

基于排队论的实时以太网缓存队列优化算法

金海波, 仲崇权*

(大连理工大学 电子信息与电气工程学部, 辽宁 大连 116024)

摘要: 提出了基于排队论的实时以太网缓存队列优化算法. 首先对数据帧在实时以太网缓存队列中的传输过程进行分析, 确定了数据帧排队延时是影响网络延时的主要因素; 然后根据随机过程理论得到数据帧进入缓存队列的过程符合 Poisson 分布. 针对该分布模型, 用排队论对数据帧排队延时及丢包建立基于通信损失代价的数学模型, 以损失代价最小为目标函数, 再利用边际法计算出目标函数取极值时的最佳队列长度. 仿真实验验证了模型的准确性和优化算法的有效性.

关键词: 实时通信; 排队论; 数学建模; 模型优化

中图分类号: TP273 **文献标志码:** A

0 引言

目前工业以太网采用全双工通信、信息优先级以及虚拟局域网等技术使其响应时间达到 5 ~ 10 ms. 然而对于复杂高精度网络控制系统该响应时间仍不能满足其要求. 因此为了提高工业以太网的实时性, 各大公司及组织在 IEEE802.3 标准的基础上, 对工业以太网进行改进和扩展, 并各自提出自己的技术方案. 按照国际电工委员会 IEC/SC65C 对实时以太网的定义, 这些技术方案都做到了与标准以太网的无缝连接且响应时间小于 1 ms. 因此它们是真正意义上的实时以太网 (real-time Ethernet, RTE). 为了规范各种 RTE 技术方案, 2003 年 5 月, 国际电工委员会 IEC/SC65C 专门成立了 WG11 实时以太网工作组, 负责制定 IEC61782-2“基于 ISO/IEC8802.3 的实时应用系统中工业通信网络行规”国际标准^[1]. 该标准收录并出台了 11 种实时以太网通信行规集 (communication profile family, CPF), 包括 Ethernet/IP(CPF2)、Profinet(CPF3)、P-net/IP(CPF4) 以及 Modbus-RTS(CPF15) 和 SERCOS-III(CPF16) 等标准. 我国 EPA(CPF14) 标准也被收入其中. 由此可见在提高以太网的实时性及对以太网进行规范方面, 国内外

的专家学者、各大公司及组织都做出了巨大贡献.

然而这些被列入标准的 RTE 技术方案在访问控制层保留了 CSMA/CD 协议, 这势必导致在解决冲突时, 排队延时具有很大不确定性. 针对这一问题, 专家学者们一直不懈努力相继提出了许多改进方法, 其中有采用实时交换机技术的硬件改进方法, 有采用改进协议和信息调度的软件方法, 也有采用将排队论用于端到端时延分析的研究方法. 硬件改进方法由于其物理元器件的限制其改进难度很大, 而改变协议和信息调度的软件方法作为以太网实时通信的主要研究方向之一取得了一定的研究成果. 如 Liu 等提出的固定优先级的单调速率 (rate monotonic, RM) 算法和最小截止期优先 (earlier deadline first, EDF) 算法^[2], 以及 Sha 等提出的 Priority Ceiling Protocol^[3], 它们都在一定程度上有效地减小了改变优先级带来的影响, 但这些确定性调度方法对实时以太网本身的总体性能指标及应用运行性能的研究还很欠缺^[4]. 排队论用于分析实时以太网通信过程具有很大的优越性, 主要体现在: (1) 容易建立通信过程 (发送、接收和处理过程) 的排队模型, 便于利用经典的排队论来处理问题; (2) 抛弃了次要因素, 一定程度上简化了通信模型, 方便建模和分

收稿日期: 2010-02-13; 修回日期: 2011-11-04.

基金项目: “八六三”国家高技术研究发展计划资助项目 (2007AA041407, 2007AA041301-6).

作者简介: 金海波 (1983-), 男, 博士生; 仲崇权* (1966-), 男, 博士, 教授, E-mail: zhongcq@dlut.edu.cn.

