Destination Reachability and BGP Convergence Time

Beichuan Zhang (UCLA)

Dan Massey (Colorado State)

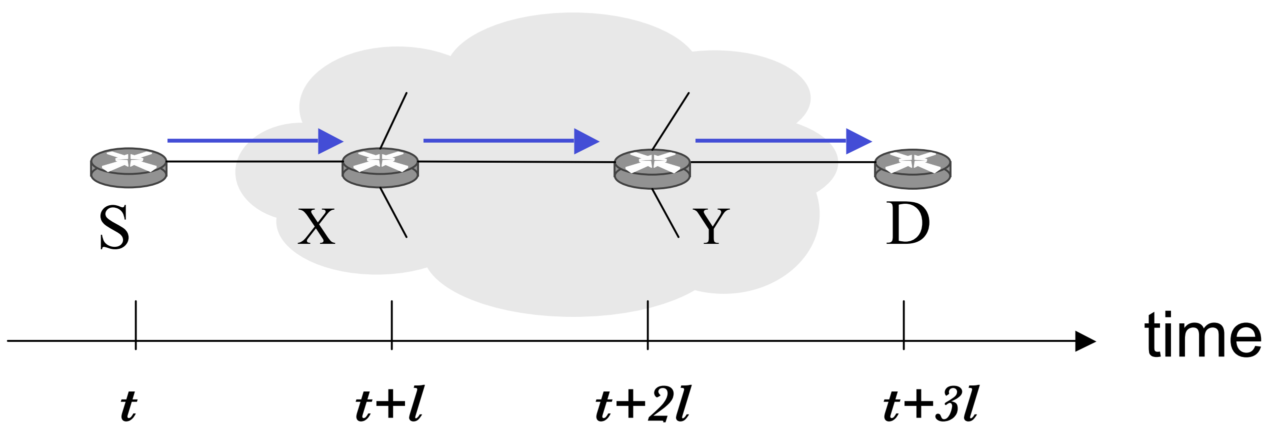
Lixia Zhang (UCLA)

Packet Delivery and Routing Dynamics

- The primary goal of routing is to ***deliver*** packets.
- Routing behaviors, such as convergence and stability, should be considered with respect to packet delivery.
- This work is one step in this direction.
  - Develop performance metrics for packet delivery
  - Analyze the impact of BGP convergence time on packet delivery in a simple case.

Destination Reachability

- D is ***connected*** at time t when there exists at least one path from S to D.
  - decided by physical topology
- D is ***reachable*** at time t when packets sent from S at time t will eventually reach D.
  - decided by both topology and routing.



Extra Downtime

- ***e(s) = downtime(s) - disconnected(s)***
- ***downtime(s):*** total time that D is unreachable from S.
- ***disconnected(s):*** total time that D is disconnected from S.
- ***e(s)*** measures the additional loss of reachability due to routing dynamics following topological changes.

False Uptime

- *f(s):* The time period during which S has a route to D, but packets sent by S will be eventually dropped in the network.
- It measures the overhead on network resources.

