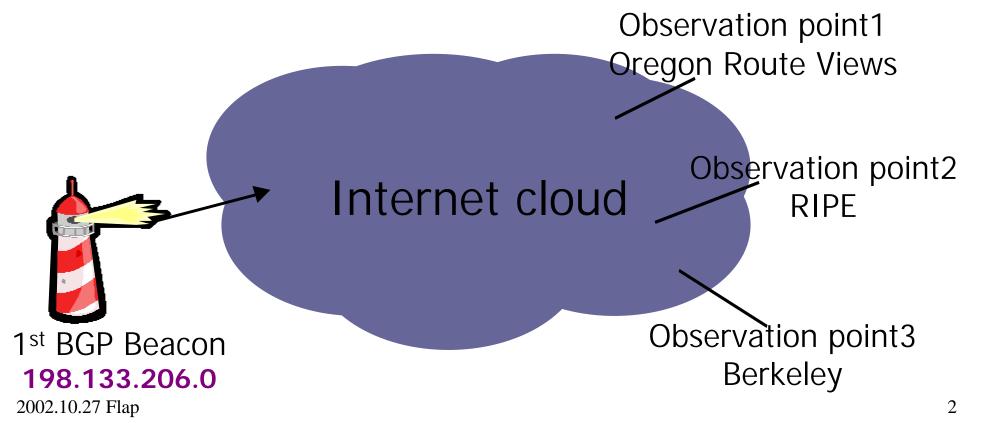# Route flap damping: harmful?

NANOG / Eugene

2002.10.28

Randy Bush (IIJ)

Tim Griffin (AT&T Research)

Z. Morley Mao (UC Berkeley)
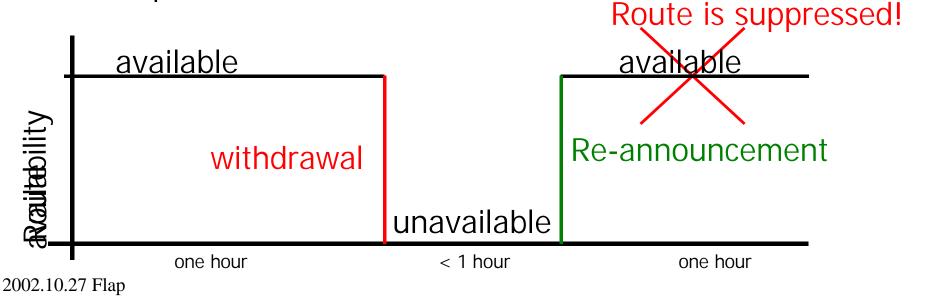
<http://psg.com/~randy/021028.zmao-nanog>

Verified using BGP Beacons

- BGP Beacon:
  - A prefix that is announced and withdrawn at **well-known** times

Observation point1
Oregon Route Views

Internet cloud

Observation point2
RIPE

Observation point3
Berkeley

1st BGP Beacon
198.133.206.0

2002.10.27 Flap

Transient instability

- Router reboot
  - Due to circuit or software upgrade, etc.
- A single link flap
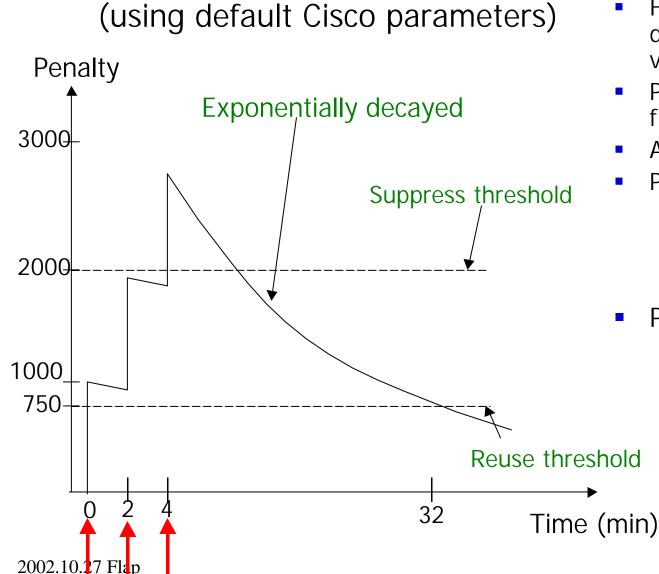 - Due to network congestion, link connectivity problems etc.