析. 因此它能揭示排队概率的规律性, 建立系统的优化方法, 对实时以太网进行准确的性能预测、分析和评估. 目前利用排队论研究实时以太网延时的文献层出不穷^[5~8]. 文献[5]给出了基于随机输入源为 Markov 状态下的实时通信优化策略. 文献[6]用随机理论对准平稳衰减信道下的速率进行了优化并给出性能分析, 但并没有考虑发送端及中继器中延时对通信性能的影响. 文献[7]构造了一种在共享缓冲区中 FIFO 和 LIFO 两种队列并行工作的回归结构. 然而这种结构需要增加额外存储单元来保证性能的稳定性. 文献[8]给出一种分析大量信息作为输入源时且队列中每个元素赋予相同权值下的整个排队系统队长分布特性的方法. 该方法只从队长的分布特性来研究排队系统的性能, 并没有对队长进行优化.

由此可见排队论应用在通信领域是近年来的研究热点, 但也存在许多问题有待解决. 针对数据帧在发送节点及中继器缓存队列中排队延时给通信性能造成一定影响这一问题, 本文用排队论建立基于损失制的以太网传输性能的数学模型, 并给出该模型的目标函数, 推导目标函数取最小时的最佳缓存队列长度.

1 以太网缓存队列模型分析

数据帧从发送节点到接收节点传输过程中共经历以下几个环节: 应用程序产生数据帧后将其发送到网络接口卡 (net interface card, NIC) 缓存队列中等待发送; 当以太网启动一个通信周期后, NIC 缓存队列中的数据帧被发送到以太网介质中传输; 经过若干个中继器 (路由器、中间节点、网关节点等) 最终到达接收节点. 该过程如图 1 所示.

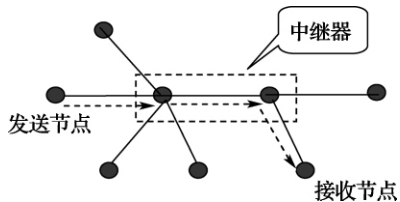


图 1 以太网传输模型

Fig. 1 Transmission model of Ethernet

从图 1 可以看出, 数据帧传输过程中网络延时包括以太网介质的传输延时和经过缓存队列时的排队延时. 由于目前工业以太网基本采用高速现场总线作为实时传输介质^[9] (100 MB 以太网)

且传输距离较小, 在以太网介质上的延时显得微乎其微. 而数据帧在缓存队列中等待发送时, 如果一次传输周期内没能及时发送需要等待下一次传输周期到来时才能发送, 就会产生较大的排队延时. 因此端到端的网络延时主要体现在排队延时上. 而产生排队延时的缓存队列包括发送节点缓存队列和中继器缓存队列. 不同物理设备上的缓存队列没有本质区别, 都是为减少数据帧丢失而开辟的存储区. 因此可以将不同设备的缓存队列作为统一的模型来分析. 下面就不同设备数据帧在 Linux 协议栈缓存队列中的发送过程进行统一分析.

图 2 中, 数据帧首先经过入口函数进入链路层, 之后由入队函数进入缓存队列. 当以太网通信周期到来时, 数据帧由出队函数出队并进入物理层入口函数得以最终发送. 以上即为数据帧在 Linux 数据链路层协议栈中的发送过程, 从发送过程也可看出, 时间延迟主要体现在数据帧进入缓存队列后的排队延时上. 因此如何减小排队延时是提高通信性能的关键.

