Routing in the Future Internet

Marcelo Yannuzzi

Graduate Course (Slideset 6)
Institute of Computer Science
University of the Republic (UdelaR)

August 24th and 27th 2012, Montevideo, Uruguay

Department of Computer Architecture
Technical University of Catalonia (UPC), Spain

Institute of Computer Science
University of the Republic (UdelaR), Uruguay

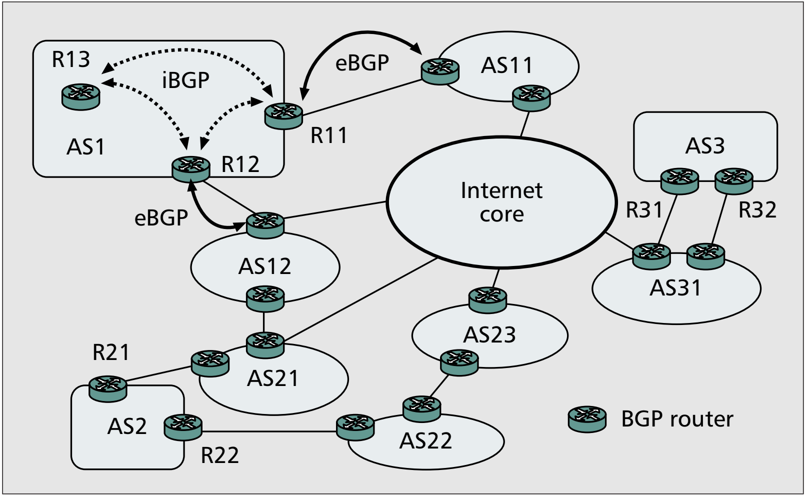
Outline

1. iBGP, eBGP, and Route Reflectors.
2. Case study: Japanese Earthquake in 2011.
3. Interdomain Traffic Engineering.
4. Research challenges in interdomain routing.
   - Rounting convergence.
   - Outline of the scalability issues.
   - Churn and its impact on the DFZ.
   - Routing Policies: policy disputes, etc.
   - Traffic Engineering: solutions and research challenges.
   - Routing Security.

Outline

1. **iBGP, eBGP, and Route Reflectors.**
2. Case study: Japanese Earthquake in 2011.
3. Interdomain Traffic Engineering.
4. Research challenges in interdomain routing.
   - Rounting convergence.
   - Outline of the scalability issues.
   - Churn and its impact on the DFZ.
   - Routing Policies: policy disputes, etc.
   - Traffic Engineering: solutions and research challenges.
   - Routing Security.

BGP flavors: iBGP and eBGP



- Source: M. Yannuzzi et al., "Open Issues in Interdomain Routing: A Survey," IEEE Network, Nov./Dec. 2005.

iBGP and Route Reflectors

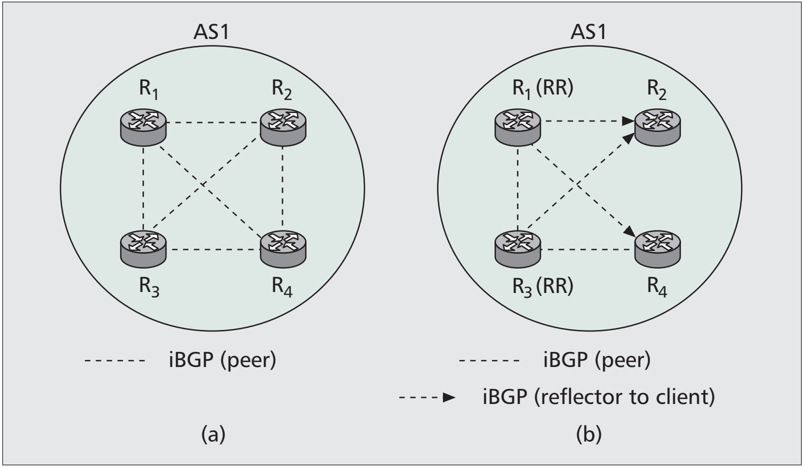


Figure 1. *Different i-BGP topologies: a) full-mesh i-BGP; b) i-BGP with route reflection.*

- Source: J. H. Park et al., “BGP Route Reflection Revisited,” IEEE Communications Magazine, July 2012.

Route Reflectors (cont.)

Basic Operation of RRs...

- Avoids the need of fully-meshed iBGP sessions, offering:
  - $\frac{N(N-1)}{2} = O(N^2) \quad \rightarrow \quad \frac{K(K-1)}{2} + \sum_{i=1}^{K} C_i = O(N)$

    N: Number of BGP routers in the AS
    K: number of RR in the AS (note the full-mesh of RRs for redundancy)
    C_i: number of client iBGP routers connected to the i-th RR ($C_i < N$)
- RRs forward reachability information learned from an i-BGP speaker to another i-BGP speaker.
- Since BGP messages travel more than a single i-BGP hop inside the AS, it is possible to create loops.
- 2 new attributes are added to BGP update messages: ORIGINATOR_ID and CLUSTER_LIST.

Route Reflectors (cont.)

Advantages of using RRs...

- From $O(N^2)$ to $O(N)$ iBGP sessions
- Reduces OPEX
- Reduces RIBs's sizes (RIB-in, Loc-RIB, and RIB-out)
  - **RIB-in:** each router maintains a RIB-in for each neighbor, which contains unprocessed routing information (i.e., before applying import policies). The total size with iBGP is $(N-1)$ x $\overline{p}_{iBGP}$ ($\overline{p}_{iBGP}$: avg. number of prefixes per neighbor). Whereas with RRs: K x $\overline{p}_{RR}$
  - **Loc-RIB:** stores the best route for each possible destination (i.e., after applying import policies across each RIB-in and running the BGP decision process).
  - **RIB-out:** contains the set of routes to be advertised to each neighbor after applying export policies (i.e., output filters)....note that the export policy to i-BGP neighbors is typically the same and that clients only need to keep K RIB-out internally.
- Reduces churn
- .....but in practice things are not that simple....

Route Reflectors (cont.)

Known issues...

- RR may:
  - Decrease the network's robustness against failures
  - Introduce delayed routing convergence
  - Reduce path diversity within the AS
  - Adopt suboptimal routes
  - And even cause data forwarding loops

Route Reflectors (cont.)

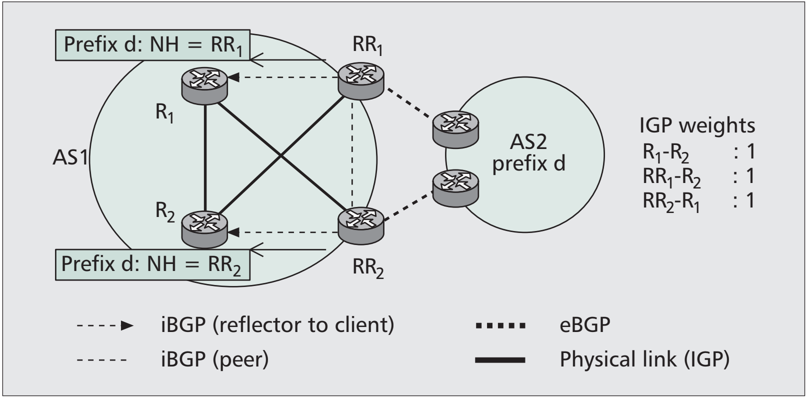


Figure 2. *Route reflection with data forwarding loop.*

- Source: J. H. Park et al., "BGP Route Reflection Revisited," IEEE Communications Magazine, July 2012.

Route Reflectors (cont.)

- Reduced path diversity, delayed convergence, and suboptimal routes

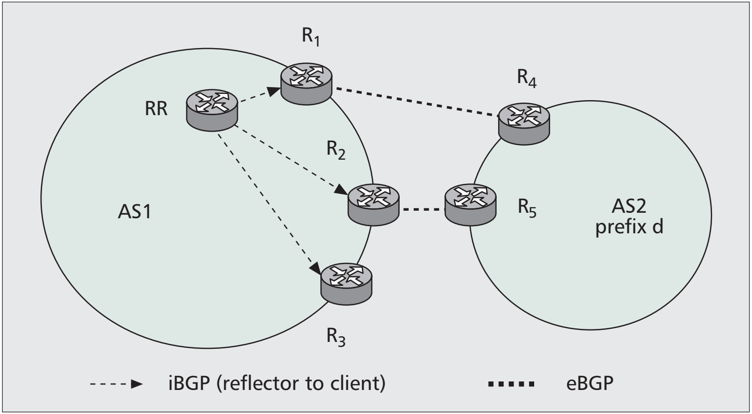


Figure 3. *RR chooses* its *best route.*

- Source: J. H. Park et al., "BGP Route Reflection Revisited," IEEE Communications Magazine, July 2012.

Route Reflectors (cont.)

- Note that R1 and R2 will use the routes through R4 and R5, respectively, since routes learned via e-BGP are typically preferred over routes learned from iBGP. However, R3 will be constrained by the RR selection.

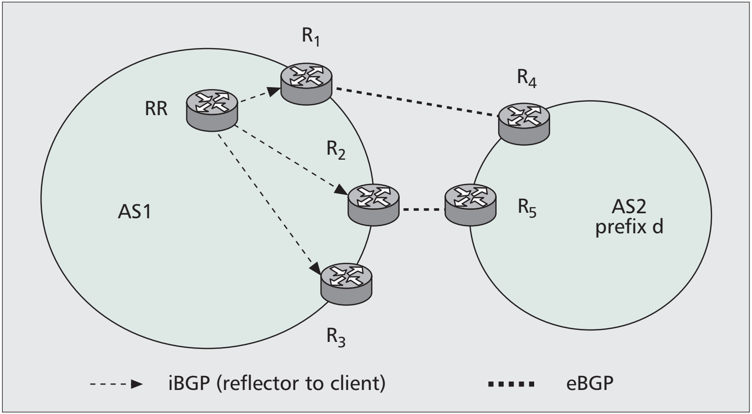
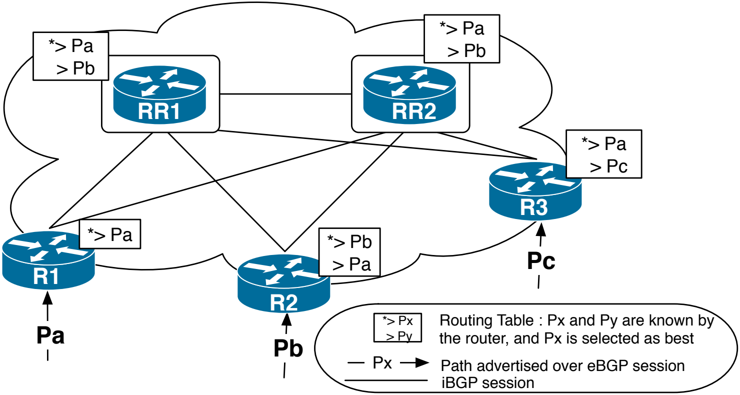


Figure 3. *RR chooses* its *best route.*

- Source: J. H. Park et al., “BGP Route Reflection Revisited,” IEEE Communications Magazine, July 2012.

Route Reflectors (cont.)

- Anbother example...note that upon failure of path Pa, router R1 cannot reach destination *d* anymore and will drop packets until the RR advertise Pb. R1 will also send eBGP withdraws on its eBGP sessions.



- Source: V. Van den Schrieck et al. “BGP Add-Paths: The Scaling/Performance Tradeoffs,” IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 28, NO. 8, OCTOBER 2010.

Route Reflectors (cont.)

- Coping with the problems through placement and RR hierarchy....though this comes at the cost of increased hop distance and path diversity reductions...

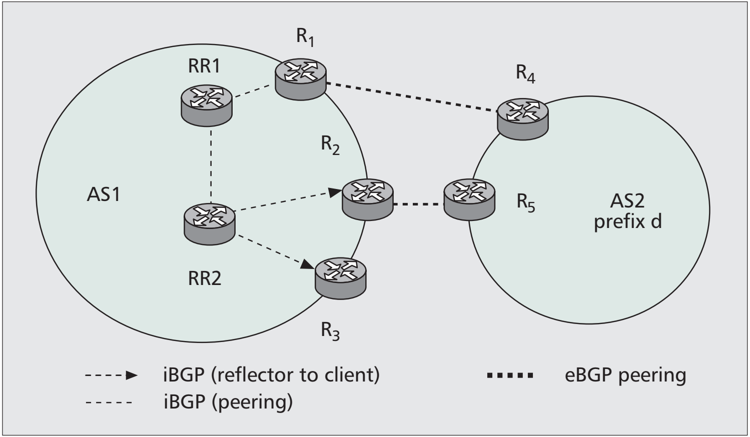
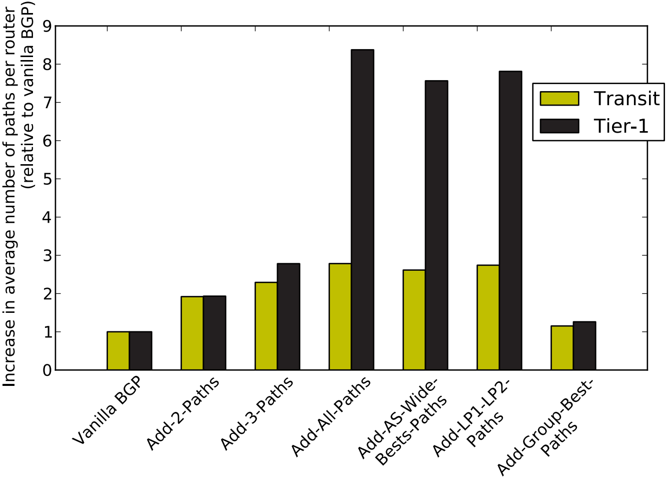


Figure 4. *POP based route reflection.*

- Source: J. H. Park et al., "BGP Route Reflection Revisited," IEEE Communications Magazine, July 2012.

BGP Add-Paths

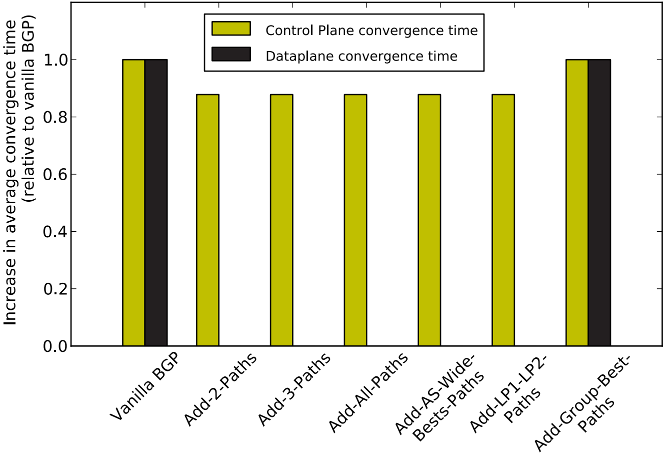
- Anbother approach: D. Walton et al. “Advertisement of Multiple Paths in BGP," IETF draft-ietf-idr-add-paths-07.txt, June 2012.



- Source: V. Van den Schrieck et al. “BGP Add-Paths: The Scaling/Performance Tradeoffs,” IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 28, NO. 8, OCTOBER 2010.

BGP Add-Paths (cont.)

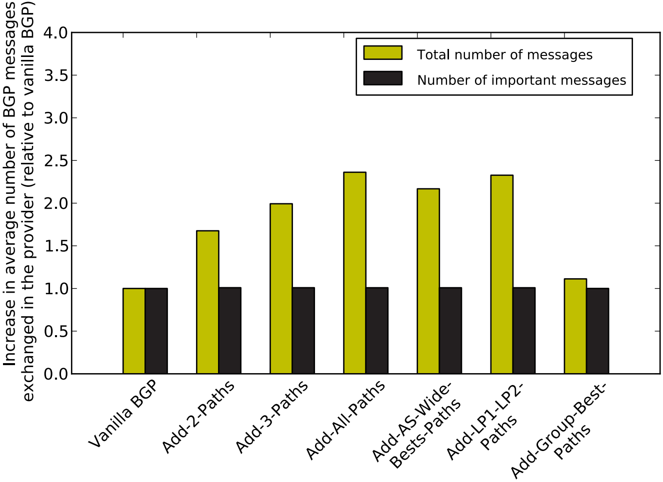
- Anbother approach: D. Walton et al. “Advertisement of Multiple Paths in BGP," IETF draft-ietf-idr-add-paths-07.txt, June 2012.



- Source: V. Van den Schrieck et al. “BGP Add-Paths: The Scaling/Performance Tradeoffs,” IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 28, NO. 8, OCTOBER 2010.

BGP Add-Paths (cont.)

- The reductions in eBGP churn come at the cost of an increase of the iBGP churn on non-best paths...



- Source: V. Van den Schrieck et al. “BGP Add-Paths: The Scaling/Performance Tradeoffs,” IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 28, NO. 8, OCTOBER 2010.

