An efficient fuzzy system for traffic management in high-speed packet-switched networks

G. Ascia, V. Catania, D. Panno

Abstract In this paper we present a novel buffer management scheme based on fuzzy logic. We deal with the problem of managing traffic flows with different priorities within the same buffer. The aim is to guarantee the QoS of high-priority traffic, and at the same time exploit unused buffer resources to accommodate best-effort traffic in order to maximize the total throughput. The scheme we propose can be applied both to ATM and IP networks. The performance evaluation of the fuzzy priority control scheme shows that it outperforms any static threshold mechanism and, as far as the total throughput is concerned, it is very close to that of the push out mechanism considered in literature as an ideal mechanism. Finally, we address some implementation issues of the control system and propose the design of a new cost-effective VLSI fuzzy processor.

Keywords Priority control, Integrated-service networks, Fuzzy logic, Hardware architecture

1 Introduction
In integrated-service, packet-switched networks, packets from different sessions belonging to different applications interact with each other when they are multiplexed at the same output link of a switch. Different applications expect different service requirements and the deployment of priority algorithms may be a key component in controlling the interaction between different traffic flows [1, 2].

This paper describes a fuzzy logic-based priority control mechanism that presents a significant improvement in performance and is more general than the one proposed in [3], as it can be applied to packet-switched networks operating with both fixed-sized and variable-sized data units, such as an ATM network and the Internet.

There are various real applications for our system, for example, in the buffer of a generic virtual path of ATM switches, when the network is operating in a CLP-significant mode, i.e. the QoS requirement in terms of cell loss probability only applies to the high-priority flow and the network can make a best-effort attempt to transmit the low-priority flow. Again, to meet the demand for QoS in the future Internet, link-sharing mechanisms allow multiple agencies or organizations to receive a guaranteed share of the resources (buffer and link bandwidth). If the real-time traffic (guaranteed flows) for an agency has little data to send, our priority control scheme allows excess bandwidth to be exploited to transmit the agency’s best-effort traffic.

The aim of the priority control function being studied is therefore to maximize network utilization without degrading the QoS for high-priority traffic. The priority control mechanisms have to be flexible so as to adapt to different types of traffic. In addition, as the priority algorithms have to act on the forwarding path of the switch or router, they cannot be excessively complex so that they can be executed as rapidly as possible. The solutions proposed so far in the literature have proved to be unsatisfactory either because they are difficult to implement (push-out) or because they do not ensure a high level of performance (fixed threshold mechanisms) [4, 5]. During the research carried out, we defined an innovative priority control scheme based on fuzzy logic, which respects the QoS of high-priority traffic and combines flexibility requirements with optimization of resource utilization. The paper is organized as follows: Sect. 2 describes the fuzzy priority control system; Sect. 3 presents an assessment of the performance of the system in an ATM and Internet environment; Sect. 4 presents some remarks on implementation of the fuzzy priority control system; Sect. 5 proposes a possible hardware implementation of the system in VLSI technology.

2 Model of fuzzy priority control scheme
In this section, we will first describe a new fuzzy priority control (FPC) mechanism which meets the requirements of performance, flexibility and cost-effective implementation of a high-speed packet-switched network based on ATM, then we will address its application to IP networks.

Let us start with ATM networks. We assume that some switch resources (i.e. buffer size, Q, and bandwidth portion, C) have been allocated to high-priority traffic to guarantee a given QoS. Our reference model is shown in Fig. 1, where low-priority cells are allowed to enter the allocated buffer to improve the exploitation of reserved resources. The aim is to prevent the low-priority traffic from degrading the QoS of high-priority traffic.

For the FPC scheme not to become a bottleneck we chose to make it work only at the buffer input to decide whether to accept or reject any new low-priority cells, and
buffer in cells, and they do not depend on the statistical traffic parameters. This is an important condition for the flexibility of our mechanism.

As the inference method chosen was the zero-order Tagaki–Sugeno method [6], the term set of the output variable ACTION is composed of 6 singleton fuzzy names: strong discard (SD), discard (D), marginal discard (MD), marginal admit (MA), admit (A) and strong admit (SA). The membership functions are shown in Fig. 5.

The membership functions and the control rules have been derived after an accurate tuning phase in which an optimization technique based on genetic algorithm has been used.

Table 1 shows the control rules; they are made up of three antecedents of the kind:

- If CLR_H is Low and N is Medium and ∆N is Negative
- then the Action is Strong Admit

except rules 1, 26 and 30 which are made up of two antecedents and do not depend on input variable ∆N value.

The fuzzy rules defined are easy to understand and are highly suitable for representation of the decision-making process a priority control mechanism has to possess. The philosophy behind the knowledge base of the FPC scheme is that of admitting low-priority cells when they do not affect the high-priority flow. This happens either when the CLR_H is considerably lower than the CLR_HI and the value of N is not high (rules 1–4) or when, although the CLR_HI is close to the CLR_HI values, the network resources are temporarily underloaded, i.e. the value of N is medium or low (rules 8–10, 17–18, 26). Vice versa, in congested conditions where the CLR_HI is very high (rules 27–30), or in risk conditions, where the state of occupation of the buffer indicates a possibility of congestion because N is high or very high and/or ∆N is positive, the action of the FPC has to be unfavorable to low-priority cells (rules 7,