

优先级调度在交换式工业以太网中的应用研究

檀 明^{1,2}, 魏 臻¹, 韩江洪¹, 钟伯成²

(1 合肥工业大学 计算机与信息学院, 安徽 合肥 230009; 2 合肥学院 网络与智能信息处理重点实验室, 安徽 合肥 230601)

摘 要: 针对交换式以太网用于工业实时通信时不能确保数据按时送达问题, 在通信站点和交换机端引入 IEEE802.1p 优先级服务机制, 通过设置实时优先级字段来区分其不同的紧急程度的实时数据帧, 并采用非抢占式固定优先级调度算法改善实时数据帧的传输时延. 通过将调度分析法和网络演算法相结合, 提出了一种混合的实时数据帧传输时延上界计算方法. 比较研究表明该方法得到的结果比网络演算法更准确.

关键词: 交换式工业以太网; 非抢占式固定优先级调度; 最大时延; 最坏响应时间

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2010)05-0148-05

Application Research of Priority Scheduling in Switched Industrial Ethernet

TAN Ming^{1,2}, WEI Zhen¹, HAN Jiang-hong¹, ZHONG Bo-cheng²

(1 School of Computer and Information, HFUT, Hefei 230009, China;

2 Key Laboratory of Network and Intelligent Information Processing, Hefei University, Hefei 230601, China)

Abstract: To overcome the drawback that the switched Ethernet does not guarantee the timely delivery of real-time data in industrial communication, IEEE802.1p queuing feature was introduced into both Ethernet switches and communication nodes to give real-time data the priority to access network. In addition, real-time priority field was introduced into real-time data frames to differentiate their emergency degree, and non-preemptive fixed priority scheduling method was applied to improve the end-to-end delay of real-time data. By combining worst-case response time analysis method and network calculus approach, a hybrid method to optimize upper bound of end-to-end delay of real-time data was also proposed. A comparison study showed the advantage of the method.

Key words: switched industrial ethernet; non-preemptive fixed priority scheduling; maximum time delay; worst-case response time

1 引言

以太网作为最为广泛的局域网通信技术, 在工业数据通信中的应用日益受到重视. 但传统共享式以太网采用的 CSMA/CD 总线仲裁方式会引起传输的不确定性问题, 因而不适于工业实时数据通信. 交换式技术的发展为从本质上解决以太网通信的“确定性”提供了可能. 但是, 交换式以太网中仍存在着端口竞争, 并没有完全消除排队等待现象^[1-3].

文献[4]在以太网数据帧中引入 IEEE802.1Q

优先级字段来区分周期性实时数据和非实时数据, 并对前者优先进行传输. 对于周期性实时数据帧采用非抢占式 EDF 调度方法来减少其端到端时延, 但未考虑到交换式工业以太网中时钟不同步的问题. 同时, 由于 EDF 属于动态调度算法, 具有开销大和实时性分析困难的缺点. 文中通过在通信站点和交换机端引入双优先级队列调度机制, 采取非抢占式固定优先级算法(Fixed Priority, FP)调度实时数据的传输, 同时将调度分析法和网络演算法(Network Calculus)相结合, 实现对交换式以太网实时性的确

收稿日期: 2009-03-20; 修回日期: 2009-06-30

基金项目: 国家自然科学基金项目(60873003, 60873195)

定性分析与评价.

2 调度模型

2.1 网络数据流模型

工业控制网络的数据流主要包括:(1)周期性实时数据.主要来源于网络控制系统中传感器站点的采样数据或控制器站点的控制数据,有着严格的时间约束;(2)突发性实时数据.用于传送报警或设备状态信息,具有随机突发性;(3)非实时数据.用于系统和站点初始化、网络控制和诊断以及高层交互操作等,一般无实时要求.实时数据(周期性实时数据和突发性实时数据)与控制系统的性能密切相关,必须保证能在截止期内完成传输.对于两类实时数据的调度优先级设置方式目前并未形成统一的观点.文献[3]指出周期性实时数据的传输应首先得到保证^[4],文献[5]则将紧急的突发实时数据设置为最高优先级.文中认为两类实时数据的优先级关系应视具体情况确定,不能简单地判定哪一类相对更重要.因此,文中对两类实时数据并不做严格区分,调度器仅根据实时数据的静态优先级采用非抢占式固定优先级算法进行处理.采用五元组(L_m, T_m, D_m, J_m, P_m)表示数据流 m .其中: L_m 表示数据帧最大长度,包括实时数据长度以及成帧的附加开销(包括 8B 的帧前同步码和 12B 的帧间隙); T_m 对于周期性数据为数据的产生周期,突发性实时数据对应为数

