

单向传输中丢包的相关性研究

高翔^a, 林宏刚^b

(成都信息工程学院 a. 计算机学院; b. 网络工程学院, 成都 610225)

摘要:随着单向传输技术在组播、安全隔离与信息交换、大数据高速传输以及卫星通信中的广泛应用,单向传输的性能和可靠性等问题成为单向传输应用中重点关注的问题。通过对单向传输中丢包数据的分析,得出丢包与传输速率、数据包负载的相关性,并对丢失数据包间相关性构建丢包模型。这对减少单向传输中的丢包,提高单向传输性能,以及单向传输中纠错编码方案的选择,具有重要的指导作用。

关键词:单向传输;丢包;负载;传输速率

中图分类号:TP393.0

文献标志码:A

引言

单向传输技术是一种基于信道单向传输数据的方法,具有传输速率快、资源占用少、效率高等特点,已在计算机的组播^[1-2]、安全隔离与信息交换^[3-4]、卫星通信^[5-6]、海量数据的高速传输^[7-8]等方面得到广泛的应用。然而单向传输技术仍然有一些亟待解决的问题与关键瓶颈,例如保证传输数据的可靠性,提高带宽的利用率,纠错编码的选取等,特别是较高的传输速率与高丢包率间的矛盾正成为单向传输技术中的瓶颈问题,因此对单向传输中丢包相关性的研究便显得尤为重要。

在单向传输中造成丢包的因素很多,本文重点研究高速网络中单向传输的丢包与传输速率、数据包负载间的关系。通过搭建测试平台,采集高速网络中单向传输的丢包数据,分别对丢包率与传输速率的相关性,丢包率与数据包负载的相关性进行统计分析;其次研究丢包间的相关性,针对丢失数据包间相关性提出丢包模型,并通过实验验证。本文研究成果可以应用于改善单向传输中的丢包问题,提高单向传输性能,以及为选择单向传输中纠错码提供参考。

1 测试环境与测试设计

搭建千兆网络测试平台(图1)作为测试环境,针对丢包率与传输速率的相关性,丢包率与负载的相关性,丢包间相关性的三方面进行相关测试。平台使用的服务器(CPU为I5双核,4G内存)采用双缓存机制发送和接收数据,因此可以忽略由于终端性能个体差异造成的影响。

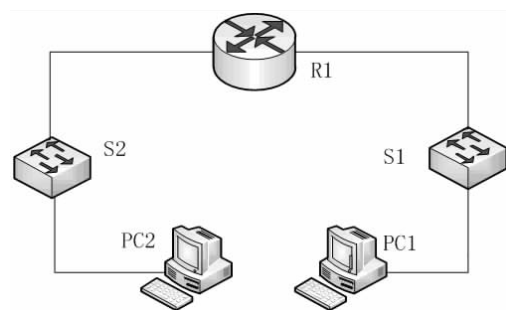


图1 单向传输系统结构图

1.1 相关参数说明

丢包率 PLR (Packet Loss Rate) 是指丢失数据包占全部发送数据包的比例,其计算方法为:丢失的数据包数

收稿日期:2013-02-21

基金项目:四川省科技厅支撑计划项目(2012FZ0016)

作者简介:高翔(1987-),男,山东青州人,硕士生,主要从事信息安全理论及应用方面的研究,(E-mail)476735206@qq.com

量 $LossPacket$ 除以发送的数据包数量 $SendPacket$ 。计算公式表示如下:

$$PLR = \frac{Losspacket}{SendPack} \quad (1)$$

传输速率指网络层单位时间内发送的数据量大小,单位为 Mbps。

负载大小指网络层 IP 数据包的负载大小,单位为 B (Byte),本文中负载大小单位均默认为 B。例如以太网的物理特性决定了数据帧的长度为 $(46 + 18) \sim (1500 + 18)$,其中的 18 是数据帧的头和尾,也就是说数据帧的内容最大为 1500 (不包括帧头和帧尾),即 MTU (Maximum Transmission Unit) 为 1500,除去头部,此时网络层数据包负载大小为 1486 字节。

