技术体系^[1]. 因此, 研究大规模流媒体业务的汇聚特性对于NGB网络的规划和建设具有重要意义^[2].

汇聚流量特性研究工作大都集中在单一业务的流量特性分析, 但是对于大规模汇聚的流媒体流量特

性的研究却并不充分. 在九十年代初期, 由美国贝尔实验室和Boston大学研究人员组成的小组采集了大量关于以太网Ethernet和可变比特率VBR视频服务^[3]的业务数据, 并利用分形思想对其进行了分析. 研究结果证明: 实际网络业务普遍存在统计上的自相似性, 随着相关研究的逐步深入, 长相关和重尾分布理论已经成为现代网络流量研究的基本工具.

因此, 文中从研究流量源特性出发, 以互联网流媒体应用广泛的MPEG-4视频为重点, 研究不同规模汇聚条件下流量汇特性的变化趋势, 重点分析流媒体大规模汇聚后的自相似特性变化情况以及相应的汇聚时延特性变化情况.

2 单一流媒体业务源特性分析

2.1 MPEG-4流媒体源自相似特性与参数估计

MPEG4是ISO/IEC的活动图像专家组(MPEG)

于 1998 年制定的视频编码标准. 最初, 针对没有或只有很少场景切换的视频会议, 研究者采用短相关模型描述其流量特性(如果随机过程 X 的自相关函数的无限和 $\sum_{k=0}^{\infty} r(k) < \infty$, 则称 X 具有短相关特性; 而如果此无限和趋于无穷, 则称 X 呈现长相关特性).

随着视频业务类型的不断丰富, 更加复杂的基于 MPEG4 的流媒体流量长相关特性逐渐被研究者发现, 因此基于分形的流量模型成为流媒体流量研究的新热点, 长相关特征已成为描述流媒体流量等网络流量的重要特征.

描述视频流量自相似性的主要参数为其 Hurst 参数(又称 H 指数), 其自相似程度越高, 则 H 指数越大. 为了估计流量的自相似程度, 人们研究了多种基于时域或频域的估计 H 指数的方法. 其中时域内估计 H 值的方法有绝对值 (ABS) 算法、方差 (Variance) 法、R/S 估计方法和方差残差 (Residuals) 算法等, 频域内有小波估计等算法. 综合考虑估计准确度和算法复杂度, 文中采用 R/S 分析法计算 MPEG-4 视频数据的 H 指数. 具体算法如下:

设时间序列 $X = \{X_k, k = 1, 2, \dots, N\}$, 记部分和函数 $X^*(t) = \sum_{u=1}^t X_u$, 定义 $S^2(t, s)$ 为子序列 $\{X_k, k = t+1, t+2, \dots, t+s\}$ 的方差, 即

$$S^2(t, s) = \frac{1}{s} \sum_{u=t+1}^{t+s} X_u^2 - \left(\frac{1}{s} \sum_{u=t+1}^{t+s} X_u \right)^2$$

定义 $R(t, s)$ 为

$$R(t, s) = \max_{0 \leq u \leq s} \{X^*(t+u) - X^*(t) - \frac{u}{s}[X^*(t+s) - X^*(t)]\} - \min_{0 \leq u \leq s} \{X^*(t+u) - X^*(t) - \frac{u}{s}[X^*(t+s) - X^*(t)]\}.$$

在数据分析时, 将长度为 N 的样本序列分为 K 块. 选取适当的 s , 计算 $\frac{R(k_i, s)}{S(k_i, s)}$, 其中 $k_1 = 1, k_2 = 1 + \frac{N}{K}, k_3 = 1 + \frac{2N}{K}, \dots$, 并且满足 $k_i - 1 + s \leq N$. 这样对于每个 s , 都可以得到一组点 $(s, \frac{R(k_i, s)}{S(k_i, s)})$, 将这些点绘于对数坐标系中. 利用最小二乘法拟和通过这些点的直线, 对于自相似过程, 其 Hurst 指数的估计值是该直线的斜率. 对于实际网络流量而言, 其 H 指数满足: $0.5 < H < 1$.