Outline

1. iBGP, eBGP, and Route Reflectors.
2. **Case study: Japanese Earthquake in 2011.**
3. Interdomain Traffic Engineering.
4. Research challenges in interdomain routing.
    - Rounting convergence.
    - Outline of the scalability issues.
    - Churn and its impact on the DFZ.
    - Routing Policies: policy disputes, etc.
    - Traffic Engineering: solutions and research challenges.
    - Routing Security.

Japanese Earthquake in 2011

- > 15,000 people dead and > 4,000 were missing even after 6 months of the disaster (90% due to the tsunami)

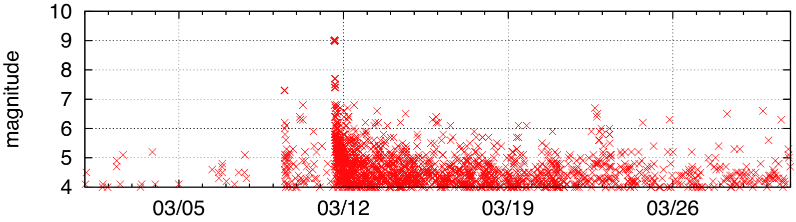


Figure 1: Earthquakes larger than Magnitude 4 in Japan for March 2011

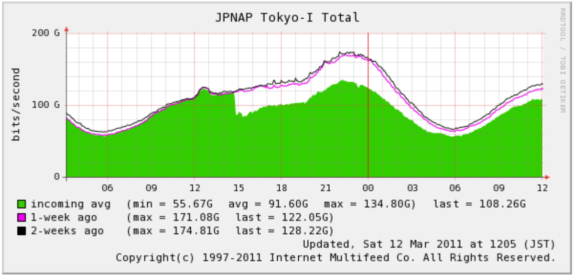
- Source: K. Cho et al. “The Japan Earthquake: the impact on traffic and routing observed by a local ISP,” ACM SWID 2011, December 2011.

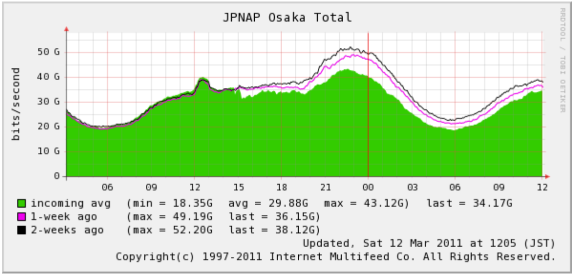
Japanese Earthquake in 2011 (cont.)

Impact on NTT

- 1.5 million circuits for fixed-line services
- 6,700 pieces of base station equipment
- 15,000 circuits for corporate data communication services.
- 65,000 telephone poles were flooded or collapsed
- 6,300km of aerial cables were lost.
- Voice calls: capacity overloads due to a surge in calls.
- ...however, the Internet was impressively resilient to the disaster.

Japanese Earthquake in 2011 (cont.)





- Source: K. Cho et al. “The Japan Earthquake: the impact on traffic and routing observed by a local ISP,” ACM SWID 2011, December 2011.

Traffic on JP-US cables of IIJ (damaged and rerouted)

- Source: K. Cho et al. “The Japan Earthquake: the impact on traffic and routing observed by a local ISP,” ACM SWID 2011, December 2011.

Link Failures and Restoration

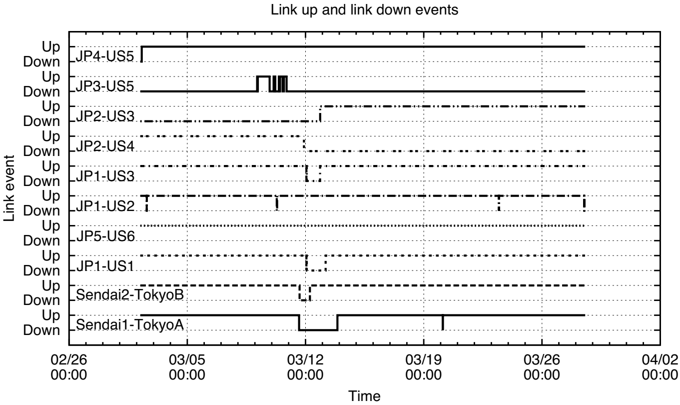
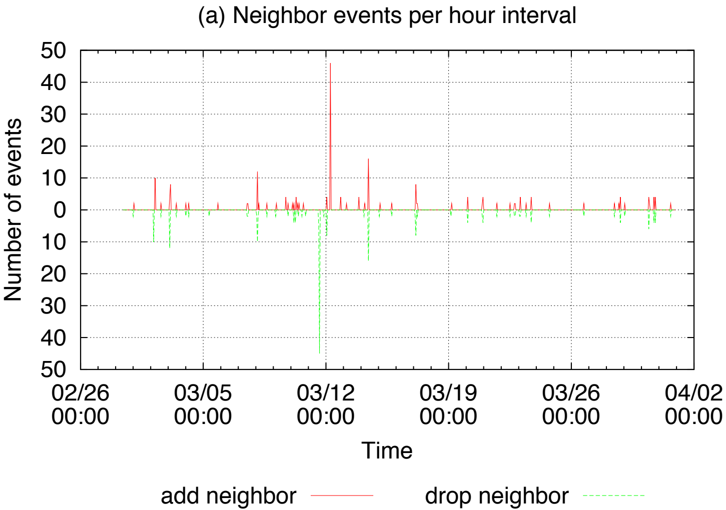


Figure 5: Quake related link failure and restoration times

- Source: K. Cho et al. “The Japan Earthquake: the impact on traffic and routing observed by a local ISP,” ACM SWID 2011, December 2011.

OSPF add and drop events



- Source: K. Cho et al. “The Japan Earthquake: the impact on traffic and routing observed by a local ISP,” ACM SWID 2011, December 2011.

BGP withdrawals seen by a neighbor AS

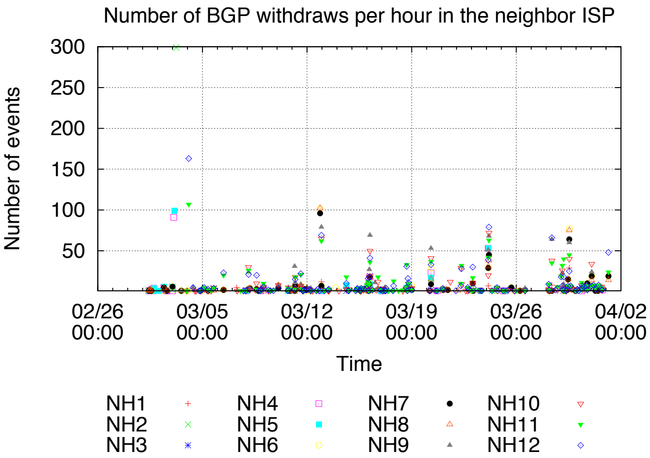


Figure 7: BGP withdrawals for our ISP in a neighboring ISP

- Source: K. Cho et al. "The Japan Earthquake: the impact on traffic and routing observed by a local ISP," ACM SWID 2011, December 2011.

Outline

1. iBGP, eBGP, and Route Reflectors.
2. Case study: Japanese Earthquake in 2011.
3. **Interdomain Traffic Engineering.**
4. Research challenges in interdomain routing.
   - Rounting convergence.
   - Outline of the scalability issues.
   - Churn and its impact on the DFZ.
   - Routing Policies: policy disputes, etc.
   - Traffic Engineering: solutions and research challenges.
   - Routing Security.

Traffic Engineering goals differ...

Transit Providers

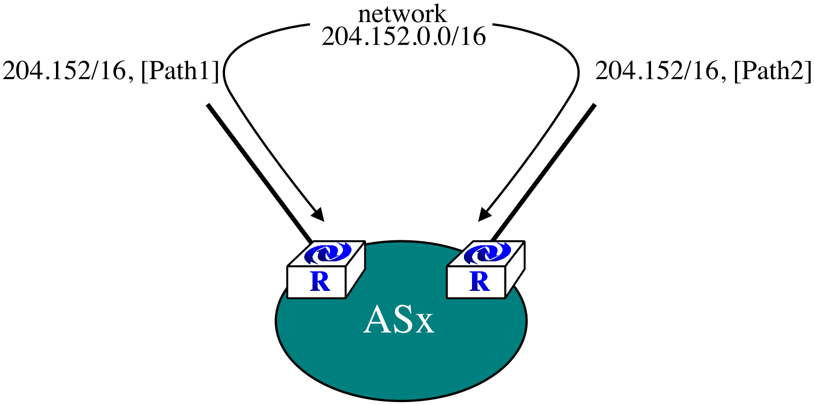
- Optimize the distribuition and exchange of large traffic volumes
- Performance differentiation for premium customers while exploiting economy of scale
- Even different goals depending on carrier's size and market niche
- Typically: considerable overprovisioning, min-max optimization cycles, optics penetration and keen on cross-layer aspects, and a lot of rule of the thumb (no real clue about the Traffic Matrix)

Non-transit Domains (e.g., enterprises)

- > 80% of the ASs in the Internet ... that means more the 32,000 ASs
- Typically: scarce overprovisioning though with sufficient redundancy, performance optimization (in general that means low delay with high service availability), and clearly, reduce as much as possible the Internet's costs.

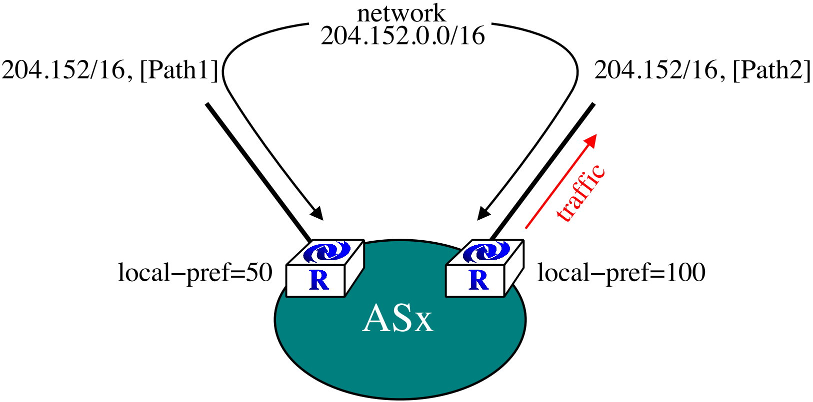
Traffic Engineering
Transit Providers
IP Layer

Egress Traffic Engineering



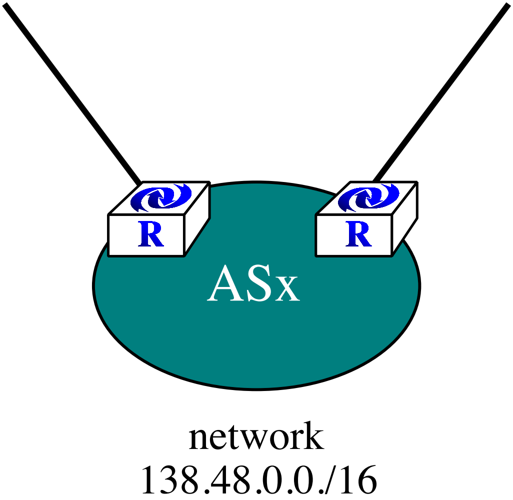
- Source: Presentation from B. Quoitin at QofIS 2002.

Egress Traffic Engineering (cont.)



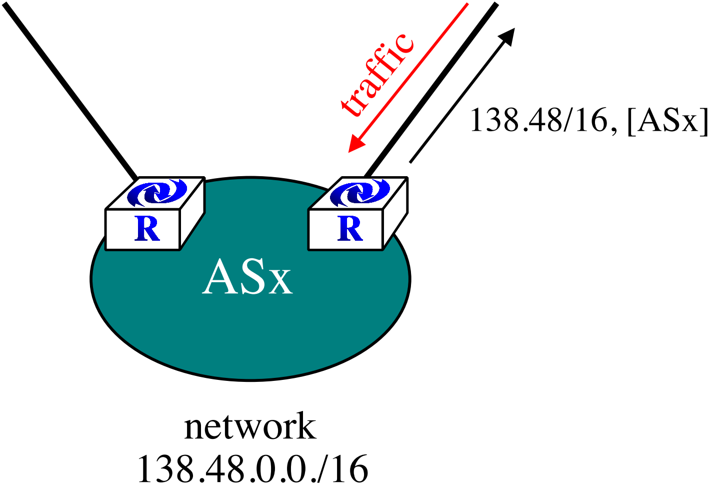
- Source: Presentation from B. Quoitin at QofIS 2002.

Ingress Traffic Engineering



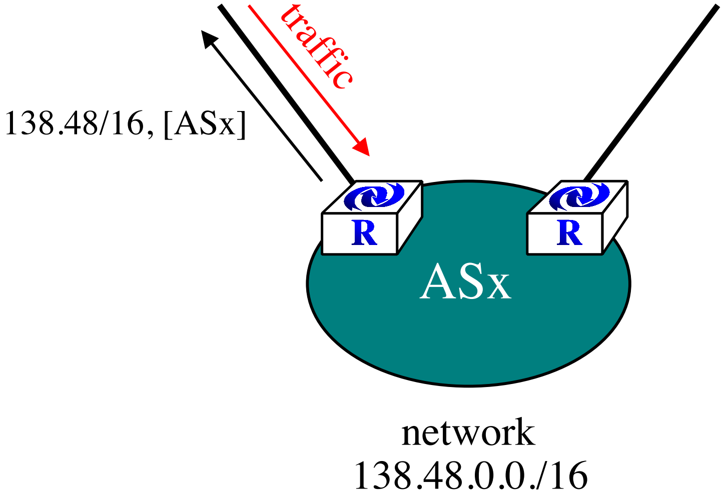
- Source: Presentation from B. Quoitin at QofIS 2002.

Ingress Traffic Engineering (cont.)

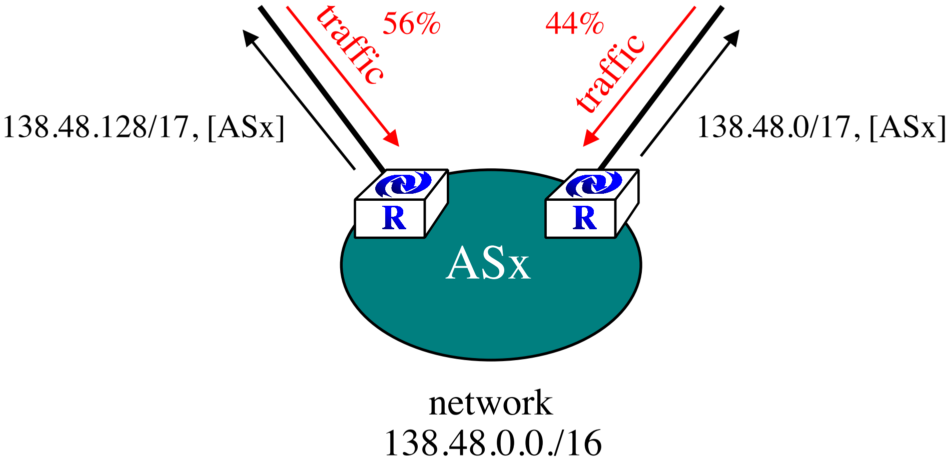


- Source: Presentation from B. Quoitin at QofIS 2002.

Ingress Traffic Engineering (cont.)



- Source: Presentation from B. Quoitin at QofIS 2002.



- Source: Presentation from B. Quoitin at QofIS 2002.

But in practice things might look quite complex...

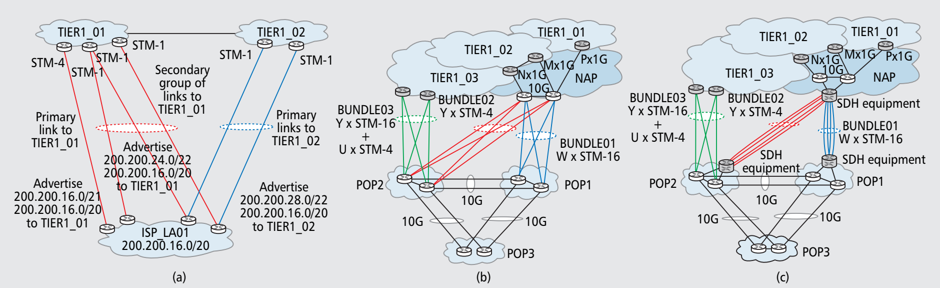


Figure 1. *a) Scenario #1: multihomed ISP with links of different capacity, load sharing, and backup routing policy; b) scenario #2: multihomed ISP with NAP presence; c) scenario #3: multihomed ISP with NAP presence and SDH multiplexers.*

- Source: M. Yannuzzi, X. Masip-Bruin, E. Grampin, R. Gagliano, A. Castro, M. German, “Managing interdomain traffic in Latin America: a new perspective based on LISP,” IEEE Communications Magazine, Vol. 47 , no. 7, July 2009.