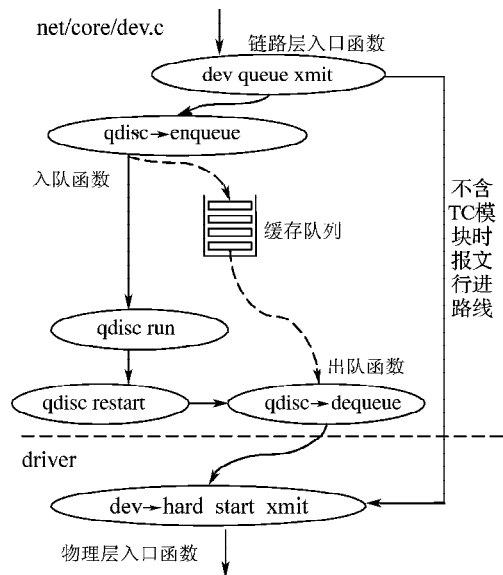


图 2 数据帧在 Linux 链路层发送流程

Fig. 2 Data frames' transmission in Linux's data link layer

2 数学模型建立及优化

对在 Linux 链路层发送数据过程分析的基础上, 建立与之对应的数学模型. 假设上位机中数据帧发送任务相互独立且发送次数没有限制, 则由排队论可知发送数据帧过程服从无限源的 Poisson 分布^[10,11]. 设单位时间内进入缓存队列的平均数据帧个数 (即数据帧的平均到达速率) 为

λ , 则到达时间间隔服从参数为 λ 的负指数分布, 其分布密度函数为

$$f(t) = \lambda e^{-\lambda t} \quad (1)$$

设实时以太网在单位时间内传输数据帧的平均个数 (即实时以太网的平均传输速率) 为 $\mu (\mu > \lambda)$, 则传输强度为

$$\rho = \lambda / \mu \quad (2)$$

当数据帧排队长度 L 过大时, 以太网处于繁忙期, 其负载过大, 势必会导致队尾的部分数据帧在一次现场总线 (实时以太网的一种) 传输周期内不能及时传输而被丢弃, 只能等到下一次传输周期到来时重新申请传输, 从而产生较大的排队延时, 对系统性能造成一定的影响. 相反如果数据帧排队长度过小, 现场总线处于闲置期, 其利用率效率很低. 因此如何确定适当的缓存队列长度使队列中数据帧都能成功发送且以太网利用率最高是本文研究的核心问题. 为了研究问题方便, 做出如下假设:

设最佳队列长度为 L_o , 当队列长度 $L > L_o$ 时, 不能及时发送而被丢弃的每个数据帧的损失代价 (即权值) 为 c_1 , 当队列长度 $L < L_o$ 时, 没有及时到达队列造成现场总线闲置的每个数据帧的损失代价为 c_2 . 第 n 个进入队列的数据帧的概率为 P_n . 由于该排队模型是 $M/G/1/\infty$ 模型, 由生灭过程的平稳分布求解公式可得

$$P_n = \frac{\sum_{i=0}^{n-1} \lambda_i}{\sum_{j=1}^n \mu_j} P_0 = \left(\frac{\lambda}{\mu} \right)^n P_0 = \rho^n P_0 \quad (3)$$

其中

$$P_0 = \left[\sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^k \right]^{-1} = 1 - \rho$$

所以得

$$P_n = (1 - \rho) \rho^n \quad (4)$$

由式 (4) 可知 P_n 服从几何分布, 且 $P_i > P_j, \forall i > j$. 当现场总线处于繁忙期时, 不能成功传输而被丢弃的平均数据帧数记为 N_d , 则

$$\begin{aligned} N_d &= \sum_{n=L_o}^{\infty} (n - L_o) P_n = \\ &= \sum_{n=L_o}^{\infty} n(1 - \rho) \rho^n - L_o \sum_{n=L_o}^{\infty} (1 - \rho) \rho^n = \\ &= \sum_{n=L_o}^{\infty} n(1 - \rho) \rho^n - L_o \rho^{L_o} \end{aligned} \quad (5)$$

当现场总线处于闲置期时, 没有进入队列而

不能及时传输的平均数据帧数记为 N_p , 则

$$\begin{aligned} N_p &= \sum_{n=0}^{L_o} (L_o - n) P_n = \\ &= L_o \sum_{n=0}^{L_o-1} (1 - \rho) \rho^n - \sum_{n=0}^{L_o-1} n(1 - \rho) \rho^n = \\ &= L_o(1 - \rho^{L_o}) - \sum_{n=0}^{\infty} n(1 - \rho) \rho^n + \\ &= \sum_{n=L_o}^{\infty} n(1 - \rho) \rho^n = \\ &= L_o(1 - \rho^{L_o}) - \frac{\rho}{1 - \rho} + \sum_{n=L_o}^{\infty} n(1 - \rho) \rho^n \end{aligned} \quad (6)$$