2.2 仿真流量的选取

鉴于直接得到汇聚的流媒体视频序列存在困

难, 文中以目前宽带网络环境下普遍使用的编码标准 MPEG-4 的影片为素材生成仿真中所用的流媒体流量.

实验中流量数据德原始视频影片等来源于 Berlin 大学的 MPEG-4 视频数据库^[4], 可通过其站点进行下载, 以《Star War IV》为例, 其采用 MPEG-4 压缩标准, 单层编码, 帧速率为 30 帧/秒, 编码模式固定为 IBBPBBPBBPBB, 量化参数固定为 10, 14, 18, 为低质量编码的码流. 其余影片采用高质量编码码流.

由于实际汇聚过程中流媒体不依照时间顺序, 并且数量巨大, 因此文中按照场景变化先将原始序列分解, 然后再随机组合来达到流量汇聚生成目的.

表 1 所示给出了流媒体流量汇聚前场景和汇聚后实验数据的 H 指数估计值的对比. 从该表可以看出, 分解组合前后自相似性基本没有变化. 因此, 考虑到 IP 网络中以 IP 包为基本传输单元的基本特性, 此后的研究将基于分组化数据开展.

表 1 源流量数据 H 指数

源名称	H(汇聚前)	H(汇聚后)
Alpinski	0.7436	0.7435
Die Hard III	0.7834	0.7836
Formular1	0.6711	0.6705
Jurassic Park I	0.7633	0.7633
Starship Troopers	0.7496	0.7499
Star War IV	0.7603	0.7604

3 流汇聚仿真环境

3.1 仿真网络拓扑环境

在图 1 所示的仿真网络环境中, 其拓扑结构可分三部分: 核心段 (3TNet)、汇聚节点 (ACR) 和用户驻地网. 用户驻地网以传送、调度和多种小流量业务处理. 汇聚节点负责流量的管理, 核心段根据把复杂推向边缘的思想, 减少其复杂度, 只负责透明传输. 汇聚节点的汇聚效果如图 2 所示.

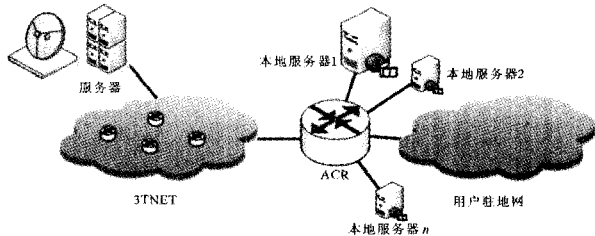


图 1 网络结构

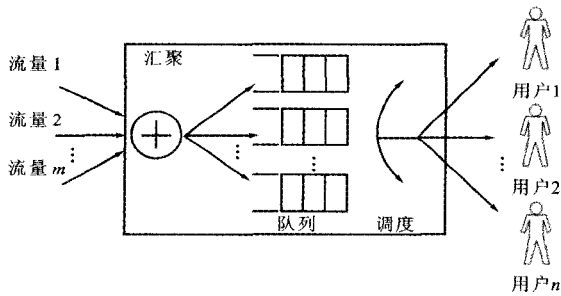


图2 汇聚方式

汇聚过程定义如下:

设 $\{X(t)\}$ 是网络中分组叠加产生的自相似流量,第 n 个流组成的分组序列为 $\{X_i(t)\}$,则由 n 个流汇聚后可简单表示为 $\{X(t)\} = \sum_{i=1}^n X_i(t)$.

文中只考虑时间复杂度和空间复杂度的主要相关因素,包括节点汇聚时延、仿真带宽、缓存管理方式、丢包率以及能够表征真实网络复杂情况的各类仿真源;但不涉及实际网络节点的各种内部复杂处理情况.各种相关参数和工作机制如下:

(1) 假设人们可以忍受的总的传输时延 200ms,如果编解码的处理时延为 80ms,那么传输时延要保证在 120ms 以下,由此假设节点汇聚时延为 20ms^[5].

(2) 仿真实验中核心链路节点链路间带宽 100MB,各个接入带宽 10MB. 缓存队列管理采用 TD(Tail Drop)方式实现,大小为 20 个分组.