Route is suppressed!

available   available

Availability / Route

withdrawal   unavailable   Re-announcement

one hour   < 1 hour   one hour

What is route flap damping?

- RFC2439
  - Supported by all major router vendors
  - Believed to be widely deployed
  - Responsible for Internet stability?

- Goals:
  - Reduce router processing load due to instability
  - Do not sacrifice convergence times for well-behaved routes (!?)

How does route flap damping work?

(using default Cisco parameters)

- For each peer, per destination, keep a penalty value
- Penalty increases for each flap
- A flap is a route change
- Penalty decays exponentially

Penalty

3000

Exponentially decayed

Suppress threshold

2000

$$P(t') = P(t)e^{-1(t'-t)}$$

- Parameters:
  - Fixed:
    Penalty increment
  - Configurable:
    half-life, suppress-, reuse-threshold, max suppressed time

1000

750

Reuse threshold

0   2   4   32

Time (min)

Router vendor default values

Parameter	Cisco	Juniper
Withdrawal penalty	1000	1000
Re-advertisement penalty	0	1000
Attributes change penalty	500	500
Suppress threshold	2000	3000
Half-life (min)	15	15
Reuse threshold	750	750
Max suppress time (min)	60	60

- Cisco
  - Three flaps can suppress route
- Juniper
  - Minimum four flaps to suppress route
- Example:
  - Three flaps with 2 min interval
  - Cisco: suppress on the third flap for more than 28 minutes

Cascaded withdrawals! (2)

- Peer: 212.47.190.1, AS=9177 from RIPE
- In response to WD-beacon at 18:00, Aug 10th.
- Using Cisco setting + RIPE229 recommendation

Time 8/10	A/W	ASPath	Penalty
18:00:15	A	9177 3320 1 2914 3130 3927	500
18:00:41	A	9177 6730 5400 2914 3130 3927	990
18:01:41	A	9177 3320 2914 3130 3927	1445
18:03:06	A	9177 3320 1239 2914 3130 3927	1853
18:03:35	W		2812
18:04:03	A	9177 6730 5400 2914 3130 3927	2752
18:04:31	W		3694

Above suppress threshold

Cascaded withdrawals! (1)

- Peer: 213.200.87.254, AS=3257 from RouteViews
- In response to WD-beacon at 01:00, Aug 20th.
- Using Cisco setting + RIPE229 recommendation
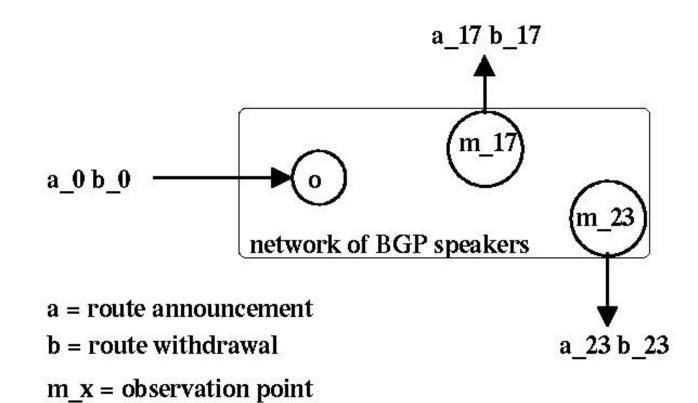 - (Note: first 2 announcements differ in community attributes)

