

EDF 调度算法在直通式交换以太网上的实现

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: 提出了一种 FTT(Flexible Time-Triggered communication paradigm)的改进调度模型. 实现了周期性硬实时消息在基于成熟商用组件技术(COTS)的交换式以太网上的实时传输. 并针对周期性实时消息在直通式交换以太网上传输提出了一种基于 EDF 的实时性调度算法. 仿真实验表明该算法在提高网络的聚合带宽方面具有明显的优点.

关键词: 交换以太网; EDF 调度算法; 周期性硬实时消息; 成熟商用组件技术

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2010)03-0033-05

Realizing of EDF Scheduling Algorithm over Switched Ethernet in the Cut-through Mode

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: An improved FTT (Flexible Time-Triggered communication) paradigm was proposed for hard real-time communications of periodic messages over a COTS-based switched Ethernet. In addition, an EDF-based real-time scheduling algorithm for periodic messages on a switched Ethernet in the cut-through mode was presented. Simulation results showed the advantages of the proposed Real-Time scheduling algorithm in terms of using the aggregated switch throughput.

Key words: switched Ethernet; EDF-based scheduling algorithm; hard real-time periodic message; COTS

1 引言

将基于 COTS 技术的交换式以太网用于工业实时通信正日益受到关注, 但交换式以太网的实时通信问题仍然没有很好解决. 因此要通过网络调度以改进交换式以太网传输实时数据的能力^[1]. 基于 Master/Slaver 结构的 FTT-Ethernet 调度模型^[2-3]利用带宽预留机制保证同步实时消息的可靠传输, 剩余的带宽用来传输异步实时与非实时消息, 以提高网络带宽利用率, 实现了时间触发与事件触发的有机结合. 同时, 通过引入在线准入控制机制协调控制整个网络的变更请求, 使 FTT-Ethernet 调度模型能

很好地满足网络动态变化的需要. 由于 FTT-Ethernet 调度模型是针对共享式以太网而提出的, 虽然可实现与交换式以太网的无缝对接, 但存在对网络带宽利用不充分的缺陷. 文献[4]在 FTT-Ethernet 调度模型的基础上, 针对交换以太网提出了 FTT-SE 调度模型, 较好地发挥了交换式以太网在同一时间存在多个并行传输链路的优势, 提升了网络的整体聚合带宽, 但仍存在不足: (1) 文献[4]中给出的链路可调度判定条件本身在实时性方面仍有改进空间; (2) 针对 Master 站点的调度器设计未给出相应实时调度算法.

针对周期性实时消息在交换以太网的调度问

题,改进了 FTT-SE 调度模型.针对 Cut-Through 交换方式的以太网,给出了一种链路调度的可行性判定条件.在此基础上,提出了一种新的周期性实时调度算法,相对已有的调度算法,文中算法能更有效地利用网络带宽,提高了工业以太网通信的实时性.

2 文中采取的网络结构和调度模型

2.1 网络结构

采取全双工、微网段的通信方式,如图 1 所示.每个站点仅与一个交换机的端口相连且具有一个发送链路(TL)和一个接收链路(RL),每条链路可同时进行数据的发送和接收.交换机采取直通(Cut-Through)工作模式,不支持优先级服务机制,即输出端口只有单个缓冲队列,排队消息按 FCFS(First Come First Served)顺序传输.