由于以太网繁忙与闲置都会给系统带来一定的损失, 在一次传输周期中, 兼顾这两方面的因素, 将通信损失代价记为

$$F(L_o) = c_1 N_d + c_2 N_p \quad (7)$$

将式 (7) 作为目标函数. 确定该目标函数取最小值时对应 L_o 的值, 即损失代价最小时的最佳队列长度. 下面推导式 (7):

$$\begin{aligned} F(L_o) &= c_1 N_d + c_2 N_p = \\ &= c_1 \left[\sum_{n=L_o}^{\infty} n(1 - \rho) \rho^n - L_o \rho^{L_o} \right] + \\ &= c_2 \left[L_o(1 - \rho^{L_o}) - \frac{\rho}{1 - \rho} + \sum_{n=L_o}^{\infty} n(1 - \rho) \rho^n \right] = \\ &= (c_1 + c_2) \sum_{n=L_o}^{\infty} n(1 - \rho) \rho^n - \\ &= c_1 L_o \rho^{L_o} + c_2 \left[L_o(1 - \rho^{L_o}) - \frac{\rho}{1 - \rho} \right] \end{aligned} \quad (8)$$

对上式右端第一项的求和进行化简:

$$\begin{aligned} \sum_{n=L_o}^{\infty} n(1 - \rho) \rho^n &= \rho^{L_o} \sum_{l=0}^{\infty} (l + L_o)(1 - \rho) \rho^l = \\ &= \rho^{L_o} \left[\sum_{l=0}^{\infty} l(1 - \rho) \rho^l + \right. \\ &= L_o \sum_{l=0}^{\infty} (1 - \rho) \rho^l \left. \right] = \\ &= \rho^{L_o} \left(\frac{\rho}{1 - \rho} + L_o \right) \end{aligned} \quad (9)$$

将式 (9) 代回式 (8) 得

$$\begin{aligned} F(L_o) &= (c_1 + c_2) \rho^{L_o} \left(\frac{\rho}{1 - \rho} + L_o \right) - c_1 L_o \rho^{L_o} + \\ &= c_2 \left[L_o(1 - \rho^{L_o}) - \frac{\rho}{1 - \rho} \right] = \\ &= (c_1 + c_2) \frac{\rho^{L_o+1}}{1 - \rho} + c_2 L_o - \frac{c_2 \rho}{1 - \rho} = \\ &= \frac{c_1 \rho^{L_o+1}}{1 - \rho} + c_2 L_o - \frac{c_2 \rho(1 - \rho^{L_o})}{1 - \rho} \end{aligned} \quad (10)$$

利用边际法求解式 (10), 使 $F(L_o)$ 取最小的

L_o 应满足

$$\begin{cases} F(L_o - 1) \geq F(L_o) \\ F(L_o + 1) \geq F(L_o) \end{cases} \quad (11)$$

解上述不等式组得

$$\left(\ln \frac{c_2}{c_1 + c_2} / \ln \rho \right) - 1 \leq L_o \leq \ln \frac{c_2}{c_1 + c_2} / \ln \rho \quad (12)$$

因为 L_o 表示队列长度, 即 L_o 取值必须是一整数, 所以取 L_o 为

$$L_o = \left\lfloor \ln \frac{c_2}{c_1 + c_2} / \ln \rho \right\rfloor \quad (13)$$

此时 L_o 即为所求, 其中 $\lfloor \cdot \rfloor$ 表示下取整。

一般认为申请发送的数据帧在缓存队列中没能及时传输而被丢弃这种损失对控制系统性能的影响较大, 即损失代价 c_1 较大. 而现场总线闲置这种损失对控制性能的影响较小, 即损失代价 c_2 较小. 所以满足 $c_1 \geq c_2$.

