Efficient Query Dispatching for Scale-Out Database Systems

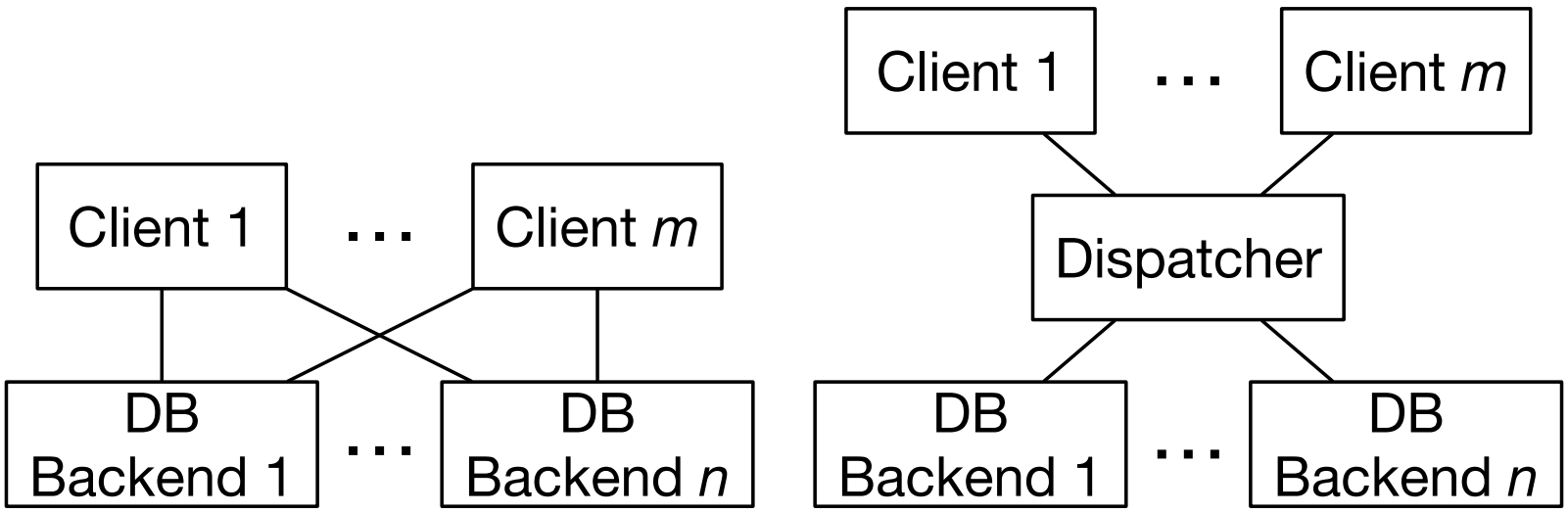
Stefan Klauck, Max Plauth, Sven Knebel
Hasso Plattner Institute, University of Potsdam

Marius Strobl, Douglas Santry, Lars Eggert
NetApp

Munich Internet Research Retreat
Raitenhaslach, Germany, November 29 – 30, 2017

SSICLOPS is funded by the EU's
Horizon2020 Programme

Problem space



(a) Direct communication. (b) Central dispatcher.

Figure: Query dispatching architectures.

SSICLOPS

Dispatcher candidates

Hyrise Dispatcher

- Query load-balancer for Hyrise [6] lazy replicating in-memory DB
- Routes based on JSON query plan using 1 thread per client
- Socket API with buffering to avoid copies, nodejs/http-parser [2]

SSICLOPS

Dispatcher candidates

Hyrise Dispatcher

- Query load-balancer for Hyrise [6] lazy replicating in-memory DB
- Routes based on JSON query plan using 1 thread per client
- Socket API with buffering to avoid copies, nodejs/http-parser [2]

HAProxy

- "The Reliable, High Performance TCP/HTTP Load Balancer" [7]
- Popular, open-source, general-purpose
- Employs socket splicing [5] on GNU/Linux

SSICLOPS

Dispatcher candidates

Hyrise Dispatcher

- Query load-balancer for Hyrise [6] lazy replicating in-memory DB
- Routes based on JSON query plan using 1 thread per client
- Socket API with buffering to avoid copies, nodejs/http-parser [2]