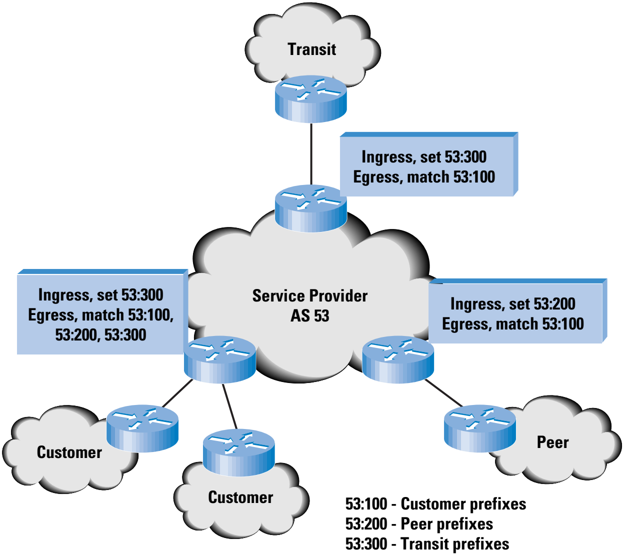
Community-based Traffic Engineering

Community-based TE

Basics of Communities

- Communities attribute (RFC1997, RFC1998): It's a "Transitive" attiribute
- Used to mark routes that share a common property and thus must undergo a specific treatment
- It is mainly used for building more scalable routing configurations
- ....some providers also allow their customers to control the redistribution of their routes by the use of communities...

Community-based TE (cont.)



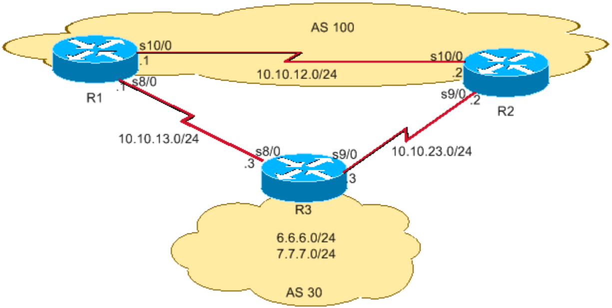
- Source: K. Foster, “Application of BGP Communities," Internet Protocol Journal, June 2003.

Community-based TE

Modus Operandi...

- Providers
  - Define their set of community values
  - And they configure specific actions, such as: "*do not announce*", "*prepend as-path*", or "*change local-pref*".
- Customers
  - Attach some of these communities to their routes to request the given treatment

Community-based TE (cont.)



To R1:

- 6.6.6.0/24 with a community attribute 100:300
- 7.7.7.0/24 with a community attribute 100:250

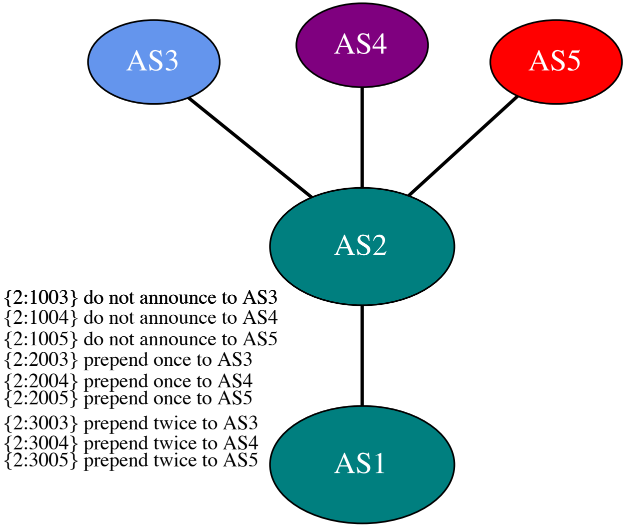
To R2:

- 6.6.6.0/24 with a community attribute 100:250
- 7.7.7.0/24 with a community attribute 100:300

Local Preference	Community Values
130	100:300
125	100:250

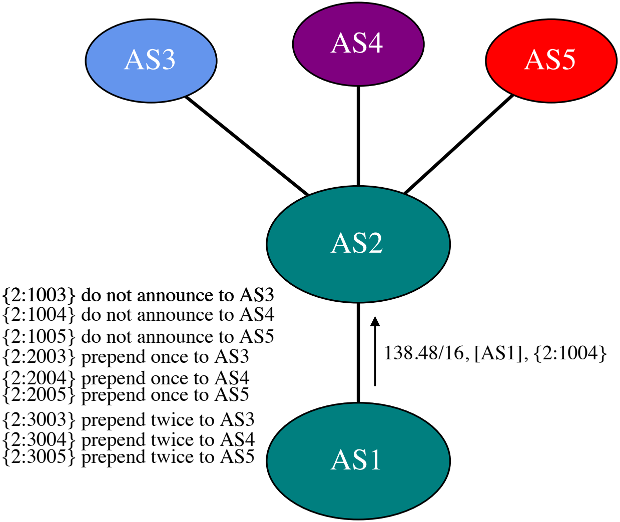
- Source: Cisco Systems, Inc.

Community-based TE...



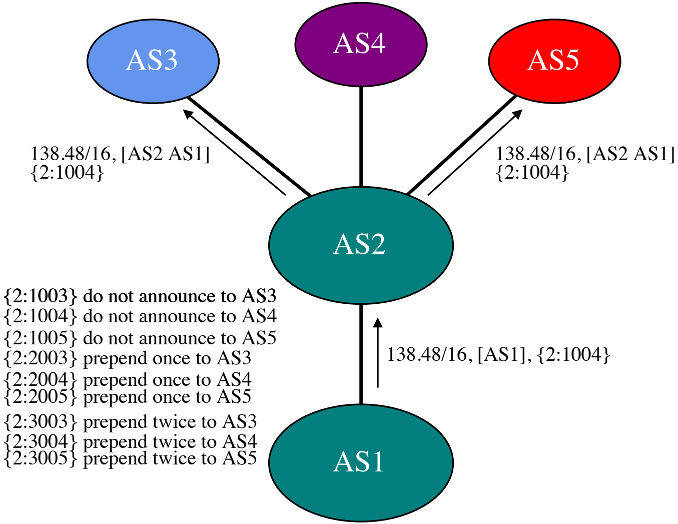
- Source: Presentation from B. Quoitin at QofIS 2002.

Community-based TE...



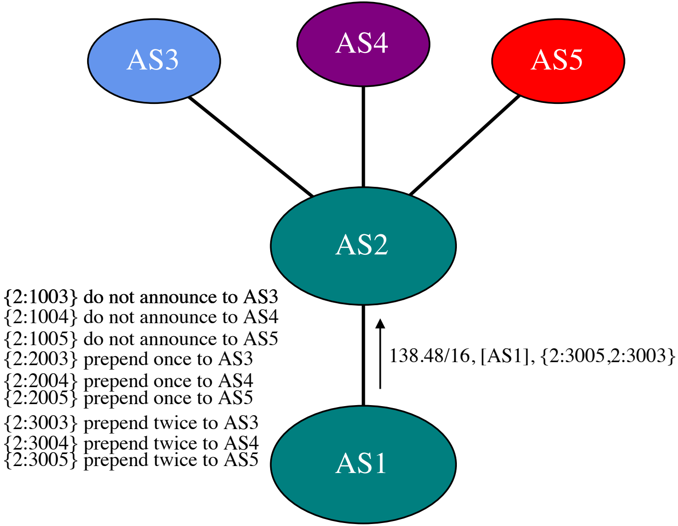
- Source: Presentation from B. Quoitin at QofIS 2002.

Community-based TE...



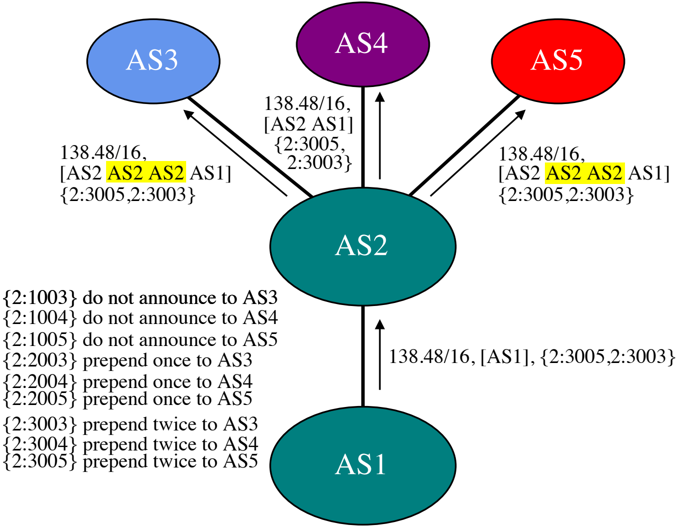
- Source: Presentation from B. Quoitin at QofIS 2002.

Community-based TE...



- Source: Presentation from B. Quoitin at QofIS 2002.

Community-based TE...



- Source: Presentation from B. Quoitin at QofIS 2002.

Limitations of Community-based TE

Main Issues

- Semantic of community values must be agreed and published
- Data models and data structure issues
- Requires manual configurations
- Transitivity contributes to additional churn

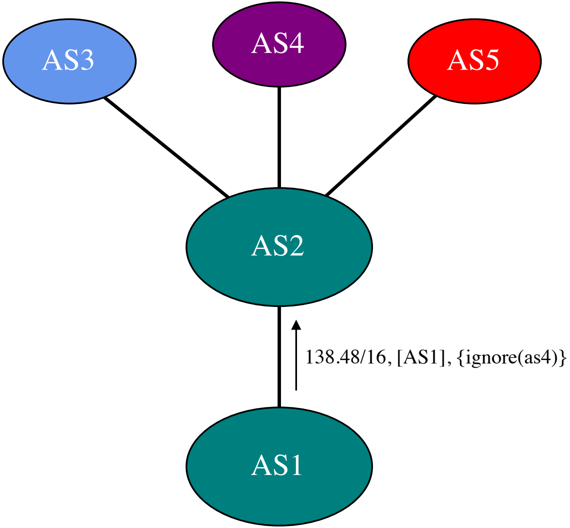
Redistribution Communities

Redistribution Communities

Proposed Modifications

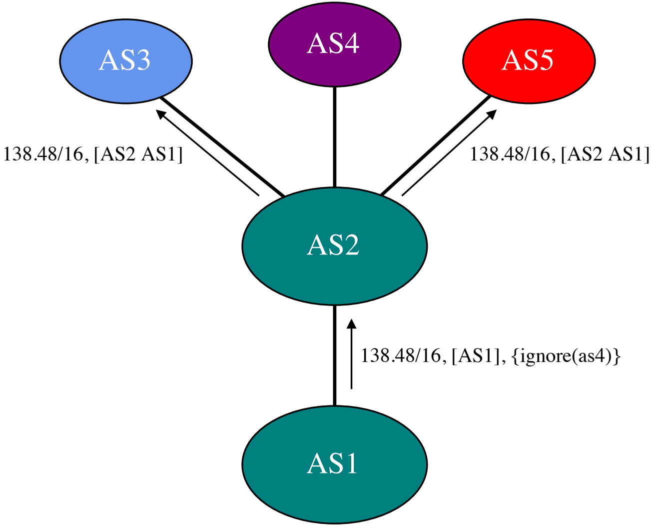
- Standardized semantics
- Actions:
  - The attached route should not be announced to the specified BGP speakers.
  - The attached route should only be announced to the specified BGP speakers.
  - The attached route should be announced with the NO_EXPORT community to the specified BGP speakers.
  - The attached route should be prepended *n* times when announced to the specified BGP speakers.

Redistribution Communities (cont.)



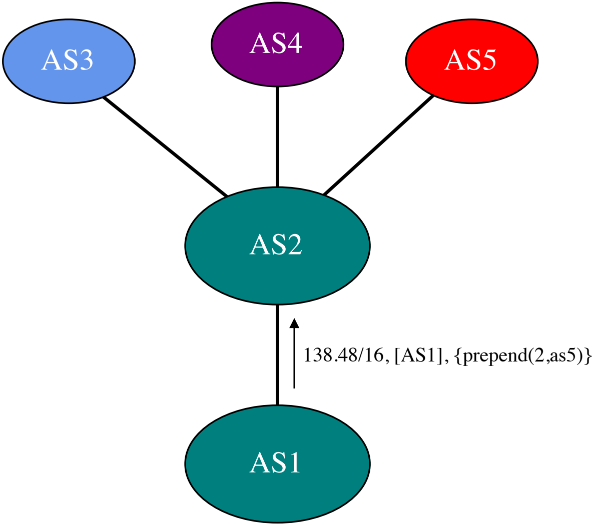
- Source: Presentation from B. Quoitin at QofIS 2002.

Redistribution Communities (cont.)



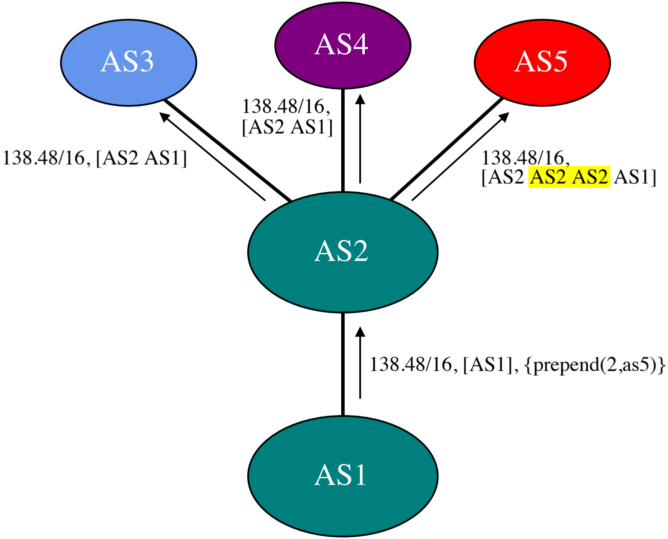
- Source: Presentation from B. Quoitin at QofIS 2002.

Redistribution Communities (cont.)



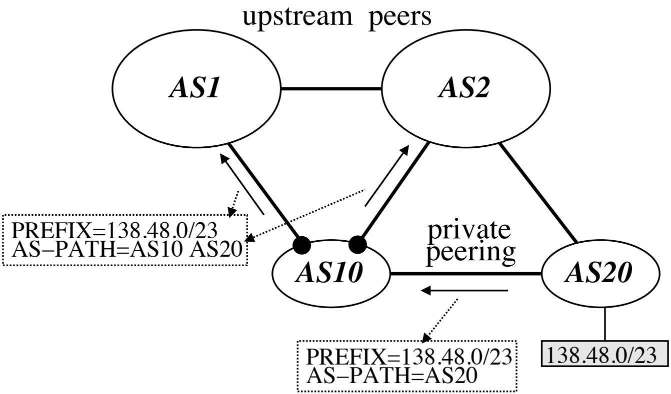
- Source: Presentation from B. Quoitin at QofIS 2002.

Redistribution Communities (cont.)

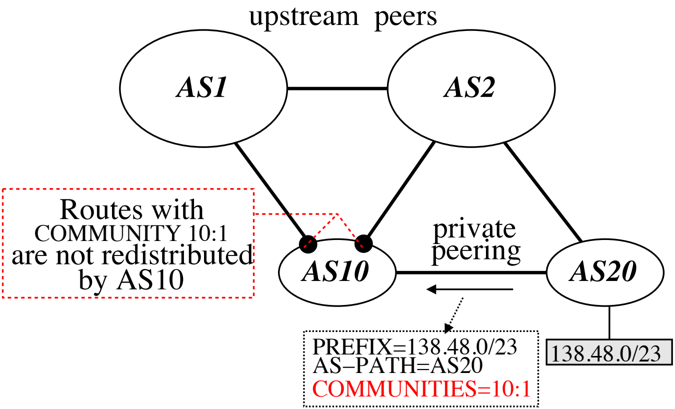


- Source: Presentation from B. Quoitin at QofIS 2002.

Community-based TE...

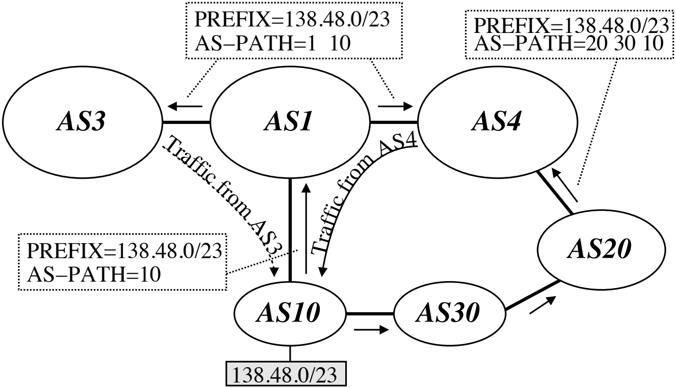


- Source: Presentation from B. Quoitin at NANOG25.



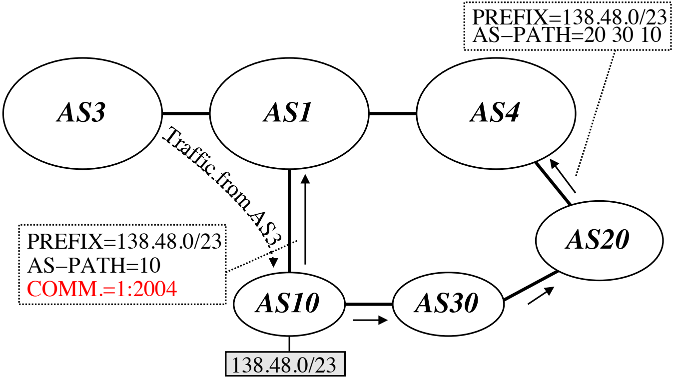
- Source: Presentation from B. Quoitin at NANOG25.