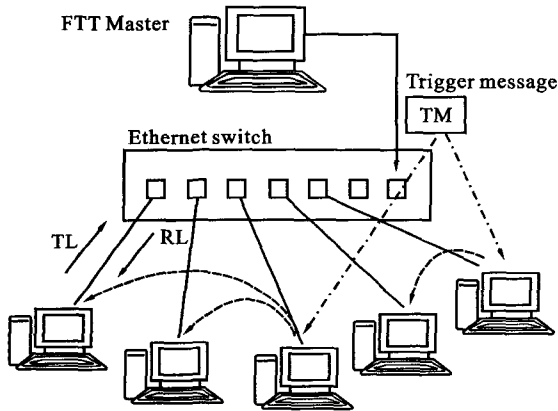


图 1 网络通信模型

2.2 调度模型

采取 FTT 同步调度模型^[3],调度的时间单位为基本周期(Elementary Cycle, EC).每个基本周期又由触发消息(Trigger Message, TM)、同步窗口(Synchronous Window, SW)和异步窗口(Asynchronous Window, AW)构成(如图 2 所示).其中 TM 用来实现站点间的同步并包含了由 Master 站点产生的实时调度信息.同步窗口用于周期性实时消息的发送,异步窗口对应非周期性实时消息及非实时消息的发送.异步窗口与同步窗口互不重叠,为防止两类窗口中的消息相互干扰,要求异步窗口与同步窗口互不重叠.

在每个基本周期内,Master 站点首先采用广播方式发送 TM,各自从站点收到 TM 后立即进行解码,按照 TM 中包含的实时调度信息实现对本地消息发送与接收.

由于典型的工业实时通信主要是针对周期性实

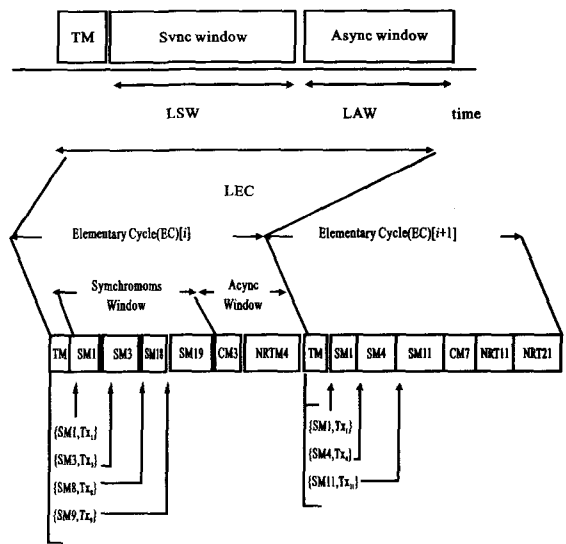


图 2 FTT-Ethernet 基本周期的组成结构

时消息,偶发实时消息的调度可通过转化为周期性实时消息加以处理,因此文中只研究周期性实时消息的调度问题.在 Master 站点中利用 SRT 数据结构表示周期性实时消息集合:

$$SRT = \{SM_k: SM_k(D_k, P_k, O_k, C_k, Src_k, Dest_k), k = 1, 2, \dots, N_s\}$$

式中, $D_k, P_k, O_k, C_k, Src_k, Dest_k$ 分别为消息 SM_k 的相对死限、产生周期、初始相位、传输时间、发送站点、接收站点, N_s 为周期性实时消息的数量.为方便分析,假定 D_k, P_k, O_k 均为基本周期 EC 长度 LEC 的整数倍,且 $D_k = P_k$.在每个 EC 内,Master 站点的调度器(Scheduler)根据 SRT 对待调度的周期性实时消息按绝对死限由小至大进行排序,形成 EDF 调度队列.由调度器依次对 EDF 调度队列中的每条消息进行链路调度的可行性判定,若其判定结果为可调度,则该消息将被纳入 TM 编码并在本 EC 被发送.与 FTT-SE 不同的是调度器将越过 EDF 调度队列中本 EC 不可调度的消息并继续后面消息的调度处理,进一步提高了网络的聚合带宽.当确定了本 EC 所有可被调度的消息后,Master 站点按 FTT-SE 方式构造 TM 并向所有从站点广播(如图 3 所示).各从站点收到 TM 后立即进行解码,确定当时 EC 本地站点应被调度传输的消息,再利用本地的 EDF 调度算法依次完成这些消息在发送链路上的传输.