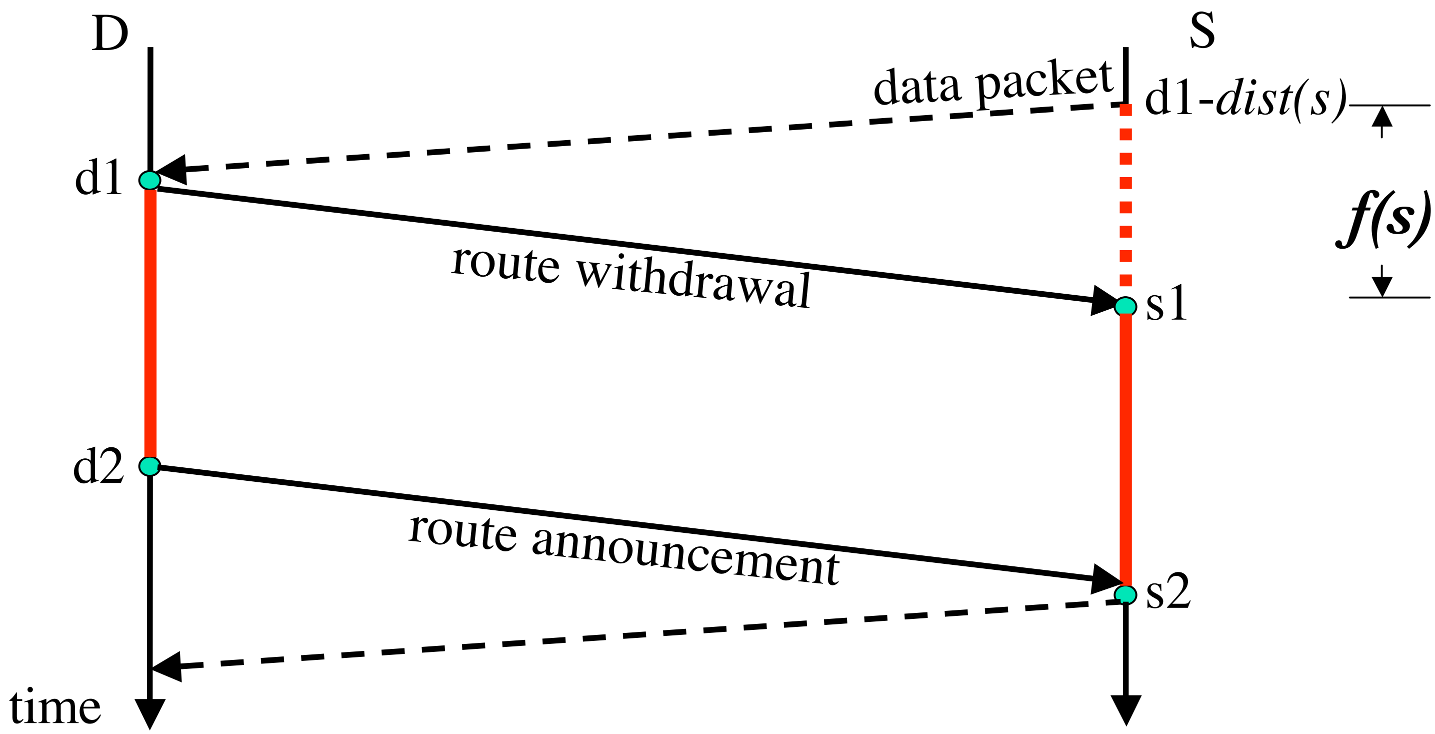
Destination Reachability

- Use $e(s)$ and $f(s)$ as metrics.
- It reflects the impact of routing dynamics on packet delivery.
- Need to know routing states in all intermediate routers over time. Only possible in analysis and simulations.