Community-based TE...



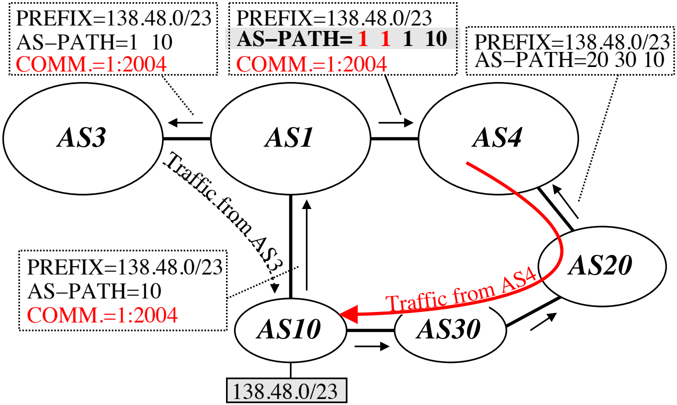
- Source: Presentation from B. Quoitin at NANOG25.

Community-based TE...



- Source: Presentation from B. Quoitin at NANOG25.

Community-based TE...



- Source: Presentation from B. Quoitin at NANOG25.

Community-based TE...

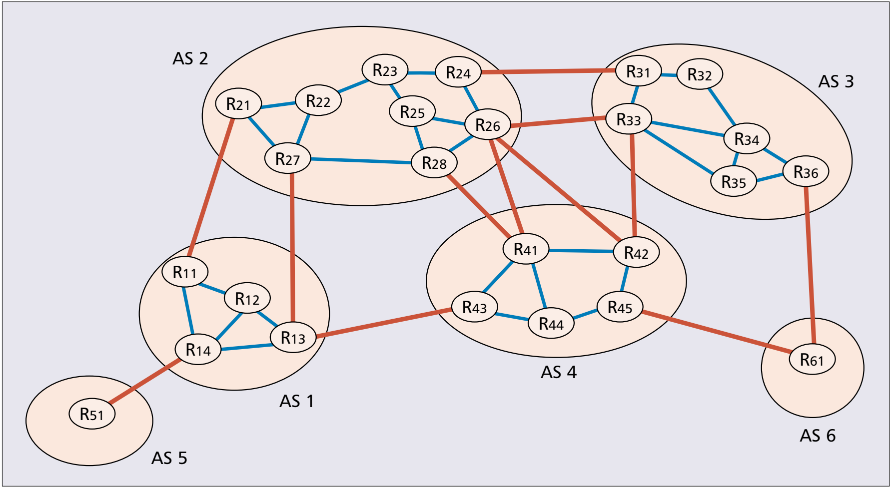


Figure 1. *A simple Internet.*

- Source: B. Quoitin, S. Uhlig, C. Pelsser, L. Swinnen, and O. Bonaventure, “Interdomain Traffic Engineering with BGP,” IEEE Communications Magazine, Vol. 41, Issue 5, May 2003.

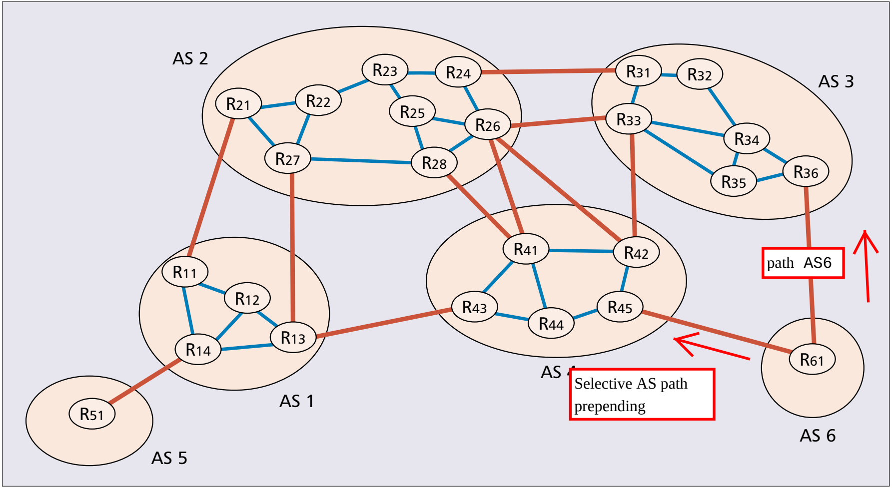


Figure 1. *A simple Internet.*

- Source: B. Quoitin, S. Uhlig, C. Pelsser, L. Swinnen, and O. Bonaventure, “Interdomain Traffic Engineering with BGP,” IEEE Communications Magazine, Vol. 41, Issue 5, May 2003.

Community-based TE...

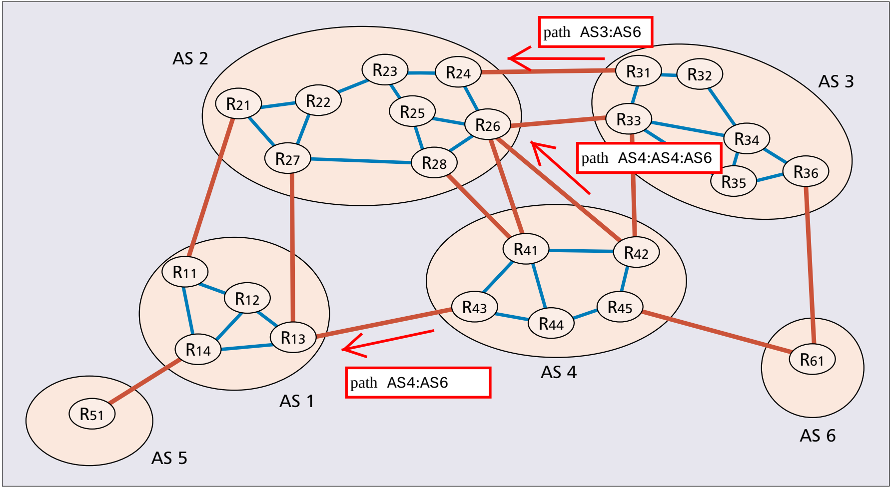


Figure 1. *A simple Internet.*

- Source: B. Quoitin, S. Uhlig, C. Pelsser, L. Swinnen, and O. Bonaventure, “Interdomain Traffic Engineering with BGP,” IEEE Communications Magazine, Vol. 41, Issue 5, May 2003.

Community-based TE...

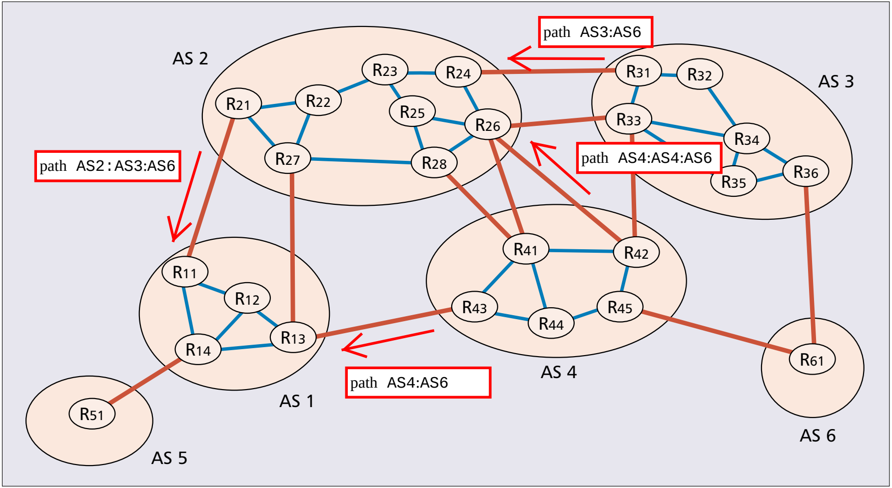


Figure 1. *A simple Internet.*

- Source: B. Quoitin, S. Uhlig, C. Pelsser, L. Swinnen, and O. Bonaventure, "Interdomain Traffic Engineering with BGP," IEEE Communications Magazine, Vol. 41, Issue 5, May 2003.

Community-based TE...

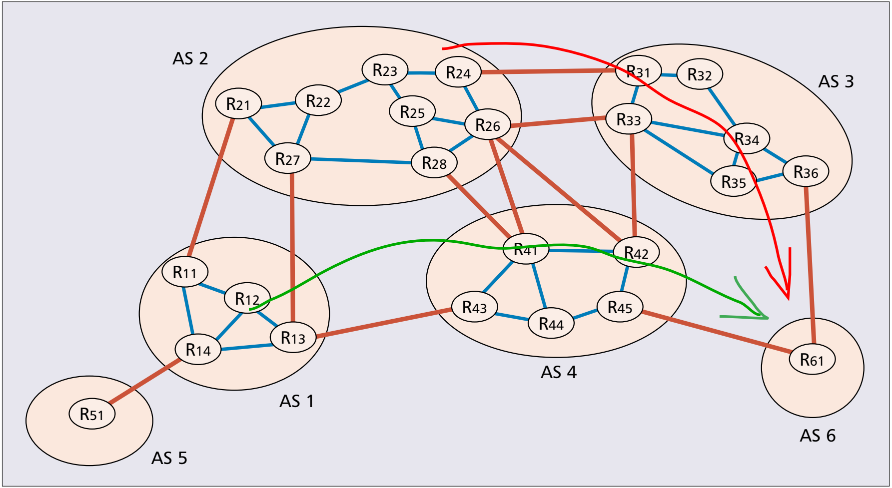
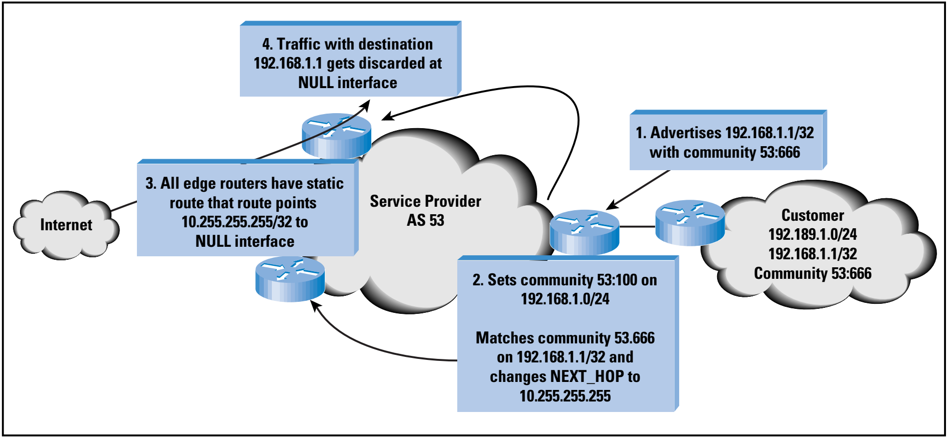


Figure 1. *A simple Internet.*

- Source: B. Quoitin, S. Uhlig, C. Pelsser, L. Swinnen, and O. Bonaventure, "Interdomain Traffic Engineering with BGP," IEEE Communications Magazine, Vol. 41, Issue 5, May 2003.

Community-based TE (cont.)

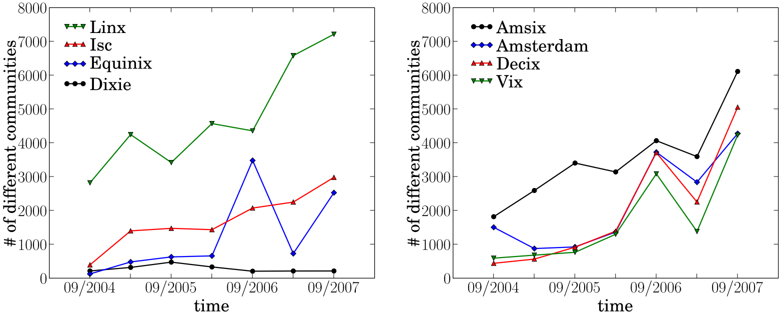
Figure 3: Customer-Initiated Black Hole to Defend Against a DoS Attack



- Source: K. Foster, “Application of BGP Communities," Internet Protocol Journal, June 2003.

Are Communities really used in Practice?

Community-based TE (cont.)

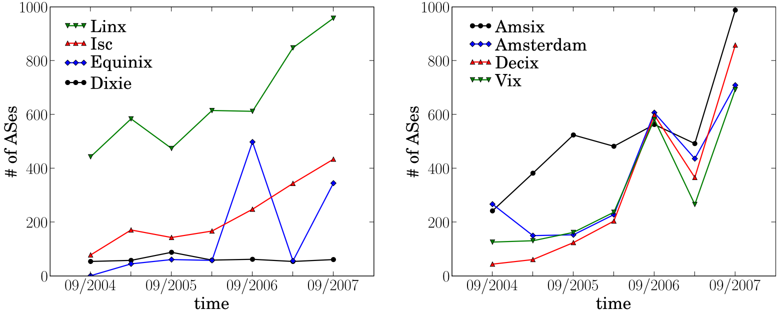


(a) Routeviews (b) RIPE

Figure 1: Evolution of BGP communities over time

- Source: B. Donnet and B. Quoitin, “On BGP Communities," ACM SIGCOMM Computer Communication Review, Volume 38 Issue 2, April 2008.

Community-based TE (cont.)

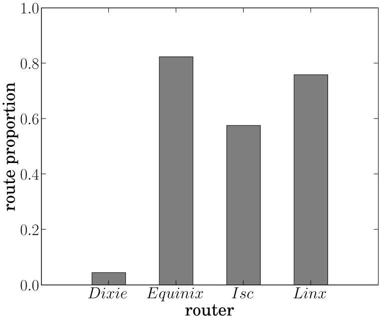


(a) Routeviews (b) RIPE

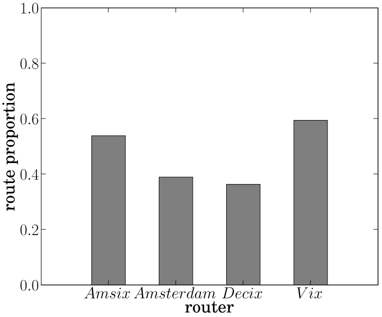
Figure 2: Evolution of ASes using BGP communities over time

- Source: B. Donnet and B. Quoitin, “On BGP Communities," ACM SIGCOMM Computer Communication Review, Volume 38 Issue 2, April 2008.

Community-based TE (cont.)



(a) Routeviews



(b) RIPE

Figure 3: Proportion of routes carrying BGP communities (Sept. 2007)

- Source: B. Donnet and B. Quoitin, "On BGP Communities," ACM SIGCOMM Computer Communication Review, Volume 38 Issue 2, April 2008.

...other TE objectives of Transit Providers

Transit Providers (hot potato routing)

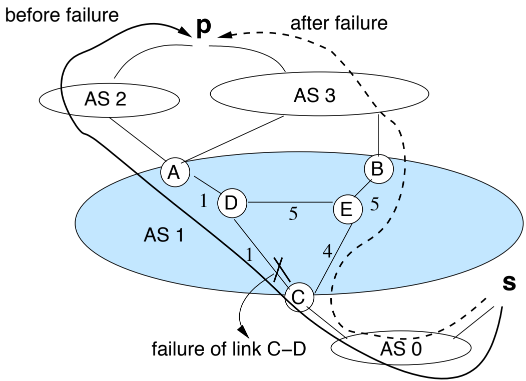
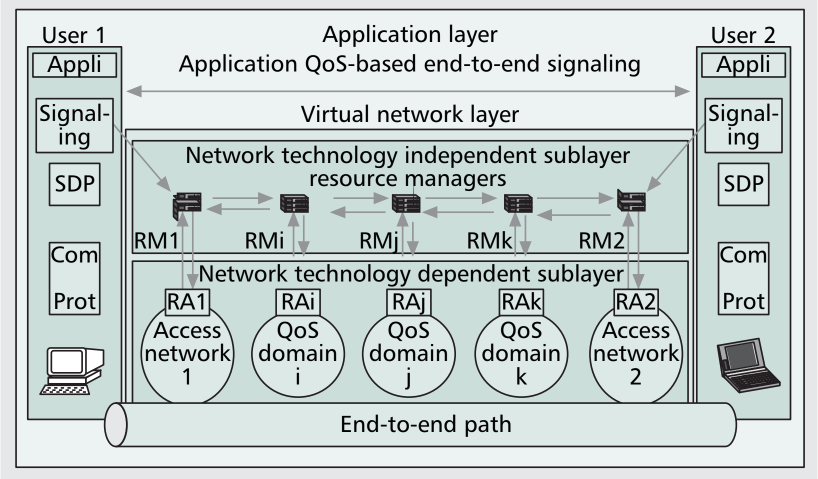


Fig. 1. Link failure causes router C to switch egress points from A to B for destination prefix p.