Time 8/20	A/W	ASPath	Penalty
01:00:16	A	3257 1299 2914 3130 3927	500
01:00:47	A	3257 1299 2914 3130 3927	988
01:00:50	W		1985
01:00:50	A	3257 1299 4200 2914 3130 3927	1985
01:01:13	A	3257 1299 701 2914 3130 3927	2451
01:02:05	W		3354

Above suppress threshold
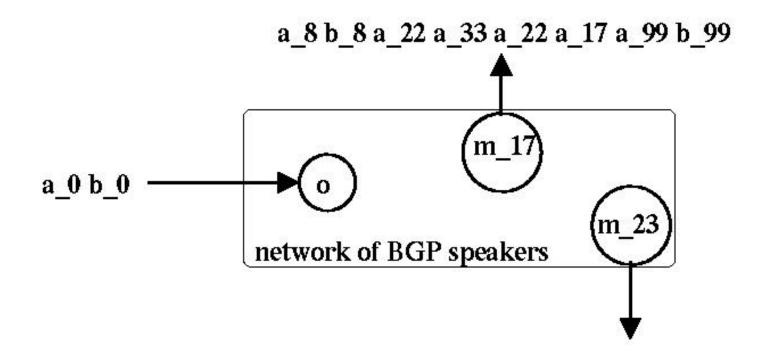
Why does this happen?

- BGP is a path vector protocol
  - Explores alternate routes before withdrawal
  - Topology dependent
- Delay in messages due to variations in
  - MinRouteAdver timer values
  - Propagation delays
  - Router processing overhead
- Route flap damping parameter setting
  - Cisco/Juniper punishes virtually all route changes
  - Default setting and RIPE-229 recommendation are too aggressive