Case Study

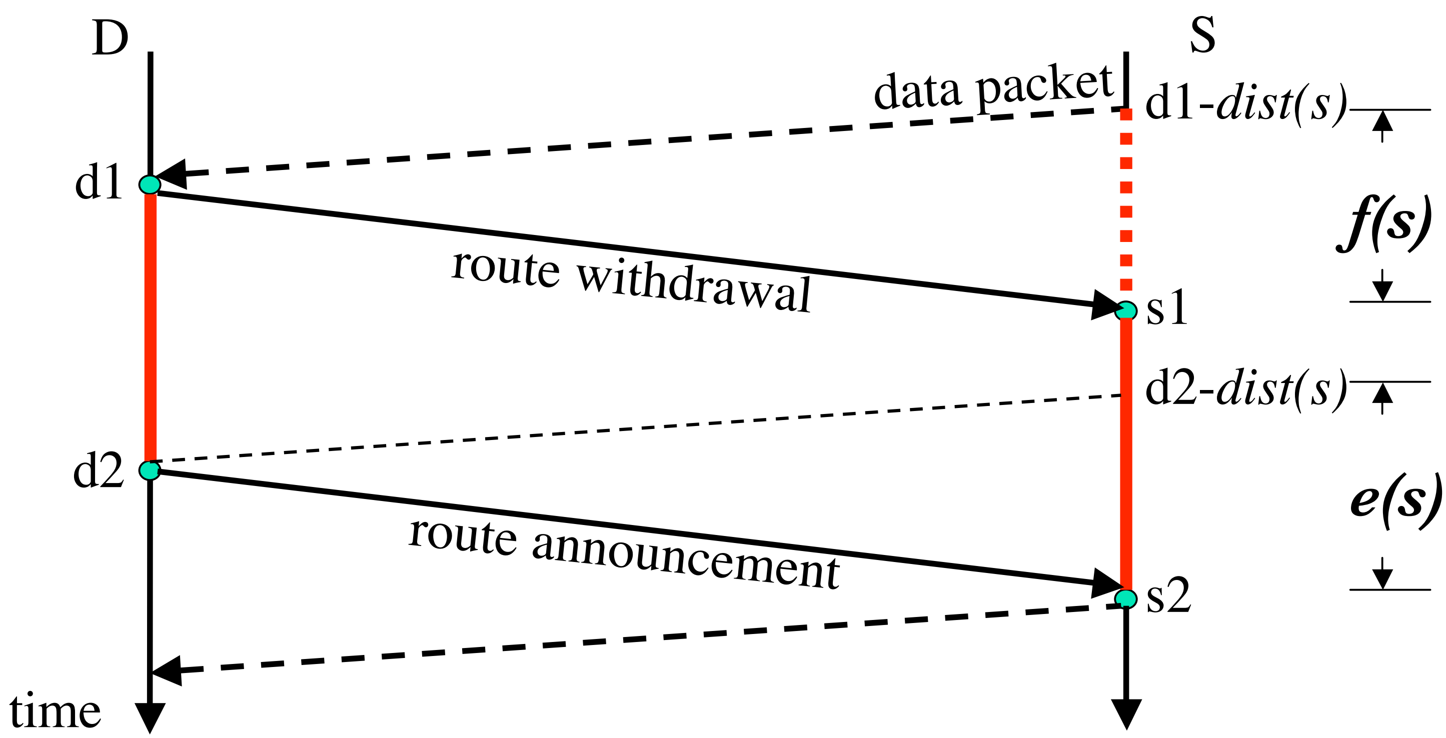
- When the destination is disconnected for a period of time, what is its reachability viewed from different sources?
  - D is disconnected at time $d1$. It takes time $T_{down}(s)$ for S to converge.
  - D is re-connected at time $d2$. It takes time $T_{up}(s)$ for S to converge.
  - Total disconnection time is $u=d2-d1$. Calculate $e(s)$ and $f(s)$.
- Use BGP as the routing protocol.
- Use simulations to verify analysis.