- Source: R. Teixeira et al., "TIE Breaking: Tunable Interdomain Egress Selection," IEEE/ACM Transactions on Networking, August 2007.

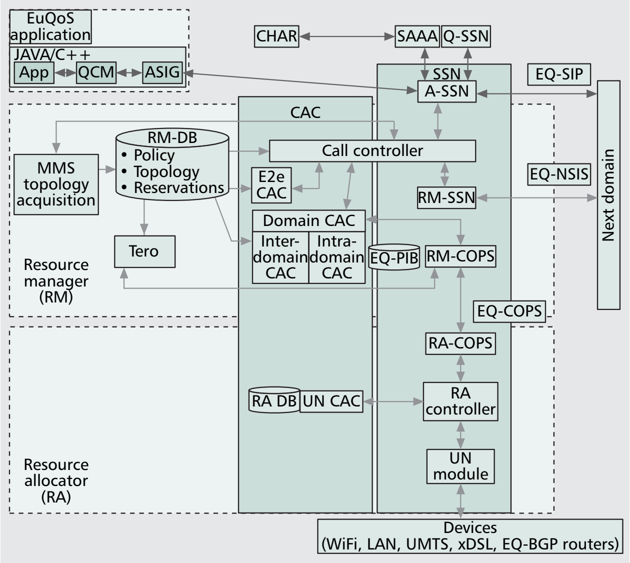
More Ambitious Approaches...

The EuQoS Approach...



- Source: X. Masip-Bruin, M. Yannuzzi et al., "The EuQoS System: A Solution for QoS Routing in Heterogeneous Networks," IEEE Communications Magazine, February 2007.

The EuQoS Approach...



- Source: X. Masip-Bruin, M. Yannuzzi et al., “The EuQoS System: A Solution for QoS Routing in Heterogeneous Networks,” IEEE Communications Magazine, February 2007.

The EuQoS Approach...

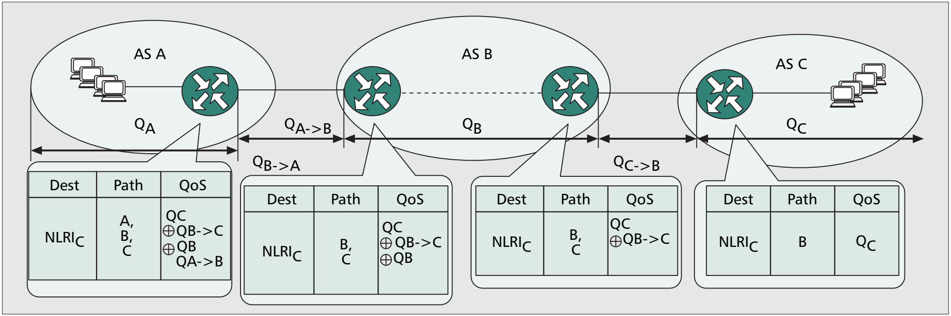


Figure 4. *Example of EQ-BGP operation.*

- Source: X. Masip-Bruin, M. Yannuzzi et al., “The EuQoS System: A Solution for QoS Routing in Heterogeneous Networks,” IEEE Communications Magazine, February 2007.

The EuQoS Approach...

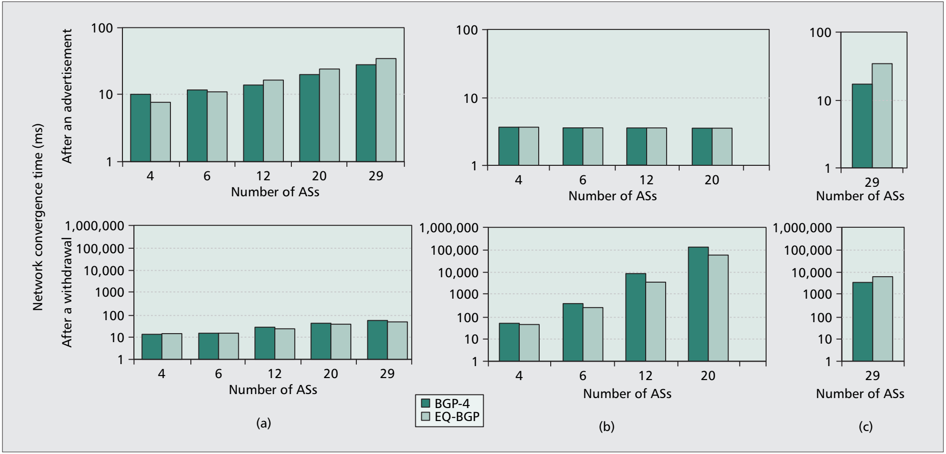


Figure 5. *Comparison of EQ-BGP and BGP-4 convergence time after a route advertisement or a route withdrawal, in the case of: a) Ring topology; b) full mesh topology; c) Internet like topology.*

- Source: X. Masip-Bruin, M. Yannuzzi et al., "The EuQoS System: A Solution for QoS Routing in Heterogeneous Networks," IEEE Communications Magazine, February 2007.

The EuQoS Approach...

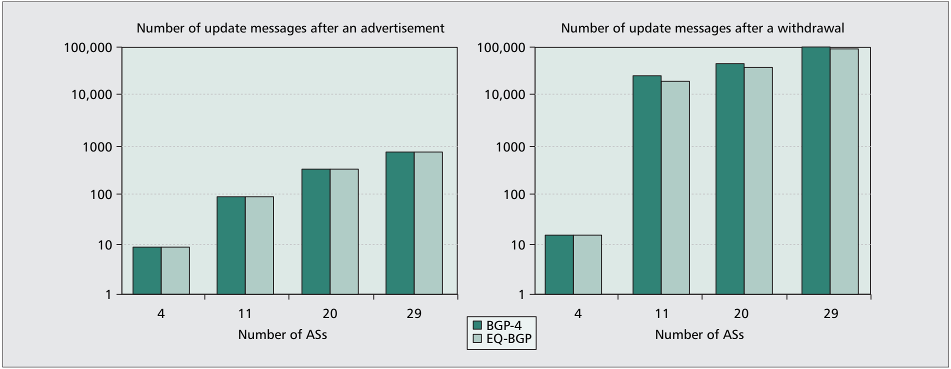
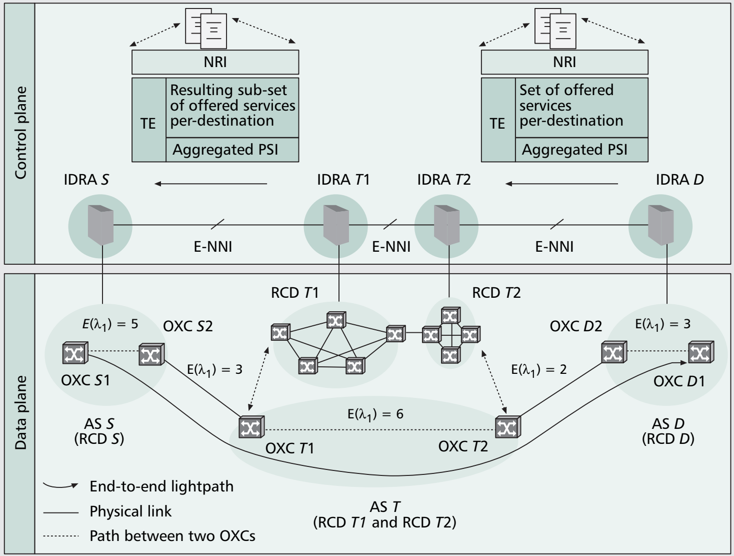


Figure 6. *Scalability of EQ-BGP vs. BGP-4.*

- Source: X. Masip-Bruin, M. Yannuzzi et al., "The EuQoS System: A Solution for QoS Routing in Heterogeneous Networks," IEEE Communications Magazine, February 2007.

Transit Providers
Traffic Engineering at the Optical Layer

OTN TE



- Source: M. Yannuzzi et al., "Toward a New Route Control Model for Multidomain Optical Networks," IEEE Communications Magazine, June 2008.

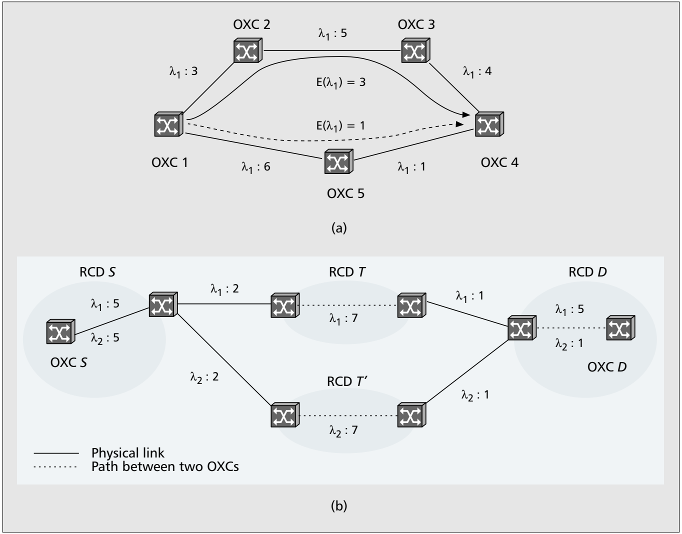


Figure 2. *a) Computation of the ENAW; b) advantage of the cost computation.*

- Source: M. Yannuzzi et al., "Toward a New Route Control Model for Multidomain Optical Networks," IEEE Communications Magazine, June 2008.

Input: NRI associated with each destination *d*
PSI between OXCs *s* and *d*

Output: The best (path, wavelength) pair between *s* and *d*

1: Choose the (path, wavelength) pair with the minimum cost
2: If the costs are equal choose the path with the highest ENAW
3: If the ENAWs are equal choose the path with the shortest number of hops *H*, and assign the wavelength λ_i with the lowest identifier *i*
4: If the hops *H* are equal prefer the path with the highest ENAW to the remote border OXC
5: If more than one path is still available run BGP tie-breaking rules [4]

Figure 3. *IDRA RWA decision process.*

- Source: M. Yannuzzi et al., "Toward a New Route Control Model for Multidomain Optical Networks," IEEE Communications Magazine, June 2008.

OTN TE

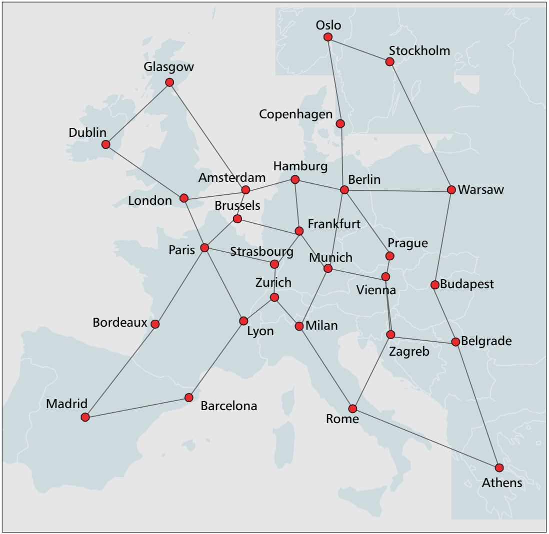
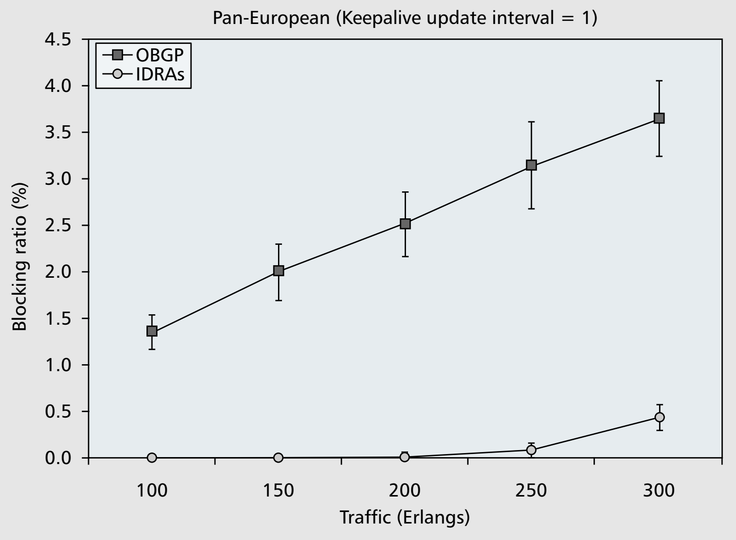


Figure 4. *Pan-European reference network topology.*

- Source: M. Yannuzzi et al., "Toward a New Route Control Model for Multidomain Optical Networks," IEEE Communications Magazine, June 2008.

OTN TE



- Source: M. Yannuzzi et al., “Toward a New Route Control Model for Multidomain Optical Networks,” IEEE Communications Magazine, June 2008.

OTN TE

	Keepalive update interval ($K_T = 1$)			Keepalive update interval ($K_T = 3$)			Keepalive update interval ($K_T = 5$)		
	200 Erlangs	250 Erlangs	300 Erlangs	200 Erlangs	250 Erlangs	300 Erlangs	200 Erlangs	250 Erlangs	300 Erlangs
IF	363.97	39.48	8.40	315.69	29.63	8.41	158.00	24.20	7.93

Traffic (Erlangs)	Routing messages OBGP	Routing messages IDRAs	Routing messages OBGP	Routing messages IDRAs	Routing messages OBGP	Routing messages IDRAs
100	6,564,525	2,819,949	5,539,285	2,771,408	4,842,449	2,764,530
150	7,907,963	3,013,904	6,544,983	2,961,622	5,574,075	2,876,943
200	8,607,917	3,141,911	6,905,969	3,041,394	5,822,980	2,946,896
250	8,992,258	3,288,572	7,033,482	3,149,322	5,864,259	3,027,520
300	9,198,274	3,661,793	7,071,856	3,393,776	5,928,454	3,179,430

Table 1. *Improvement factor in the blocking requests for 200, 250, and 300 Erlangs, and overall number of routing messages exchanged.*

- Source: M. Yannuzzi et al., “Toward a New Route Control Model for Multidomain Optical Networks,” IEEE Communications Magazine, June 2008.

Traffic Engineering
Non-transit Providers

Non-transit Domains: Intelligent Route Control (IRC)

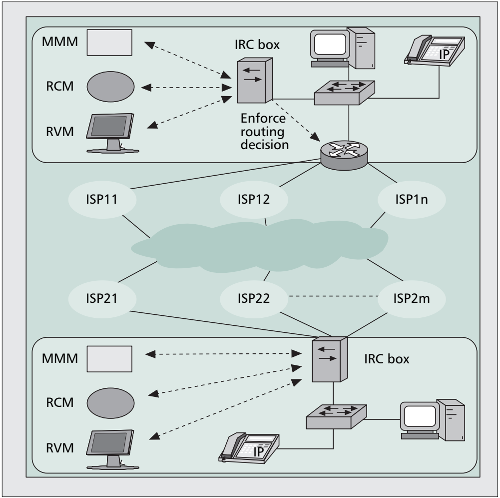


Figure 1. *The IRC model. IRC systems are composed of three modules: the monitoring and measurement module (MMM), the route control module (RCM), and a reporting and viewer module (RVM).*

- Source: M. Yannuzzi et al., “Improving the Performance of Route Control Middleboxes in a Competitive Environment,” IEEE Network, Sept./Oct. 2008.

Non-transit Domains: Intelligent Route Control (IRC)

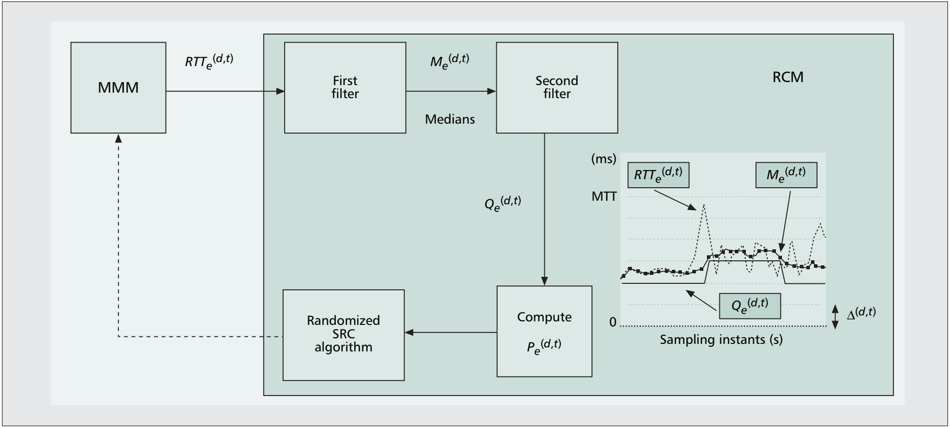


Figure 2. *Filtering process and interaction between the monitoring and measurement module (MMM) and the route control module (RCM) of a sociable route controller. The Randomized SRC Algorithm within the RCM is outlined in Algorithm 1.*

- Source: M. Yannuzzi et al., "Improving the Performance of Route Control Middleboxes in a Competitive Environment," IEEE Network, Sept./Oct. 2008.

