P.O.Box 8718, Beijing 100080, China | Journal of Software May 2003,14(5):1011-1022
E-mail: jos@iscas.ac.cn | ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn | Copyright © 2003 by The Editorial Department of Journal of Software

高速IP路由器中输入排队调度算法综述

庞 斌, 贺思敏, 高 文

Full-Text PDF Submission Back

庞 斌1, 贺思敏1, 高 文1,2,3 1(中国科学院 计算技术研究所,北京 100080) 2(哈尔滨工业大学 计算机科学与工程系,黑龙江 哈尔滨 150001)3(中国科学院 研究生院,北京 100039)
第一作者: 庞斌(1971－),男,山东临沂人,博士生,主要研究领域为计算机网络,多媒体通信技术.
联系人: 庞斌 Telephone: 86-10-82649316, Fax: 86-10-82649298, E-mail: bpang@jdl.ac.cn

Received 2002-06-06; Accepted 2002-12-10

Abstract

Most high-speed IP routers exploit cell-based switching fabrics, whose scalability and performance are mainly affected by queuing scheme and scheduling algorithm. Input-queued router is referred to as an ideal structure in terms of scalability. However, it needs an efficient scheduling algorithm to guarantee throughput and delay. Several input-queued scheduling algorithms are surveyed in this paper. The scheduling algorithms are classified into four classes: maximum size matching, maximum weight matching, stable marriage matching, and deterministic scheduling algorithm. The similarities and the difference of different algorithms in mechanisms of each class are described, and their performances are compared. Finally, the future directions and possible open problems are discussed.

Pang B, He SM, Gao W. A survey on input-queued scheduling algorithms in high-speed IP routers. *Journal of Software*, 2003,14(5):1011~1022.
http://www.jos.org.cn/1000-9825/14/1011.htm

摘要

高速IP路由器一般采用基于定长信元的交换结构,其可扩展性和性能分别受排队策略和调度算法的影响.基于输入排队策略的路由器具有良好的可扩展性,但需要一个有效的调度算法的支持,才能保证吞吐率和延迟等性能.主要讨论输入排队调度算法,将现有的调度算法分为4类:最大(无权重)匹配、最大权重匹配、稳定婚姻匹配和确定型调度.对每一类算法,从技术特点和性能指标两个方面进行比较和分析.最后给出了输入排队调度算法的发展趋势.

基金项目：Supported by the National Natural Science Foundation of China under Grant No.69983008 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2001AA112100 (国家高技术研究发展计划(863)); the Knowledge Innovation Program in CAS under Grant No.KGCXZ-103 (中国科学院知识创新工程)

References:

[1] Bux W, Denzel WE, Engberson T, Herkersdorf A, Luijten RP. Technologies and building blocks for fast packet forwarding. IEEE Communication Magazine, 2001,39(1):70~77.

[2] Nong G, Hamdi M. On the provision of quality-of-service guarantees for input queued switches. IEEE Communications Magazine, 2000,38(12):62~69.

[3] Javidi T, Magill R, Hrabik T. A high-throughput scheduling algorithm for a buffered crossbar switch fabric. In: Neuvo Y, ed.Proceedings of the IEEE International Conference on Communications (ICC). Helsinki: IEEE Communications Society, 2001. 1586~1591.

[4] Prabhakar B, McKeown N, Ahuja R. Multicast scheduling for input-queued switches. IEEE Journal on Selected Areas in Communications, 1997,15(5):855~866.

[5] Parekh AK, Gallager RG. A generalized processor sharing approach to flow control in integrated service networks: the single-node case. IEEE/ACM Transactions on Networking, 1993,1(3):344~357.

[6] Karol M, Hluchyj M, Morgan S. Input versus output queueing on a space division switch. IEEE Transactions on Communication, 1988,35 (12):1347~1356.

[7] Tamir Y, Frazier G. Dynamically-Allocated multi-queue buffer for VLSI communication switches. IEEE Transactions on Computers, 1992,41(6):725~737.

[8] Hopcroft J E, Karp RM. An n5/2 algorithm for maximum matching in bipartite graphs. SIAM Journal on Computing, 1973,1.2: 225~231.

[9] McKeown N, Mekkittikui A, Anantharam V, Walrand J. Achieving 100% throughput in an input-queued switch. IEEE Transactions on Communication, 1999,47(8):1260~1267.

