Delayed Internet Routing

Convergence

Labovitz00

Bill Cheng

http://merlot.usc.edu/cs551-f12

But poorly understood

BGP widely deployed in the Internet

Context

ISP's don't tell you what they are doing

Convergence time takes longer than we expected

Simulation to study worst case behavior

Robustness

Why Is Convergence Important?

Internet fail-over times are in 10s of seconds

Open problem: how can Internet do much better?

Robustness

Key Idea

measurement

similation

study and understand BGP convergence time

observe 2.5 minute convergence times (ex: longer than expected)

Internet failure times takes longer than we expected

ISP1

ISP2

ISP3

ISP4

ISP5

ISP6

Stub AS

Fault Injection Server

Upstream

ISP2

ISP3

ISP4

ISP5

ISP6

Upstream

ISP1

BGP Fault

BGP Fault

Stub AS

RouteViews

Data

Connection

Probe

BGP

BGP

BGP

BGP

ICMP

echos

Internet-scale experimentation

What kind of complexities/errors can arise?

In general, good news to propagate fast, bad news to propagate slow.

Troubleshooting time to switch from a longer route to a shorter one.

Troubleshooting time to switch from a shorter route to a longer one.

TR: How long does it take for a node to fail-over?

Measurement:

Troubleshooting, reports, fail-over only for own talk, not other

Internetartin ability fails across Internet

Tdown:

time for good news to propagate

Measure:

failures, repairs, fail-over

Analysis: helps understand worst case bounds

Tshort:

time to switch from a longer route to a shorter one

Simulation to study worst case behavior

Tlong:

time to switch from a shorter route to a longer one
Observed Convergence Latency

Cumulative Percentage of Events

Less than half of $T_{down}$ events converge within two minutes

Long tailed distribution (up to 15 minutes)

$T_{up}$, $T_{short}$, $T_{long}$

Failure

Short $\rightarrow$ Long Fail-Over

New Route

Long $\rightarrow$ Short Fail-over

$T_{up}$ and $T_{short}$

In general, want bad news to travel fast and good news to travel slowly

No correlation between network distance (latency, router, or AS hops) and convergence times

Why is long convergence bad?

Impact on Traffic

Percentage Packet Loss

One Minute Bins Before and After Fault

Fault

0

5

10

15

20

25

30

35

Why does loss go up?

Because there are route loops in the net causing packet drops.

Figure 4a of [Labovitz00a]

Simulate BGP

Model one router per AS

How To Tell What’s Going On?

Simple model that captures key details

Synchronous processing via global queue

Ignore latency

Assume all routers are identical

Model one router per AS

Shim routing

BGP Convergence Example

AS0

R

via 3

via 0 3

via 1 3

via 2 3

AS1

B

AS2

B

B

R

R

R

R

AS0

AS1

AS2

What’s Going On?

BGP can try all paths of length 2, then 3, then 4, etc.

and it takes a long time with BGP to figure out that none of them work

There are many possible routes (indirect through other ASs)

Even with MinRouteAdver timers it still can take $O(n)$ steps (13 steps vs. 48 steps originally)

Why is long convergence bad?

Good news to travel slowly

In general, want bad news to travel fast and less than half of true events converge within two minutes

Correlation of Events

Observed Convergence Latency

Other Observations
Why Does This Happen?

O(n!) such paths in BGP, the theoretical worst case occurs when all possible alternate paths are explored, explaining pathological convergence time.

What About MinRouteAdver?

Designed to limit updates and to encourage aggregation by delaying announcements, routers figure out the pain sooner, so that the path length and therefore are quickly accepted.

Discussion

Written when the Internet was a large infrastructure and some problems were only hypothetical. But until then, BGP’s path length was a large infrastructure with the Internet.

Other Observations

 andraught and can preempt some of the thrashing. Tlong actually sometimes goes quicker if it’s not long enough. Thing deathy sometimes goes quicker if it’s not long enough. Tdown converges slowly because BGP tries hard to find the paths and then the paths are quickly accepted. Tup-Tdown converges quickly because they shorten path length.

Does This Explain Measurements?

Could do loop detection at sender side and not just receiver. Explain pathologic convergence time. (C)in (C)sect 5.2, race the input. How does it affect convergence? Explanation pathologic convergence time. It’s not long enough. Could do loop detection at sender side and not just receiver.