Non-transit Domains: Intelligent Route Control (IRC)

Input: d – A target destination of network S
$\{e\}$ – Set of egress links of network S
$P_e^{(d,t)}$ – Performance function to reach d through e at time t

Output: e^{best} – The best egress link to reach target destination d

1: Wait for changes in $P_{ebest}^{(d,t)}$
2: **if** $P_{ebest}^{(d,t)} - P_e^{(d,t)} < R_{th}\ \forall e \neq e^{best}$ **then** go to Step 1
3: /* Egress link selection process for d */
4: Choose e′ as $P_{e'}^{(d,t)} = min\{P_e^{(d,t)}\}$
5: Estimate the performance after switching the traffic
6: **if** $P_{ebest}^{(d,t)} - P_{e'}^{(d,t)}{}_{Estimate} \geq R_{th}$ **then**
7: Wait until T_H =0 /* Hysteresis Switching Timer */
8: Switch traffic toward d from e^{best} to e'
9: $e^{best} \leftarrow e'$
10: $P_{ebest}^{(d,t)} \leftarrow P_{e'}^{(d,t)}$
11: **end if**
12: /* End of egress link selection process for d */
13: Go to Step 1

Algorithm 1. *Randomized SRC algorithm.*

- Source: M. Yannuzzi et al., "Improving the Performance of Route Control Middleboxes in a Competitive Environment," IEEE Network, Sept./Oct. 2008.

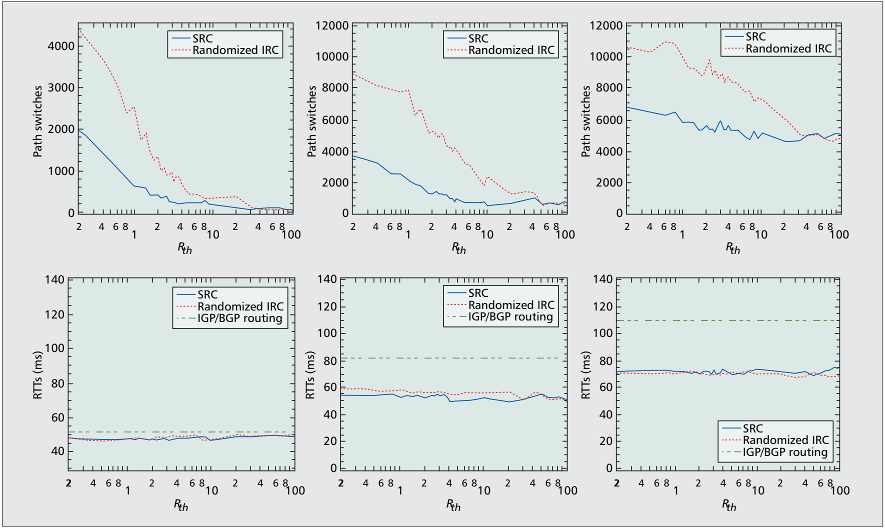


Figure 3. *Number of path switches (top) and <RTTs> (bottom) for* L = *0.450 (left),* L = *0.675 (center), and* L = *0.900 (right).*

- Source: M. Yannuzzi et al., “Improving the Performance of Route Control Middleboxes in a Competitive Environment,” IEEE Network, Sept./Oct. 2008.

Non-transit Domains: Intelligent Route Control (IRC)

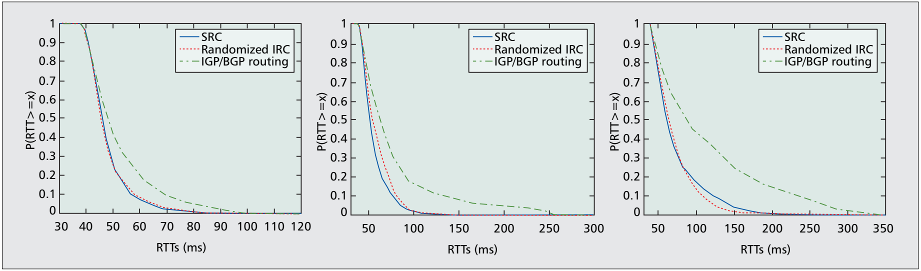


Figure 4. *Complementary cumulative distribution function (CCDF) of the RTTs for the 300 competing IRC flows, for* $R_{th} = 1$*, and for* $L = 0.450$ *(left),* $L = 0.675$ *(center), and* $L = 0.900$ *(right).*

- Source: M. Yannuzzi et al., "Improving the Performance of Route Control Middleboxes in a Competitive Environment," IEEE Network, Sept./Oct. 2008.

Non-transit Domains: Intelligent Route Control (IRC)

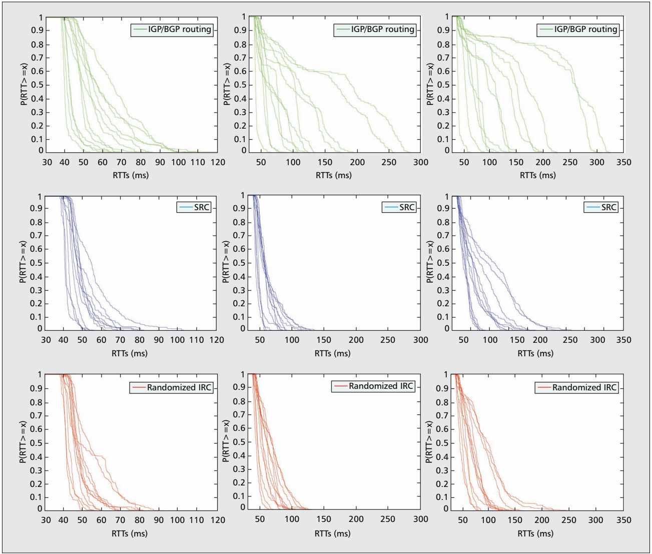


Figure 5. *CCDFs for IGP/BGP routing (top), SRC (center), and randomized IRC (bottom), for* L = *0.450 (left),* L = *0.675 (center), and* L = *0.900 (right).*

- Source: M. Yannuzzi et al., “Improving the Performance of Route Control Middleboxes in a Competitive Environment,” IEEE Network, Sept./Oct. 2008.

Outline

1. iBGP, eBGP, and Route Reflectors
2. Case study: Japanese Earthquake in 2011.
3. Interdomain Traffic Engineering
4. **Research challenges in interdomain routing**

Well-known issues in BGP ...

1. Slow convergence
2. Scalability Issues
3. High churn rate of route advertisements
4. Limited expressiveness of routing policies and TE control
5. Security vulnerabilities
6. ...

These are due:

- ... in part to the utilization of path vectors
- ... in part to implementation decisions made in BGP

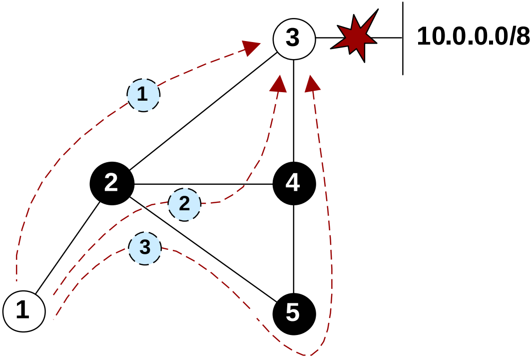
Slow Convergence

Slow Convergence

- Depending on the location of the origin of an event and where the observation is made, a BGP convergence might vary between tens and several hundreds of seconds [C. Labovitz et al. 1999, 2001].
- This slow convergence is mainly caused by the *path hunting* performed by BGP.

Slow Convergence (cont.)

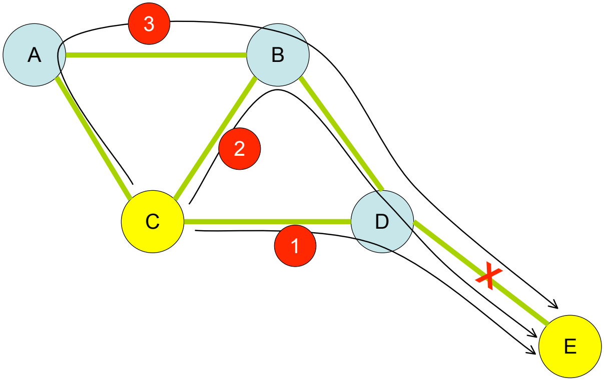
- The path exploration or path hunting phenomenon



- Source: M. Yannuzzi, R. Serral-Gracia, and X. Masip-Bruin, ”Chapter 3: Distance and Path Vector Routing Models,” to be published in the book "MULTI-DOMAIN NETWORKS: A PRACTICAL PERSPECTIVE," Springer Series, Series Ed.: B. Mukherjee, Eds: N. Ghani, M. Peng, and I. Monga.

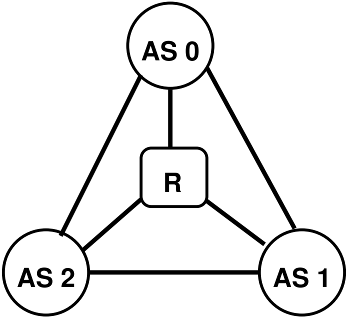
Slow Convergence (cont.)

- Another example of path-exploration:



- Source: R. Oliveira, B. Zhang, D. Pei, and L. Zhang, “Quantifying Path Exploration in the Internet,” IEEE/ACM Trans. on Networking, Vol. 17, No. 2., pp. 445-458, April 2009.

A detailed example of path exploration



- Source: C. Labovitz, A. Ahuja, A. Abose, and F. Jahanian, “Delayed Internet routing convergence,” IEEE/ACM Trans. on Networking, vol. 9, no. 3, pp. 293–306, June 2001.

A detailed example of path exploration (cont.)

Stage	Routing Tables	Messages Processed	Messages queued or delivered though not processed yet					
0	Steady State 0(*R, 1R, 2R) 1(0R, *R, 2R) 2(0R, 1R, *R)							
1	R withdraws its route 0(- , *1R, 2R) 1(*0R, - , 2R) 2(*0R, 1R, -)	R → 0 (w) R → 1 (w) R → 2 (w)	0 → 1 (01R) 0 → 2 (01R)	1 → 0 (10R) 1 → 2 (10R)	2 → 0 (20R) 2 → 1 (20R)			
2	1 and 2 receive the updates from 0 0(- , *1R, 2R) 1(- , - , *2R) 2(01R, *1R, -)	0 → 1 (01R) 0 → 2 (01R)	1 → 0 (10R) 1 → 2 (10R)	2 → 0 (20R) 2 → 1 (20R)	1 → 0 (12R) 1 → 2 (12R)	2 → 0 (21R) 2 → 1 (21R)		
3	0 and 2 receive the updates from 1 0(- , - , *2R) 1(- , - , *2R) 2(*01R, 10R, -)	1 → 0 (10R) 1 → 2 (10R)	2 → 0 (20R) 2 → 1 (20R)	1 → 0 (12R) 1 → 2 (12R)	2 → 0 (21R) 2 → 1 (21R)	0 → 1 (02R) 0 → 2 (02R)	2 → 0 (201R) 2 → 1 (201R)	
4	0 and 1 receive the updates from 2 0(- , - , -) 1(- , - , *20R) 2(*01R, 10R, -)	2 → 0 (20R) 2 → 1 (20R)	1 → 0 (12R) 1 → 2 (12R)	2 → 0 (21R) 2 → 1 (21R)	0 → 1 (02R) 0 → 2 (02R)	2 → 0 (201R) 2 → 1 (201R)	0 → 1 (w) 0 → 2 (w)	1 → 0 (120R) 1 → 2 (120R)
5	0 and 2 receive the updates from 1 0(- , *12R , -) 1(- , - , *20R) 2(*01R, - , -)	1 → 0 (12R) 1 → 2 (12R)	2 → 0 (21R) 2 → 1 (21R)	0 → 1 (02R) 0 → 2 (02R)	2 → 0 (201R) 2 → 1 (201R)	0 → 1 (w) 0 → 2 (w)	1 → 0 (120R) 1 → 2 (120R)	0 → 1 (012R) 0 → 2 (012R)
6	0 and 1 receive the updates from 2 0(- , *12R , 21R) 1(- , - , -) 2(*01R, - , -)	2 → 0 (21R) 2 → 1 (21R)	0 → 1 (02R) 0 → 2 (02R)	2 → 0 (201R) 2 → 1 (201R)	0 → 1 (w) 0 → 2 (w)	1 → 0 (120R) 1 → 2 (120R)	0 → 1 (012R) 0 → 2 (012R)	1 → 0 (w) 1 → 2 (w)
...	....	....			....	....		

- Source: C. Labovitz, A. Ahuja, A. Abose, and F. Jahanian, “Delayed Internet routing convergence,” IEEE/ACM Trans. on Networking, vol. 9, no. 3, pp. 293–306, June 2001.

A detailed example of path exploration (cont.)

Number of paths that can be potentially explored

For a complete graph of n nodes there exist $O((n-1)!)$ distinct paths to reach a destination.

$$P(n) = (n-1) + (n-1)(n-2) + \cdots + (n-1)!$$

$$P(n) = (n-1)! \left[1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{(n-2)!}\right] \approx (n-1)!$$

In slide 95: $n = 4 \Rightarrow P(4) = 3 + 3.2 + 3.2.1 = 15$ (15 different paths in total in the bad gadget)

- Source: C. Labovitz, A. Ahuja, A. Abose, and F. Jahanian, “Delayed Internet routing convergence,” IEEE/ACM Trans. on Networking, vol. 9, no. 3, pp. 293–306, June 2001.

Convergence Time (usual metrics)

Time Metrics

- T_{up}: A previously unreachable destination becomes reachable through a path by the end of the event.
- T_{down}: A previously reachable destination becomes unreachable by the end of the event.
- T_{short}: A reachable destination has changed the path to a more preferred one by the end of the event.
- T_{long}: A reachable destination has changed the path to a less preferred one by the end of the event.
- T_{equal}: A reachable destination has changed the path by the end of the event, but the starting and ending paths have the same preference.
- T_{pdist}: The AS path is the same before and after the event, with some transient change(s) during the event.

- Source: R. Oliveira, B. Zhang, D. Pei, and L. Zhang, "Quantifying Path Exploration in the Internet," IEEE/ACM Trans. on Networking, Vol. 17, No. 2., pp. 445-458, April 2009. 2001.

The effects of path exploration

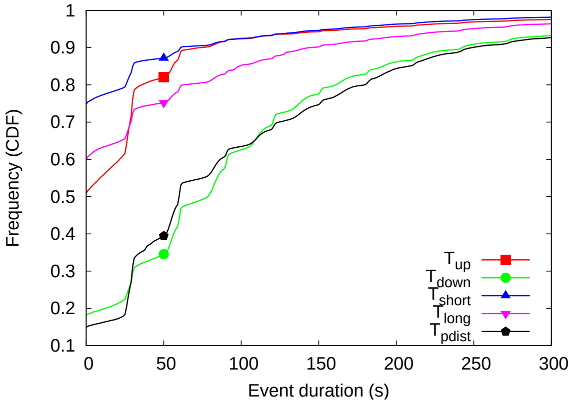


Figure 10: Duration of Events.

- Source: R. Oliveira, B. Zhang, D. Pei, and L. Zhang, "Quantifying Path Exploration in the Internet," IEEE/ACM Trans. on Networking, Vol. 17, No. 2., pp. 445-458, April 2009.

The effects of path exploration (cont.)

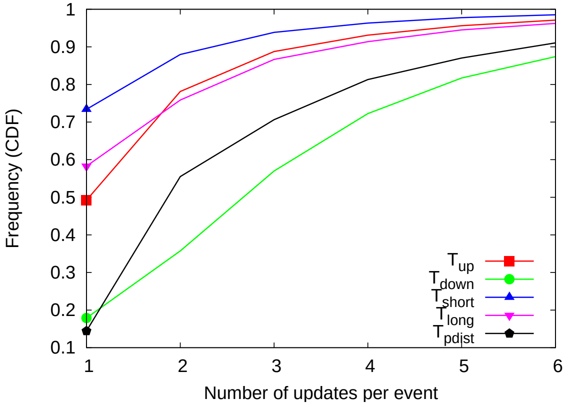


Figure 11: Number of Updates per Event.

- Source: R. Oliveira, B. Zhang, D. Pei, and L. Zhang, "Quantifying Path Exploration in the Internet," IEEE/ACM Trans. on Networking, Vol. 17, No. 2., pp. 445-458, April 2009.

The effects of path exploration (cont.)

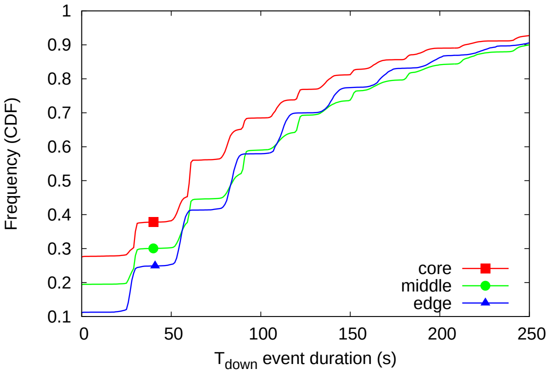


Figure 14: Duration of T_{down} events as seen by monitors at different tiers.