[10] Anderson T, Owicki S, Saxes J, Thacker C. High speed switch scheduling for local area networks. ACM Transactions on Computer Systems, 1993,11(4):319~352.

[11] McKeown N. The iSLIP scheduling algorithm for input-queued switches. IEEE/ACM Transactions on Networking, 1999,7(2): 188~201.

[12] McKeown N, Anderson TE. A quantitative comparison of scheduling algorithms for input-queued switches. Computer Networks and ISDN Systems, 1998,30(24):2309~2326.

[13] Serpanos DN, Antoniadis PI. FIRM: A class of distributed scheduling algorithms for high-speed ATM switches with multiple input queues. In: Sidi M, ed. Proceedings of the IEEE INFOCOM. Tel Aviv: IEEE Communications Society, 2000. 548~555.

[14] Chao HJ. Saturn: A terabit packet switch using dual round robin. IEEE Communications Magazine, 2000,38(12):78~84.

[15] Goudreau MW, Kolliopoulos SG, Rao SB. Scheduling algorithms for input-queued switches: Randomized techniques and experimental evaluation. In: Sidi M, ed. Proceedings of IEEE INFOCOM. Tel Aviv: IEEE Communications Society, 2000. 1634~ 1643.

[16] Tarjan RE. Data structures and network algorithms. SIAM, 1983.

[17] Mekkittikui A, McKeown N. A starvation-free algorithm for achieving 100% throughput in input-queued switches. In: Lee D, ed. Proceedings of the IEEE International Conference on Computer Communications and Networks (ICCCN). Rockville, MA: IEEE Communications Society, 1996. 226~231.

[18] Mekkittikui A, McKeown N. A practical scheduling algorithm to achieve 100% throughput in input-queued switches. In: Akyildiz I, ed. Proceedings of the IEEE INFOCOM. San Francisco: IEEE Communications Society, 1998. 792~799.

[19] McKeown N. Scheduling algorithms for input-queued switches [Ph.D. Thesis]. University of California at Berkeley, 1995.

[20] Marsan MA, Bianco A, Leonardi E, Milla L. RPA: A flexible scheduling algorithm for input buffered switches. IEEE Transactions on Communications, 1999,47(12):1921~1933.

[21] Duan H, Lockwood JW, Kang SM, Will JD. A high performance OC12/OC48 queue design prototype for input buffered ATM switches. In: Hasegawa T, ed. Proceedings of the IEEE INFOCOM. Kobe: IEEE Communications Society, 1997. 20~28.

[22] Marsan MA, Bianco A, Giaccone P, Leonardi E, Neri F. Input-Queued router architectures exploiting cell-based switching fabrics. Computer Networks, 2001,37(5):541~559.

[23] Tassiulas T. Linear complexity algorithms for maximum throughput in radio networks and input queued switches. In: Akyildiz I, ed. Proceedings of the IEEE INFOCOM. New York: IEEE Communications Society, 1998. 533~539.

[24] Giaccone P, Prabhakar B, Shah D. Towards simple, high-performance schedulers for high-aggregate bandwidth switches. In: Kermani P, ed. Proceedings of the IEEE INFOCOM. New York: IEEE Communications Society, 2002. 1160~1169.

[25] Marsan MA, Bianco A, Giaccone P, Neri F. Packet scheduling in input-queued cell-based switches. In: Sengupta B, ed. Proceedings of the IEEE INFOCOM. Anchorage: IEEE Communications Society, 2001. 1085~1094.

[26] Ganjali K, Keshavarzian A, Shah D. Input queued switches: Cell switching vs. packet switching. In: Bauer T, ed. Proceedings of the IEEE INFOCOM. San Francisco: IEEE Communications Society, 2003. http://www.stanford.edu/~yganjali/#Publications.

[27] Andrews M, Zhang L. Achieving stability in networks of input-queued switches. In: Sengupta B, ed. Proceedings of the IEEE INFOCOM. Anchorage: IEEE Communications Society, 2001. 1673~1679.

[28] Leonardi E, Mellia M, Marsan MA, Neri F. On the throughput achievable by isolated interconnected input-queueing switches under multiclass traffic. In: Kermani P, ed. Proceedings of the IEEE INFOCOM. New York: IEEE Communications Society, 2002. 1605~1614.

[29] Leonardi E, Mellia M, Neri F, Marsan MA. On the stability of input-queued switches with speedup. IEEE/ACM Transactions on Networking, 2001,9(1):104~118.