The Naïve View

a_17 b_17

m_17

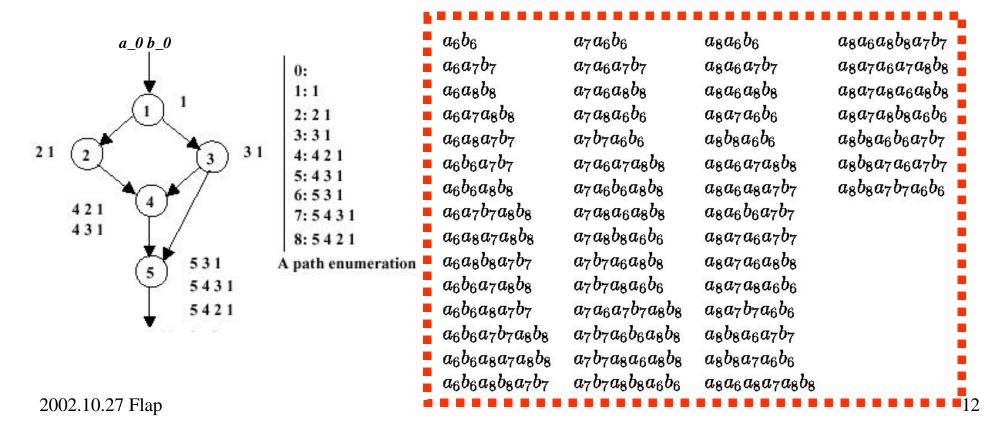a_0 b_0 → o

network of BGP speakers

m_23

a_23 b_23

a = route announcement

b = route withdrawal

m_x = observation point

... reality is a wonderful thing

a_8 b_8 a_22 a_33 a_22 a_17 a_99 b_99

m_17

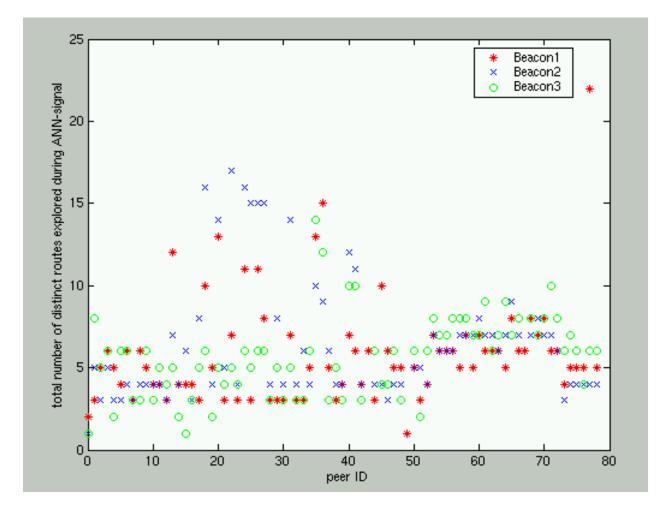a_0 b_0 → o

m_23

network of BGP speakers

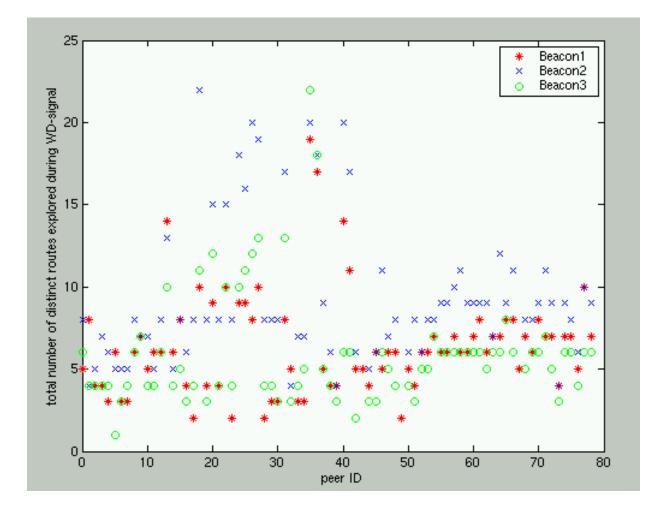# No wonder I'm going crazy trying to interpret those BGP updates ....

It is easy to construct a 5 node BGP system where a simple Announce/Withdraw signal (a_0 b_0) at one node can produce any of these 52 output signals at another...