Ideal Routing Convergence



- $T_{down}(s) = s1 - d1$
- $f(s) = T_{down}(s) + dist(s)$

Ideal Routing Convergence

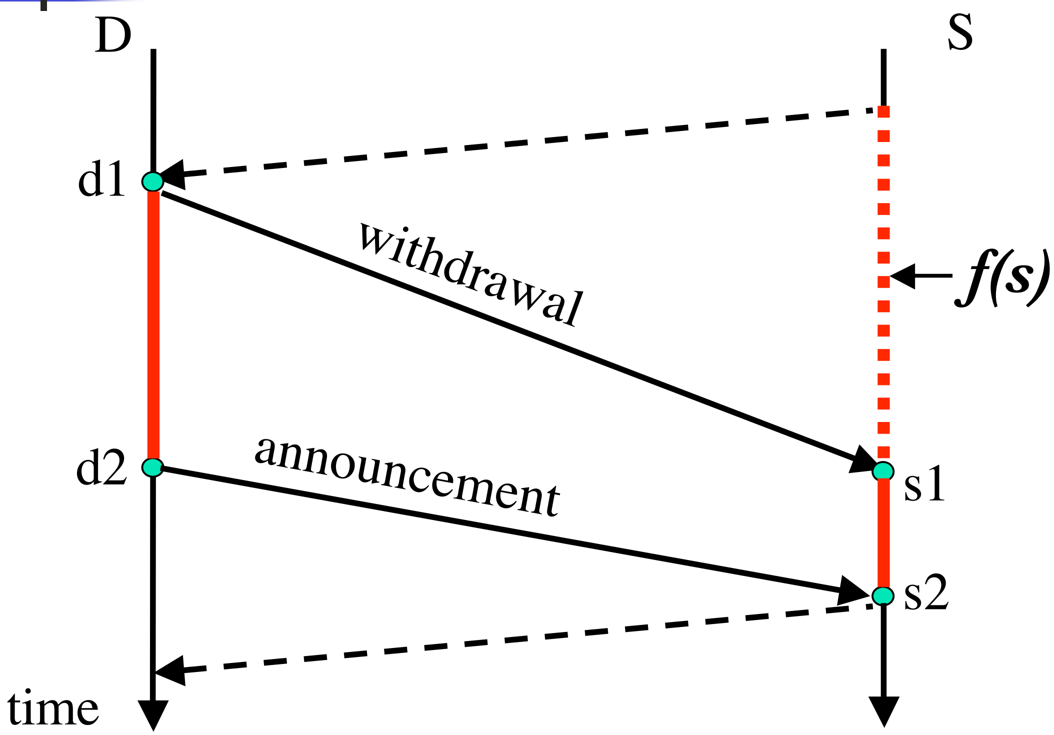


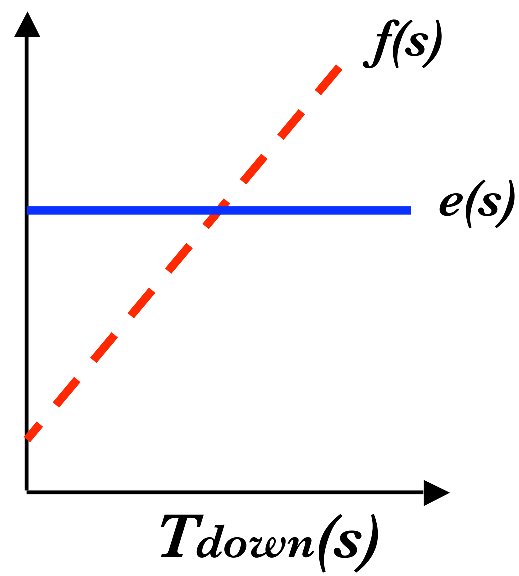
- $T_{up}(s) = s2 - d2$
- $e(s) = T_{up}(s) + dist(s)$

BGP Convergence

- Path Vector Protocol for inter-domain Routing
- BGP Slow convergence
  - Path exploration slows down T_{down} significantly.
  - MRAI slows down T_{up}.
  - Overall $T_{down} >> T_{up}$
- Convergence Improvement Proposals
  - Reduce or eliminate path exploration, therefore $T_{down} << T_{up}$
- What's the impact of varying T_{down} on packet delivery?

