

交换式以太网 NC 建模与实时性能分析

王晓欣, 刘鲁源, 刘 昆, 马津艳

(天津大学电气与自动化工程学院, 天津 300072)

[摘要] 交换式以太网的实时性能分析是工业以太网研究领域的一个热点课题, 目前还没有一个有效的数学模型来对其进行描述。通过对典型交换机结构的分析, 应用网络微积分理论建立了交换机的服务曲线模型, 该模型将交换机中的数据流分为实时数据和非实时数据, 并得到了最大延迟值和最大缓存需求值等, 最大延迟为交换机的实时应用提供了理论边界值, 最大缓存需求值为在实时工业环境中的交换机构建提供了理论参考。

[关键词] 网络微积分; 交换机; 到达曲线; 服务曲线; 延迟

[中图分类号] TP393 **[文献标识码]** A **[文章编号]** 1009-1742(2006)06-0055-05

1 引言

交换式以太网采用微冲突域划分技术, 解决了共享式以太网中 CSMA/CD 仲裁机制所带来的冲突延时不确定性问题。此外, 802.1p, 802.1q、全双工等功能的引入也使交换式以太网应用到实时性较强的工业控制领域成为可能^[1, 2]。尽管已经有很多将交换式以太网应用到实时数据传输的实例, 但其传输机制多采用令牌驱动、时间驱动或以高带宽来换取高实时性等方式, 带宽浪费严重。交换式以太网采用星型拓扑结构, 其核心功能由基于 802.1D 的交换机完成。由于不同厂商生产的交换机体系结构存在着很大的差异, 而且网络中的数据流具有非线性特性, 为交换式以太网的建模和实时性能分析带来了困难, 目前, 以太网交换机的建模和实时性能分析仍然是个有待解决的问题。

将采用网络微积分理论对以太网交换机进行建模, 并通过对模型的分析得出量化结论。

2 以太网交换机结构

交换机的交换结构 (switch fabric) 类型主要包括总线交换、共享内存交换和阵列交换 (crossbar)

3种, 其中阵列交换是目前以太网交换机的主流交换结构^[3], 如图1所示。阵列交换机允许多个端口的数据流同时进行交换, 大大减少了总线交换和共享内存交换结构中资源分时复用所带来的等待延迟, 而且, 输入缓存和输出缓存的采用降低了数据包的存取次数, 因此, 具有背板带宽高、交换延迟小的优点, 非常适合构建高速、低延迟的交换网络, 在企业 and 学术组织中都得到了广泛关注^[4-7]。此外优先级机制 IEEE802.1p 的采用使得区分服务 DiffSev 得以在交换式以太网中实现, 交换机的每个输入端口根据实时性要求的不同将输入数据流分成最多可达 8 个优先级队列。产生数据流的网络节点在发送数据前将优先级标签打入数据帧的 VLAN tag 字段, 然后, 交换机的输入端口逻辑分类器对其进行分类。经过阵列交换结构后, 输出端口逻辑根据不同数据流的优先级进行调度输出。将以网络微积分为数学工具对采用阵列交换结构的以太网交换机的实时性能进行分析。

3 网络微积分 NC (network calculus)

网络微积分是 20 世纪末逐渐成形的一门针对计算机网络的确定性排队系统进行分析的系统理论

[收稿日期] 2005-07-01; 修回日期 2005-08-25

[作者简介] 王晓欣 (1977-), 男 (满族), 河北遵化市人, 天津大学博士研究生, 主要研究方向工业网络建模及其实时性能

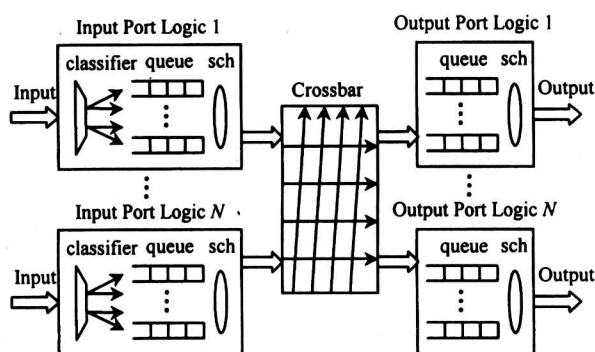


图 1 支持 802.1q 功能的以太网交换机逻辑结构

Fig.1 The logical structure of Ethernet switch supporting 802.1q

论^[8-11]，其以最小加代数（min-plus algebra）为数学基础。网络微积分系统理论将目标系统的非线性特性转化到线性域进行处理，即以非线性系统特征参数的边界值的集合作为网络的特征曲线进行分析。因此，本质上是一种最糟糕响应分析（the worst case response analysis），这对于分析硬实时系统非常有利。采用网络微积分对计算机网络进行分析时，首先要建立数据流的到达曲线（arrival curve）和传输设备的服务曲线（service curve），通过推导它们之间的内在关系得到所需要的量化参数。