HAProxy

- "The Reliable, High Performance TCP/HTTP Load Balancer" [7]
- Popular, open-source, general-purpose
- Employs socket splicing [5] on GNU/Linux

Prism

- Splits single TCP connections across servers [3]
- Controller reprograms SDN switch (P4 [1] or mSwitch [4])
- Eliminates controller/dispatcher as central bottleneck

SSICLOPS

Prism

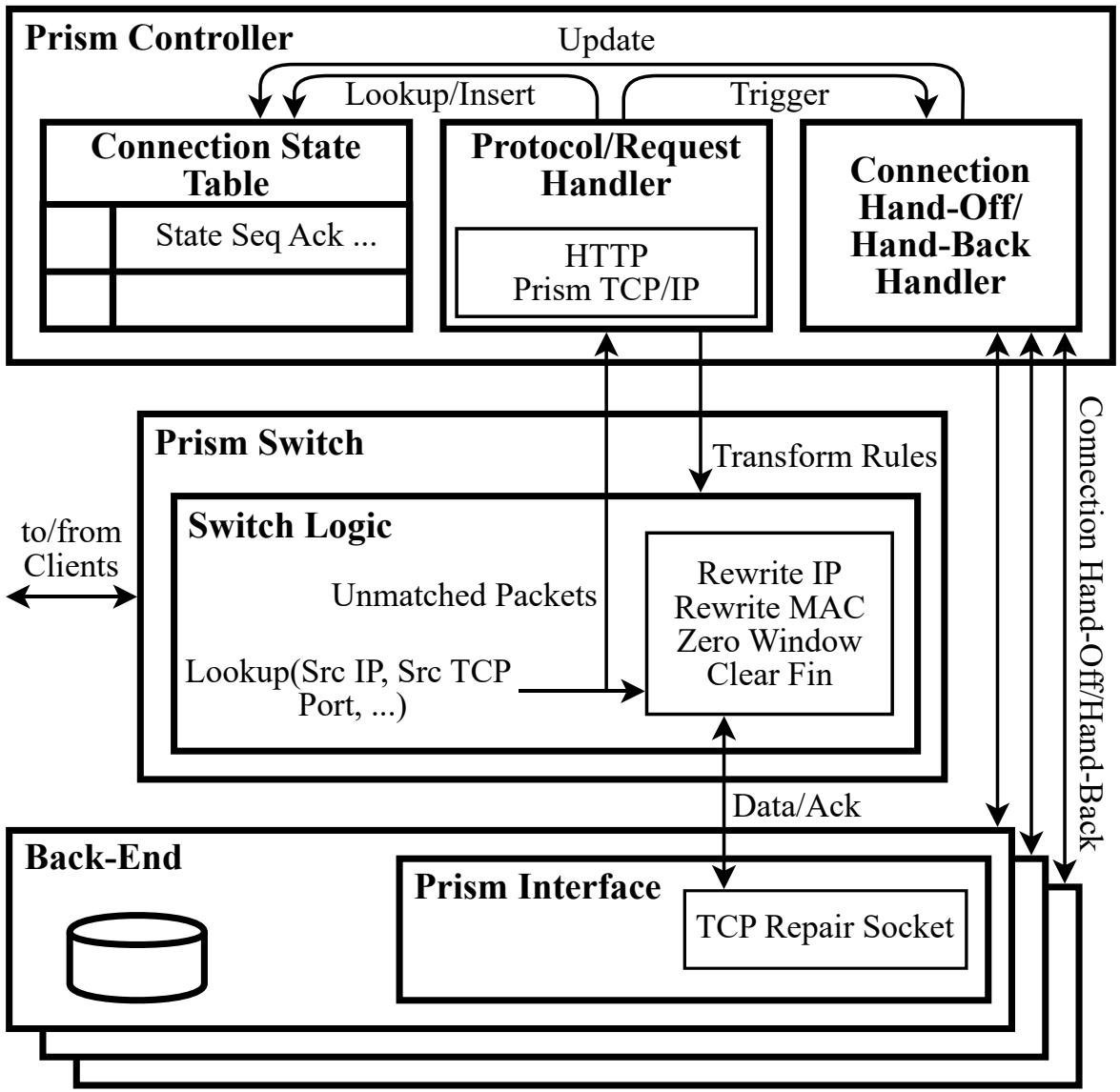
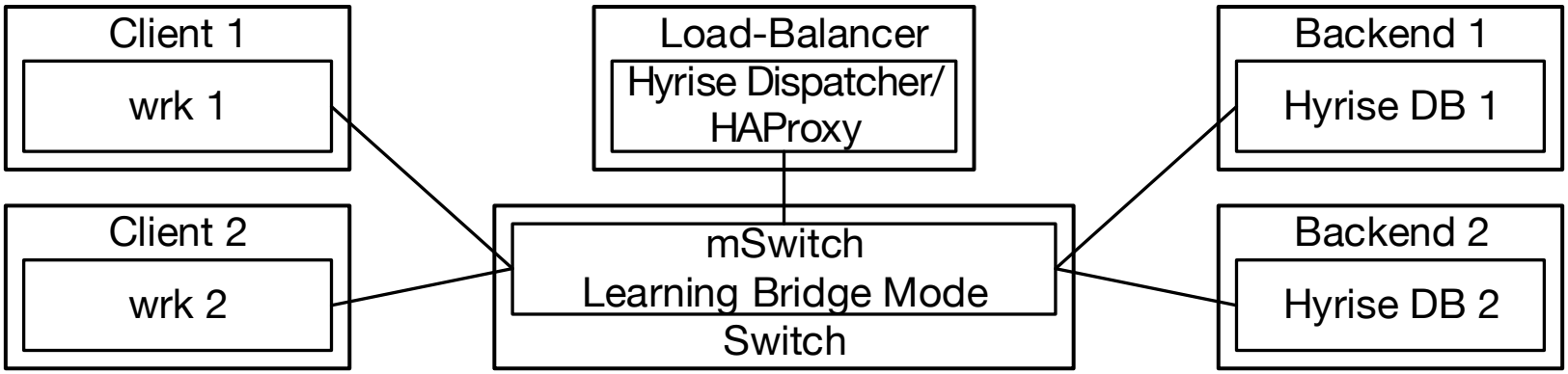