$a\_0\ b\_0$

0:
1: 1
2: 2 1
3: 3 1
4: 4 2 1
5: 4 3 1
6: 5 3 1
7: 5 4 3 1
8: 5 4 2 1

A path enumeration

1

2 1

3 1

4 2 1
4 3 1

5 3 1

5 4 3 1

5 4 2 1

$a_6 b_6$	$a_7 a_6 b_6$	$a_8 a_6 b_6$	$a_8 a_6 a_8 b_8 a_7 b_7$
$a_6 a_7 b_7$	$a_7 a_6 a_7 b_7$	$a_8 a_6 a_7 b_7$	$a_8 a_7 a_6 a_7 a_8 b_8$
$a_6 a_8 b_8$	$a_7 a_6 a_8 b_8$	$a_8 a_6 a_8 b_8$	$a_8 a_7 a_8 a_6 a_8 b_8$
$a_6 a_7 a_8 b_8$	$a_7 a_8 a_6 b_6$	$a_8 a_7 a_6 b_6$	$a_8 a_7 a_8 b_8 a_6 b_6$
$a_6 a_8 a_7 b_7$	$a_7 b_7 a_6 b_6$	$a_8 b_8 a_6 b_6$	$a_8 b_8 a_6 b_6 a_7 b_7$
$a_6 b_6 a_7 b_7$	$a_7 a_6 a_7 a_8 b_8$	$a_8 a_6 a_7 a_8 b_8$	$a_8 b_8 a_7 a_6 a_7 b_7$
$a_6 b_6 a_8 b_8$	$a_7 a_6 b_6 a_8 b_8$	$a_8 a_6 a_8 a_7 b_7$	$a_8 b_8 a_7 b_7 a_6 b_6$
$a_6 a_7 b_7 a_8 b_8$	$a_7 a_8 a_6 a_8 b_8$	$a_8 a_6 b_6 a_7 b_7$	
$a_6 a_8 a_7 a_8 b_8$	$a_7 a_8 b_8 a_6 b_6$	$a_8 a_7 a_6 a_7 b_7$	
$a_6 a_8 b_8 a_7 b_7$	$a_7 b_7 a_6 a_8 b_8$	$a_8 a_7 a_6 a_8 b_8$	
$a_6 b_6 a_7 a_8 b_8$	$a_7 b_7 a_8 a_6 b_6$	$a_8 a_7 a_8 a_6 b_6$	
$a_6 b_6 a_8 a_7 b_7$	$a_7 a_6 a_7 b_7 a_8 b_8$	$a_8 a_7 b_7 a_6 b_6$	
$a_6 b_6 a_7 b_7 a_8 b_8$	$a_7 b_7 a_6 b_6 a_8 b_8$	$a_8 b_8 a_6 a_7 b_7$	
$a_6 b_6 a_8 a_7 a_8 b_8$	$a_7 b_7 a_8 a_6 a_8 b_8$	$a_8 b_8 a_7 a_6 b_6$	
$a_6 b_6 a_8 b_8 a_7 b_7$	$a_7 b_7 a_8 b_8 a_6 b_6$	$a_8 a_6 a_8 a_7 a_8 b_8$	

Max no. updates for ANN-signal

Max no. updates for WD-signal

Multihoming Won't Save You

- Damping is not per path
- Damping is per prefix!
- Sure, multi-homing gives you more paths
- But damping just **loves** more paths

Preliminary ideas to improve Route flap damping

- Change the constants
  - Descrease suppression threshold
  - Decrease penalty increment, especially for attribute changes
- Change the Algorithms
- Do we really need damping at all?
- Note that damping
  - is opaque
  - happens on the other side of the internet
  - and is hard to debug

Existing BGP Beacons

Prefix	Source AS	Start date	Upstream Providers	Beacon Host
198.133.206.0/24	3927	8/10/02	AS2914, AS1	Randy Bush
192.135.183.0/24	5637	9/4/02	AS3701, AS2914	David Meyer
203.10.63.0/24	1221	9/25/02	AS1221	Geoff Huston
198.32.7.0/24	3944	10/24/02	AS2914, AS8001	Andrew Partan

- Announced and withdrawn with a fixed period (2 hours) between updates
- 1st daily ANN: 3:00AM GMT
- 1st daily WD: 1:00AM GMT

Reference

- RIPE-229: "RIPE Routing-WG Recommendations for Coordinated Route-flap Damping Parameters", 2001
- RFC 2439: "BGP Route Flap Damping", 1998
- Randy Bush, Tim Griffin, Z. Mao, "Route flap Damping: harmful?" RIPE 43, 2002
- Tim Griffin, "What is the sound of one route flapping" Dartmouth talk slides, June 2002
- C. Labovitz, G. R. Malan, F. Jahanian, "Internet Routing Instability", TON 1998
- C. Labovitz, R. Malan, F. Jahanian, "Origins of Internet Routing Instability", Infocom 1999

More references

- C. Labovitz, A. Ahuja, F. Jahanian, "Experimental Study of Internet Stability and Wide-Area Network Failures", FTCS 1999

- C. Labovitz, A. Ahuja, A. Bose, F. Jahanian, "Delayed Internet Routing Convergence" Sigcomm 2000

- C. Labovtiz, A. Ahuja, R. Wattenhofer, S. Venkatachary, "The Impact of Internet Policy and Topology on Delayed Routing Convergence", Infocom 2001

- Z. Mao, R. Govindan, G. Varghese, R. Katz "Route Flap Damping Exacerbates Internet Routing Convergence" Sigcomm 2002