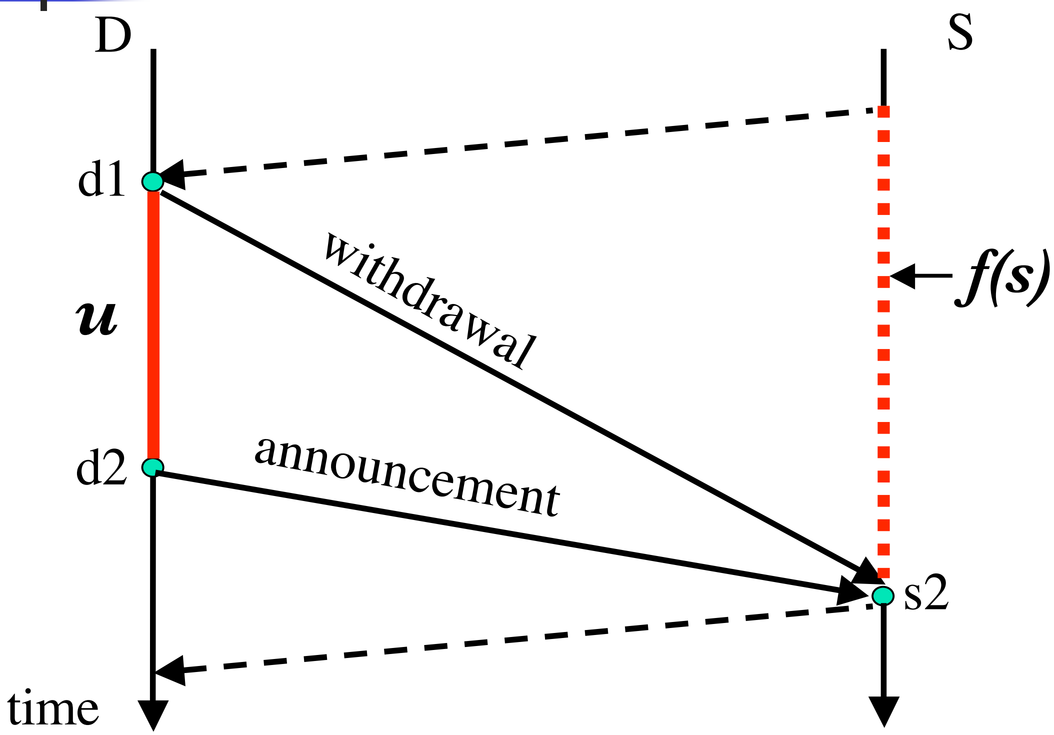
Increase T_{down}

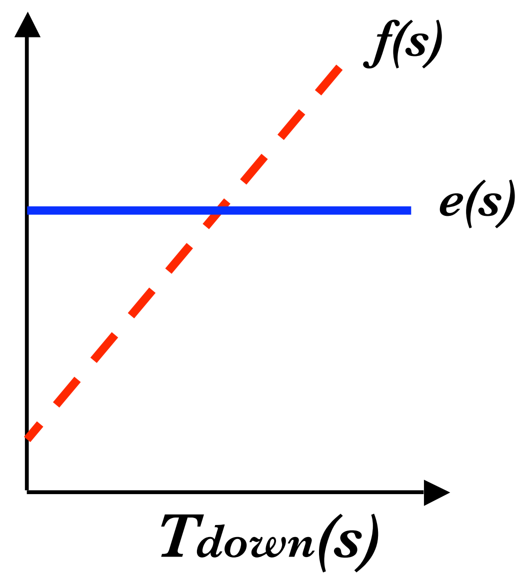




- $e(s) = T_{up}(s) + dist(s)$
- $f(s) = T_{down}(s) + dist(s)$

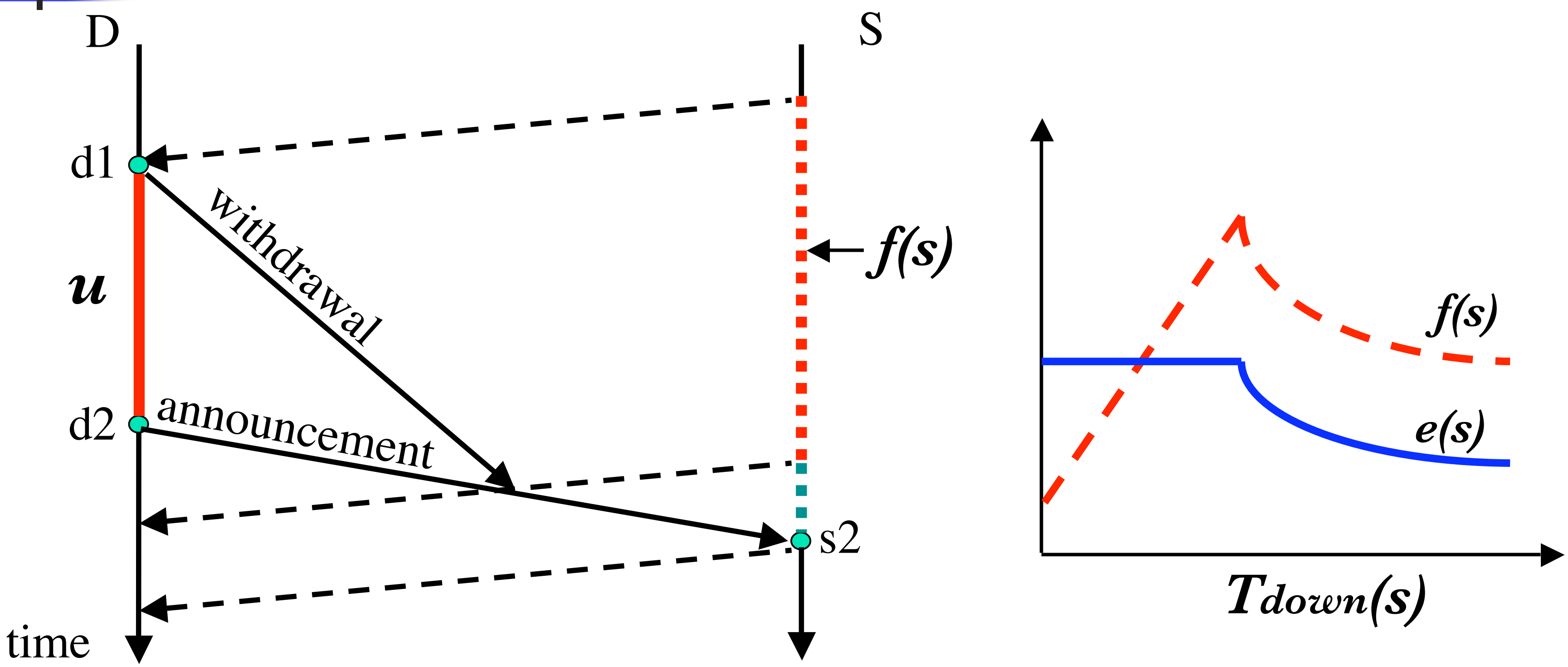
Worst Case





- $T_{down}(s) = s2 - d1$, $T_{up}(s) = s2 - d2$
- $T_{down}(s) = T_{up}(s) + u$

Long T_{down}



- $T_{down}(s) > T_{up}(s) + u$
- Longer $T_{down}(s)$ results in shorter $e(s)$ and $f(s)$!