(3) 可以接受的丢包率取决于,编码方式和用户体验要求.无法给出一个有效的参考数值.文中选择万分之一作为折衷.

(4) 为了能表达丰富的网络流量,文中选取了具有自相似特性的两类背景流量^[6-7],一类包括 www 应用的 get 请求和应答、FTP 和 SMTP 的流量,采用 M/Pareto 流量模型($\alpha=1.2$),平均文件大小 3KB,平均包长选择 1500 和 576 字节.

(5) 另一类代表网络中音频流量部分,采用 CBR 模型,平均包长 8KB,同样采用 Pareto 模型.同时按照视频处理的方式对两类数据进行 IP 分组化处理.

3.2 仿真数据处理

首先是数据流生成后 IP 分组到达间隔时间,具体分为两种情况:一种是连续数据包,另外一种是具有时间间隔的数据包.在链路重负载的情况下,流量是由连续的数据包组成,因此间隔时间能够用数据包的长度描述;当链路负载较轻时,包到达时间间隔

主要是数据包之间的间隙.在此,用互补累积分布函数(CCDF)来拟合包到达时间间隔的分布特性.

通过标记出需要汇聚和汇聚后的流,并进行相关统计后,得到图 3 为生成流量的包到达时间间隔的 CCDF.包到达间隔时间都可以用直线较好的拟合,由于图中的纵坐标采用了对数形式,由此可见指数分布可以对包到达时间间隔的分布进行拟合. CCDF 分别从分布的尾部与小时间间隔两个方面揭示了其具有的自相似特性.

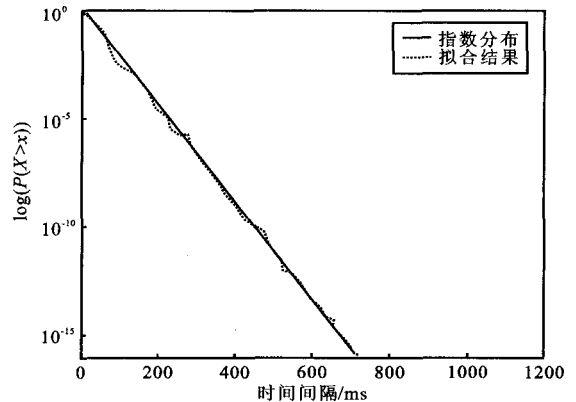


图3 汇聚后的包间隔

需要说明的是,视频源如何发送数据到网络接入对于汇聚节点的性能有着重要的影响.如果每帧的数据先缓存起来,然后在下一帧开始时就以峰值速率发送,这时会在分组之间引入难以处理的相关性.为了消除这种相关性,客观的反映汇聚过程,在下一帧的时间间隔内随机地发送数据.生成 IP 分组数据到达时间间隔与数据包长之间相互独立如图 4 所示.

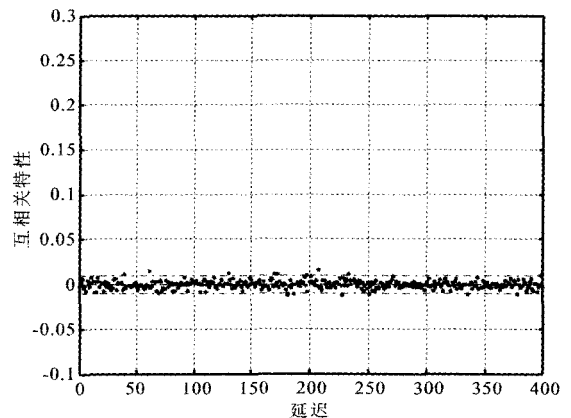


图4 包长和包间隔的相关性

4 仿真结果与分析

H 作为描述自相似特性的唯一参数,其取值范围是 0.5~1, H 越大,序列的自相似程度越高,对于

$H=0.5$, 序列对应于 Poisson 过程. 如图 5 所示, 经过源流量策略汇聚后自相似程度明显减弱, 同时随着汇聚流量的不断增加自相似性有逐步减弱的趋势, 当链路利用率较高时, 由于仿真中汇聚节点的限制所呈现的调制效应使 H 趋近于 0.5.

