



Efficient Switches with QoS Support for Clusters

Alejandro Martínez Vicente
Francisco José Alfaro Cortés
José Luis Sánchez García
José Duato Marín

I3A 
Instituto de Investigación
en Informática de Albacete



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Table of Contents

- Introduction
- Related Work
- Architecture for QoS support
 - Providing Full QoS Support with 2 VCs
- Performance Evaluation
 - Simulation conditions
 - Simulation results
- Conclusions

- **Introduction**

- Related Work

- Architecture for QoS support

- Performance Evaluation

- Conclusions

Introduction

- High-performance interconnection networks are a key component of a variety of systems.
 - Clusters
 - Parallel Machines
 - Routers
- Quality of Service (QoS) is useful when several traffic classes share the network.
 - Guaranteeing performance in terms of latency, jitter and throughput.
- For instance, QoS is important when there is multimedia traffic involved.

•Introduction

- Related Work
- Architecture for QoS support
- Performance Evaluation
- Conclusions

Introduction

- Proposals to provide QoS usually rely on Virtual Channels (VCs).
 - One VC per traffic class.
- However, most of the final implementations do not offer so many VCs.
 - The trend is to increase the number of ports instead of increasing the number of VCs per port.
- If VCs are available, using a lot of VCs for QoS prevents the use of these VCs for other purposes.
 - Fault tolerance
 - Adaptive routing
 - Congestion management

•Introduction

•Related Work

•Architecture for QoS support

•Performance Evaluation

•Conclusions

Introduction

- Our aim is to provide full QoS support using simple switches.

A. Martínez, F. J. Alfaro, J. L. Sánchez, and J. Duato. **Providing full QoS support in clusters using only two VCs at the switches.** In *Proceedings of the 12th International Conference on High Performance Computing (HiPC 2005)*

- In our previous work we obtained promising results.
- In this work, we study:
 - The use of WRR table-based scheduling at the switches, eliminating the necessity of strict priorities.
 - The scalability of the new architecture, compared with traditional solutions.

•Introduction

•Related Work

•Architecture for QoS support

•Performance Evaluation

•Conclusions

Table of Contents

- Introduction
- Related Work
- Architecture for QoS support
 - Providing Full QoS Support with 2 VCs
- Performance Evaluation
 - Simulation conditions
 - Simulation results
- Conclusions

•Introduction

•**Related Work**

•Architecture for
QoS support

•Performance
Evaluation

•Conclusions

Related Work

- During the last decade, many switch designs with QoS support have been proposed:
 - Multimedia Router (MMR)
 - MediaWorm
 - InfiniBand
 - PCI Express Advanced Switching
- The usual solution is to provide a VC for each service class.
 - Complex architectures
 - Few implementations

•Introduction

•**Related Work**

•Architecture for QoS support

•Performance Evaluation

•Conclusions

Related Work

- Another trend is simpler designs with a few broad traffic classes
 - Avici TSR, Regular and Premium
 - ATLAS I, three categories
- This is not enough, a more fine-grained classification is more useful.
- Our contribution:
 - We use only two VCs in the switches but,
 - the global behavior of the network is very similar as if the switches were using many more VCs.

•Introduction

•**Related Work**

•Architecture for QoS support

•Performance Evaluation

•Conclusions

Table of Contents

- Introduction
- Related Work
- Architecture for QoS support
 - Providing Full QoS Support with 2 VCs
- Performance Evaluation
 - Simulation conditions
 - Simulation results
- Conclusions

•Introduction

•Related Work

•**Arch. for QoS support**

•Performance Evaluation

•Conclusions

Providing Full QoS Support with 2 VCs

- Obvious strategy to provide QoS support: one VC per traffic class.
- However, this is not enough for complete QoS support:
 - Strict guarantees on throughput and latency demand traffic. A connection admission control (CAC) is necessary.
 - Providing QoS with the scheduling at switches is not enough, there must be some scheduling at the output of the network interfaces as well.
 - Head-of-line (HOL) blocking and buffer hogging must be dealt with. This can be solved using virtual output queuing (VOQ), usually at the switch level.
- Once the previous conditions are met, traffic flows seamlessly through the network.
 - Therefore, to devote a different VC to each traffic class might be redundant.