据产生的最小时间间隔,非周期的非实时数据可忽略该项; D_m 为数据的相对时限,表示数据从产生到被正确接收的最大允许时间,对于实时数据假定 $D_m = P_m$,非实时数据可忽略该项; J_m 为队列抖动时间,表示数据到达缓冲队列的最早和最晚时间差; P_m 为实时数据的调度优先级,非实时数据忽略该项.为了方便分析,忽略队列抖动时间,从而实时数据可简化表示为(L_m, T_m, D_m, P_m).

2.2 基于优先级的实时调度模型

2.2.1 数据帧结构

利用 IEEE802.1p 的队列特性,采用 IEEE802.1Q 以太帧封装实时数据和非实时数据.在以太帧的源地址和帧类型/数据长度字段之间增加 4 个字节作为分组标签.前两个字节设定为 0x8100,称为 802.1Q 标签类型;后两个字节中前面 3 个比特是“用户优先级”(User Priority)域,接下来 1 个比特 CFI(Canonical Format Indicator)代表“规范的格式指示”,最后的 12 个比特是虚拟局域网 VLAN(Virtual Local Area Network)标识符.通过用户优先级这 3 个比特,可将用户数据分为 8 个不同的级别,优先级 0 的级别最低,7 的优先级别最高.文中将实时数据帧与非实时数据帧的用户优先级字段值分别设置为 7 和 6.另外,在实时数据帧的数据单元附加实时优先级字段,以区分不同实时数据帧的紧急程度,作为对实时数据进行优先级调度的依据,如图 1 所示.

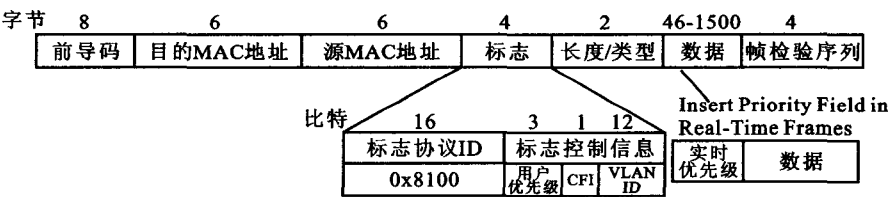


图 1 带优先级字段的 IEEE802.1Q 数据帧格式

2.2.2 双优先级队列调度模型

工业数据通信中按产生方式的不同可分为周期性数据和非周期性数据.为了利用网络演算理论进行网络实时性分析,每个通信站点对所发送的数据流采用参数为(r_m, b_m)的漏桶算法进行流量管制(如图 2).经过流量管制后得到的数据流 m 的到达曲线可以表示为 $a_m(t) = b_m + r_m \times t$,其中 r_m, b_m 分别为数据流 m 的持续平均传输速率和数据最大突发尺寸.与文献[3-4]中站点仅对非周期性数据流采用漏桶流量管制不同,文中对周期性实时数据也使用漏桶算法进行流量管制,以平滑周期性实时数据受队列抖动的影响.对于实时数据流($L_m, T_m,$

D_m, P_m),设置: $b_m = L_m, r_m = L_m/T_m$.为防止网络过载,要求满足稳定性条件:(1)网络的负荷小于等于交换机的容量;(2)交换机的任一端口的聚合数据流量不能超过端口带宽或对应链路速率.