到达曲线用来描述系统输入数据流特征，设输入数据流 $R(t)$ 为累积函数，当其满足 $R(t) \leq C_a(R)(t)$ ，则称 α 为系统的到达曲线，其中 $C_a(R)(t) = (\alpha \otimes R)(t)$ ，符号 \otimes 表示最小加卷积，其定义为 $h(t) \otimes x(t) = \inf_{s, 0 \leq s \leq t} \{h(t-s) + x(s)\}$ 。典型的到达曲线有漏斗模型和通用信元速率算法模型 GERA（generic cell rate algorithm），它们分别对应于精练函数（affine function）和梯形函数（staircase function）。与到达曲线不同，服务曲线则用来描述网络传输单元的服务特性，当输入数据流 $R(t)$ 进入网络系统 S 时，经传输单元中的调度器后输出数据流为 $R^*(t)$ ，当且仅当广义增函数 $\beta(0) = 0$ 且 $R^*(t) \geq (R \otimes \beta)(t)$ 时，则称系统 S 提供给数据流 $R(t)$ 的服务曲线为 $\beta(t)$ ，典型的服务曲线有通用处理器共享曲线 GPS 和 Greedy 整形器曲线。

4 交换式以太网的到达曲线和服务曲线

由文献[3]可知，阵列交换机内部存在着输入逻辑、阵列交换结构和输出逻辑 3 个串连的服务单

元，如图 1 所示，其延迟分别为输入逻辑延迟 t_{input} 、阵列交换延迟 t_s 和输出逻辑延迟 t_{output} 。阵列交换延迟 t_s 是很小的确定值，而端口逻辑中存在的调度器延迟和排队延迟，是交换机中主要的延迟，其计算最为关键。设各个单元所提供的服务曲线为 $\beta_1(t), \beta_2(t), \dots, \beta_n(t)$ ，则由文献[11]可知，这些串连服务单元所能提供的服务曲线 $\beta(t)$ 为 $\beta(t) = \beta_1(t) \otimes \beta_2(t) \dots \beta_n(t)$ ，当各个服务曲线为速率延迟函数 LR（latency-rate）时，则

$$\beta(t) = (\beta_{r_1, \theta_1} \otimes \beta_{r_2, \theta_2} \dots \beta_{r_m, \theta_m})(t) = \beta_{\min(r_1, 2, \dots, m), \theta_1 + \theta_2 + \dots + \theta_m}(t) \quad (1)$$

4.1 到达曲线

将分类后的每一个优先级队列作为集成服务数据流，则由文献[12]的到达曲线 $\alpha(t)$ 满足四元组 T-SPEC (p, M, r, b) ，即 $\alpha(t) = \min(M + pt, rt + b)$ ，其中 M 为最大数据帧长度， p 为尖峰速率， b 为最大迸发速率， r 为平均到达速率，如图 2 所示，实线为到达曲线 $\alpha(t)$ ，即：

$$\alpha(t) = \begin{cases} pt + M & 0 < t < (M - b)/(r - p) \\ rt + b & t > (M - b)/(r - p) \end{cases} \quad (2)$$

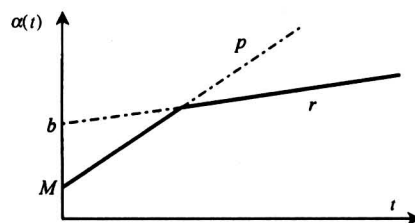


图 2 交换机中一个优先级数据流的到达曲线

Fig.2 The arrival curve of one priority data flow in switch

4.2 服务曲线

服务曲线特性由交换机内部的调度器来决定，传统的交换机内部缺乏灵活的调度策略，随着人们对网络 QoS 要求不断提高，越来越多的厂商开始在其产品内部集成了先进的调度策略，如速率 PQ，WRR，WFQ，DWRR 等，这里以 PQ 为例来建立交换机的服务曲线。工业以太网环境中通常含有实时和非实时两种数据类型，因此，这里将以双队列缓存交换为基础建立以太网交换机的服务曲线模型，并对其实时性能进行评估。如图 3 所示，数据流经过交换机的分类器分类，在输出端口缓存中形成高优先级的实时消息队列 $R_H(t)$ 和低优先级的非实