•Introduction

•Related Work

•**Arch. for QoS support**

•Performance Evaluation

•Conclusions

Providing Full QoS Support with 2 VCs

•Introduction

•Related Work

•Arch. for QoS support

•Performance Evaluation

•Conclusions

- When using *store and forward* or *virtual cut-through* switching, the minimum buffer space that is needed to achieve maximum throughput is:
 - Credit-based flow control: $1 \text{ MTU} + (1 \text{ RTT} \times \text{BW})$
 - Stop and Go flow control: $4 \text{ RTT} \times \text{BW}$
- However, depending on the characteristics of traffic, like burstiness or locality (hot-spots), more memory at the buffers is necessary to obtain acceptable performance.
- VCs produce a static partition of buffer memory.
 - Traffic of one VC cannot use space devoted to another VC, even if it is available.
- For that reason, although VCs provide traffic isolation, they may degrade overall performance under bursty traffic.
 - We will see this at the performance evaluation section.

Providing Full QoS Support with 2 VCs

- Traffic coming from network interfaces has already been scheduled.
- Implicit in the order of arrival, there is information about the priority of packets.
 - Packets arriving earlier have more priority.
- We can take advantage of the scheduling at network interfaces.
 - Flows from different input ports going to the same output port are interleaved attending to their arrival order.
- The network interfaces implement the full number of VCs.

•Introduction

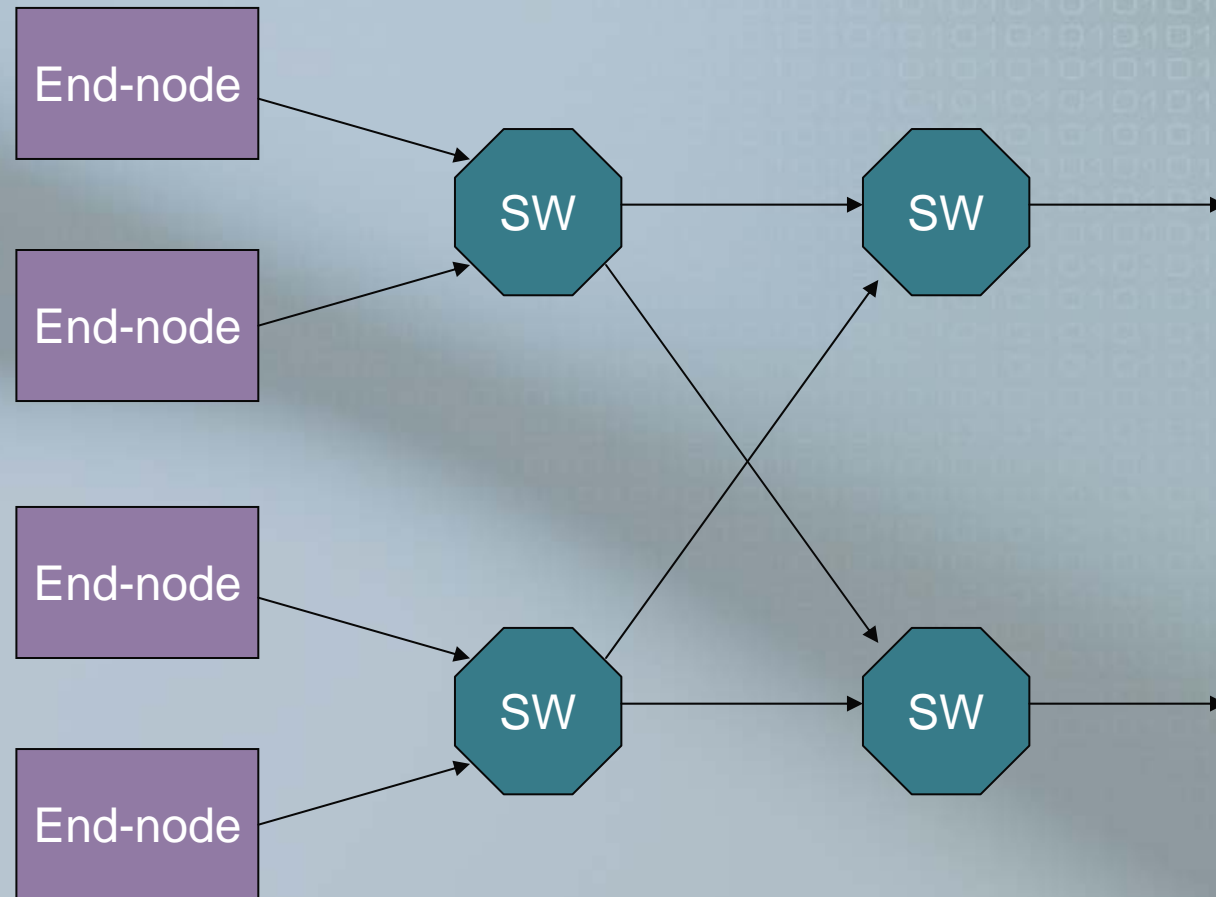
•Related Work

•**Arch. for QoS support**

•Performance Evaluation

•Conclusions

Providing Full QoS Support with 2 VCs



- Introduction
- Related Work
- **Arch. for QoS support**
- Performance Evaluation
- Conclusions

