Prism: A Proxy Architecture for Datacenter Networks

Yutaro Hayakawa (Keio University)
Lars Eggert (NetApp)
Michio Honda (NEC Laboratories Europe)
Douglas Santry (NetApp)
Background

- TCP proxy plays important roles in modern cloud services
- Web cache
- Application level firewall
- Application level load balancer
Problem

• Aggregated bandwidth of backends are always constrained by proxy bandwidth
Problem

- Aggregated bandwidth of backends are always constrained by proxy bandwidth
Problem

- Aggregated bandwidth of backends are always constrained by proxy bandwidth
Problem

• Aggregated bandwidth of backends are always constrained by proxy bandwidth
Problem

- Aggregated bandwidth of backends are always constrained by proxy bandwidth
Problem

- Aggregated bandwidth of backends are always constrained by proxy bandwidth

Total Bandwidth for Backends == 10G
Why does it matter?

• This problem is significant especially at the edge of today’s datacenter topology
Example: HTTP/1.1

- Simple web system
  - One frontend proxy and multiple backend servers
Example: HTTP/1.1

Client’s HTTP GET

40Gbps Links

10Gbps Links

Backend

Frontend Proxy

Backend

Backend
Example: HTTP/1.1

Diagram showing the following:
- Backend
- Frontend
- Proxy

Connections:
- 40Gbps Links
- 10Gbps Links

Forward To backend
Example: HTTP/1.1

HTTP OK

40Gbps Links

10Gbps Links

Backend

Frontend Proxy

Backend

Backend
Example: HTTP/1.1

- Backend
- Frontend Proxy
- Backend
- Backend

10Gbps Links
40Gbps Links

Forward it to client
Example: HTTP/1.1
Example: HTTP/1.1

Problem:
All incoming/outgoing traffic go through here
How can we solve this problem?

Outgoing traffic from backend bypass frontend proxy
Related work

• L4 load balancer
  • Maglev[1], Ananta[2], Duet[3]


Balance TCP connect request
Prism Architecture

- Legacy Proxy
- Client
- L2 Switch
- 40Gbps Links
- 10Gbps Links
- Back end
- Proxy
- Back end
- Back end
Prism Architecture

Legacy Proxy
- Client
- L2 Switch
  - 10Gbps Links
  - 40Gbps Links
  - Backend
  - Proxy
  - Backend

Prism
- Client
- Programmable Switch (e.g. P4)
- 40Gbps Links
- 10Gbps Links
- Custom Backend
- Switch Controller (Proxy)
- Custom Backend
- Custom Backend
Prism Architecture

Basic idea: Offloading forwarding of application protocol payload to a switch
Prism Architecture

Basic idea: Offloading application protocol payload forwarding to a switch
=> the Switch Controller (Proxy) only handles header exchange
How does it work?

Example Behavior
HTTP/1.1

Client

Programmable Switch

40Gbps Links

- Custom Backend
- Switch Controller (Proxy)
- Custom Backend
- Custom Backend
- Custom Backend

10Gbps Links
How does it work?

Client and Switch Controller establish TCP connection

Programmable Switch

40Gbps Links

10Gbps Links

Custom Backend

Switch Controller (Proxy)

Custom Backend

Custom Backend

...
How does it work?

Client sends request to switch controller

GET /example/index.html

Programmable Switch

40Gbps Links

10Gbps Links

Custom Backend

Switch Controller (Proxy)

Custom Backend

Custom Backend
How does it work?

Switch controller selects backend

Programmable Switch

Client

40Gbps Links

10Gbps Links

Custom Backend

Switch Controller (Proxy)

Custom Backend

Custom Backend

...
How does it work?

Switch controller injects a rule to the switch which rewrites fields of packets from backend.

Switch controller injects a rule to the switch which rewrites fields of packets from backend.
How does it work?

Switch controller sends client request to backend with some TCP metadata.

Client

Programmable Switch

Custom Backend

Switch Controller (Proxy)

Custom Backend

Custom Backend

Rule

Sequence Number
ACK Number
Client’s request
TCP options…

40Gbps Links

10Gbps Links
How does it work?

Backend migrate TCP connection from information given by Switch controller.
How does it work?

Backend and client communicate directly via migrated TCP connection

Client

Programmable Switch

HTTP OK and Payload data

Rule

Custom Backend

Switch Controller (Proxy)

Custom Backend

Custom Backend

40Gbps Links

10Gbps Links
How does it work?

Backend send number of bytes sent to client to switch controller

Client

Programmable Switch

Rule

I sent N bytes

40Gbps Links

10Gbps Links

Custom Backend

Switch Controller (Proxy)

Custom Backend

Custom Backend
How does it work?

Switch controller removes the rule from the switch.

Client

Rule

Programmable Switch

40Gbps Links

10Gbps Links

Custom Backend

Switch Controller (Proxy)

Custom Backend

Custom Backend
How does it work?

Switch controller waits for next request or FIN packet

Client

Programmable Switch

Waiting for FIN or next request

Custom Backend

Switch Controller (Proxy)

Custom Backend

Custom Backend
Challenge

- Why do we need to use programmable switch like P4?

- How can we stop pipelined HTTP request from client during TCP connection is migrated in backend?

- What happen when the client resets the connection?

- More details are in paper!
Evaluation

• We setup two experiments

1. End-to-End throughput comparison between legacy proxy and Prism for HTTP/1.1

2. Packet transformation performance measurement for Prism Switch
Evaluation: End-to-End Throughput

- 2 clients and 2 servers
Evaluation: End-to-End Throughput

• 2 clients and 2 servers
Evaluation: End-to-End Throughput

- Total throughput of proxy’s clients are limited to about 9Gbps in maximum
- Total throughput of Prism’s clients exceeds 10Gbps in 2MB data transfer
- Finally reached to about 18.7Gbps in maximum

Prism can use bandwidth of uplinks which proxy’s clients can’t use
• Total throughput of proxy’s clients are limited to about 9Gbps in maximum

• Total throughput of Prism’s clients exceeds 10Gbps in 2MB data transfer

• Finally reached to about 18.7Gbps in maximum

Prism can use bandwidth of uplinks which proxy’s clients can’t use
Evaluation: Prism Switch Performance

- Measure performance of the Prism switch packet transformation
- Prism switch achieves
  - 63Gbps in 1514-byte packet using 1 core
  - 127Gbps in 2 cores

The Prism switch can achieve high throughput even if it is implemented as software.
Conclusion

- Legacy proxy architecture can’t utilize full bandwidth of switch uplinks in datacenters.

- To solve this problem, we designed and implemented Prism which connects clients and backend servers directly during payload transaction of application protocol.

- Prism improves bandwidth utilization.
Question?