根据文献[5],同步窗口可实际用于消息传输的有效长度 $LSW' = LEC - 2 * (2 * t_p + t_{sw}) - LTM - LAW$.其中 LTM 为触发消息 TM 的传输时间; LEC、LAW 分别为基本周期长度和异步窗口长度;

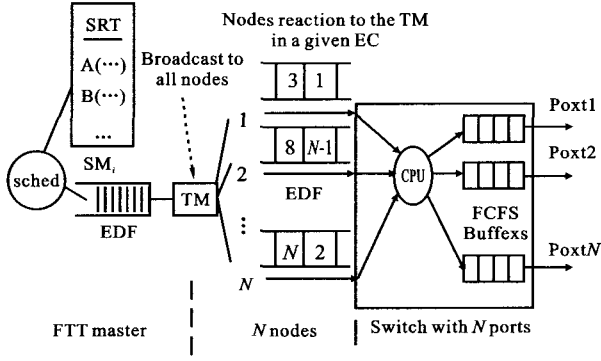


图3 周期性实时消息调度模型

t_p 为消息在发送链路上的信号传播时延; t_{sw} 为交换机基本转发时间, 包含目标端口的查找时间和交换单元的建立时间. 当使用 LSW' 进行同步窗口边界判定时, 已将 t_p 和 t_{sw} 时延从同步窗口长度 LSW 中扣除, 故可不考虑 t_p 和 t_{sw} .

3 周期性实时消息调度算法

3.1 链路调度的可行性判定

由2.2节中的实时消息调度模型, 若站点 i 向 j 发送的消息 $SM_k(D_k, P_k, O_k, C_k, Src_k, Dest_k)$ 在 EC_n (第 n 个 EC) 可被调度, 则要求 SM_k 在发送链路 (TL_i) 和接收链路 (RL_j) 上的传输能于 EC_n 的同步窗口内完成. 为方便链路调度可行性分析, Master 站点要维护各条链路在 EC_n 带宽已分配情况, 为此引入:

$T_i(n)$: EC_n 中待调度消息 SM_k 在发送链路 TL_i 传输开始时间;

$F_j(n)$: EC_n 中已调度的消息在接收链路 RL_j 传输结束时间.

当满足:

$$\text{Max}\{F_j(n), T_i(n)\} + C_k \leq LSW' \quad (1)$$

则 SM_k 在 EC_n 可被调度, $T_i(n), F_j(n)$ 按式(2)和式(3)更新:

$$T_i(n) = T_i(n) + C_k \quad (2)$$

$$F_j(n) = \text{Max}\{F_j(n), T_i(n)\} + C_k \quad (3)$$

图4中给出了 EC_1 中 SM_3 (如表1所示) 在接收链路 RL_2 的可调度性判定实例, 不失一般性, 假定 $LAW = 0, LEC = LSW' = 1$. 根据 EDF 调度原则,

表1 周期性实时消息举例1 ($LEC = LSW' = 1$)

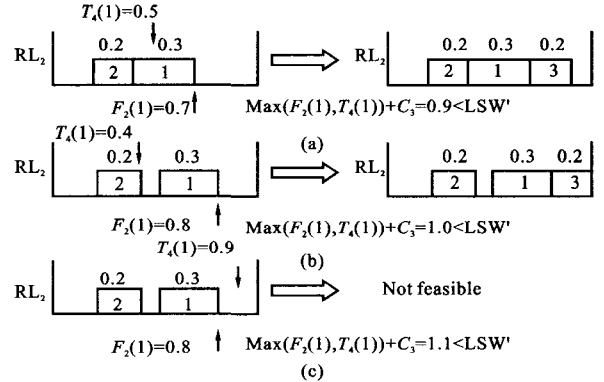
SM_k	C_k	O_k	P_k	D_k	Src_k	$Dest_k$
1	0.3	1	3	3	1	2
2	0.2	1	2	2	3	2
3	0.2	1	4	4	4	2

由于 $O_1 = O_2 = O_3$ 且 $D_2 < D_1 < D_3$, 消息调度顺序为 2, 1, 3.