Architecture for QoS Support

- Only one queue is necessary
- Problem 1: unregulated traffic may degrade performance of regulated traffic.
- Solution: We use two VCs at the switch ports:
 - One for Video: regulated traffic – Connection Admission Control
 - The other for best-effort: unregulated traffic
- Problem 2: switches are not able to schedule packets as freely as when having many VCs.
- Limited impact: bandwidth guarantee means that regulated traffic will flow.

•Introduction

•Related Work

•Arch. for QoS support

•Performance Evaluation

•Conclusions

Table of Contents

- Introduction
- Related Work
- Architecture for QoS support
 - Providing Full QoS Support with 2 VCs
- Performance Evaluation
 - Simulation conditions
 - Simulation results
- Conclusions

•Introduction

•Related Work

•Architecture for
QoS support

•**Performance
Evaluation**

•Conclusions

Simulation Conditions

- The simulated model was the next:
 - Multi-stage network with 64 ports
 - 16 port switches, combined input and output queuing, and a crossbar for internal interconnection.
 - Virtual Output Queuing at switch level.
 - Credit-based flow control.
 - Maximum packet size of 2 KBytes (8 bytes header).
 - 64 KBytes buffers per port (to share between VCs).
 - 8 Gb/s links, the crossbar has a 2.0 speed-up (16 Gb/s).

•Introduction

•Related Work

•Architecture for
QoS support

•**Performance
Evaluation**

•Conclusions

Simulation Conditions

- The Network Processing Forum Switch Fabric Benchmark Specifications.
- The input load was
 - 8 service levels.
 - Zipf's law-based destination distribution.
 - The TC 7 has the highest priority, the TC 0 the lowest.
 - TCs 3 to 0 are BE and do not have any reserved resources.
- Multimedia traffic:
 - Audio traffic composed of CBR, point-to-point connections of 64 Kb/s.
 - Video traffic composed of MPEG-4 750 Kb/s sequences. Packet size varies up to 2 KBytes.

Name	%	P.Size	Notes
Network Control	1	64 B	Self-similar
Audio	15	128 B	CBR 64 Kb/s
Video	15	2 KB	Trs. MPEG-4
Controlled Load	15	2 KB	CBR 1 Mb/s
Excellent-effort	13	2 KB	Self-similar
Pref. Best-effort	13	2 KB	Self-similar
Best-effort	13	2 KB	Self-similar
Background	13	2 KB	Self-similar

- Introduction
- Related Work
- Architecture for QoS support
- Performance Evaluation**
- Conclusions

Simulation Conditions

- We compare the performance of four architectures
 - Traditional designs with 8 VCs
 - Our new proposed switch designs with 2 VC
 - Each alternative with either 8 or 16 ports/switch
 - In all the cases, the same total memory/switch.
- We have considered the next QoS indices:
 - Latency
 - Throughput
- We also consider cumulative distribution function of latency and jitter results.