在通信站点的 MAC 层和交换机的输出端口分别设置两个输出缓冲队列,如图 2 和图 3.一个为实时数据队列,采取非抢占式固定优先级顺序调度传输;另一个非实时数据队列用于非周期数据的传输,按 FCFS 顺序调度传输.到达的数据帧根据其不同类型(由用户优先级字段标识)分别进入实时和非实时队列进行等待,实时数据队列相对非实时数据队列具有更高的传输优先级,即只有实时数据队列为

空,非实时数据队列才能开始发送数据,避免非实时性数据阻塞实时数据的传输.对于不同紧急程度的实时数据帧,通过读取实时优先级字段,实时优先级高的数据先发送.相对使用动态调度算法(如EDF),实时数据队列采用非抢占式固定优先级调度算法具有实现简单、额外开销小、实时性评价方便等优点.

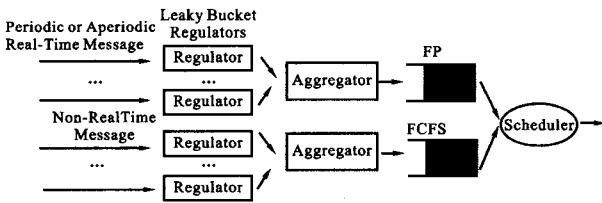


图2 通信站点的消息调度模型

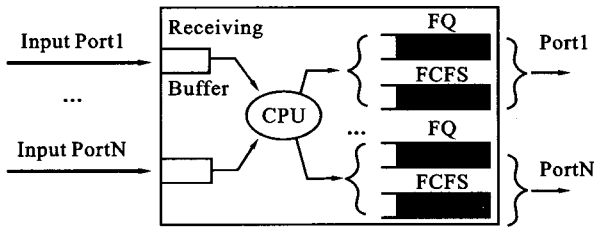


图3 双队列交换机调度模型

3 实时数据传输时延分析

针对图4所示的基于两级树型拓扑结构的全双工、微网段的网络模型,设置交换机为存储-转发(Store-Forward)工作模式,假设交换机的基本转发时延为 T_{sw} ,通信站点的发送和接收速率、链路容量与交换机的端口带宽均为 C bit/s.

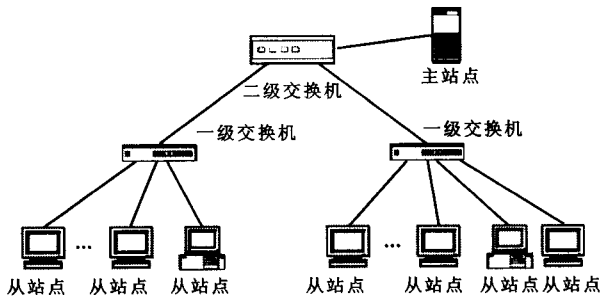


图4 两级树型拓扑网络模型

3.1 实时数据传输时延

实时数据传输时延 Delay 定义为实时数据从进入缓冲队列到被目标站点接收所经历的延迟:

$$\text{Delay} = \text{DSRC} + \text{DSW} + \text{DPROP}$$

其中,DSRC 为数据帧在发送站点所经历的延迟,包括基本的发送时间和排队时间,排队时间取决于调度方式;DSW 为数据帧经过交换机转发时所经历可

变延迟,包括基本转发时延和排队时间,排队时间取决于交换方式和调度方式;DPROP 为数据帧在传输线路上的传播延迟,它正比于传播所经电缆的长度,数据帧在物理介质上的传播延迟是容易计算的.取电缆线的速度为 $2 \times 10^8 \text{ m/s}$,那么在 20m 的电缆线上的固定延迟为 $0.1 \mu\text{s}$.为了对交换式以太网的实时能力进行分析与评价,需要计算出实时数据传输时延上界,两种主要方法分别为调度分析法和网络演算法.文献[6]表明调度分析法得到的数据传输时延上界相对更准确,但其缺点是计算量比网络演算法稍大,且不能像网络演算法那样可以方便地用于推导经过级联交换机后的数据传输时延上界.因此,文中将这两种方法相结合,即用调度分析法得到数据在通信站点的发送时延上界,而将网络演算法用于计算数据通过多级交换机的时延,实现对交换式以太网实时性的确定性分析与评价.