(1) 由于 $\text{Max}\{F_2(1), T_4(1)\} + C_3 = 0.9 < LSW'$, 因此 SM_3 在 EC_1 可被调度;

(2) 由于 $\text{Max}\{F_2(1), T_4(1)\} + C_3 = 1.0 = LSW'$, 因此 SM_3 在 EC_1 可被调度;

(3) 由于 $\text{Max}\{F_2(1), T_4(1)\} + C_3 = 1.1 > LSW'$, 因此 SM_3 在 EC_1 不可被调度. 对(1)和(2)两种可调度情况, 按式(2)、式(3)更新 $T_4(1), F_2(1)$.

图4 EC_1 中 SM_3 在接收链路 RL_2 的可调度性判定实例

3.2 EDF 调度算法实现

EDF 属于一种动态优先级调度算法. 消息 SM_k 第 m 个实例的优先级取决于其绝对死限 $\text{abs_}D_k(m) = A_k(m) + D_k$, 其中 $A_k(m)$ 为实例的到达时间, 由第2.2节可知 $A_k(n)$ 为 EC 长度的整数倍. 由于 $A_k(n)$ 随着调度时间的推移而无限增大, 对 $\text{abs_}D_k$ 的计算和排序均不方便. 为此引入消息 SM_k 第 m 个实例在当前 EC_n 对应的松弛周期数 $\text{Slack}(k, m) = \lfloor D_k - (n * LEC - A_k(m)) \rfloor / LEC$, 显然 $\text{slack}(k, m) \leq D_k$. 容易证明, 使用 $\text{Slack}(k, m)$ 代替 $\text{abs_}D_k$ 对于 EDF 调度算法是等效的, 即 $\text{Slack}(k, m)$ 越小, 则消息的 EDF 优先级越高. $\text{Slack}(k, m)$ 的计算和更新步骤如下: (1) 消息实例产生时: $\text{Slack}(k, m) = D_k / LEC$; (2) 若某一周期 SM_k 未被调度: $\text{Slack}(k, m) = \text{Slack}(k, m) - 1$.

3.3 调度算法

// N_s : 周期性实时消息的数量

// LSW' : 同步窗口有效长度

// LEC : 基本周期长度

// $T_i(n)$: EC_n 中待调度消息 SM_k 在发送链路 TL_i 传输开始时间

// $F_j(n)$: EC_n 中已调度的消息在接收链路 RL_j 传输结束时间

// $S_k(n)$: 消息 SM_k 在 EC_n 中是否能被调度

```

// $r_k(n)$ :ECn 中是否存在消息 SMk 的实例等待被调度
//Src(SMk):返回消息 SMk 对应的发送站点
//Dest(SMk):返回消息 SMk 对应的接收站点
//QSMk:EDF 调度消息队列
//C(QSMk),P(QSMk),O(QSMk) 分别返回 消息
QSMk 的传输时间,产生周期和初始相位
1. for( $k = 1; k \leq N_s$ ;
 $k++$ ) {if( $O(SM_k)/LEC == 1$ ) $r_k(1) = 1$ ;|
2. for( $n = 1; n \leq LCM(P(SM_k)); n++$ ) {
//LCM:Least Common Multiple
3.  $T_i(n) = 0$  for all  $TL_i$ ;  $F_j(n) = 0$  for all  $RL_j$ ;
4. 对消息 SMk 集合按松弛周期数升序排列得到 EDF 调
度队列 QSMk,  $k = 1, 2, \dots, N_s$ ;
5. for ( $k = 1; k \leq N_s; k++$ ) {
6.  $r_k(n+1) = r_k(n)$ ;
7.  $S_k(n) = 0$ ; // 初始化消息 QSMk 在当前 EC 不被发
送
8. if( $r_k(n) == 1$ ) {
9.  $i = \text{Src}(QSM_k), j = \text{Dest}(QSM_k)$ ;
10. if( $\text{Max}\{F_j(n), T_i(n)\} + C(QSM_k) \leq LSW'$ ) {
11.  $T_i(n) = T_i(n) + C(QSM_k)$ ;
12.  $F_j(n) = F_j(n) + C(QSM_k)$ ;
13.  $r_k(n+1) = 0$ ;
14.  $S_k(n) = 1$ ;
15. }
16. }
17. |
18. 根据  $S_k(n), k = 1, 2, \dots, N_s$  按 FTT-SE 方式构造
TM 并向所有从站点广播;
19. 将 EDF 调度消息队列不能在本周期被调度的消息
的松弛周期数减 1;
20. if( $((n+1) \bmod P(QSM_k))/LEC == O(QSM_k)$ )
 $r_k(n+1) = 1$ ;
21. |