•Introduction

•Related Work

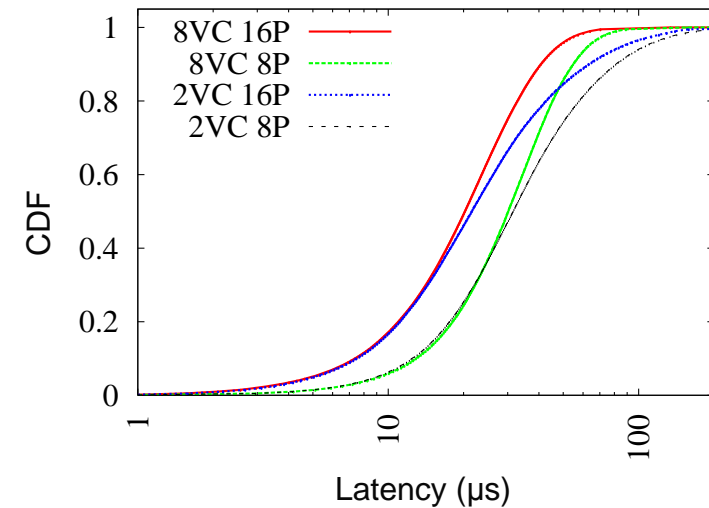
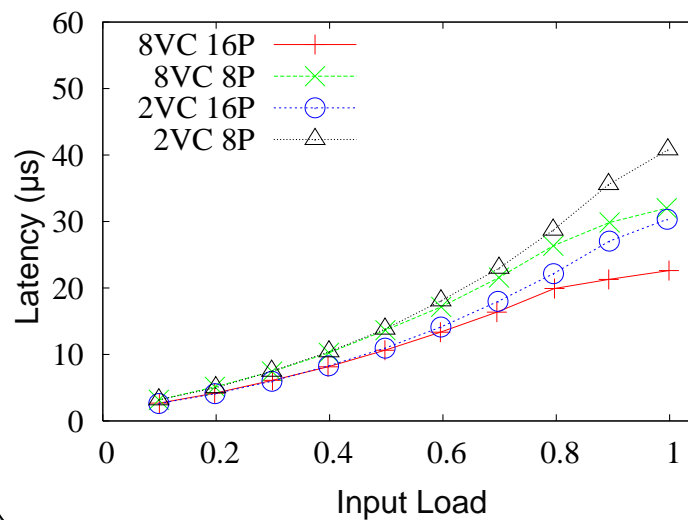
•Architecture for
QoS support

•**Performance
Evaluation**

•Conclusions

Simulation Results

Network Control traffic

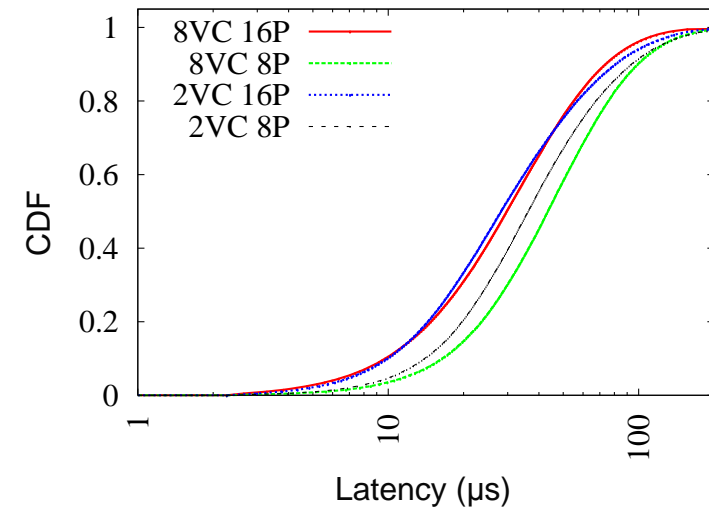
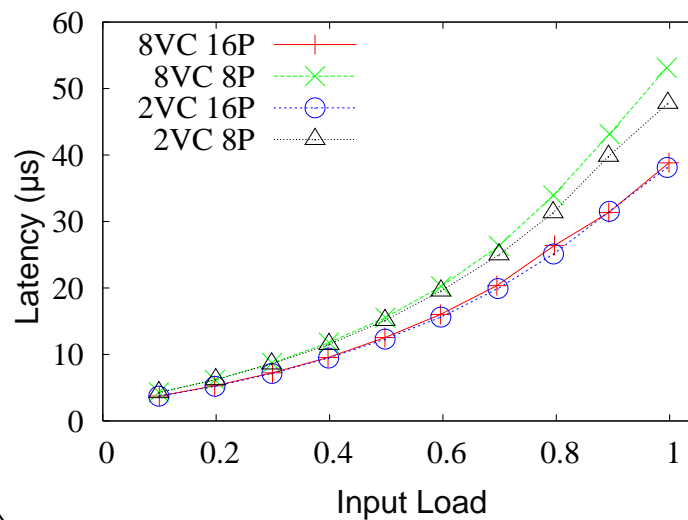


Average and CDF latency

- Introduction
- Related Work
- Architecture for QoS support
- **Performance Evaluation**
- Conclusions