- Source: R. Oliveira, B. Zhang, D. Pei, and L. Zhang, “Quantifying Path Exploration in the Internet,” IEEE/ACM Trans. on Networking, Vol. 17, No. 2., pp. 445-458, April 2009.

The effects of path exploration (cont.)

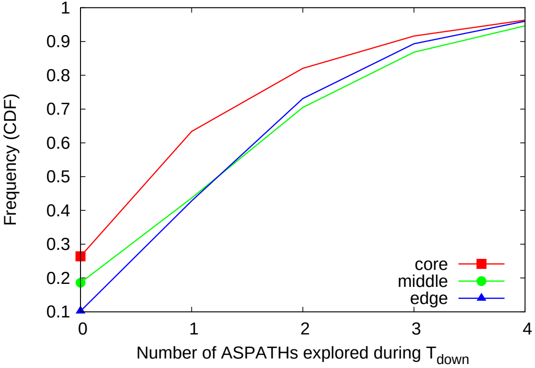


Figure 15: Number of unique paths explored during T_{down} as seen by monitors at different tiers.

- Source: R. Oliveira, B. Zhang, D. Pei, and L. Zhang, “Quantifying Path Exploration in the Internet,” IEEE/ACM Trans. on Networking, Vol. 17, No. 2., pp. 445-458, April 2009.

The effects of path exploration (cont.)

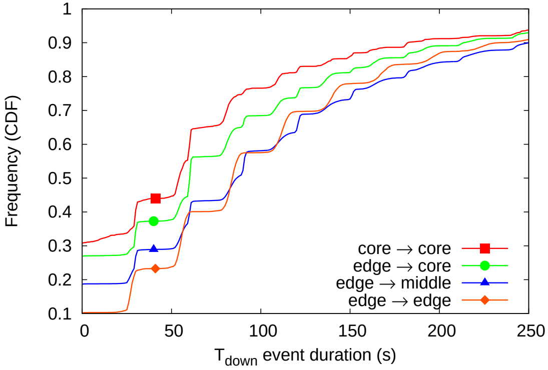


Figure 17: Duration of T_{down} events observed and originated in different tiers.

- Source: R. Oliveira, B. Zhang, D. Pei, and L. Zhang, "Quantifying Path Exploration in the Internet," IEEE/ACM Trans. on Networking, Vol. 17, No. 2., pp. 445-458, April 2009.

- ...almost 3 T_{down} events per minute....

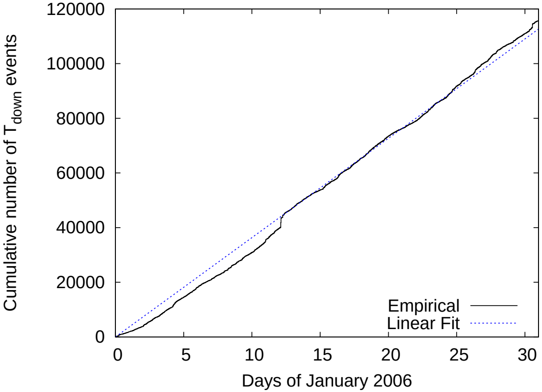


Figure 19: Number of T_{down} events over time.

- Source: R. Oliveira, B. Zhang, D. Pei, and L. Zhang, "Quantifying Path Exploration in the Internet," IEEE/ACM Trans. on Networking, Vol. 17, No. 2., pp. 445-458, April 2009.

Some proposals from the literature

- A. Bremler-Barr, Y. Afek and S. Schwarz, “Improved BGP Convergence via Ghost Flushing,” in Proceedings of IEEE INFOCOM, 2003.
- A. Feldmann, O. Maennel, Z. M. Mao, A. Berger, and B. Maggs “Locating Internet routing instabilities,” in Proc. ACM SIGCOMM, Portland, USA, September 2004.
- D. Pei, M. Azuma, D. Massey, and L. Zhang, “BGP-RCN: improving BGP convergence through root cause notification,” Computer Networks, Volume 48, Issue 2, pp 175-194, 2005.
- J. Chandrashekar, Z. Duan, Z.-L. Zhang, and J. Krasky, “Limiting path exploration in BGP,” in Proceedings of INFOCOM, Miami, USA, 2005.
- ...

Industry ... the MRAI timers

T3 + 2.MRAI: 9 { 7,8,6,4,1}

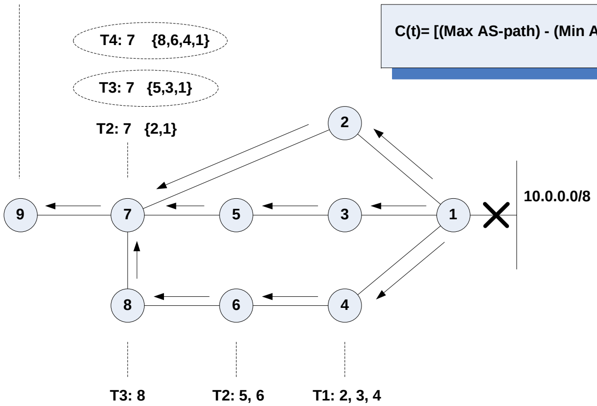
T3 + MRAI: 9 { 7,5,3,1}

T3: 9 {7, 2,1}

Min AS-path: {7,2,1} → 3

Max AS-path: {7,8,6,4,1} → 5

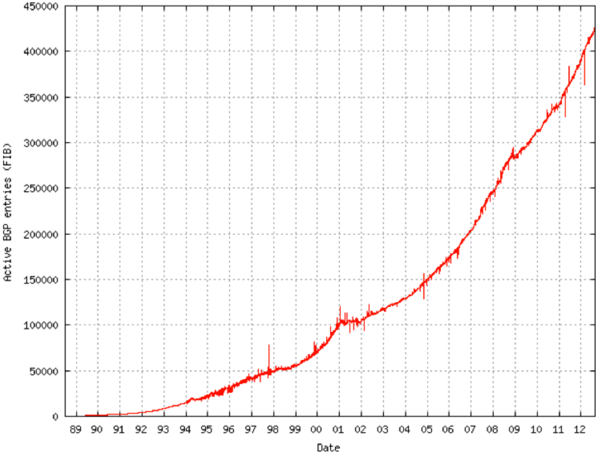
$$C(t) = [(\text{Max AS-path}) - (\text{Min AS-path})].\text{MRAI} + \sum_{\text{Min AS-path}} D_i$$



Scalability Issues

Scalability Issues

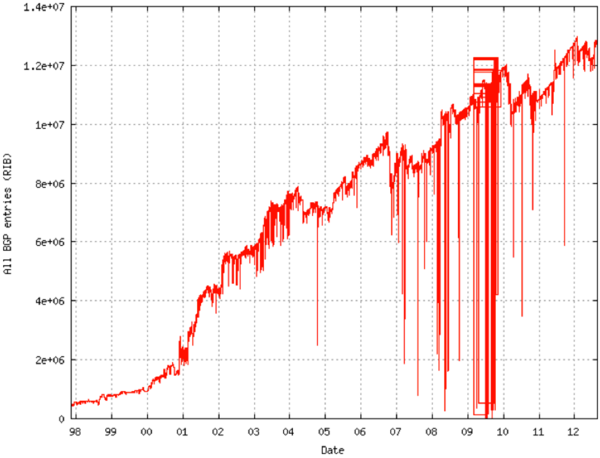
- FIB Evolution



- Source: CIDR Report.

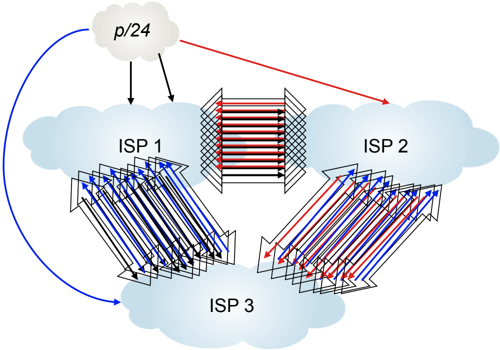
Scalability Issues

- RIB Evolution



- Source: CIDR Report.

A study from ARBOR Networks (2010)

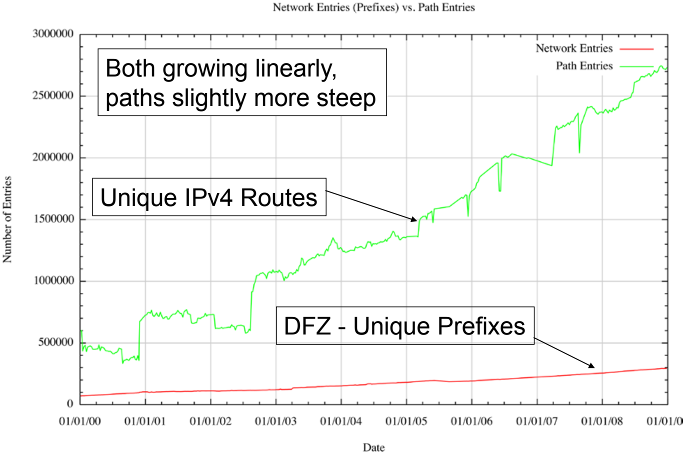


ISP1 - one unique prefix (*p*), 22 routes total on PE routers, without intra-domain BGP effects

- Consider N ASes: if an edge AS E connects to one of the N ASes, each AS has (N-1) paths to each prefix *p* announced by E
- When E connects to n of N ASes, each AS has at least n*N routes to *p*
  - In general the total number of routes to p can grow super-linearly with n
  - Edge AS multi-homing n times to the same ISP does NOT have this effect on adjacent ISPs
- It's common for ISPs to have 10 or more interconnects with other ISPs
  - when E connects to n ISPs, each ISP likely to see n*10 routes for p announced by E
- New ISPs in core, or nested transit relationships, often exacerbate the problem

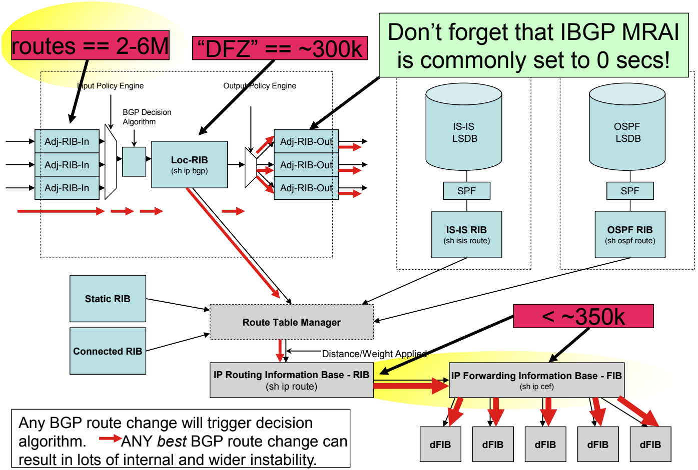
- Source: Danny McPherson (ARBOR Networks) "Prefixes, Paths & Internet Routing System Scalability," ARIN 25, April 2010.

A study from ARBOR Networks (2010) (cont.)



- Source: Danny McPherson (ARBOR Networks) "Prefixes, Paths & Internet Routing System Scalability," ARIN 25, April 2010.

A study from ARBOR Networks (2010) (cont.)



- Source: Danny McPherson (ARBOR Networks) "Prefixes, Paths & Internet Routing System Scalability," ARIN 25, April 2010.

Scalability Issues (cont.)

Region	IXPs	# of prefixes	De-aggregation factor (DF)
Africa	21	5K	3.46
Asia & Pacific	73	66K	2.81
Europe & Mid. East	123	67K	1.74
LA & Caribbean	24	26K	4.38
North America	88	124K	1.87
		Global BGP table	**Global average**
		288K	2.12

Table 1. *Statistics by region (data of April 2009, extracted from [1] and APNIC [7]).*

$$DF = \left(\frac{\text{Prefixes in the Global Routing Table}}{\text{Aggregatable Prefixes}} \right)$$

- Source: M. Yannuzzi, X. Masip-Bruin, E. Grampin, R. Gagliano, A. Castro, M. German, “Managing interdomain traffic in Latin America: a new perspective based on LISP,” IEEE Communications Magazine, Vol. 47 , no. 7, July 2009.

Scalability Issues (cont.)

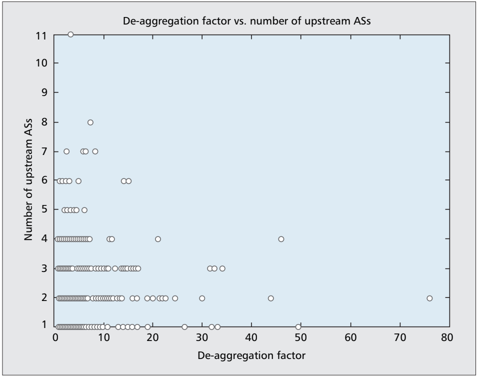


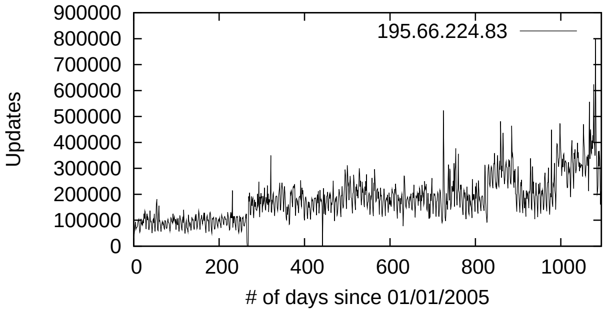
Figure 2. *Distribution of the de-aggregation factor as a function of the number of upstream providers in Latin America (data of April 2009, extracted from [1] and APNIC [7]).*

- Source: M. Yannuzzi, X. Masip-Bruin, E. Grampín, R. Gagliano, A. Castro, M. Germán, "Managing interdomain traffic in Latin America: a new perspective based on LISP," IEEE Communications Magazine, Vol. 47 , no. 7, July 2009.

Churn

Problems with BGP-4: Churn

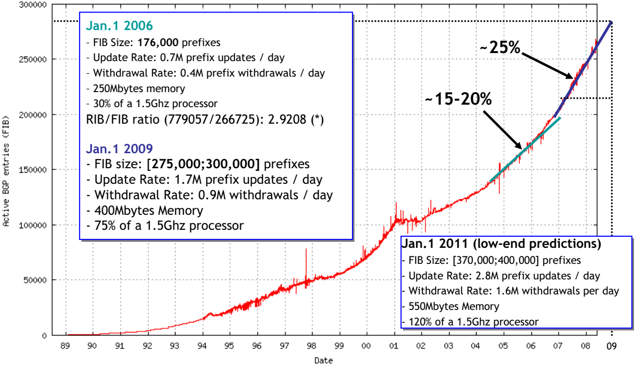
- The number of updates grew approximately by 200% over three years (2005–2007).



- Growth in churn from a monitor in France Telecom's network.

  Source: A. Elmokashfi, A. Kvalbein, C. Dovrolis, "On the scalability of BGP: the roles of topology growth and update rate-limiting," ACM CoNEXT 2008, Madrid, Spain, December 2008.

Growth of Active BGP Entries
(from Jan'89 to Mar'08)

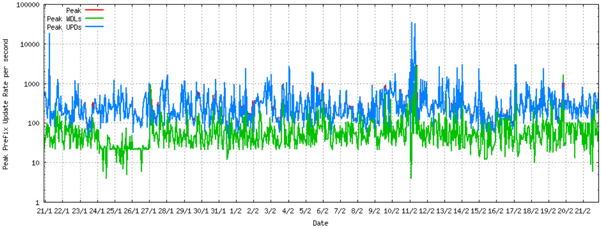


(*) RIB/FIB ratio can vary from ~3 to 30 (function of number of BGP peering sessions at sample point)

Source: BGP Routing Table Analysis Reports - http://bgp.potaroo.net/index-bgp.html

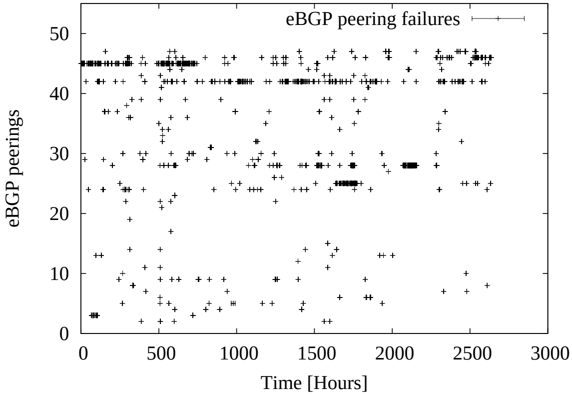
In practice...