时消息队列 $R_L(t)$, 相应的输出数据流分别为 $R_H^*(t)$ 和 $R_L^*(t)$, 高优先级的消息队列的到达曲线为 $\alpha(t)$, 调度器的服务速率为 C 。

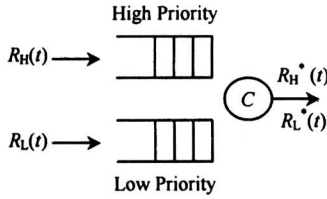


图 3 双队列交换机服务模型

Fig.3 The service model of double queues switch

4.2.1 实时消息队列服务曲线 设 s 为高优先级消息队列的沉积开始时间, 即 $R_H^*(s) = R_H(s)$, 对于大于 s 的时刻 t 来说有 $R_H^*(t) - R_H^*(s) \geq C(t - s) - M$, 则 $R_H^*(t) \geq R_H(s) + C(t - s) - M = R_H(t) + [C(t - s) - M]^+$, 其中 $[Ct - M]^+$ 为速率为 C , 延迟值为 M/C 的速率延迟函数^[13] $\beta_{C, M/C}(t)$ 。因此 $R_H^*(t) \geq (R_H \otimes \beta_{C, M/C})(t)$, 即高优先级消息队列的服务曲线为 $\beta_{C, M/C}(t)$ 。

4.2.2 非实时消息队列服务曲线 设当高优先级消息的到达曲线为 $\alpha_H(t)$ 满足公式(1)时, 低优先级消息队列不会出现由于高优先级的消息到达速率过快而产生“饥饿”现象。当低优先级消息队列开始在时间 s' 沉积时, 则有 $R_H^*(t) = R_H(s')$ 和 $R_L^*(t) = R_L(s')$, 对于大于 s' 的时间 t :

$$R_L^*(t) - R_L^*(s') = C(t - s') - [R_H^*(t) - R_H^*(s')] \quad (3)$$

又因为 $R_H^*(t) - R_H^*(s) = R_H^*(t) - R_H(s') \leq R_H(t) - R_H(s') \leq \alpha_H(t - s')$, 且 $R_H^*(t) - R_H^*(s') \geq 0$, 则:

$$R_L^*(t) - R_L(s') = R_L^*(t) - R_L^*(s') \geq [C(t - s') - \alpha_H(t - s')]^+ \quad (4)$$

当 $0 < t < (M - b)/(r - p)$ 时有

$$[C(t - s') - \alpha_H(t - s')]^+ = [(C - p)(t - s') - M]^+ = \beta_{C-p, M/(C-p)}(t - s') \quad (5)$$

当 $t > (M - b)/(r - p)$ 时有

$$[C(t - s') - \alpha_H(t - s')]^+ = [(C - r)(t - s') - b]^+ = \beta_{C-r, M/(b-r)}(t - s') \quad (6)$$

由式 (5) 和式 (6), 设

$$S(t) = \begin{cases} \beta_{C-p, M/(C-p)}(t) & 0 < t < (M - b)/(r - p) \\ \beta_{C-r, b/(C-r)}(t) & t > (M - b)/(r - p) \end{cases} \quad (7)$$

则结合公式 (4) 有 $R_L^*(t) \geq (R_L \otimes s)(t)$, 即调度器为低优先级消息队列提供的服务曲线 $S(t)$ 是分段的速率延迟函数。

5 交换式以太网的实时性能评估

交换式以太网的实时特性是由网络中所传输数据的到达曲线和交换机的服务曲线决定的, 此外, 交换机内部为消息队列所提供的缓存大小也直接影响着传输的 QoS。

实时数据流的到达曲线 $\alpha(t) = \min(p_1 t + M, r_1 t + b_1)$ 与服务曲线 $S_1(t) = (\beta_{C, M/C})(t)$ 如图 4 所示, 其中 $t_{max} = (M - b_1)/(r_1 - b_1)$ 。根据网络微积分理论中延迟边界的计算方法^[10], 当系统的到达曲线和服务曲线分别为 $\alpha(t)$ 和 $\beta(t)$ 时, 数据流 $R(t)$ 的延迟边界 t_d 为 $\alpha(t)$ 与 $\beta(t)$ 的垂直距离最大值, 即 $t_d = \max_{t, \tau \geq 0} \{ \min \{ \Delta : \Delta \geq 0 \text{ and } \alpha(t) \leq \beta(t + \Delta) \} \}$ 。因此, 实时数据的最大延迟为