Simulation Results

Audio traffic

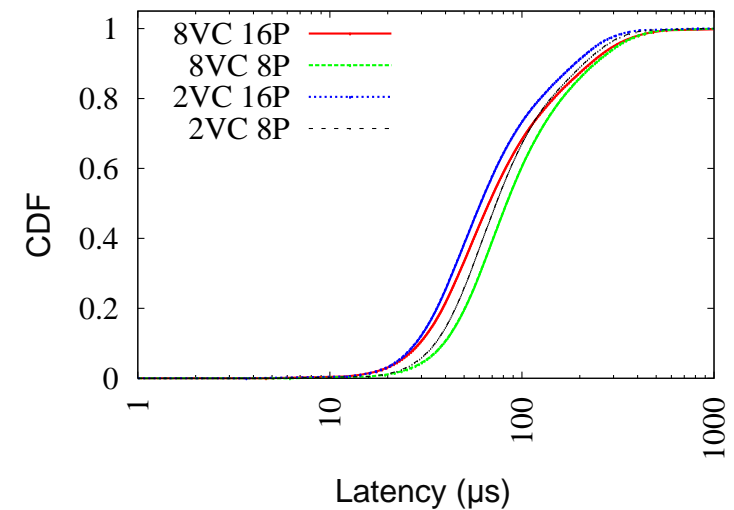
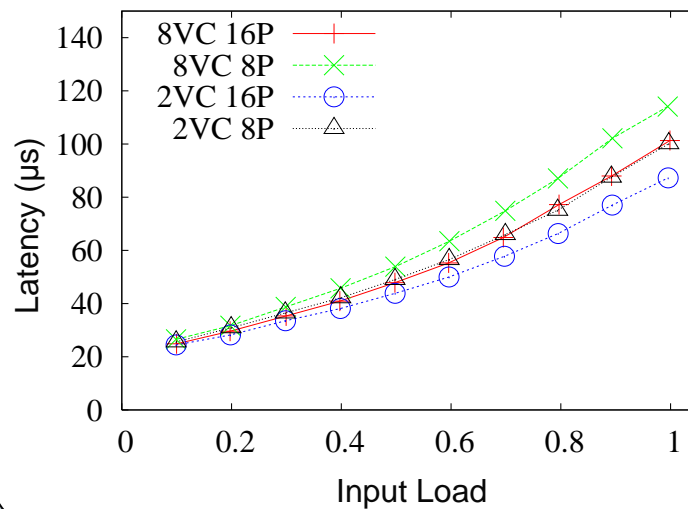


Average and CDF latency

- Introduction
- Related Work
- Architecture for QoS support
- **Performance Evaluation**
- Conclusions

Simulation Results

Video traffic

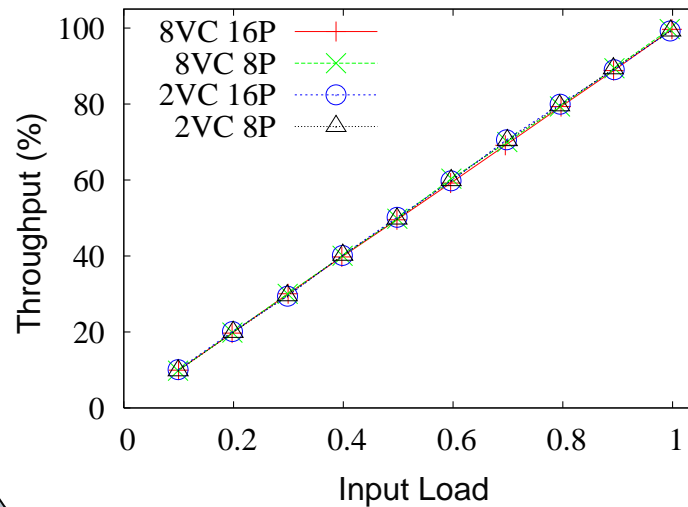


Average and CDF latency

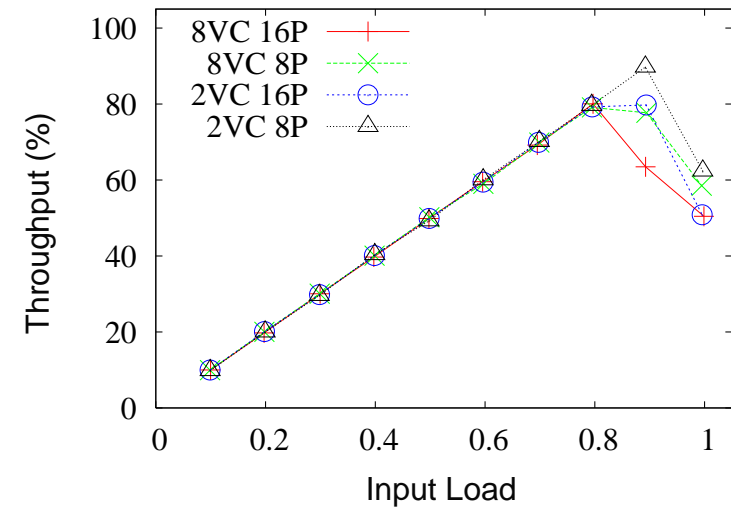
- Introduction
- Related Work
- Architecture for QoS support
- **Performance Evaluation**
- Conclusions

