Hierarchical Scheduling Algorithm for QoS Guarantee in the MIQ Switches

Hakyong Kim, Kiseon Kim, and Yongtak Lee

Indexing terms: hierarchical scheduling, MIQ switch, VC scheduling, MIQ scheduling, QoS guarantee

---

A hierarchical scheduling concept is proposed to provide both the QoS guarantee and enhanced switching throughput in the multiple input-queued switch under a bursty environment.

Introduction: One of the most important and imperative issues in the future high-speed packet switching and routing is to provide QoS guarantees of diversified services. However, scheduling algorithms developed for the multiple input-queued (MIQ) switches have focused either on the throughput enhancement by finding maximal input-output matchings or on the methodology of finding the matchings, rather than QoS guarantees. In this Letter, we present a scheduling approach which provisions both the QoS guarantees and high switch throughput in the MIQ switches.

Classification of scheduling strategies: The scheduling algorithms developed for the packet switches with multiple number of queues or virtual connection (VC) queues in each input or each output can be classified into two categories: *the dynamic scheduling algorithm* and *the static scheduling algorithm.* With the dynamic algorithm, the scheduler makes decisions for scheduling every time slot to find out maximal input-output matching and, therefore, it is also called cell-level scheduling. In general, dynamic scheduling is used in the MIQ switches in order to improve the switch throughput. Examples are parallel iterative matching (PIM) [1], SLIP [2], and parallel solitary-request-first (PSRF) [3] algorithms. With the static scheduling algorithm, meanwhile, the scheduler determines the service order of packets by assigning or reserving time slots for specific VC connections, whenever changes occur in the number of connections, such as connection setups/takedowns. In general, static scheduling, also called flow-level or connection-level scheduling, is used in the output-queued switches to guarantee the QoS requirements of connections. Examples include weighted round-robin (WRR) and scheduling algorithms based on the priority scheme.

Hierarchical scheduling: For a scheduling algorithm to have a combined nature of providing QoS guarantees as well as high throughput, it is required to mix or unite a dynamic algorithm with a static algorithm. The idea proposed in this Letter is to mix dynamic and static algorithms hierarchically, where we employ a dynamic scheduler in finding input-output matchings and a static scheduler in selecting a VC connection in the input port selected by the dynamic scheduler.

Fig. 1 depicts an example of the two-level hierarchical scheduling in the MIQ switch used in our study, and S1 and S2 correspond to the dynamic scheduler and the static scheduler, respectively. That is, scheduler or server S1 is responsible for finding a set of optimal input-output matchings. In the MIQ switch, since a queue is dedicated to a specific output port or a group of output ports, for S1 to select an input-output pair implies to select a specific queue in the corresponding input. Therefore, we name the scheduler S1 *the MIQ scheduler.* It is usual that the MIQ scheduling is performed in a centralized manner to find optimal input-output matchings by coordinating transmission requests from multiple inputs. Scheduler S2 is responsible for scheduling packets from VC queues within the queue selected by the MIQ scheduler, naming it *the VC scheduler.* The VC scheduler with the call admission controller (CAC) manages the QoS of connections. Usually, the VC scheduler is distributed in each input where a selected queue uses it. The example shown in Fig. 1 depicts that S1 selects queue 2 and S2 selects VC n within the queue 2.

For providing QoS guarantees, we have to consider as many classes of traffic as possible. In the ATM arena, for example, there are CBR, VBR, ABR, UBR, and so on. However, we consider here only two types of traffic, real-time (RT) traffic and non-real-time (NRT) traffic, since the objective is to provide potentially different service support for different type of traffic. For providing QoS guarantees, in effect, it is a minimum requirement to prioritize RT traffic over NRT traffic or best-effort services [4].

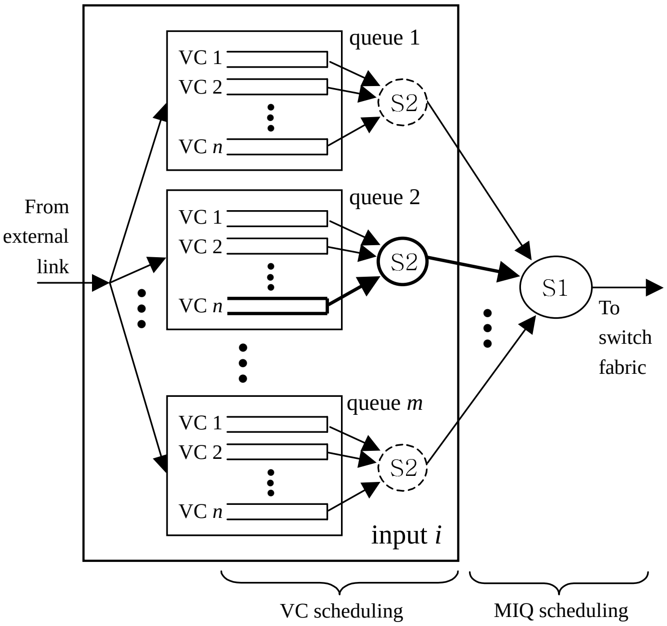

Figure 1: Two-level hierarchical scheduling in the MIQ switch.