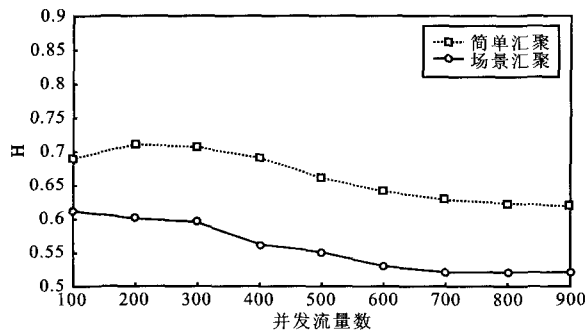


图 5 自相似特性比较

分组等待时延是决定流媒体用户体验质量的主要因素之一. 考虑到传输时延主要由链路固有特性决定, 分析了汇聚节点处理时延, 如图 6 所示. 经过场景变化进行分类, 再进行汇聚操作的流媒体流量的等待时延, 要好于简单的先入先出的汇聚方式. 因此依据业务源特性分类, 再进入各自排队系统可优化汇聚节点性能.

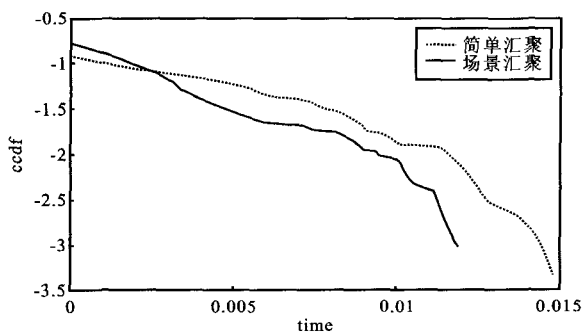


图 6 时延比较

5 结束语

基于对大规模汇聚流媒体业务流量研究的结果可以推断, NGB 网络中的汇聚流量具有两大基本特

征即自相似特性和近似的泊松特性. 这两个特征在汇聚节点环境中共存, 也可在一定的条件下相互转变, 从自相似特性向近似的泊松特性转换与网络结构和链路的负载情况有着密切的关系. 在 NGB 网络整体设计与资源规划以及关键设备开发中, 合理的利用上述结果可以有效提高网络性能和资源利用效率.

参考文献:

- [1] 科技部, 广电总局. 广电总局解读国家高性能宽带信息网等合作协议 [EB/OL]. (2008-12-05) [2010-01-01]. http://www.gov.cn/zwhd/2008-12/17/content_1180630.htm.
- [2] 郭江兴. 中国高性能宽带信息网(3TNet)综述 [J/OL]. (2002-12-16) [2010-01-01]. <http://www.c114.net/ZHANHUI/863/jx.htm>.
- [3] Leland W, Murad S Taqqu, Walter Willinger, et al. On the self-similar nature of ethernet traffic (extended version) [J]. IEEE/ACM Trans. on Networking, 1994, 2(1): 1-15.
- [4] MPEG-4 trace file [EB/OL]. (2009-04-01) [2010-01-01]. <http://www.tkn.tu-berlin.de/research/trace/ltvt.html>.
- [5] 马维旻, 李忠诚, 王俊峰. 基于仿真的宽带网络流量特征分析 [J]. 系统仿真学报, 2002, 16(4): 681-684.
- [6] 张校辉, 赵昕, 谈满堂. 流量特性在降低线卡并行处理需求中的应用 [J]. 微电子学与计算机, 2009, 26(9): 9-13.
- [7] 梁卫芳, 黄建华. 基于混合 P2P 的可运营流媒体服务体系的研究 [J]. 微电子学与计算机, 2007, 24(9): 188-191.

作者简介:

赵昕 男, (1981-), 博士研究生. 研究方向为下一代网络技术.

陈庶樵 男, (1973-), 博士, 副教授. 研究方向为高性能路由交换技术.

张校辉 男, (1979-), 博士. 研究方向为下一代网络技术.