[30] Dai JG, Prabhakar, B. The throughput of data switches with and without speedup. In: Sidi M, ed. Proceedings of the IEEE INFOCOM. Tel Aviv: IEEE Communications Society, 2000. 556~564.

[31] Leonardi E, Mellia M, Neri F, Marsan MA. Bounds on average delays and queue size averages and variances in input-queued cell-based switches. In: Sengupta B, ed. Proceedings of the IEEE INFOCOM. Anchorage: IEEE Communications Society, 2001. 1095~1103.

[32] Shah D, Kopikare M. Delay bounds for the approximate maximum weight matching algorithm for input queued switches. In: Kermani P, ed. Proceedings of the IEEE INFOCOM. New York: IEEE Communications Society, 2002. 1024~1031.

[33] Gale D, Shapley LS. College admission and the stability of marriage. American Mathematical Monthly, 1962,69:9~15.

[34] Gusfield D, Irving R. The Stable Marriage Problem: Structure and Algorithms. The MIT Press, 1989.

[35] Prabhakar P, Mckeown N. On the speedup required for combined input and output queued switching. Technical Report, Stanford CSL-TR-97-738, 1997.

[36] Stoica I, Zhang H. Exact emulation of an output queueing switch by a combined input and output queueing switch. In: Knightly E, ed. Proceedings of the IEEE IWQoS. Napa: IEEE Communications Society, 1998. 218~224.

[37] Chuang ST, Goel A, McKeown N. Matching output queueing with a combined input/output-queued switch. IEEE Journal on Selected Areas in Communications, 1999,17(6):1030~1039.

[38] Krishna P, Patel NS, Charny A, Simcoe RJ. On the speedup required for work-conserving crossbar switches. IEEE Journal on Selected Areas in Communications, 1999,17(6):1057~1066.

[39] Kam AC, Siu KY. Linear-Complexity algorithms for QOS support in input-queued switches with no speedup. IEEE Journal on Selected Areas in Communications, 1999,17(6):1040~1056.

[40] Weller T, Hajek B. Scheduling nonuniform traffic in a packet-switching system with small propagation delay. IEEE/ACM Transactions on Networking, 1997,5(6):813~823.

[41] Chang CS, Chen WJ, Huang HY. Birkhoff-von Neumann input buffered crossbar switches. In: Sidi M, ed. Proceedings of the IEEE INFOCOM. Tel Aviv: IEEE Communications Society, 2000. 1614~1623.

[42] Chang CS, Lee DS, Jou YS. Load balanced Birkhoff-von Neumann switches Part I: One-stage buffering. Computer Communications, 2002,25(6):611~622.

[43] Chang CS, Lee DS, Lien CM. Load balanced Birkhoff-von Neumann switches Part II: Multi-Stage buffering. Computer Communications, 2002,25(6):623~634.

[44] Keslassy I, McKeown N. Maintaining packet order in two-stage switches. In: Kermani P, ed. Proceedings of the IEEE INFOCOM. New York: IEEE Communications Society, 2002. 1032~1041.

[45] Wang W, Dong L, Wolf W. A distributed switch architecture with dynamic load-balancing and parallel input-queued crossbars for terabit switch fabrics. In: Proceedings of IEEE INFOCOM. New York: IEEE Communications Society, 2002. 352~361.

[46] Iyer S, Awadallah A, McKeown N. Analysis of a packet switch with memories running slower than the line-rate. In: Sidi M, ed. Proceedings of the IEEE INFOCOM. Tel Aviv: IEEE Communications Society, 2000. 529~537.

[47] Iyer S, McKeown N. Making parallel packet switches practical. In: Sengupta B, ed. Proceedings of the IEEE INFOCOM. Anchorage: IEEE Communications Society, 2001. 1680~1687.

[48] Blake S, Black D, Carison M, Davies E, Wang Z, Weiss W. An architecture for differentiated services. IETF RFC 2475, 1998.

[49] Bianco A, Franceschinis M, Ghisolfi S, Hill AM, Leonardi E, Neri F, Webb R. Frame-Based matching algorithms for input-queued switches. In: Aoyama T, ed. Proceedings of the IEEE Workshop on High Performance Switching and Routing (HPSR). Kobe: IEEE Communications Society, 2002.

[50] Kar K, Lakshman TV, Stiliadis D, Tassiulas L. Reduced complexity input buffered switches. In: Proceedings of the Hot Interconnects VIII. 2000. http://www.bell-labs.com/user/stiliadi/publications.html.