3 仿真实验

由于实时以太网的传输速度可达 100 Mbit/s, 设平均传输速率为 12.5 f/ms, 即 $\mu = 12.5$. 节点的平均申请速率为 10 f/ms, 即 $\lambda = 10$. 则传输强度 $\rho = 0.8$. 对于损失代价 c_1, c_2 给出 3 种不同的组合并分别代入式(13)得到对应的 L_o , 如表 1 所示.

表 1 损失权值与最佳队列长度关系

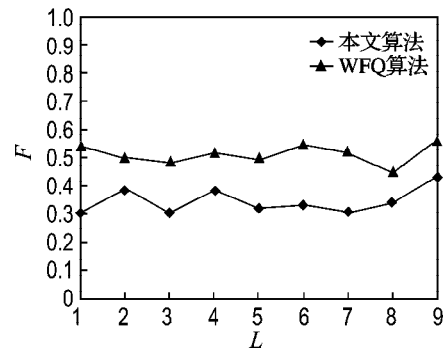
Tab. 1 Relationship of optimal queue length and loss weight

组别	c_1	c_2	L_o
第 1 组	0.5	0.5	3
第 2 组	0.7	0.3	5
第 3 组	0.8	0.2	7

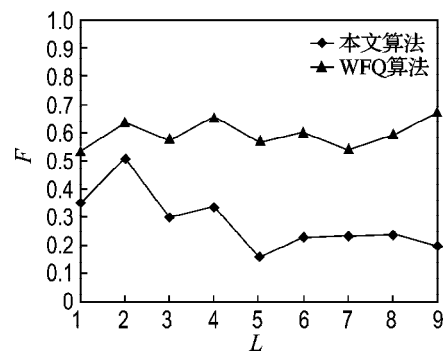
针对 3 组最佳队列长度, 用蒙特卡罗方法在 Matlab2006a 环境下进行了仿真实验, 并在相同参数下给出文献[8]WFQ(weighted fair queuing)算法结果. 实验结果如图 3 所示.

从理论结果和实验结果都可以看出, 当对丢弃的数据帧赋较大权值时(即 c_1 较大时), 说明以太网处于繁忙期, 进入以太网的数据帧较多, 为了减少数据帧不能及时传输而被丢弃就需要更多的缓冲区来临时存储, 从而避免对通信性能造成更大的影响. 而当 c_1 较小时, 说明网络处于闲置期, 进入以太网的数据帧较少, 这就不需要很大的缓冲区. 然而这种情况虽然没有丢包现象, 但是网络利用率不高. 另外从 3 组结果也能看出, 理论分析

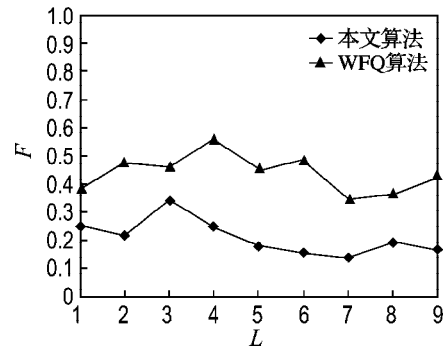
得到的最优队列长度与实验结果完全吻合.



(a) 第 1 组



(b) 第 2 组



(c) 第 3 组

图 3 系统损失值随队列长度变化关系

Fig. 3 Relationship between system loss and queue length

从对比结果中可以看出, 第 1 组权值下, 本文得到的系统损失值比 WFQ 算法平均低 0.15. 第 2 组平均低 0.3. 第 3 组平均低 0.21. 这说明对于系统损失值方面, 本文优化方案优于 WFQ 算法. 原因在于文献[8]只分析了在发送端发送数据时间足够长的情况下, 赋予相同权值队列的队长分布情况, 并没有对队列长度和系统通信性能之间的关系进行深入研究. 而本文在分析发送端、中继器中数据帧排队延时及丢包的基础上, 对缓存队列长度对系统通信性能的影响进行了深入研究, 并对队列长度进行了优化.