3.2 发送站点时延上界

利用调度分析法推导实时数据流 m 在发送站点的传输时延上界 DSRC_m ,即计算最坏情况下调度实时数据流 m 的最坏响应时间 R_m .最坏情况是指:当实时数据流 m 到达通信站点的 MAC 层发送队列时,一个低优先级且最大长度的数据刚好开始传输,同时,所有优先级比 m 高的数据也同时到达,造成对 m 的调度传输被阻塞.不失一般性,假定实时数据流 m 在 $t = 0$ 时到达,则 R_m 包括消息的排队时延 q_m 和传输时间 L_m/C 两部分,即 $R_m = q_m + L_m/C$.由于采取非抢占式固定优先级调度算法,当实时数据流 m 开始发送后就不会被中断传输^[7],因此只有在时间区间 $[0, R_m - L_m/C]$ 才能被更高优先级的实时数据所阻塞,从而有:

$$q_m = B_m + \sum_{j=1}^{m-1} \left\lceil \frac{R_m - L_m/C}{T_j} \right\rceil \times \frac{L_j}{C} \quad (1)$$

式中, $B_m = \max(L_{m+1}, \dots, L_M)/C$, B_m 为实时数据流 m 被发送站点上的低优先级数据流阻塞的最大时间; $L_1, \dots, L_{m-1}$ 为发送站点上比数据流 m 优先级高的实时数据帧长度,从而有:

$$R_m = B_m + \sum_{j=1}^{m-1} \left\lceil \frac{R_m - L_m/C}{T_j} \right\rceil \times \frac{L_j}{C} + \frac{L_m}{C} \quad (2)$$

为计算 R_m ,构造迭代式:

$$R_m^{(n+1)} = B_m + \sum_{j=1}^{m-1} \left\lceil \frac{R_m^{(n)} - L_m/C}{T_j} \right\rceil \times \frac{L_j}{C} + \frac{L_m}{C} \quad (3)$$

将 $R_m^{(0)} = L_m/C$ 代入式(3)进行迭代,若

$R_m^{(n+1)} = R_m^{(n)}, i = 1, 2, \dots$, 则 $R_m = R_m^{(n+1)} = \text{DSRC}_m$.

3.3 交换机时延上界

网络演算以最小加代数(Min-Plus Algebra)为数学基础,主要概念包括到达曲线、服务曲线以及最小加代数下的卷积和反卷积运算.其中到达曲线限制了进入进程,而服务曲线则限制了网络站点的输入和输出行为.利用网络演算法通过推导交换机提供的服务曲线和离开曲线,从而可以依次计算出数据流在各交换机上的延迟上界.对于网络演算文中只给出相关的定理^[8].

定理1 若数据流通过某服务节点的最大延迟为 $\bar{D} < +\infty$, r_{in} 和 r_{out} 分别代表数据流的输入和输出速率, b_{in} 和 b_{out} 分别代表数据流的输入突发尺寸和输出突发尺寸,数据流的到达曲线可表示为 $a_{\text{in}}(t) = b_{\text{in}} + r_{\text{in}} \times t$, 对于所有 $t \geq 0$, 则输出数据流的到达曲线为 $a_{\text{out}}(x) = a_{\text{in}}(t + \bar{D})$, 即 $r_{\text{out}} = r_{\text{in}}$, $b_{\text{out}} = b_{\text{in}} + r_{\text{in}} \times \bar{D}$.