Figure: Prism software architecture, based on [3].

SSICLOPS

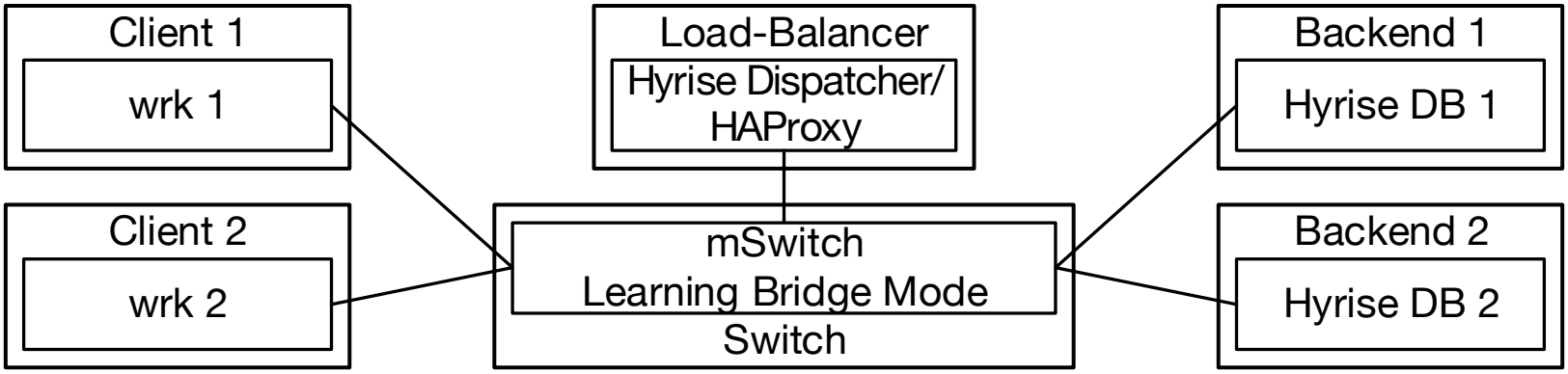
Experimental setup



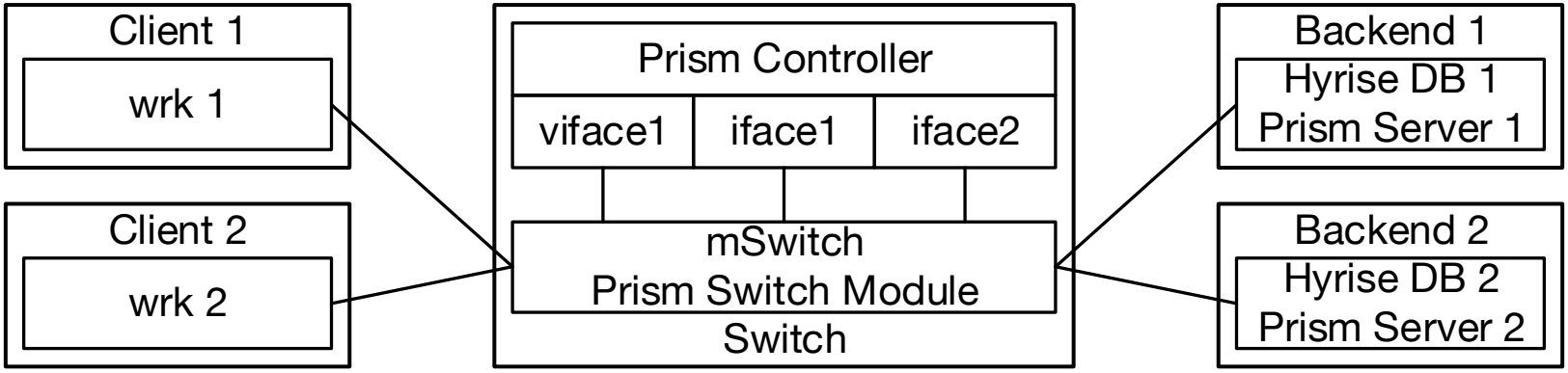
(a) Hyrise dispatcher and HAProxy topology.

SSICLOPS

Experimental setup



(a) Hyrise dispatcher and HAProxy topology.



(b) Prism topology.

Figure: Topologies for the evaluations, based on [3].