Simulation Results

Best-effort traffic



Excellent-effort



Background

- Introduction
- Related Work
- Architecture for QoS support
- Performance Evaluation
- Conclusions

Scalability Test

- In the next test we vary the amount of QoS traffic injected.
- We want to see how much QoS traffic we can inject in each case, while results are still acceptable.
- The input load is 100% link capacity in all cases
 - Proportion of QoS traffic varies from 10% to 90%

•Introduction

•Related Work

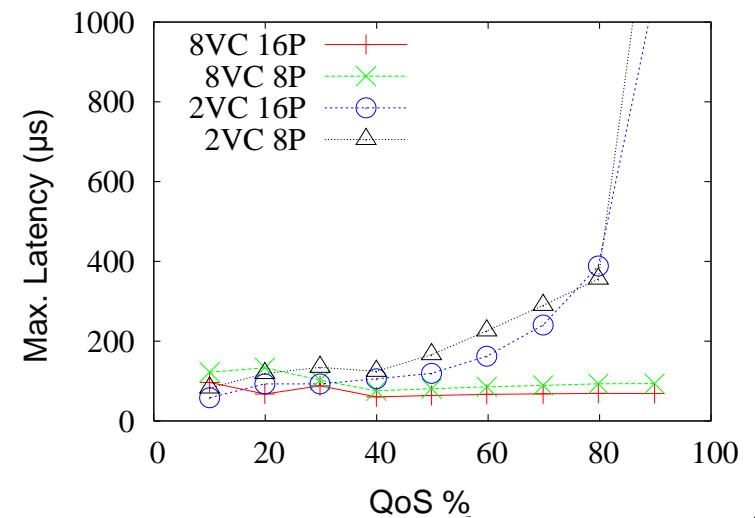
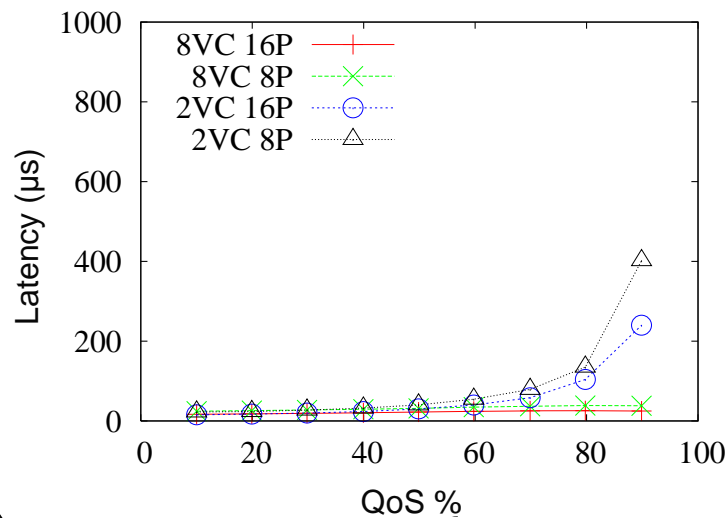
•Architecture for QoS support

