

RouteScout: Performance-Driven Routing

Maria Apostolaki



BGP does not take performance into consideration.

protocol used to calculate paths to route traffic across the Internet.

Multiple (and increasing) alternative next hops per network.

- ◆ 95p of alternatives next hops per AS has doubled

Performance difference among alternative next hops.

- ◆ 30–80 of BGP path selection is suboptimal [1]
- ◆ for 30% of the tested destination, there is a next hop with 67% lower performance

New powerful & programmable hardware.

- ◆ flexible pipeline & 6.5 Tbps of traffic in the backplane[2]

Monitoring Performance

Delay

- ◆ approximate delay with SYN-ACK
- ◆ no need for bi-directional traffic
- ◆ minimum application influence

Loss

- ◆ approximate loss with retransmission rate
- ◆ no need for bi-directional traffic
- ◆ one retransmission per flow to disambiguate network/receiver problems

Data-Plane Accuracy

- ◆ 85% of cases, the delay ranking is correct.
- ◆ median absolute delay differs by 142 ms.
- ◆ 96% of cases, the loss estimate is accurate.
- ◆ 4% of cases, loss estimates are off by 12–50%.

Data-Plane Memory requirements

Monitoring & Forwarding Table

#prefs	# flows/pref	Memory
511	30	0.18MB
898	20	2MB
1504	20	3MB

Loss Monitor

#prefs	# flows/pref	Memory
511	30	0.44 MB
1063	20	0.8 MB
2624	30	2MB

Delay Monitor

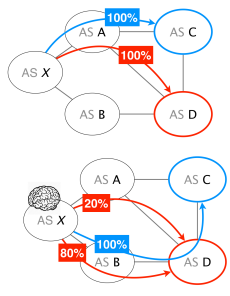
#prefs	% Coverage	Memory
511	73%	8KB
898	82%	19KB
2624	93%	115KB

Objectives

prefA opportunistic
 delay $\leq 200\text{ms}$
 loss $\leq 10\text{e-}6$
 1% traffic for testing
 single next hops in use

prefB: reactive
 delay $\leq 80\text{ms}$
 loss $\leq 0.001\%$
 0.1% traffic for testing

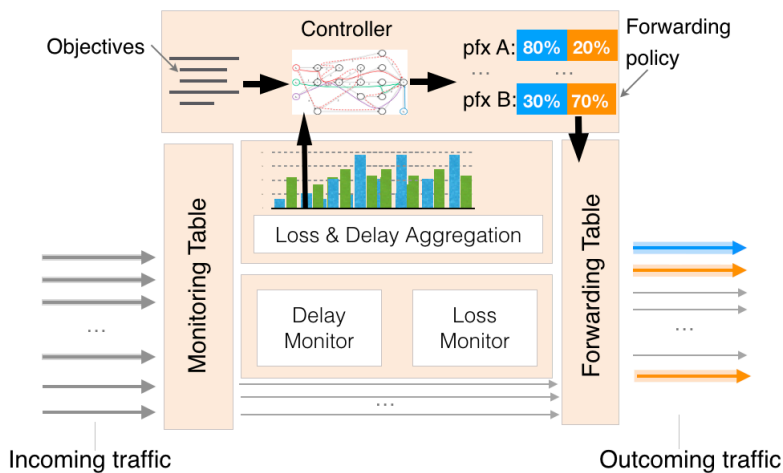
Solution



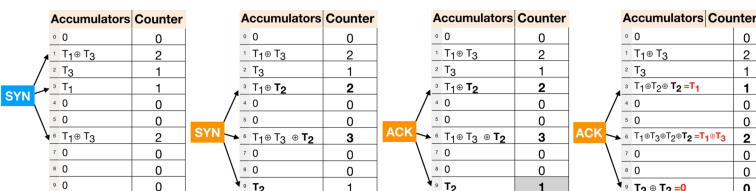
Challenges

- Stability
- Optimality
- Scalability
- Deployability
- Correctness

System Overview



Delay Monitor



References

- [1] S. Savage et al., Detour: Informed Internet routing and transport
- [2] Barefoot Tofino Switches: The Technology, <https://www.barefootnetworks.com/technology>