```

LCM(P_k/LEC) ($k = 1, 2, \dots, N_s$) 个基本周期构成一个宏周期(Macro Cycle, MC), 在每个宏周期内, 每条消息至少被调度一次. Master 站点以宏周期为一个主调度周期, 在宏周期的每个 EC 内, 调度器对待调度的周期性实时消息按松弛周期数由小至大进行排序, 形成 EDF 调度队列. 并对队列中的每条消息按式(1) 判定条件依次进行检查, 不可调度的消息被跳过并等待在后续的 EC 中被调度. 对于被调度消息, 根据式(2)、式(3) 更新 $T_i(n)$, $F_j(n)$. 当确定了本 EC 所有可被调度的消息后, Master 站点按 FTT-SE 方式构造 TM 并向所有从站点广播.

3.4 调度准入控制(Admission Control) 算法

为实现周期性实时消息的在线调度, 利用

Master 站点的调度准入控制算法协调整个网络周期消息的变更请求. 首先计算最坏情况下(所有周期性实时消息在 $t = 0$ 同时到达) 各消息的最坏完成时间 WC_k , 根据实时调度理论^[6], 当满足 $WC_k < D_k = P_k =, k = 1, 2, \dots, N_s$ 时, 则消息集合必可被调度.

```

1. for( $k = 1; k \leq N_s; k++$ ) { $WC_k = 0; r_k(1) = 1$ ;|
2. for( $n = 1; n \leq \text{Max}(P(SM_k)); n++$ ) {
3.  $T_i(n) = 0$  for all  $TL_i$ ;  $F_j(n) = 0$  for all  $RL_j$ ;
4. 对消息 SMk 集合按松弛周期数升序排列得到 EDF 调
度队列 QSMk,  $k = 1, 2, \dots, N_s$ ;
5. for( $k = 1; k \leq N_s; k++$ ) {
6.  $r_k(n+1) = r_k(n)$ ;
7. if( $r_k(n) == 1$ ) {
8.  $i = \text{Src}(QSM_k), j = \text{Dest}(QSM_k)$ ;
9. if( $\text{Max}\{F_j(n), T_i(n)\} + C(QSM_k) \leq LSW'$ ) {
10.  $T_i(n) = T_i(n) + C(QSM_k)$ ;
11.  $F_j(n) = F_j(n) + C(QSM_k)$ ;
12.  $r_k(n+1) = 0$ ;
13. if( $RWC_k = 0$ ) $RWC_k = n$ ;
14. }
15. }
16. |
17. 将 EDF 调度消息队列不能在本周期被调度的消息
的松弛周期数减 1;
18. if( $((n+1) \bmod P(QSM_k))/LEC == O(QSM_k)$ )
 $r_k(n+1) = 1$ ;
19. |

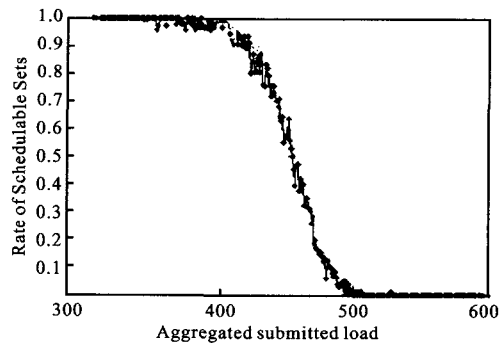
```