定理2 对于采用非优抢占式固定优先级调度方法的多路复用器(Non-preemptive priority multiplexer),假设到达多路复用器的 M 个输入数据流的到达曲线为 $a_k(t) = b_k + r_k \times t, k = 1, 2, \dots, M$, 其中数据流 1 具有最高优先级,数据流 M 优先级最低,数据流长度的最大值为 L , 则数据流 m ($1 \leq m \leq M$) 经过多路复用器的最大延迟上界 $\bar{D}$ 为:

$$\bar{D}_m = \frac{\sum_{j=1}^m b_j + L}{C - \sum_{j=1}^{m-1} r_j} \quad (4)$$

根据第 2.2.2 节,当采用参数为 (r_m, b_m) 的漏桶算法对发送站点上实时数据流 m 进行流量管制后,其到达曲线可以表示为 $a_m(t) = b_m + r_m \times t$, 由式(3)可得实时数据流 m 的在发送站点的最大时延 DSRC_m . 根据定理 1, 数据流 m 到达一级交换机的最大突发尺寸为 $b_m^{(1)} = b_m + r_m \times \text{DSRC}_m$, 根据式(4)得到实时数据流 m 在第一级交换机的最大时延为:

$$\text{DSW}_m^{(1)} = \frac{\sum_{j=1}^m b_j^{(1)} + L_{\max}}{C - \sum_{j=1}^{m-1} r_j} + T_{sw} \quad (5)$$

式中, $L_{\max} = \max(L_{m+1}, \dots, L_M)$, 即到达交换机端口优先级比 m 低的数据流最大长度; $r_1, \dots, r_{m-1}$ 为到达交换机端口优先级比 m 高的实时数据流到达

速率.

类似可得实时数据流 m 在第二级交换机的最大时延为:

$$\text{DSW}_m^{(2)} = \frac{\sum_{j=1}^m b_j^{(2)} + L_{\max}}{C - \sum_{j=1}^{m-1} r_j} + T_{sw} \quad (6)$$

式中, $b_m^{(2)} = b_m^{(1)} + r_m \times \text{DSW}_m^{(1)}$.

4 仿真结果

网络仿真软件采用 OPNET Modeler, 仿真网络模型采用了图 4 所示的两级树型拓扑, 用主站点模拟工业以太网中控制器的功能, 用从站点模拟现场设备, 交换机的端口的速率为 100Mb/s. 二级交换机共连接 8 个一级交换机, 每个一级交换机均连接 n 个从站点; 连接从站、交换机和主站的网线长度都为 20m, 因此数据帧在单个传输线路上的传播延迟固定为 $0.1\mu\text{s}$. 假定实时数据帧长度均为 84B (其中 64B 的数据, 8B 的帧前同步码和 12B 的帧间隙), 从站点每 10ms 向主站点发送一个实时数据帧, 按站点编号设置该实时数据帧的实时优先级字段. 同时从站点每 100ms 还向主站点发送长度为 1538B 的非实时数据帧.

图 5 为当 n 分别取 2、4、5、10 和 20 时分别利用网络演算法、文中方法以及仿真所得到的最低优先级实时数据流对应的最大传输时延. 可以看出, 随着网络负载的增大, 实时数据的最大时延也在增加, 当网络达到较高的利用率时 ($n = 20$), 仿真所得结果说明实时数据传输的最大传输时延仍然可以满足 10ms 级控制系统的实时性能需求. 由于是按最坏情况即所有数据流同时到达进行传输延时上界的计算, 因此计算值比仿真值要大. 同时, 图 5 也表明文中方法求得的实时数据传输时延上界比网络演算法计算的结果要小.

5 结束语

文中利用 IEEE802.1Q 的用户优先级字段来区分网络中实时数据帧和非实时数据帧, 对实时数据帧利用附加的实时优先级字段区分其紧急程度, 通过在通信站点和交换机端引入双优先级队列调度机制, 保证实时信息的及时传输, 仿真结果证明方法的有效性. 另外, 将调度分析法和网络演算法相结合, 利用相对精确但计算量较大的调度分析法确定数据在发送站点的发送时延上界, 而将计算量小的网络