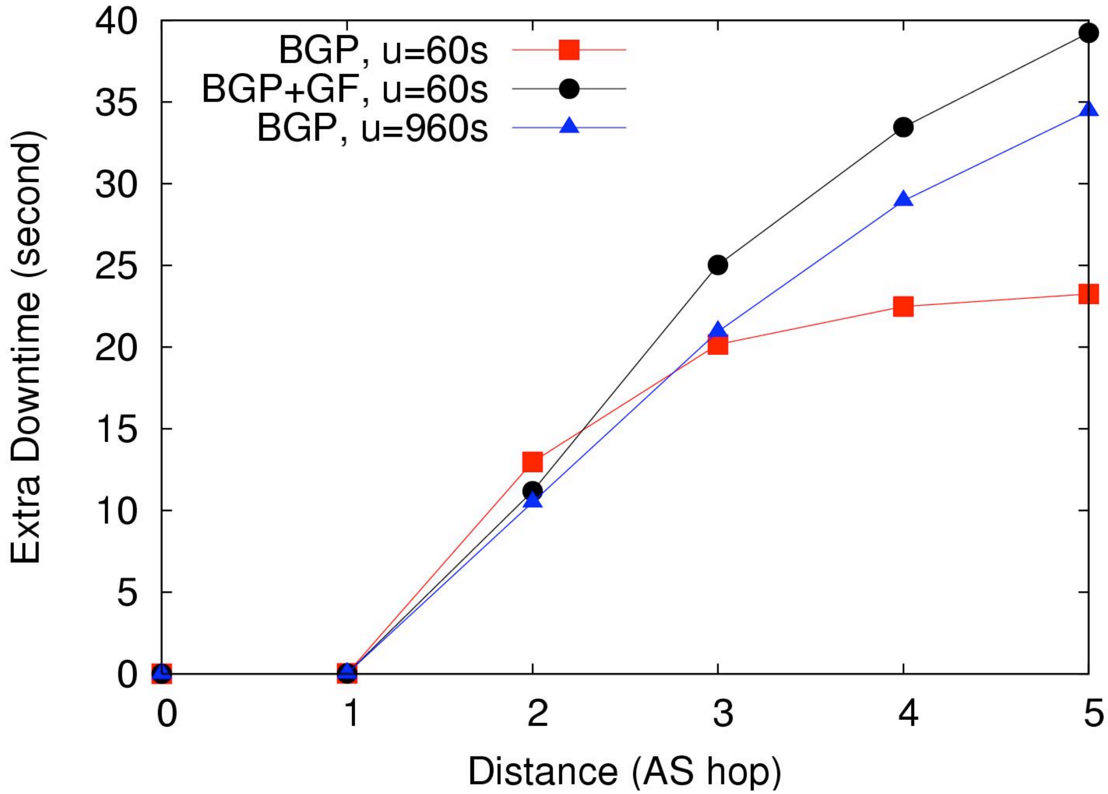
Long T_{down}

- When the destination's failure is short, not adapt to the failure allows the source to avoid the T_{up} delay.
  - $T_{down}(s) > T_{up}(s) + u$
- BGP convergence improvement proposals could have negative impact on packet delivery during transient failures.

Simulation

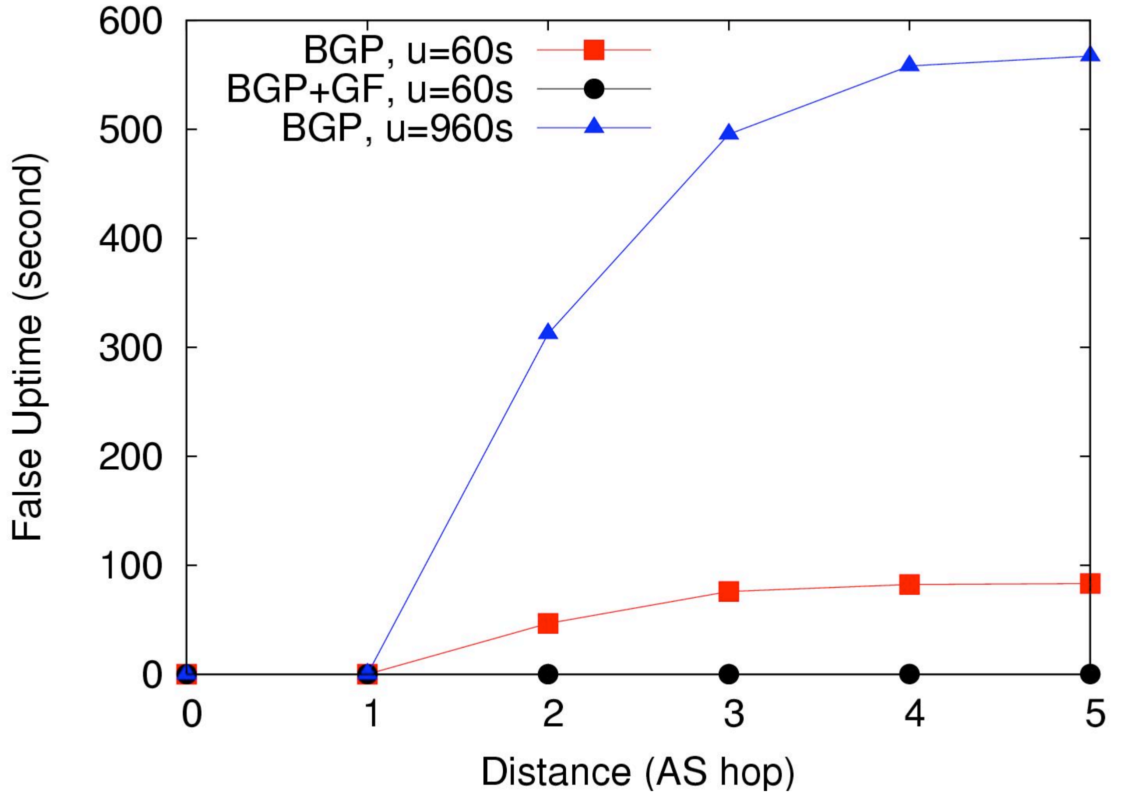
- Use SSFNet to simulate BGP
- Topology of 110 nodes, derived from Internet AS topology.
- Use "Ghost Flushing" to represent convergence improvement proposals.
- Three scenarios:
  - BGP (u = $60s$)
  - BGP + GF (u = $60s$)
  - BGP (u = $960s$)

