Hot Potatoes Heat Up BGP Routing

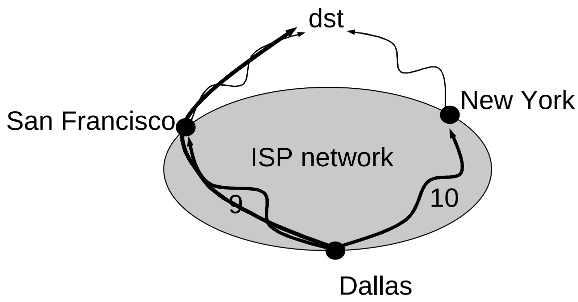
Renata Teixeira
(*UC San Diego*)
http://www-cse.ucsd.edu/~teixeira
with
Aman Shaikh (AT&T), Tim Griffin(Intel), and Jennifer Rexford(AT&T)

30th NANOG – Miami, Florida

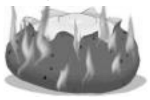
BGP Routing inside an AS

- Most people think…
  - BGP = inter-domain routing
  - Routing inside an AS = OSPF/IS-IS
- But in large transit ASes…
  - Most traffic is routed using BGP
  - IGP changes affect BGP decisions

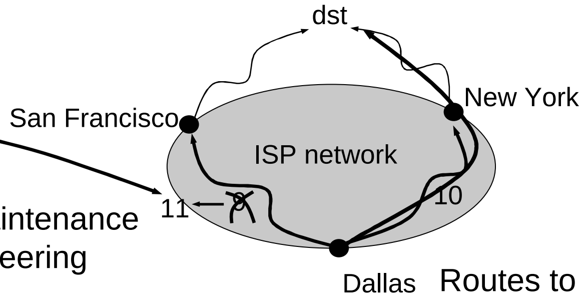
Hot-Potato Routing



Hot-potato routing = route to closest exit point when there is more than one route to destination



Hot-Potato Routing



failure
planned maintenance
traffic engineering

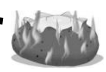
Routes to thousands of destinations switch exit point!!!

Consequences:
- Transient forwarding instability
- Traffic shift
- Inter-domain routing changes

BGP Decision Process

"equally good"

- **Ignore if exit point unreachable**
- Highest local preference
- Lowest AS path length
- Lowest origin type
- Lowest MED (with same next hop AS)
- **Lowest IGP cost to next hop**
- Lowest router ID of BGP speaker

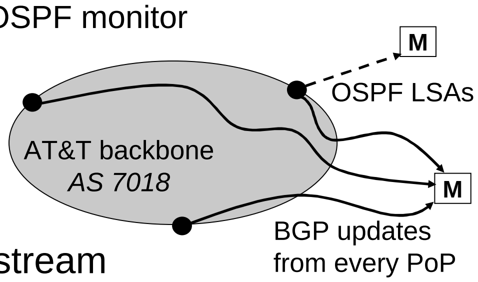


Why Care about Hot Potatoes?

- Large number of routes potentially affected
  - All routes from peers and multi-homed customers
- Understanding of impact in real networks
  - How often hot-potato changes happen in a real network and how many prefixes do they affect?
  - What are the convergence delays?
- Avoiding routing instability
  - Operators: avoid hot-potato changes
  - Vendors: reduce impact of hot-potato changes

Measuring Hot-Potato Routing

- Collect measurement of both protocols
  - BGP monitor and OSPF monitor



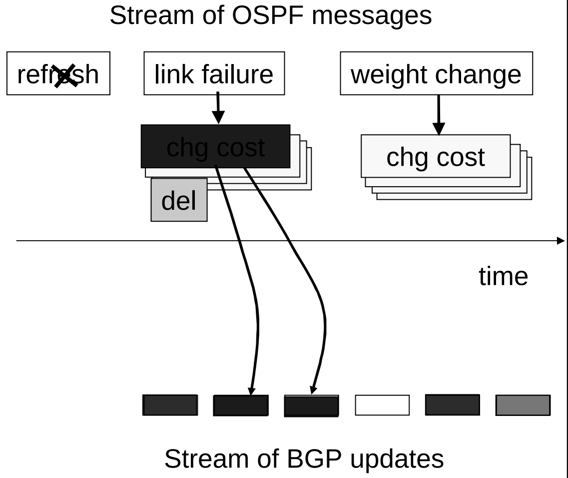
- Pre-process each stream
  - Clustering related messages
- Correlate the two streams of data
  - Match BGP updates with OSPF events