演算法用于推导数据经过多级交换机的时延,实现了对交换式以太网实时性的确定性分析与评价,使用本文所给出的方法求得的时延上界比直接使用网络演算法计算的结果更准确。

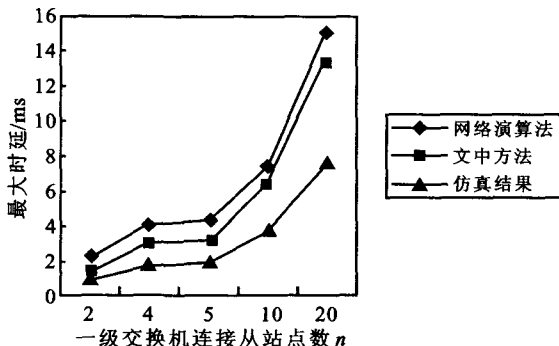


图5 最低优先级实时数据的最大时延

参考文献:

- [1] 张奇智,尹汝波.交换式工业以太网的现状和研究[J].传感器世界,2005,11(2):35-39.
- [2] Wang Zhi, Song Yeqiong, Chen Jiming, et al. Real-time characteristics of ethernet and its improvement [C]// Proc 4th World Congress on Intelligent Control and Automation. China, Shanghai, 2002:311-318.
- [3] Zhang Qizhi, Zhang Weidong. Priority scheduling in switched industrial ethernet [C]// proceedings of 2005 American Control Conference. Portland, 2005: 3366-3370.
- [4] 张奇智,曹春生,张卫东.EDF调度方法在交换式工业以太网中的实现[J].化工自动化及仪表,2004,31(6):41-43.
- [5] Jean-Philippe Georges, Thierry Divoux, Eric Rondeau. Strict priority versus weighted air queuing in switched Ethernet networks for time critical applications [C] //Proc 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05) 2 Papers. Los Angeles, 2005: 1530-2075.
- [6] Koubaa A Y Q Song. (σ, ρ) -calculus based analysis vs. worst-case response time analysis, which bound is better? [R]. [S. L]: Internal Report TRIO-LORIAINRIA, 2002.
- [7] 胡修林,李喜林,胡远斌.一种基于非抢占RM算法的可行性判决算法及应用[J].计算机与数字工程,2005,34(11):34-37.
- [8] Cruz R L. A calculus for network delay, Part I [J]. IEEE Transactions on Information Theory, 1991, 37(1): 114-131.

作者简介:

- 檀明男,(1974-),博士研究生,讲师.研究方向为工业控制网络、数据库与多媒体技术。
- 魏臻男,(1965-),博士,教授,博士生导师.研究方向为分布式控制、计算机控制等。
- 韩江洪男,(1954-),教授,博士生导师.研究方向为分布式控制、计算机控制、计算机网络等。
- 钟伯成男,(1964-),博士,教授.研究方向为计算机网络。

(上接第147页)

- [3] 李凡.无线通信中的混合ARQ技术研究[D].合肥:中国科学技术大学,2006.
- [4] 陈彦萍,李增智.基于Intserv和Diffserv的端到端QoS研究[J].微电子学与计算机,2003,20(11):24-26.
- [5] 万润泽,刘河.一种基于区分服务的适应性带宽控制策略[J].微电子学与计算机,2006,23(z1):231-233.
- [6] 武悦.关于灰色系统模型GM(1,1)适用范围的讨论[J].北京工业大学学报,1999,29(1):72-75.
- [7] StarWarsIV[EB/OL]. [2008-01-10]. <http://www.tkn.tu-berlin.de/research/trace/ltvt.html>.
- [8] Aarmvith S, Lin C W, Roy S, et al. Wireless video transport using conditional retransmissions and low-delay interleaving[J]. IEEE Transaction on Circuits and Systems of Video Technology, Special issue on Wierless Video, 2002 (1):558-565.
- [9] 梅峥.视频流化自适应错误控制技术研究[D].北京:中国科学院计算技术研究所,2005.
- [10] RFC3580, IEEE 802.1X remote authentication dial in user service(RADIUS) usage guidelines[S]. 2003.

作者简介:

- 靳勇男,硕士,助教.研究方向为无线多媒体网络、差错控制等。
- 白光伟男,博士,教授.研究方向为无线ad hoc网络、无线多媒体网络、QoS、网络体系结构和协议、网络系统性能分析、评价等。