可见本文通过优化缓存队列长度来减小排队延时,提高通信性能的方案是行之有效的。

4 结 论

本文对实时以太网发送端及中继器缓存队列的实际模型即在 Linux 数据链路层协议栈中发送数据帧的过程进行了深入分析,得出数据帧排队延时及丢包是影响通信性能的主要因素。根据数据帧排队延时及实时以太网传输延时的分布特性,确立了二者组成的通信系统为 $M/G/1/\infty$ 排队模型。在该模型基础上,给出了在一次传输周期中实时以太网通信效率最优即通信损失代价最小这一目标函数。结合排队论对该目标函数进行简化并最终推导出通信损失代价最小时的最佳队列长度,从而对队列缓冲区的长度进行了优化。实验结果表明本文方案在一定程度上改善了以太网总体通信性能,对实际系统缓冲区大小的设计具有一定指导意义。

参考文献：

- [1] 缪学勤. 基于国际标准的十一种工业实时以太网体系结构研究(上)[J]. 仪器仪表标准化与计量, 2009(3):14-18
- [2] LIU C L, LAYLAND J W. Scheduling algorithms for multiprogramming in a hard-real-time environment [J]. *Journal of the ACM*, 1973, **20**(1):46-61
- [3] SHA L, RAJKUMAR R, LEHOCZKY J P. Priority inheritance protocols: an approach to real-time synchronization [J]. *IEEE Transactions on Computers*, 1990, **39**(9):1175-1185
- [4] 雷 攀,王行刚. 应用网络仿真技术进行网络性能评价[J]. 计算机应用, 2001, **12**(2):54-56
- [5] MAHAJAN A, TENEKETZIS D. Optimal design of sequential real-time communication systems [J]. *IEEE Transactions on Information Theory*, 2009, **55**(11):5317-5338
- [6] SHEN C, LIU T, FITZ M P. On the average rate performance of hybrid-ARQ in quasi-static fading channels [J]. *IEEE Transactions on Communications*, 2009, **57**(11):3339-3352
- [7] HUANG Po-kai, CHANG Cheng-shang, CHENG Jay, *et al.* Recursive constructions of parallel FIFO and LIFO queues with switched delay lines [J]. *IEEE Transactions on Information Theory*, 2007, **53**(5):1778-1798
- [8] ASHOUR M, LE-NGOC T. Performance of weighted fair queuing systems with long range dependent traffic inputs [C] // *Canadian Conference on Electrical and Computer Engineering 2005*. New York:IEEE, 2005:1002-1005
- [9] 王 智,王天然,宋叶琼,等. 工业实时通信网络(现场总线)的基础理论研究现状(上)[J]. 信息与控制, 2002, **31**(2):146-152
- [10] GUO Y, GONG W B, TOWSLEY D. Time-stepped hybrid simulation(TSHS) for large scale networks [J]. *Proceedings-IEEE INFOCOM*, 2000, **2**:441-450
- [11] KWEON Seok-kyu, SHIN K G. Statistical real-time communication over Ethernet [J]. *IEEE Transactions on Parallel and Distributed Systems*, 2003, **14**(3):322-335

An optimal algorithm for real-time Ethernet buffer queue based on queuing theory

JIN Hai-bo, ZHONG Chong-quan*

(Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: An optimal algorithm for a real-time Ethernet buffer queue using queuing theory is proposed. In this algorithm, frames' waiting delay is identified as the key factor in network delay through analyzing frames transmitting process in real-time Ethernet's buffer queue. The process of data frames into the buffer queue is substantially Poisson process according to stochastic process theory so that the mathematical model of transmission delay for frames queue and packet-loss based on cost is established and minimum cost in this model is considered as objective function. At last, the optimal length of queue is obtained by calculating the objective function's minimum value using marginal method. Experimental results show that the proposed model and optimal algorithm are correct and effective.

Key words: real-time transmission; queuing theory; mathematical modeling; model optimization