Heuristic for Matching

Transform stream of OSPF messages into routing changes

Match BGP updates with OSPF events that happen close in time

Classify BGP updates by possible OSPF causes



Impact of OSPF/BGP Interaction (June 2003)

- High variability according to location and day
  - Impact on external BGP measurements and customers

location	% updates		days > 10% of hot-potato changes
	min	max	
rich peering	0%	5.3%	0
no peering	0%	46.5%	5

- One LSA can have a big impact

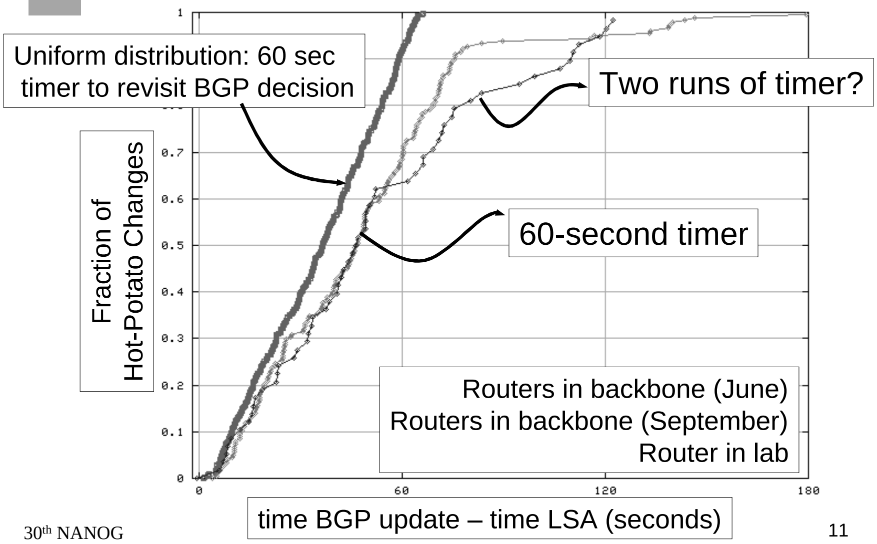
location	LSAs with no impact	prefixes impacted by an LSA
rich peering	95.1%	less than 1%
no peering	94.8%	59%

Delay for BGP Routing Change

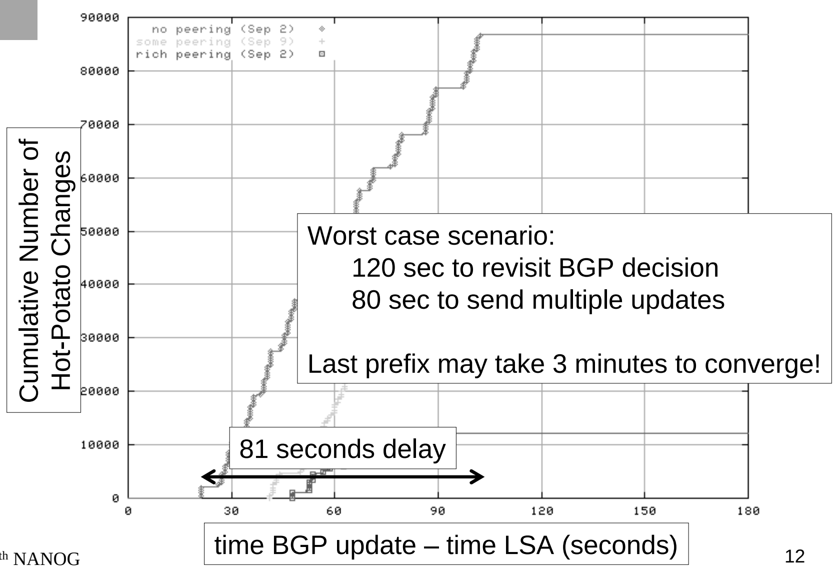
- Router rerunning BGP decision process
  - Implementation specific
    - Timer driven (tunable parameter)
    - Event driven
- Internal BGP route propagation delay
  - Internal BGP hierarchy (route reflectors)
    - Wait for another router to change best route
- Transmitting many BGP messages
  - Latency for transferring the data

Delays very sensitive to router implementation decisions and network design!