丢包事件概率 p 指丢包事件发生次数 PLN (Packet Loss Number) 的接收到的数据包数 RPN (Receive Packet Number) 的比例,若一次连续丢若干数据包,只记为一次丢包事件。 p 具体计算公式如下:

$$p = \frac{PLN}{RPN} \quad (2)$$

1.2 丢包与传输速率间的相关性

本文使用实际测试得出的数据分析说明丢包与传输速率间相关性。为简化研究的复杂程度,仅将传输速率作为可变参数,研究在其变化下丢包率 PLR 的变化情况,并设计如下测试:使用图 1 中所构建的传输系统,通过专用单向传输软件,以不同的传输速率 (200、300、350、400、480、500、600,单位:Mbps) 传输 1.5G、3G、5G、10G 四种不同大小的文件,记录测试数据并根据式 (1) 计算不同传输速率下的丢包率 PLR 。因为在实际环境中大文件需要写入磁盘,不可能放在内存中,故将磁盘的写入速度上限 600 Mbps 作为测试最大值。为排除出 IP 数据包分片重组对 PLR 影响,负载选择了数据包分片的临界值 1486 字节,具体参数见表 1。

表 1 速率与数据丢包间的相关性研究参数

参数	数值
传输速率 (Mbps)	200, 300, 350, 400, 480, 500, 600
负载 (B)	1486
发送协议	UDP
发送端缓存 (软件) (B)	32M
接收端缓存 (软件) (B)	128M
带宽 (bps)	1G
测试文件大小 (B)	1.5G 3G 5G 10G
测试数据量 (B)	2000G

1.3 丢包与数据包负载的相关性

本文使用实际测试得出的数据分析说明丢包与数据包负载间的相关性。为简化研究的复杂度,仅将数

据包负载作为可变参数,研究在其变化下丢包率 PLR 的变化情况,并设计如下测试。测试使用图 1 中所构建的传输系统,并选取不同的负载大小 (512、1024、1331、1434、1486、2K、4K、8K、16K、32K、64K),通过专用单向传输软件以相同的传输速率 (300 Mbps) 发送 1.5G、3G、5G、10G 四种不同大小的文件。记录测试数据并根据式 (1) 计算不同传输速率下的 PLR 。因为 IP 数据包的包头只有两字节表示数据包长度,即 IP 数据包最大长度为 $2^{16} - 1 = 65\,535$,所以负载大小只能测试到 64K。为保证测试数据的可用性,我们将传输速率设定为 300 Mbps,因为若传输速率过小,无论多大的负载都很少丢包,若传输过大,丢包率较大得出的测试数据区分性较小,故我们采用适中值 300 Mbps。具体参数见表 2。

表 2 丢包与数据包负载间的相关性研究参数

参数	数值
传输速率 (Mbps)	300
负载大小 (B)	512, 1024, 1331, 1434, 1486, 2K, 4K, 8K, 16K
带宽 (Gbps)	1
发送协议	UDP
发送端缓存 (软件) (B)	32M
接收端缓存 (软件) (B)	128M
测试文件大小 (B)	1.5G, 3G, 5G, 10G
测试数据量 (B)	2000GB

1.4 丢包间相关性

在网络数据传输中丢包分为随机丢包与拥塞丢包。随机丢包是由于链路传输出错,造成数据包出现比特差错,上层协议将数据包简单丢弃,但其产生的丢包数较少 (对于光缆传输系统,误码率通常在 $10^{-9} \sim 10^{-12}$ 之间);拥塞丢包是由于网络节点来不及处理达到的数据包,缓存将被耗尽,这时到达的数据包将被丢弃。单向传输属于网络数据传输中的一种,与传统网络数据传输使用的传输介质基本一致,故本节借鉴传统的丢包研究方法对单向传输中丢包间的相关性进行研究。

单向传输中随机丢包事件发生主要是由于链路或设备的错误造成的,丢包是相互独立的事件,故他们呈离散分布,相互间没有相关性。而拥塞丢包是由于网络拥塞造成的,网络节点使用的 FIFO (先进先出) 以及其“丢尾”性,使两个丢包间存在短暂的相关性。

为研究拥塞丢包中丢包间的相关性,设计测试如下:为保证发生的丢包绝大多数为拥塞丢包,故选用丢包率较高的实验环境,使用图 1 所示平台进行单向传输测试,利用测试所得数据根据式 (2) 计算不同参数下的 p 值。具体测试参数见表 3。