$$t_{dH} = (r_1 - C)(M - b_1)/(r_1 - p_1) + (b_1 + M) \quad (8)$$

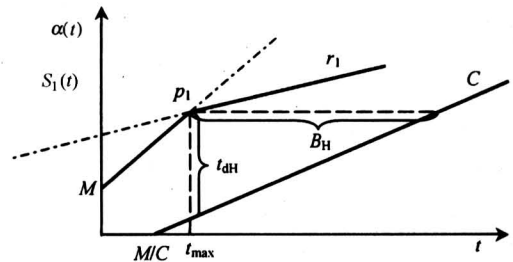


图 4 实时消息队列延迟边界和沉积边界

Fig.4 The delay and backlog bounds of real-time queue

实时数据所需的最大缓存由其沉积边界决定, 即到达曲线与服务曲线的横向最大距离

$$B_H = (M(p_1 + 2r_1 + C) + b_1 C)/C(r_1 - p_1) \quad (9)$$

由四元组 T -SPEC (p, M, r, b) 的特性可知 $C - r_1 > C - p_1$, 因此, 非实时数据流到达曲线 $\alpha_2(t) = \min(p_2 t + M, r_2 t + b_2)$ 和服务曲线 $S_2(t)$ 如图 5 所示, 其中 $t_{max} = (M - b_2)/(r_2 - p_2)$, t_{dL} 存在 t_{dL} 和 t'_{dL} 两种情况, 而由文献[11]可知, 在实际网络中通常有 $C - p_1 \leq r_2$, 因此得到非实时数据流的最大延迟值为

$$t_{dl} = t'_{d2} = (r_1 + r_2 - C)(M - b_1) / ((r_1 - p_1) + (b_1 + b_2)) \quad (10)$$

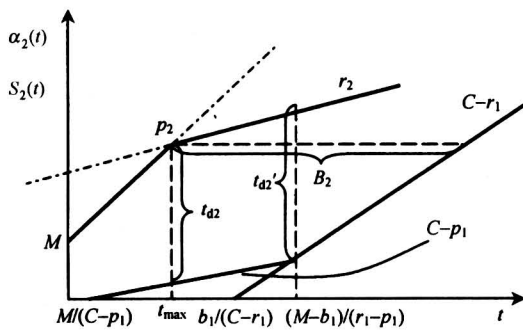


图 5 非实时消息队列延迟边界和沉积边界

Fig.5 The delay and backlog bounds of non real-time queue

由 $p_2 t_{\max} + M = (C - r_1)t - b_1$ 可得非实时数据所需最大缓存为

$$B_L = p_1(M - b_2) / (C - r_1)(r_2 - p_2) + (M + b_1) / (C - r_1) - (M - b_2) / (r_2 - p_2) \quad (11)$$

当交换机的输入端口逻辑与输出端口逻辑采用相同的调度器时, 其服务曲线相同, 均为速率延迟函数。因此, 根据式 (1), 当实时和非实时数据流的队列缓存的大小分别满足式 (9) 和式 (11) 时, 其交换延迟的最大值分别为 $t_{\text{switch}} = 2t_{dH} + t_s$ 和 $t_{\text{switch}} = 2t_{dL} + t_s$, 其中 t_{dH} 和 t_{dL} 分别满足式 (8) 和式 (10)。此外, 由 GPS 特性可知, 每个队列的平均到达速率和应小于最大服务速率 C , 即 $\sum_{k=1}^N r_k \leq C$ 。

6 总结

采用了网络微积分的确定性理论建立了具有实时和非实时双队列的交换机的到达曲线和服务曲线模型, 得到了交换机的最大延迟值和最大缓存需求值, 为交换机以太网的实时性能分析提供了有效理论指导。网络微积分理论在不断发展, 已经能够进行网络性能的确定性分析, 其随机性理论还处在起步阶段。下一步的工作是研究应用网络微积分的随机性理论对交换机中的非实时数据流进行建模和分析, 以提高非实时数据服务曲线的软延迟特性。

参考文献

[1] Hoang H, Jonsson M, Hagstrom U. Switched real-time ethernet with earliest deadline first scheduling protocols and

traffic handling [A]. Proceedings of the International Parallel and Distributed Processing Symposium [C]. 2002. 94 ~ 99