•**Performance Evaluation**

•Conclusions

Simulation Results

Network control traffic

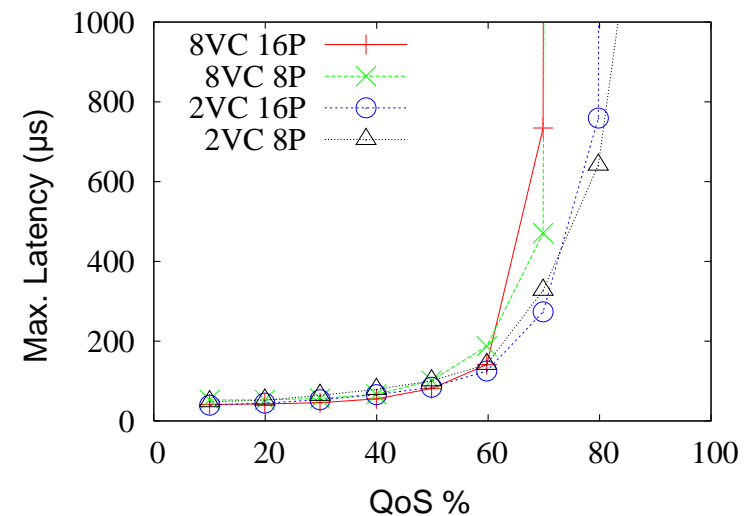
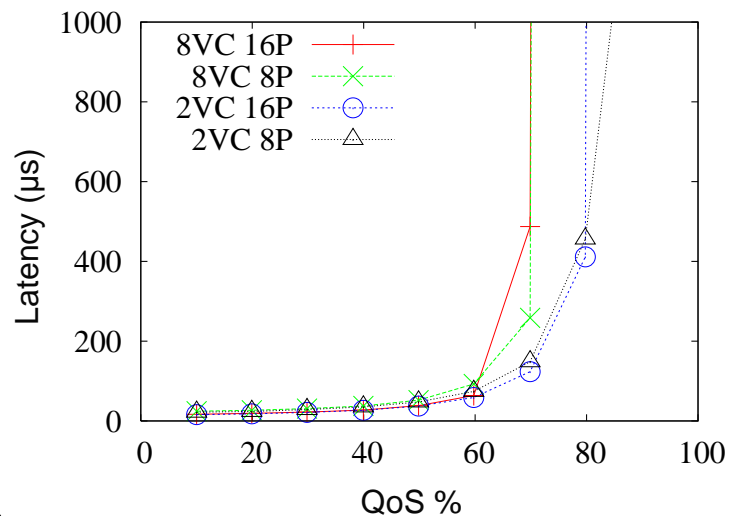


Average and maximum latency

- Introduction
- Related Work
- Architecture for QoS support
- **Performance Evaluation**
- Conclusions

Simulation Results

Audio traffic

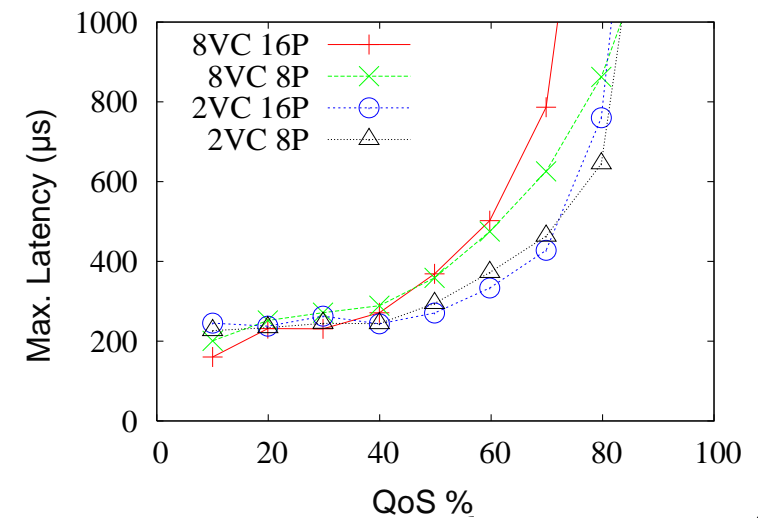
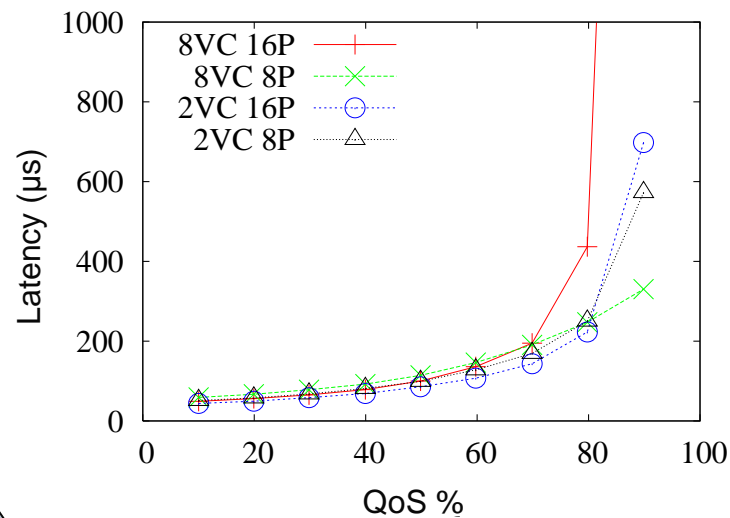


Average and maximum latency

- Introduction
- Related Work
- Architecture for QoS support
- **Performance Evaluation**
- Conclusions