For those traffic, we employed the known PIM scheduling algorithm as the dynamic scheduler (S1 in Fig. 1). The PIM algorithm is composed of three phases of Request, Grant, and Accept, and uses parallelism, randomness, and iteration to find a maximal matching between inputs and outputs. Maximal matching is used to determine which inputs transmit cells to which outputs in the next time slot. Details of PIM can be found in reference [1]. Based on the result of MIQ scheduling, the static scheduler (S2) selects a VC connection based on a priority scheme. In the priority scheme employed in our study, RT traffic is given preemption for transmission request since we expect that RT traffic has more stringent QoS requirements, in particular packet delay. However, RT traffic does not have strict priority over NRT traffic in the request phase. When the queue length of a NRT-traffic connection becomes larger than the predefined threshold value (Th_{nrt}), the connection has higher priority over RT traffic and, thereby, meet the packet loss requirements.

Performance evaluation: In order to validate that the hierarchical scheduling can guarantee QoS requirements as well as provide high switching throughput, we performed computer simulations under following assumptions: The switch size used is 16 × 16 and each input has 16 queues, viz. virtual output-queued (VOQ) switch. Up to 2 VC connections, one RT and one NRT traffic, are allowed per each input in a specific instance. RT and NRT traffics are generated with the same probability of 0.5. Each connection is bursty with geometrically distributed ON and OFF periods of mean λ_{ON} and λ_{OFF}, respectively. During the ON period, cells arrive continuously and destine for the same output, while during the OFF period, no cell arrives. The destination of ON periods distributes uniformly for output destination. We restrict that ON and OFF periods be larger than one cell time. In the simulation, we fixed λ_{OFF} to 15 cells, but λ_{ON} takes on one value among 6, 7, 8, and 9 cells. It corresponds to the average offered load (ρ) of 0.571, 0.636, 0.696 and 0.75 since there exist two connections on average for each input. Buffer space of 10 cells and 1000 cells are assigned to RT and NRT traffic, respectively, and the threshold value of the buffer for NRT traffic is set to 950 cells. The running

Table 1: Cell loss ratio (CLR), average delay time (ADT), and normalized throughput (T_{norm}) for RT and NRT traffic. (switch size =16, number of queues in an input = 16, $BS_{rt} = 10$, $BS_{nrt} = 1000$, $Th_{nrt} = 950$ cell units)

Load	RT traffic			NRT traffic		
	CLR	ADT	T_{norm}	CLR	ADT	T_{norm}
0.571	N/A	7.09	0.99	N/A	8.34	0.99
0.636	$3.93e^{-7}$	9.39	0.99	N/A	11.5	0.99
0.696	$1.76e^{-5}$	12.8	0.99	N/A	16.7	0.99
0.750	$1.89e^{-4}$	18.5	0.99	N/A	26.0	0.99

time is 10^7 cell time slots.

Table 1 shows the simulation result when we iterate the dynamic algorithm (PIM) twice during a time slot. As shown in the table, RT traffic has less queueing delay than NRT traffic, even though it experiences higher cell loss due to buffer overflow. For NRT traffic, we can not observe any cell loss in 10^7 cell time slots even though the offered load is relatively high. In Table 1, CLR for RT traffic and ADT for NRT traffic is relatively higher, compared to results in other works, since we used bursty traffic patterns and one of the simplest priority mechanisms in the static level. When we applied non-bursty traffic, CLR of RT traffic was not available in 10^7 cell time slots, neither, and ADT of NRT traffic was much smaller. The normalized throughput (T_{norm}) is 0.99 for both types of traffic irrespective of the offered load.

Conclusions: A hierarchical scheduling prototype for supporting both RT and NRT traffic streams in high-speed multiple input-queued packet switching systems has been presented. By deploying a dynamic scheduler and a static scheduler in different scheduling layer, respectively, the hierarchical scheduling approach can ensure both high switch throughput and QoS guarantees of connections. Besides the performance, the hierarchical approach in scheduling provides flexibility in that optimal scheduling algorithms can be derived in different scheduling level. Furthermore, since the VC scheduler handles not all VC connections in the same input but only VC connections within the queue selected by the MIQ scheduler, the hierarchical scheduling can allows much larger number of VC connections per input.

Day Month. 2000
Electronics Letters Online No: 19990000

Hakyong Kim, Yongtak Lee and Kiseon Kim (*Department of Information and Communications, Kwang-Ju Institute of Science and Technology (K-JIST), 1 Oryong-dong, Puk-gu, Kwangju, 500-712, S. Korea*)

E-mail: hykim@ieee.org

References

1 ANDERSON, T.E., OWICKI, S.S., SAXE, J.B., and THACKER, C.P.: 'High-speed switch scheduling for local-area networks', *ACM Tr. Computer Systems*, Nov. 1993, 11 (4), pp.319-352.

2 MCKEOWN, N.: 'The *i*SLIP scheduling algorithm for input-queued switches', *IEEE/ ACM Tr. Networking*, April 1999, 7 (2), pp. 188-201.

3 KIM, H., KIM, K. YOON, H., and LEE, Y.: 'A Throughput-Enhanced Parallel Scheduling Algorithm for the MIQ Switch with a Moderate Number of Queues', *IEE Electronics Letters*, 2nd March 2000, 36 (5), pp.477-478.

4 ROTHERMEL, K.: 'Priority mechanisms in ATM Networks', *IEEE GLOBECOM '90*, Dec. 1990, vol. 2, pp.847-851.