4 仿真实验

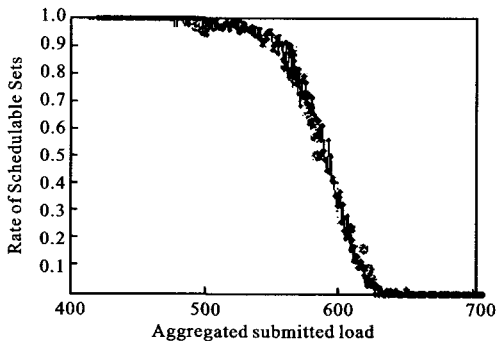
网络仿真软件采用 OPNET Modeler, 对于有 8 个站点, 交换机端口速率为 100Mb/s 的实验网络, 当 EC 为 5ms, LSW 设置为 EC 长度 85% (4.25ms) 时, 调度消息的产生周期服从 [1, 5] LEC 均匀分布, 消息的长度服从 [1 000, 1 500] bytes 的均匀分布, 消息的相对死限与产生周期相等且为 EC 长度的整数倍, 在保证交换机的缓冲区不会溢出和消息集合可调度的前提下, 通过逐渐向网络中添加消息的方式直到网络的聚合带宽达到最大值, 图 5 说明使用 FTT-SE 模型(基于 EDF 调度算法)可以将使得网络的实际聚合带宽约可达到交换机理论聚合带宽值(800Mb/s)的 51%, 而使用文中所提出的优化调度算法则约可达到 67%.

5 结束语

文中在 FTT-SE 调度模型的基础上, 利用网络



(a) FTT-SE(EDF-base)仿真结果



(a) 文中算法仿真结果

图5 仿真结果

调度改进交换式以太网传输实时数据的能力. 主要工作包括:(1) 改进了 FTT-SE 调度模型. 使对 EC_n 内消息调度不会因某个消息的不可调度而中止, 提高了网络的带宽利用率;(2) 针对 Cut-Through 交换方式的以太网, 给出了一种链路调度的可行性判定条件, 并采用消息的松弛周期数代替绝对死限方式简化 EDF 调度算法的实现;(3) 提出了一种基于 EDF 的实时性调度算法, 仿真实验表明该算法使网络的聚合带宽较 FTT-SE 调度模型有了进一步的提高.

参考文献:

[1] Tor Skeie, Svein Johannessen, Holmeide yvind. Timeli-

ness of real-time IP communication in switched industrial ethernet networks[J]. IEEE Transactions on industrial informatics, 2006,2(1):25-39.

[2] Pedreiras P, Gai P, Almeida L, et al. FTT-ethernet: a flexible real-time communication protocol that supports dynamic QoS management on ethernet-based systems [J]. IEEE Trans on Industrial Informatics, 2005,1(3): 162-172.

[3] Paulo Pedreiras, Luis Almeida. The flexible time-triggered (FTT) paradigm: an approach to QoS management in distributed real-time systems[C]// Proceedings of the International Parallel and Distributed Processing Symposium (IPDPS'03). France: University of Aveiro, 2003: 124-182.

[4] Ricardo Marau, Luis Almeida, Paulo Pedreiras. Enhancing real-time communication over COTS ethernet switches [C]// IEEE International Workshop on Factory Communication Systems(WFCS'06). Italy, Turin, 2006:295-302.

[5] Dao Manh Cuong, Myung Kyun Kim, Hee Chan Lee. Supporting hard realtime communication of periodic messages over switched ethernet[C]// IEEE Trans on Industrial Informatics. Ulsan, 2006:419-422.

[6] Almeida Luis, Paulo Pedreiras, Fonseca José Alberto G. The FTT-CAN protocol: why and how[J]. IEEE transactions on industrial electronics, 2002, 49(6): 1189-1201.

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.

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.

钟伯成 男,(1964-),博士,教授.研究方向为计算机网络.