表3 拥塞丢包测试参数

传输速率(Mbps)	负载(B)
320	8K
480	1486
560	1486
560	1331

2 测试结果及分析

2.1 丢包与传输速率相关性测试结果及分析

表4为丢包与传输速率相关性研究所得的主要数据,以及在此基础上的分析计算结果。表中的参数包括负载大小,数据传输速率数据包丢失数目以及PLR。

表4 丢包与传输速率相关性数据

负载(B)	传输速率(Mbps)	丢包数(个)	PLR
1486	200	0	0.0%
1486	300	0	0.0%
1486	350	7	4.0×10^{-9}
1486	400	4745	3.1×10^{-6}
1486	480	6545806	4.3×10^{-3}
1486	600	38209238	2.5×10^{-2}

分析统计各传输速率下的丢包数目等研究数据,通过式(1)得到了各传输速率下的PLR。将传输速率作为X轴,PLR作为Y轴得出了图2丢包率与传输速率间的关系图。

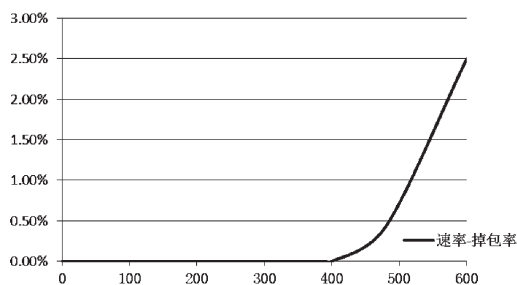


图2 丢包率-传输速率关系图

图2是负载1486 B下,PLR在传输速率0~600 Mbps间的变化图。结合表4与图2我们得出,传输速率在0~300 Mbps间PLR为零,即没有发生丢包;传输率在300~350 Mbps间发生丢包,PLR保持在 10^{-9} 数量级;350~400 Mbps时PLR开始快速增长,由 10^{-9} 数量级增长到 10^{-6} 数量级,在传输速率400~600 Mbps间时PLR由 10^{-6} 数量级增长到 10^{-2} 数量级。综上所述,传输速率在350 Mbps之后随着传输速度的增加,PLR发生数量级式的增长。

2.2 丢包与负载数据包相关性研究结果及分析

表5是丢包与数据包负载的相关性研究所得的测试数据,以及在此基础上的分析计算结果。表中的参数包括负载大小,各负载大小下的数据包丢失数目以及丢

包率PLR。

表5 丢包与负载数据包研究数据表

负载大小(B)	丢包数(个)	传输数据总量(B)	PLR
512	84	2000G	2.12×10^{-8}
1024	39	2000G	1.97×10^{-8}
1331	6	2000G	3.94×10^{-9}
1486	11	2000G	7.78×10^{-9}
1550	5189	2000G	3.93×10^{-6}
2K	8746	2000G	8.84×10^{-6}
4K	24535	2000G	4.97×10^{-5}
8K	203934	2000G	8.09×10^{-4}
16K	48984	2000G	1.6×10^{-2}
32K	6083	2000G	9.83×10^{-5}
64K	1924	2000G	6.22×10^{-5}

