Introduction to the course
Scope:

- High performance networking design and usage.

Course structure:

- Lectures – 6 hours – FS09
- Supervised Labs – 10 hours - SW02 (ACS lab), tutorials in FS09

Assessment:

- Practical Assignment (100%) – 21/04/2020 12:00
## Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Lecture</th>
<th>Lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General architecture of high performance network devices</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Programmable devices</td>
<td>Introduction to NetFPGA (FS09)</td>
</tr>
<tr>
<td>3</td>
<td>High throughput devices – Part I</td>
<td>Introduction to P4 (FS09) Project selection</td>
</tr>
<tr>
<td>4</td>
<td>High throughput devices – Part II</td>
<td>Project architecture</td>
</tr>
<tr>
<td>5</td>
<td>Low latency devices - Part I</td>
<td>Performance profile</td>
</tr>
<tr>
<td>6</td>
<td>Low latency devices - Part II</td>
<td>Evaluation plan and testing</td>
</tr>
</tbody>
</table>
Project

• Starting point: a reference design of a network device
• Goal: Improving data-delivery in the presence of network congestion
• Examples:
  • Fast TCP retransmit
  • Shared output buffer
  • More examples on the website
• Projects done in pairs
• More information in Lab 1
Some logistics for 2018-19

Web page: http://www.cl.cam.ac.uk/teaching/current/P51/

Mailing list: cl-acs-p51-announce@cam.ac.uk

Grades:

Mphil (ACS) – Pass / Fail - based on a mark out of 100

All others (DTC) – Mark out of 100
Next steps

• Explore the web page
  http://www.cl.cam.ac.uk/teaching/current/P51/

• Decide if you still want to take the class – promptly

• Project:
  • Pair with a classmate – at least one must have taken ECAD!
  • Register to NetFPGA repository
    http://netfpga.org/site/#/SUME_reg_form/
  • Register to the P4-NetFPGA repository
    https://goo.gl/forms/h7RbYmKZL7H4EaUf1
Introductions
General architecture of high performance network devices
What Is a Switch?

We use switches all the time!

ON / OFF

Left / Right
What Is a Network Switch?

Conceptually, a left / right switch…

• Receives a packet through port <N>
• Decides through which port to send it
  • A *forwarding* decision

+ Some “real world” considerations
Real World Switches

- High Throughput Switch Silicon: 6.4Tbps (64x100G) – 25Tbps (64x400G)
  - Top of Rack Switches
    - E.g. Broadcom Tomahawk 4, Barefoot Tofino 2, Mellanox spectrum II
  - High Throughput Core Switch System: ~ 1 Petabit/sec
    - E.g. Arista 7500R series, Huawei NE5000E, Cisco CRS Multishelf
Real World Switches

• Low latency switch (Layer 1): ~5ns fan-out, ~55ns aggregation

• Low latency switch (Layer 2): 95ns - 300ns
  • Examples: g. Mellanox spectrum II, Exablaze Fusion

• Low latency NIC: <1us (loopback)
  • E.g. Mellanox Connect-X, Solarflare 8000, Chelsio T6, Exablaze ExaNIC

• Low latency switches don’t always support full line rate!
Cool numbers, what do they mean?

• Streaming data at 25Tbps:
  • Game of Thrones (Entire series, FHD, 237GB) – 76 milliseconds
  • Wikipedia (text, 161GB) – 52 milliseconds
  • ImageNet (ML dataset, 150GB) – 48 milliseconds
  • Wikimedia (232TB) – 74 seconds

• 100ns latency is equivalent to:
  • Travelling at the speed of light 0.037% of the distance between Cambridge and London (30m)
  • Traversing 20m of fibre
Real World Switch Silicon in Numbers

- Over 20 Billion Transistors
- Manufacturing process of down to 7nm
- Silicon size: 400 to 600 square mm
- Clock Rate: ~1.25GHz (typical)
- Packet Rate: ~10 Billion packets per second
- Buffer Memory: ~16MB-30MB on-chip
- Ports: Up to 256
- Power: ~100W-300W
- 2019 Numbers
What Drives The Architecture of a Switch?

- Cost
- Manufacturing limitations (e.g. maximum silicon size)
- Power consumption
- General purpose or user specific?
- I/O on the package
- Number of ports:
  - Front panel size (24, 32, 48 ports in 19 inch rack)
  - MAC area
Packet Rate as a Performance Metric

- Bandwidth is misleading
  - For example: full line rate for 1024B packets but not for 64B packets…

- Packet Rate: how many packets can be processed every second?
  - Unit: packets per second (PPS)

- An easy way to calculate the packet rate:
  \[ \frac{\text{Clock Frequency}}{\text{Number of Clock Cycles per Packet}} \]
Switch Internals 101

What defines the architecture of a switch?
Input Ports
Header Processing
Network Interfaces
Switching
Output Queues
Scheduling
Is This A Real Switch?
Recall What Drives Real World Switches

- Cost
- Power
- Area
Sharing Resources Is Good!

• Single header processor (if possible)
• Shared memories
• No concurrency problems
  • Also no need to synchronise tables, no need to send updates, ….
Rethinking The Switch Architecture
Rethinking The Switch Architecture
Where Is The Switching?
Output Queueing

Schedule & Rate limit
Input Queueing
Virtual Output Queueing
Virtual Output Queueing
Virtual Output Queueing
Deep Buffers

Queues

Queues Manager

External Memory Controller

External Memory PHY

External Memory
Scheduling

- Different operations within the switch:
  - Arbitration
  - Scheduling
  - Rate limiting
  - Shaping
  - Policing
- Many different scheduling algorithms
  - Strict priority, Round robin, weighted round robin, deficit round robin, weighted fair queueing...
Scheduling Hierarchies

- SCH (Priority)
- SCH (RR)
- SCH (WFQ)

SP – Strict Priority
Pn – Priority <n>
BE – Best Effort
RL – Rate Limiting
WFQ – Weighted Fair Queueing
RR – Round Robin
Software Defined Networking (SDN)

Key Idea: Separation of Data and Control Planes

(a) Classical-Router Network

(b) SDN Network
Switch Architecture and SDN