Simulation Results

Video traffic

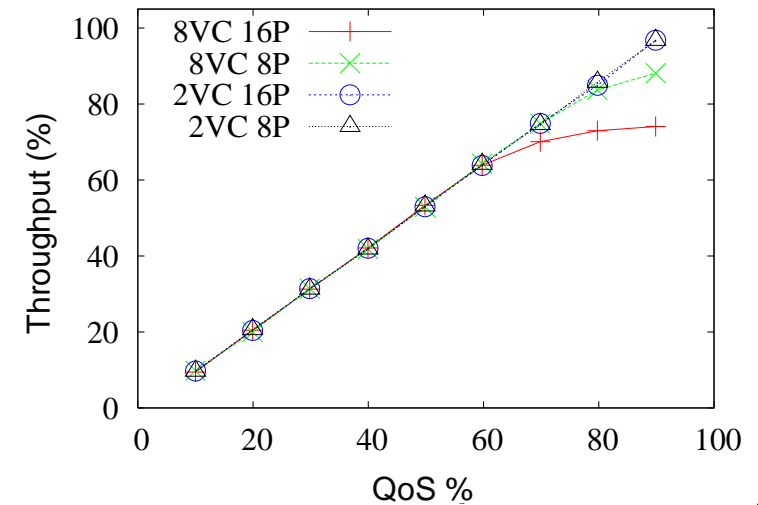
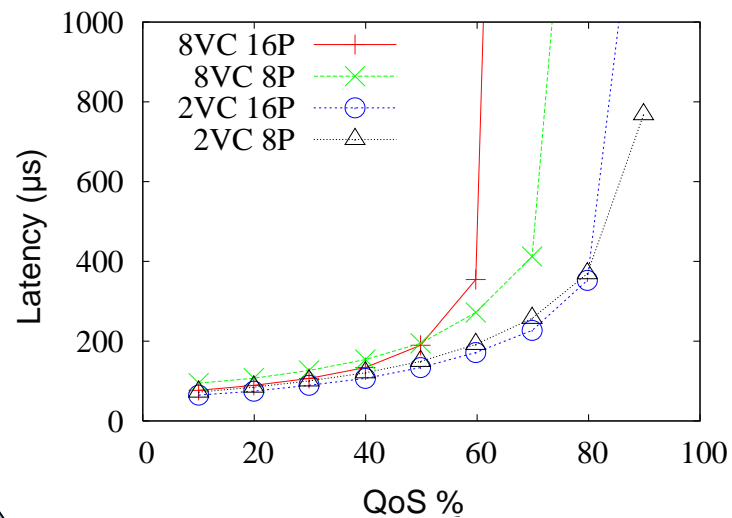


Average and maximum latency

- Introduction
- Related Work
- Architecture for QoS support
- **Performance Evaluation**
- Conclusions

Simulation Results

Controlled Load traffic



Average latency and throughput

- Introduction
- Related Work
- Architecture for QoS support
- **Performance Evaluation**
- Conclusions

Results Summary

- The following table summarizes the QoS traffic acceptance of the four architectures
- The best results are for our architectures, due to the more efficient buffers.

Traffic class	8VC 16P	8VC 8P	2VC 16P	2VC 8P
Network Control	90%	90%	80%	80%
Video	70%	70%	80%	80%
Best-effort	70%	80%	80%	80%
Background	60%	70%	80%	90%
All QoS	60%	70%	80%	80%

•Introduction

•Related Work

•Architecture for QoS support

•**Performance Evaluation**

•Conclusions

Table of Contents

- Introduction
- Related Work
- Architecture for QoS support
 - Providing Full QoS Support with 2 VCs
- Performance Evaluation
 - Simulation conditions
 - Simulation results
- Conclusions

•Introduction

•Related Work

•Architecture for
QoS support

•Performance
Evaluation

•**Conclusions**

Conclusions

- We presented a proposal to use only two VCs at each switch port to provide QoS support.
- We obtain results similar to those of more complex architectures.
- We improve performance when there is a lot of bursty traffic because buffer assignation is more flexible.

•Introduction

•Related Work

•Architecture for QoS support

•Performance Evaluation

•**Conclusions**

Efficient Switches with QoS Support for Clusters

Alejandro Martínez Vicente
Francisco José Alfaro Cortés
José Luis Sánchez García
José Duato Marín

I3A 
Instituto de Investigación
en Informática de Albacete



UNIVERSIDAD
POLITECNICA
DE VALENCIA