分析统计各负载下的丢包数等数据,通过式(1)得到各负载下的PLR。将负载作为X轴,PLR作为Y轴得出丢包率与负载间的关系图,如图3与图4所示。

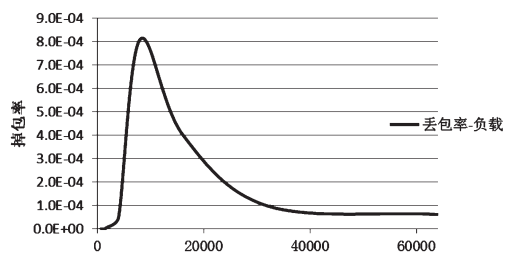


图3 丢包率负载关系图

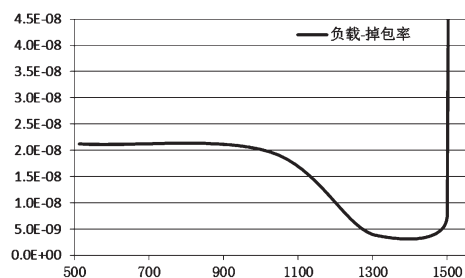


图4 丢包率负载关系子图

图3是负载在0~64 KB的完整图(IP数据包最大为64K),图4是图3中负载在500~1486 B区间图,即图4是图3中的一段曲线。综合两图与数据表我们可以得出负载在0~1486间PLR都保持在 10^{-8} 数量级,在负载1331时PLR达到最小。负载在0~1331间PLR随负载增长而减小,负载在1331~1486间PLR随负载增长而增加。在负载1486以后PLR随负载增长而成指数性增长,负载到8K时PLR达到极大值,然后PLR随负载下降。但总体来说负载大于1486时的PLR远远大于负载小于1486时的PLR。由于负载1486的数据包大小正好是IP数据包分片的数值。所以我们有理由认为IP数据包的分片重组可造成丢包率的升高。

2.3 丢包间相关性分析

根据测试数据我们可以知道当数据负载在 1486 以下,传输速率在 350 Mbps 以下时 PLR 保持在 10^{-9} 数量级,丢包事件发生数极少,因为光缆传输系统误码率通常在 $10^{-9} \sim 10^{-12}$ 之间,所以我们认为此时发生的丢包为随机丢包,丢包呈离散分布,相互间没有关系。但是随着传输速率的增长 PLR 快速增长,此时拥塞丢包发生。

本文使用两种状态的马尔可夫 (Markov) 模型,即吉尔伯特 (Gilbert) 模型来对单向传输中的拥塞丢包进行研究。Gilbert 丢包模型如图 5 所示。

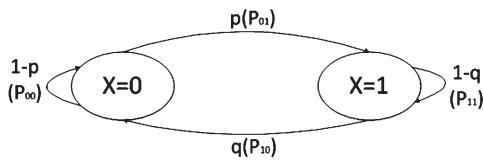


图 5 Gilbert 丢包模型

图 5 所示是在单向传输中拥塞丢包的 Gilbert 丢包模型。随机变量 X 表示丢包事件, $X=0$ 代表没有丢包, $X=1$ 表示丢包; P_{01} 表示 X 由 0 转为 1 状态的概率 p ,即发生丢包事件的概率。 P_{10} 表示 X 由 1 状态到 0 状态的概率 q ,即丢包事件占丢包总数的比例。 P_{00} 表示不发生丢包事件概率 $1-p$ 。 P_{11} 表示发生连续丢包事件的概率 $1-q$ 又记为 CLP 。根据模型 p, q 关系推理得出式 (6)。具体计算公式如下所示,其中公式 (3) ~ (4) 中的参数都是由数据统计得知,而公式 (5) ~ (6) 是由图 5 所示 Gilbert 丢包模型推导出的结果。

$$P(X = 0 | X = 1) = p \tag{3}$$

$$q = P(X = 0 | X = 1) = \frac{\text{发生丢包事件次数}}{\text{丢包总数}} \tag{4}$$

$$CLP = P(X = 1 | X = 1) = 1 - q \tag{5}$$

$$p = \frac{PLR * (1 - CLP)}{1 - PLR} \tag{6}$$

表 6 是利用测试结果数据,分别根据式 (1)、式 (4) 和式 (2) 计算不同测试环境下的 PLR, CLP 与 p 值。

表 6 网络测试结果

测试环境	负载 8K 传输速率 320 Mbps	负载 1486 传输速率 480 Mbps	负载 1486 传输速率 600 Mbps	负载 1331 传输速率 480 Mbps
测试结果	PLR = $8.09 * 10^{-4}$ CLP = $0.1p = 7.3 * 10^{-4}$	PLR = $4.30 * 10^{-3}$ CLP = $0.2p = 3.4 * 10^{-3}$	PLR = $2.51 * 10^{-2}$ CLP = $0.8p = 5.1 * 10^{-3}$	PLR = $1.97 * 10^{-2}$ CLP = $0.9p = 2.0 * 10^{-3}$

