Network Processors

CSE 291E / ECE 260C
Spring 2002
Outline

• What is the network processing problem?
• Examples
  • Broadcom/SiByte
  • Intel IXP
  • Cognigine
Network Processing

- Ethernet switching
- IP routing
- TCP offload
- L3/L4 load balancing
- Content/URL-aware load balancing
- ACL (access control)
- Stateful firewalls
- VPN
- SSL acceleration
- IPSec acceleration
- ...
Network Processing

• Stateless processing
  • Decisions don’t require knowledge of previous packets
  • e.g. IP routing: just look at destination address for forwarding decisions

• Stateful processing
  • Decisions require knowledge of other packets processed
  • e.g. TCP offload: needs to keep track of sequence numbers & acknowledged packets ...

• Ethernet switching
• IP routing
• TCP offload
• L3/L4 load balancing
• Content/URL-aware load balancing
• ACL (access control)
• Stateful firewalls
• VPN
• SSL acceleration
• IPSec acceleration
• ...
Linux (or Win2K) as Networking Appliance

- Linux (or Win2K) OS on Intel PC + NICs can execute all these networking functions
  - Ethernet switching
  - IP routing
  - TCP offload
  - L3/L4 load balancing
  - Content/URL-aware load balancing
  - ACL (access control)
  - Stateful firewalls
  - VPN
  - SSL acceleration
  - IPSec acceleration
  - …
But …

- Linux (or Win2K) on PC as appliance has limited performance (perhaps up to 100Mbps bandwidth)
Gigabit/Terabit Router Architecture Today

Control plane
- Implements routing protocols (e.g. RIP, BGP, OSPF) to update routing tables

Data plane
- **Line card**: lookup routing table (e.g. in SRAM) to decide on output port
- **Fabric**: switch packets to output port
Gigabit/Terabit Router Architecture Today

- **Line cards**: 10 Gbps emerging, 40 Gbps to come ...
- **Fabric**: 320 Gbps single-IC fabric chips, aggregate-able to many terabits emerging
The IP Lookup Problem

- Example Routing Table

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) 00001*</td>
<td>2</td>
</tr>
<tr>
<td>b) 00010*</td>
<td>3</td>
</tr>
<tr>
<td>c) 00011*</td>
<td>2</td>
</tr>
<tr>
<td>d) 001*</td>
<td>1</td>
</tr>
<tr>
<td>e) 0101*</td>
<td>0</td>
</tr>
<tr>
<td>f) 011*</td>
<td>2</td>
</tr>
<tr>
<td>g) 100*</td>
<td>1</td>
</tr>
<tr>
<td>h) 1010*</td>
<td>3</td>
</tr>
<tr>
<td>i) 1100*</td>
<td>0</td>
</tr>
<tr>
<td>j) 11100000*</td>
<td>3</td>
</tr>
</tbody>
</table>

- Longest Prefix Match: IP (32 bits) = 10100010... matches (h)
IP Lookup using Binary Trie

Example Prefixes

- a) 00001*
- b) 00010*
- c) 00011*
- d) 001*
- e) 0101*
- f) 011*
- g) 100*
- h) 1010*
- i) 1100*
- j) 11110000*

- e.g. Implementation = SRAM data structures + ASIC/NP to implement lookup algorithm
More General Packet Classification Problem

- Instead of just looking at destination IP, look at multiple packet fields
  - e.g. **src IP** (32 bits), **dest IP** (32 bits), **protocol field** (8 bits), **TCP/UDP src port** (16 bits), **TCP/UDP dest port** (16 bits), **TCP flags** (8 bits)
- **e.g. applications**
  - **Packet filtering**: e.g. deny traffic from an ISP (recognized by src IP) to destination (recognized by dest IP)
  - **Policy routing**: send voice-over-IP traffic (recognized by TCP/UDP port) using different priority queues and/or different output trunk
  - **Traffic shaping**: make sure traffic from an ISP (recognized by src IP) is limited to 50Mbps
Broadcom/SiByte 1250

- 2 x 1GHz quad-issue 64-bit MIPS, 128 Gbps internal bus
- Built-in DDR memory controller, 3 x GE, PCI/Hypertransport I/O
- Targets/claims 1Gbps line speed, 10M packets/sec IP forwarding, 5-10 watts
Architectures for 10-40 Gbps Line Speeds

• 1-2 RISC cores not enough for 10-40Gbps line speeds

• Major problems
  • Need enormous raw computing power
  • Need enormous memory I/O to store packets and data structures externally during algorithm execution
  • Specialized computations like CRC checks, classification ... more efficient in specialized hardware/datapaths
Intel IXP2800 (10Gbps)
Intel IXP2800

Microengine
CSE 291E/ECE 290C, Spring 2002 - Copyright © 2002, Bill Lin
Multiple LAN/WAN configurations with the IXP2800 network processor
Cognigine

RCU: Reconfigurable Communications Unit
RSF: RCU Switch Fabric
Cognigine (10-40Gbps)

- Distributed arbitration
- Split transaction
- Hierarchical partitioning
- Memory mapped
- “Narrow” 64 bit links

RCU: Reconfigurable Communications Unit
RSF: RCU Switch Fabric
Cognigine

- 5 Stage Pipeline
- 4-way multi-threaded
- Hardware RSF synchronization
- 128 bit reconfigurable address path
- 256 bit reconfigurable data path
Closing Comments

- Many Network Processor vendors
- Interesting killer app to push deep-pipelined & chip-multiprocessor architectures
Questions?