SSICLOPS

Experimental results

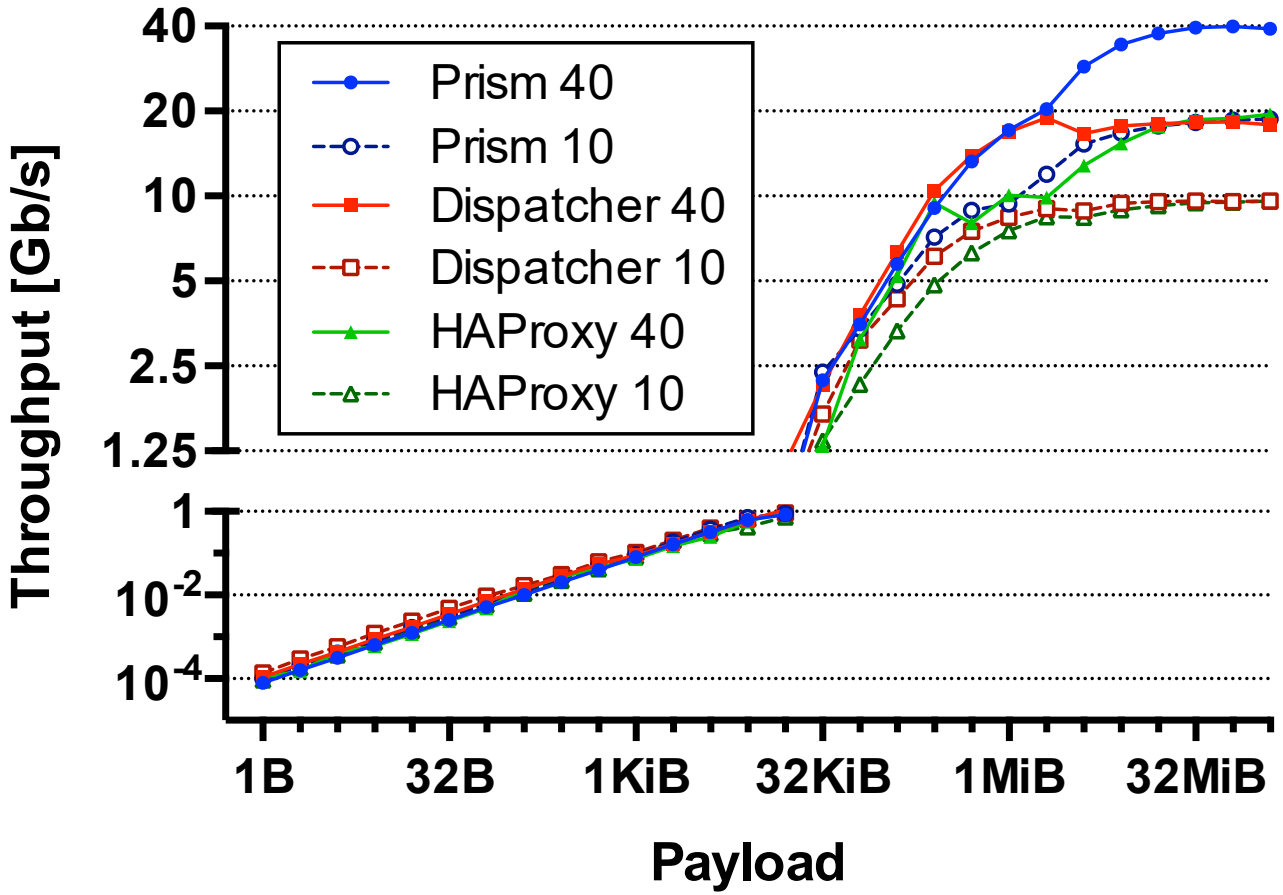


Figure: Dispatcher throughputs for varying payloads.

SSICLOPS

Thank you for your attention!

Disclaimer:

~~No hardware was harmed in the making of this presentation.~~
Two Mellanox ConnectX-3 cards died in the making of this presentation.

SSICLOPS

References

[1] Pat Bosshart et al. 2014. P4: Programming Protocol-independent Packet Processors. *SIGCOMM Comput. Commun. Rev.* 44, 3 (July 2014), 87–95.

[2] Node.js Foundation. [n. d.]. HTTP Parser. `https://github.com/nodejs/http-parser`. ([n. d.]).

[3] Yutaro Hayakawa, Lars Eggert, Michio Honda, and Douglas Santry. 2017. Prism: A Proxy Architecture for Datacenter Networks. In *Proceedings of the 2017 Symposium on Cloud Computing (SoCC '17)*. ACM, New York, NY, USA, 181–188.

[4] Michio Honda, Felipe Huici, Giuseppe Lettieri, and Luigi Rizzo. 2015. mSwitch: A Highly-scalable, Modular Software Switch. In *Proceedings of the 1st ACM SIGCOMM Symposium on Software Defined Networking Research (SOSR '15)*. ACM, New York, NY, USA, Article 1, 13 pages.

[5] David A. Maltz and Pravin Bhagwat. 2000. TCP Splice Application Layer Proxy Performance. *J. High Speed Netw.* 8, 3 (Jan. 2000), 225–240.

[6] David Schwalb et al. 2015. Hyrise-R: Scale-out and Hot-Standby Through Lazy Master Replication for Enterprise Applications. In *Proceedings of the 3rd VLDB Workshop on In-Memory Data Mangement and Analytics (IMDM '15)*. ACM, New York, NY, USA, Article 7, 7 pages.

[7] Willy Tarreau. [n. d.]. The Reliable, High Performance TCP/HTTP Load Balancer. `https://www.haproxy.org`. ([n. d.]).