Extra Downtime



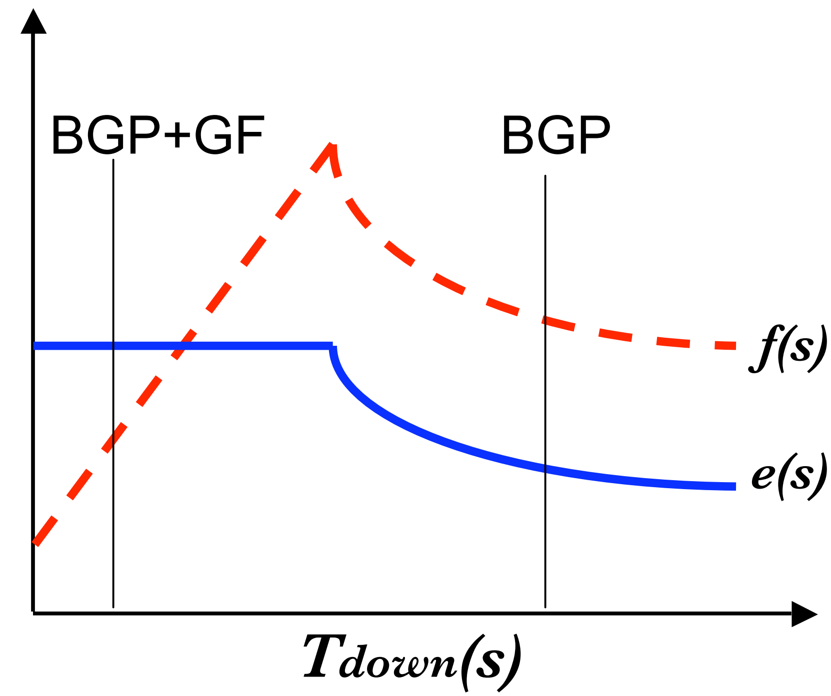
- $T_{down}(s) > T_{up}(s) + u$ for sources more than 3 hops away when $u=60s$.

False Uptime



- BGP-GF has shorter *f(s)* than BGP

Impact of Shortening T_{down}



- By reducing T_{down}, BGP-GF has *mixed* impact on packet delivery, shorter $f(s)$ but longer $e(s)$.

Summary

- Packet delivery is the primary goal of routing.
- Extra downtime and false uptime reflect the impact of routing on packet delivery.
- Current BGP convergence improvement proposals could have negative impact on packet delivery during transient failures.
- Possible Solutions
  - Mask transient failures
  - Shorten T_{up}.

Thanks !

Some Numbers

- 40% failures last less than 1 minute, 80% failures last less than 15 minutes.
  - Iannaccone *et al.* on Sprint network
- BGP T_{down} can be as many as several minutes longer than T_{up}.
  - Labovitz *et al.* from Internet measurement.
- Therefore, the case of $T_{down}(s) > T_{up}(s) + u$ may indeed exist in operational Internet.

Possible Solutions

- Shorten T_{up} too.
  - Need reduce MRAI, which may affect other BGP behaviors.
- Mask transient failures
  - Don’t send withdrawal if the failure is short, but how to predict the failure’s duration?