Design Issues for High-Performance Active Routers

T. Wolf and J. S. Turner, WU
IEEE JSAC Mar. 2001

K. C. Shih, 2001/11/1
Outline

- Introduction
- System Organization
- Router Port Design
- Scaling Issues
- Benchmarking
- Commercial Products
- Conclusion
Introduction

- Active and Programmable Networks
  - Customizing routing, open signaling, fully programmable control plane, support new protocols without changes in the underlying hardware
  - Competition among exiting and future ISPs may hinge on the speed at which one service provider can respond to new market demands over another
The problems are …

- Inherently lower performance of software processing compared to hardwired logic
- Active networks implement more complex services than just plain forwarding
- Terabit routers with 10 Gbps links are commercially available now. Can active routers keep pace with the rapid growth?
- Moore’s Law is losing
  - Transmission bandwidth has been growing faster than processing bandwidth
For example,

- A single processor

<table>
<thead>
<tr>
<th></th>
<th>500 MIPS processor</th>
<th>1 GIPS processor</th>
</tr>
</thead>
<tbody>
<tr>
<td>150Mbps link</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>1.5 Gbps link</td>
<td>12.5</td>
<td><strong>25</strong> Instr./byte</td>
</tr>
</tbody>
</table>

- JPEG: **81** instr./byte
Focus on

1. Multiple network processors on a single application specific integrated circuit (ASIC)
2. Design a scalable hardware for processing packets at wire speeds of 2.4 Gbps and higher
3. Benchmarking
System Organization - 1

Processing engine at each router port
System Organization - 2

Shared pool of processing engines
Router Port Design

Port Processor (PP)

Processing Engine (PE)
Port Processor

- Packet classification and queuing (PCQ) chip
- The headers are passed to the packet classifier and the entire packet is passed to the queue controller (QCTL)
- The queues can be rate-controlled to provide QoS guarantees
Processing Engine

- Active processing is provided by one or more *active processor chips* (APC)
- Multiple APCs are arranged in a daisy-chain configuration to eliminate the need for multiple interfaces to the PCQ.
Multiple APCs

I/O Channel

APC1 → APC2 → APC3 → PCQ

* Support priority naturally
Design of APCs

- Using 0.25-um (deep submicron) CMOS technology
Design of APCs

- 200mm²
- 8mm²
- 2mm² RISC
- 30mm², 1MB DRAM, 32kB cache
I/O Channel Bandwidth

- 32-bit interface in each direction at a clock rate of 250 MHz
  - 8 Gbps
  - packet fragmentation effects
  - bus (I/O channel) contention
  - > 2.4 Gbps (links speed)
More other consideration…

- High-performance memory interface
- Size of embedded memory
  - Operating system kernel
  - Active application codes
  - Per flow state (few hundred flows)
  - Packets currently being processed
- Size of Queue Memory
### Scaling Issues (1/2)

**Technology scaling**

#### APC Technology Scaling

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature size (μm)</td>
<td>0.25</td>
<td>0.18</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>No. of APUs</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Cache size (kB)</td>
<td>32</td>
<td>64</td>
<td>64</td>
<td>128</td>
</tr>
<tr>
<td>DRAM size (MB)</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Proc + MC area (mm²)</td>
<td>10</td>
<td>5.2</td>
<td>2.3</td>
<td>1.3</td>
</tr>
<tr>
<td>SRAM area per MB</td>
<td>175</td>
<td>90</td>
<td>40</td>
<td>23</td>
</tr>
<tr>
<td>DRAM area per MB</td>
<td>25</td>
<td>13</td>
<td>5.8</td>
<td>3.2</td>
</tr>
<tr>
<td>Total APU area (mm²)</td>
<td>162</td>
<td>148</td>
<td>131</td>
<td>137</td>
</tr>
<tr>
<td>Processor clock frequency (MHz)</td>
<td>400</td>
<td>556</td>
<td>833</td>
<td>1,111</td>
</tr>
<tr>
<td>External memory bandwidth (MB/s)</td>
<td>500</td>
<td>694</td>
<td>2,083</td>
<td>2,778</td>
</tr>
<tr>
<td>Instructions per byte for 2.4 Gb/s link</td>
<td>5.3</td>
<td>7.4</td>
<td>22</td>
<td>30</td>
</tr>
</tbody>
</table>
Scaling Issues (2/2)

- Multiple APCs
  - Each interface that connects to another APCs acts as a gateway and routes data to other APCs
# Benchmarking

## SIZE AND COMPUTATIONAL COMPLEXITY OF BENCHMARK APPLICATIONS

<table>
<thead>
<tr>
<th>Application</th>
<th>Description</th>
<th>Object Code (bytes)</th>
<th>Executed Code (bytes)</th>
<th>Complexity (instr. / byte)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTR</td>
<td>routing table lookup</td>
<td>16,000</td>
<td>15,220</td>
<td>2.1</td>
</tr>
<tr>
<td>DRR</td>
<td>packet scheduling</td>
<td>2,500</td>
<td>5,412</td>
<td>4.1</td>
</tr>
<tr>
<td>FRAG</td>
<td>packet fragmentation</td>
<td>2,400</td>
<td>5,032</td>
<td>7.7</td>
</tr>
<tr>
<td>TCP</td>
<td>traffic monitoring</td>
<td>352,000</td>
<td>29,028</td>
<td>10</td>
</tr>
<tr>
<td>JPEG</td>
<td>image compression</td>
<td>260,000</td>
<td>24,620</td>
<td>81</td>
</tr>
<tr>
<td>CAST</td>
<td>data encryption</td>
<td>19,500</td>
<td>10,116</td>
<td>104</td>
</tr>
<tr>
<td>ZIP</td>
<td>data compression</td>
<td>117,000</td>
<td>14,152</td>
<td>226</td>
</tr>
<tr>
<td>REED</td>
<td>forward error correction</td>
<td>6,900</td>
<td>6,040</td>
<td>603</td>
</tr>
</tbody>
</table>
Suggestions

- The large difference of code size (6 – 30 kB) suggests that it may be necessary to specialize the different APUs.
- For example, CAST, a single APC will be able to encrypt all the data on a 2.4 Gbps link only sometime after 2005 (16 APUs).
Commercial Products

- Programmable packet processing engines for routers

Intel IPX1200
- Six processing microengines (6 APUs), one control processor, 200-MHz clock rate, 2.6-Gbps line speed, 6.26-Gbps I/O bus, four threads per processor
Conclusion

- Active networking is an important new direction in networking research and potentially for commercial networks.
- Active routers should keep pace with the link speed.
- The paper proposed a fundamental design of active routers and related design issues.