- [2] Hoang H. Real-Time Communication for Industrial Embedded Systems Using Switched Ethernet [A]. Parallel and Distributed Processing Symposium [C]. 2004. 127 ~ 130
- [3] Seifert R. The Switch Book: The Complete Guide to LAN Switching Technology [M], Wiley, 2000. 587 ~ 647
- [4] Minkenber C, Engbersen T. A combined input and output queued packet-switched system based on PRIZMA switch-on-a-chip technology [J], Communications Magazine, 2000, 38(12): 70 ~ 77
- [5] Li Yihan, Panwar S, Chao H J. On the performance of a dual round-robin switch [A], Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies [C]. 2001. 1688 ~ 1697
- [6] Javidi T, Magill R, Hrabik T. A high-throughput scheduling algorithm for a buffered crossbar switch fabric [A], IEEE International Conference on Communication, Vol 5 [C]. 2001. 1586 ~ 1591
- [7] Lin Mingjie, McKeown N. The Throughput of a buffered crossbar switch [J]. IEEE Communication Letters, 2005, 9(5): 465 ~ 467
- [8] Agrawal R, Cruz R L, Okino C, Rajan R. Performance bounds for flow control protocols [J]. IEEE/ACM Transactions on Networking, 1999, (7)3: 310 ~ 323
- [9] Chang Chengshang, Cruz R L, Boudec J-Y L, Thiran P. A min-plus system theory for constrained traffic regulation and dynamic service guarantees [J]. IEEE/ACM Transactions on Networking, 2002, 10(6): 805 ~ 817
- [10] Boudec J-Y L, Thiran P, Giordano S. A short tutorial on network calculus II: min-plux system theory applied to communication networks [A]. The IEEE International Symposium on Circuits and Systems, Vol 4 [C]. 2000. 365 ~ 368
- [11] Boudec J-Y L, Thiran P. Network Calculus: a Theory of Deterministic Queuing Systems for the Internet [M]. LNCS2050, Springer Verlag, 2004
- [12] Blake S, Black D, Carlson M, et al. An Architecture for Differentiated Services [S]. IETF Standards Track RFC-2475, December 1998
- [13] Stiliadis D, Varma A. Rate latency servers: a general model for analysis of traffic scheduling algorithms [J]. IEEE/ACM Transactions on Networking, 1998, 6(5). 611 ~ 624

Modeling and Real-time Performance Analysis of Switched Ethernet by Network Calculus

Wang Xiaoxin, Liu Luyuan, Liu Kun, Ma Jinyan

(*School of Electrical & Automation Engineering, Tianjin University, Tianjin 300072, China*)

[**Abstract**] The real-time performance of switched Ethernet is a hot research field. However, there isn't an effective modeling to describe it until now. Through analyzing the structure of classical switch, a switch model by network calculus is proposed, in which the data flow is divided into real-time flow and non-real-time flow and the arrival curve, service curve, the max delay and max storage needed are got. The max delay provides theory direction for transmitting real-time data in switched Ethernet and the max storage provides the theory reference for constructing switch in real-time industrial Ethernet.

[**Key words**] network calculus; switch; arrival curve; service curve; delay

(cont. from p.43)

Effects of Structure of Opening Hole in Gas Distributor on Detained Mass of Material for Fluidized Drying

Liu Wei^{1,2}, Tang Wencheng¹

(1. *Department of Mechanical Engineering, Southeast University, Nanjing 210096, China*;

2. *Department of Mechanical Electronics Engineering, Nanjing Normal University, Nanjing 210042, China*)

[**Abstract**] The value of detained mass of material is important for researching the dead bed point of fluidized drying. In order to acquire the effect law of structure of opening hole in gas distributor on detained mass of material, the drying processes of detergent suspension liquor were investigated in the inert particle fluidized bed drier. The curves of detained mass of material for drier with vertical-hole distributor and tilted-hole distributor were measured and compared under the same experimental parameters, such as feed volume, diameter of inert particles, gas temperature of entrance, height of static bed, gas rate of entrance, and concentration of material. Based on the result, the relationship data of drier with tilted-hole distributor were determined between the production capacity and the hole ratio of distributor. The results show that the detained mass of material increase with adding the feed volume or concentration of material, and decrease with adding the diameter of inert particles, gas temperature of entrance, height of static bed, or gas rate of entrance. In addition, changing the hole shape of distributor from vertical-hole to tilted-hole and adding the hole ratio of distributor, can help drier reduce the detained mass of material and enhance the production capacity.

[**Key words**] inert particles; fluidized drying; gas distributor; hole shape; hole ratio; detained mass of material