表 7 是根据表 6 中的 PLR 与 CLP 参数值利用 Gilbert 模型推导出的公式 (6) 计算出的 p 值。

表 7 模型计算结果

测试环境	负载 8K 传输速率 320 Mbps	负载 1486 传输速率 480 Mbps	负载 1486 传输速率 600 Mbps	负载 1331 传输速率 480 Mbps
环境参数	PLR = $8.09 * 10^{-4}$ CLP = 0.1	PLR = $4.30 * 10^{-3}$ CLP = 0.2	PLR = $2.51 * 10^{-2}$ CLP = 0.8	PLR = $1.97 * 10^{-2}$ CLP = 0.9
理论结果	$p = 7.281 * 10^{-4}$	$p = 3.440 * 10^{-3}$	$p = 5.149 * 10^{-3}$	$p = 2.010 * 10^{-3}$

比较两张表格 p 值,发现根据 Gilbert 模型计算得出的 p 值与实际测量结果基本一致,与文献 [9] 所得到结果相符。所以我们认为在单向传输中的拥塞丢包时,丢包间的存在如 Gilbert 模型表示的相关性,即当丢包发生时,很大程度上伴随着连续的丢包,发生连续丢包的概率为 60% ~ 90%,并且每一次拥塞增大到一定程度时就发生连续丢包,然后慢慢恢复到正常,进入一个非拥塞的时期,等待下一次拥塞的到来,这样周期性的重复。

3 结论

通过对高速网络中单向传输丢包数据的分析,可以得到以下结论:丢包率与数据包的传输速率有明显的相关性,存在某传输速率(因硬件性能而定)当传输速率超过这一数值,丢包率随传输速率增加而急剧上升。丢包率与数据包负载有明显的相关性,IP 数据包的分片重组对丢包率有影响;丢包分为随机丢包与拥塞丢包,丢包

事件为离散事件,相互间无关联性;拥塞丢包时丢包间存在相关性可用 Gilbert 模型描述,当一个数据包发生丢失时,发生连续丢包的概率较大。这些结论可以用于单向传输的应用中,减少单向传输中的丢包,提高单向传输性能,以及单向传输中纠错编码的选择,具有重要的意义。

参考文献:

[1] Zhu B, Setia S, Jajodia S, et al. Localized multicast efficient and distributed replica detection in Large-Scale sensor networks[J]. IEEE Transactions on Mobile Computing, 2010, 9(7): 913-926.

[2] Nie P P, Yao Q Q, Lu J D. Design and implementation of IP multicast video conferencing system based on JAIN SIP/JMF[J]. Computer Applications and Software, 2008,

- 25(12):189-192.
- [3] Omella M, Polo V, Lazaro J, et al. 10 Gb/s RSOA transmission by direct duobinary modulation[C]// Peter V D. Proc. of 34th European Conference on Optical Communication, Brussels, Belgium, September 21-25, 2008:1-2.
- [4] 黄胜召, 赵辉, 鲍忠贵. 网间安全隔离技术分析研究[J]. 通信技术, 2010, 40(5):100-102.
- [5] Louise D, Michael F, Alexei L, et al. Satellite control using rational agent programming[J]. IEEE Intelligent Systems, 2010, 25(3):92-97.
- [6] Wang S Q, Hou Y. Key technique on satellite communication system and its future development [J]. Modern Electronics Technique, 2009, 32(17):63-65.
- [7] 任勇毛, 唐海娜, 李俊, 等. 高速长距离网络传输协议[J]. 软件学报, 2010, 27(7):1576-1588.
- [8] Yang C T, Shih W C, Chen G H, et al. Implementation of a cloud computing environment for hiding huge amounts of data[C]// Wu C H. Proc. of 2010 International Symposium on Parallel and Distributed Processing with Applications (ISPA), Taipei, Taiwan, September 6-9, 2010:1-7.
- [9] 纪其进, 董育宁. IP 网络性能特征模型分析[J]. 通信学报, 2003, 24(3):151-159.

Research on Correlation of Packet Loss in Unidirectional Transmission

GAO Xiang^a, LIN Hong-gang^b

(a. College of Computer; b. College of Network Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: With the application of unidirectional transmission technology in multicast, safety isolation and information exchange, high-speed transmission of bulk data and satellite communication, performance and reliability become a major concern in unidirectional transmission. By analysis of the packet loss data, the correlation between packet loss and transmission rate, packet load is found. It is of significant guidance in reducing packet loss, improving the performance and selecting an error correction coding scheme in unidirectional transmission.

Key words: unidirectional transmission; packet loss; packet load; transmission speed