Delay for First BGP Update



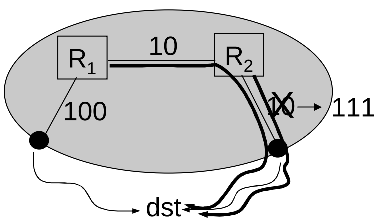
Transferring Multiple Prefixes



Implications on Packet Forwarding

- Forwarding plane convergence
  - FIB update only when both IGP and BGP converge
- Measuring the performance impact
  - Do we need new measurement techniques?

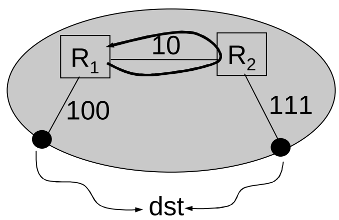
Forwarding Plane Convergence



1 - Scan process runs in R_2

2 - R_2 starts using R_1 to reach dst

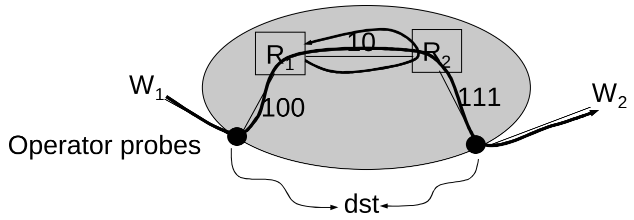
3 - R_1's scan process can take up to 60 seconds to run



Packets to dst may be caught in a loop for 60 seconds!

Detecting Loops is Hard

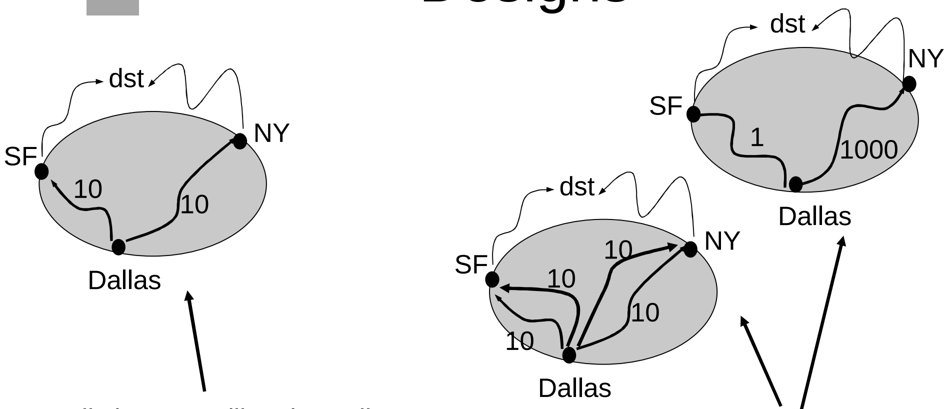
- Active probing of forwarding path
  - Probing machines just have one exit point



Avoiding Unnecessary Hot-Potato Changes

- Router connectivity to exit points
  - Likelihood of hot-potato routing changes
- Cost in/out of links during maintenance

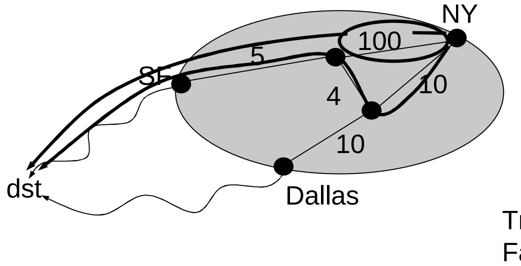
Comparison of Network Designs



Small changes will make Dallas switch exit points to dst

More robust to intra-domain routing changes

Careful Cost in/out Links



Traffic is more predictable
Faster convergence
Less impact on neighbors

Conclusions

- Hot potatoes can affect a lot of prefixes and have high convergence delays
- Vendors:
  - Event-driven implementations
  - Understand network-wide effect
- Operators:
  - Tuning timer
  - Network design
  - Maintenance practices

Future Work

- Impact of the IGP-triggered BGP updates
  - Sudden shifts in the flow of traffic
  - Forwarding loops during convergence
  - Externally visible BGP routing changes