- **Static**: DFZ routing tables
  - 300.000 prefix entries (growing at ~20-25% per year)
  - 30.000 ASs (growing ~15-20% per year)
- **Dynamics BGP updates** (routing convergence)
  - Average: 2-3 per sec. – Peak: O(1000) per sec.
  - BGP suffers from churn which increases load on BGP routers (due to link/nodes failures and traffic engineering)
  - BGP's path vector amplifies these problems



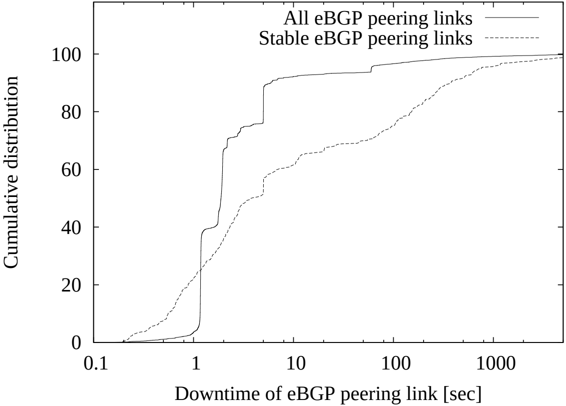
28-08-2009 19

One of the causes → failures of eBGP peerings



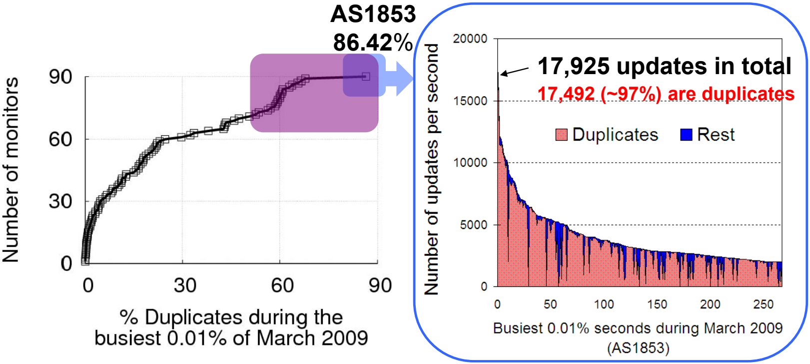
- Source: O. Bonaventure, C. Filsfils, and P. Francois, "Achieving Sub-50 Milliseconds Recovery Upon BGP Peering Link Failures," ACM CoNEXT 2005, Toulouse, France, October 2005.

Downtime of eBGP peering links



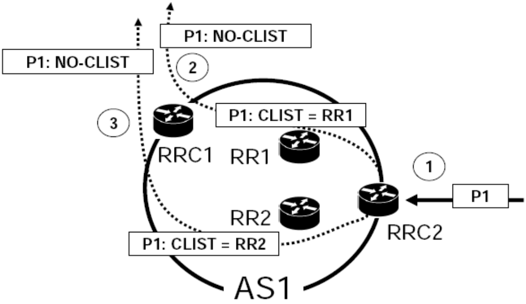
- Source: O. Bonaventure, C. Filsfils, and P. Francois, “Achieving Sub-50 Milliseconds Recovery Upon BGP Peering Link Failures,” ACM CoNEXT 2005, Toulouse, France, October 2005.

Other causes ... duplicates ...



- Source: Danny McPherson (ARBOR Networks) “Prefixes, Paths & Internet Routing System Scalability,” ARIN 25, April 2010.

Other causes ... duplicates ... (cont.)



- Source: J. H. Park, D. Jen, M. Lad, S. Amante, D. McPherson, and L. Zhang "Investigating occurrence of duplicate updates in BGP announcements," Passive and Active Measurement Conference (PAM), Zurich, Switzerland, 2010.

Cross dependencies

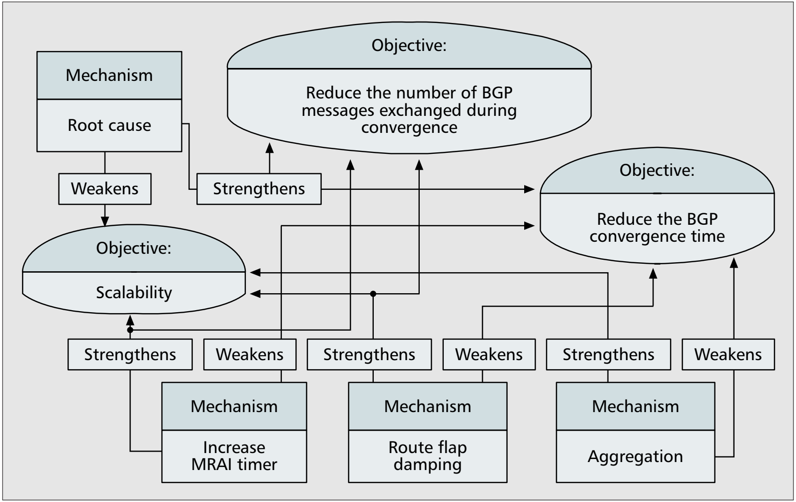


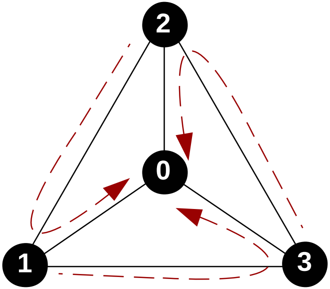
Figure 3. *The complex and still unsolved balance between three interdomain routing objectives.*

- Source: M. Yannuzzi et al. "Open issues in interdomain routing: a survey," IEEE Network, Nov./Dec. 2005.

Routing Policies and Traffic Engineering Limitations

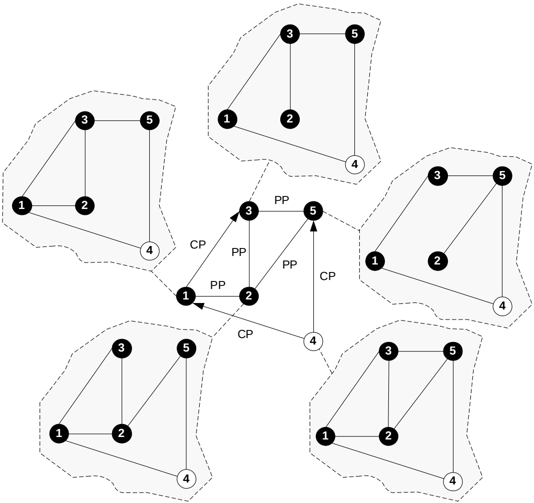
The effects of routing policies

- The example of the bad gadget



- Source: T. Griffin T, and G. Wilfong G, “An Analysis of BGP Convergence Properties,” ACM/SIGCOMM, Cambridge MA, USA, 1999.

The effects of routing policies (cont.)



- Source: M. Yannuzzi, R. Serral-Gracia, and X. Masip-Bruin, ”Chapter 3: Distance and Path Vector Routing Models,” to be published in the book "MULTI-DOMAIN NETWORKS: A PRACTICAL PERSPECTIVE," Springer Series, Series Ed.: B. Mukherjee, Eds: N. Ghani, M. Peng, and I. Monga.

Limited Traffic Engineering (TE) functionality

Limited Traffic Engineering (TE) functionality

Limited Control

Limited Traffic Engineering (TE) functionality

Limited Control

- BGP only offers a limited set of TE functionalities, whose effects are rarely predictable beyond the local domain.

Limited Traffic Engineering (TE) functionality

Limited Control

- BGP only offers a limited set of TE functionalities, whose effects are rarely predictable beyond the local domain.
- Basic TE requirements, such as route control remain unsolved in practice.

Limited Traffic Engineering (TE) functionality

Limited Control

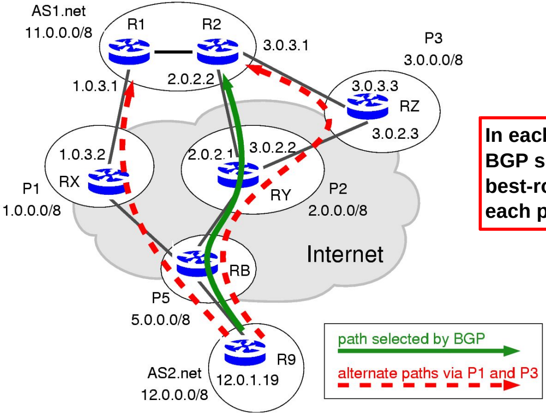
- BGP only offers a limited set of TE functionalities, whose effects are rarely predictable beyond the local domain.
- Basic TE requirements, such as route control remain unsolved in practice.
- A BGP router only advertises its best path toward a destination, i.e., the path contained in its FIB which the one used by the router to forward traffic to the destination. Clearly, this improves the overall scalability of the routing system, but adversely reduces the number of paths that can be used for improving the performance and reliability of inter-domain traffic.

Limited Traffic Engineering (TE) functionality

Limited Control

- BGP only offers a limited set of TE functionalities, whose effects are rarely predictable beyond the local domain.
- Basic TE requirements, such as route control remain unsolved in practice.
- A BGP router only advertises its best path toward a destination, i.e., the path contained in its FIB which the one used by the router to forward traffic to the destination. Clearly, this improves the overall scalability of the routing system, but adversely reduces the number of paths that can be used for improving the performance and reliability of inter-domain traffic.
- Business-driven competition between domains together with the potentially conflicting nature of routing policies, make the accurate control of inter-domain routing an extremely hard problem to solve.

Limited TE control



In each router, BGP selects a single best-route towards each prefix.

- Source: B. Quoitin and O. Bonaventure, “A Cooperative Approach to Interdomain Traffic Engineering,” 1st Conference on Next Generation Internet Networks Traffic Engineering (NGI 2005), Rome, Italy, April 18-20th 2005.

Routing limitations of Path Vector Protocols

Routing limitations of Path Vector Protocols

Routing limitations of Path Vector Protocols

- With the current implementation of BGP, a router has no means to find inter-domain paths subject to constraints, such as:

Routing limitations of Path Vector Protocols

- With the current implementation of BGP, a router has no means to find inter-domain paths subject to constraints, such as:
    - paths with a certain amount of available bandwidth

Routing limitations of Path Vector Protocols

- With the current implementation of BGP, a router has no means to find inter-domain paths subject to constraints, such as:
  - paths with a certain amount of available bandwidth
  - with bounded delay

Routing limitations of Path Vector Protocols

- With the current implementation of BGP, a router has no means to find inter-domain paths subject to constraints, such as:
  - paths with a certain amount of available bandwidth
  - with bounded delay
  - bounded losses

Routing limitations of Path Vector Protocols

- With the current implementation of BGP, a router has no means to find inter-domain paths subject to constraints, such as:
  - paths with a certain amount of available bandwidth
  - with bounded delay
  - bounded losses
  - ... or combinations of these.

Routing limitations of Path Vector Protocols

- With the current implementation of BGP, a router has no means to find inter-domain paths subject to constraints, such as:
  - paths with a certain amount of available bandwidth
  - with bounded delay
  - bounded losses
  - ... or combinations of these.
- The protocol also lacks multi-path routing capabilities, and therefore, the traffic cannot be balanced among different paths—except for specific settings and vendor implementations.

Routing limitations of Path Vector Protocols

- With the current implementation of BGP, a router has no means to find inter-domain paths subject to constraints, such as:
  - paths with a certain amount of available bandwidth
  - with bounded delay
  - bounded losses
  - ... or combinations of these.
- The protocol also lacks multi-path routing capabilities, and therefore, the traffic cannot be balanced among different paths—except for specific settings and vendor implementations.
- This restriction also disables the possibility of finding and establishing primary and protection paths for critical communications.

Routing limitations of Path Vector Protocols

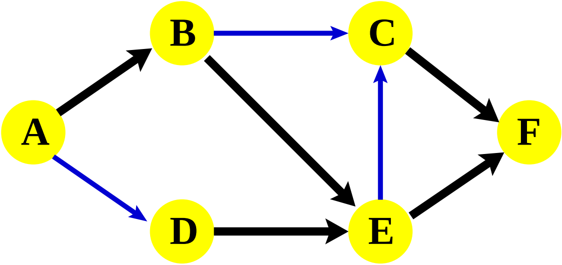
- With the current implementation of BGP, a router has no means to find inter-domain paths subject to constraints, such as:
  - paths with a certain amount of available bandwidth
  - with bounded delay
  - bounded losses
  - ... or combinations of these.
- The protocol also lacks multi-path routing capabilities, and therefore, the traffic cannot be balanced among different paths—except for specific settings and vendor implementations.
- This restriction also disables the possibility of finding and establishing primary and protection paths for critical communications.

Routing limitations of Path Vector Protocols

- With the current implementation of BGP, a router has no means to find inter-domain paths subject to constraints, such as:
  - paths with a certain amount of available bandwidth
  - with bounded delay
  - bounded losses
  - ... or combinations of these.
- The protocol also lacks multi-path routing capabilities, and therefore, the traffic cannot be balanced among different paths—except for specific settings and vendor implementations.
- This restriction also disables the possibility of finding and establishing primary and protection paths for critical communications.

- QoS Routing (QoSR): cumbersome and expensive both in CAPEX and OPEX .... providers have preferred to simplify the operation and maintenance of their networks and relied on capacity overprovisioning for improving the performance and reliability of their services.

Single path in the FIB (... and the only one advertised)



- Source: W. Xu and J. Rexford, “MIRO: Multi-path Interdomain ROuting,” ACM SIGCOMM, Pisa, Italy, September 2006.

Limited expressiveness of policies and TE control

	Traffic	Scope	Predictability	Scalability	Robustness	Efficiency
BGP-based approaches						
Local-Pref	Out	Domain	✓	✓	✓	
IGP weights	Out	Domain	✓	(✓)	✓	
Sel. announcements	In	Internet	✓			Not robust to access link failure.
More spec. prefixes	In	Internet			✓	Sensitive to filtering
MED	In	Neighbor(s)	✓	✓	(✓)	Requires bilateral agreement(s)
AS-Path prepending	In	Internet		✓	✓	Limited granularity (given the diameter of the Internet). Impact difficult to predict.
Communities	In	Internet		✓	✓	Impact difficult to predict. Large search space.
Non BGP-based approaches						
RON, Detours	In/Out	Internet	✓		✓	Require modifications to end-systems. Rely on a large number of IP tunnels.
NAT	In	Internet	✓			Target multi-homed enterprise networks. Poses problem when one access link fails.
New architectures	In/Out	Internet	✓	✓	✓	Difficult to deploy in the current Internet.

- Source: B. Quoitin, “BGP-based Interdomain Traffic Engineering,” Doctoral Thesis, Louvain-la-Neuve, Belgium, 2006.

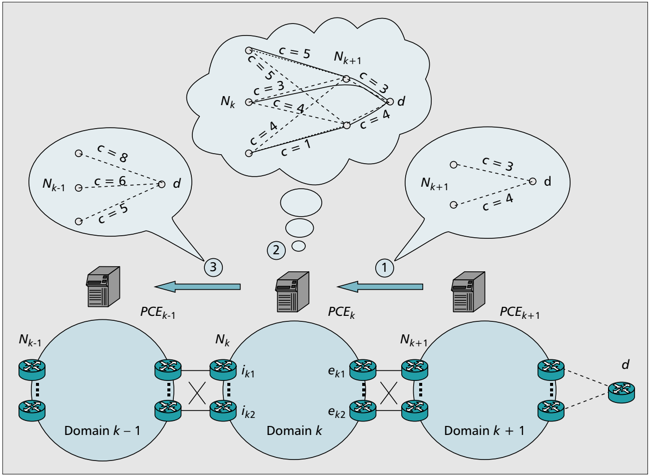


Figure 1. *Reference interdomain scenario and PCE-based computation scheme.*

- F. Ricciato, U. Monaco, and D. Alì, “Distributed Schemes for Diverse Path Computation in Multidomain MPLS Networks,” IEEE Communications Magazine, vol. 43, no. 6, pp. 138 - 146, June 2005.

Related Work: Ricciato et al. (2005) (cont.)

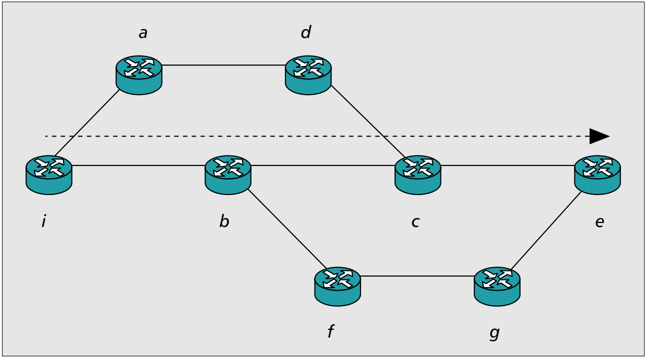


Figure 2. *Trap topology: The shortest path from i to e across b–c leaves the residual graph disconnected. Therefore, sequential computation fails to compute the diverse pair (i-a-d-c-e and i-b-f-g-e).*

- F. Ricciato, U. Monaco, and D. Alì, "Distributed Schemes for Diverse Path Computation in Multidomain MPLS Networks," IEEE Communications Magazine, vol. 43, no. 6, pp. 138 - 146, June 2005.

Related Work: IEEE Infocom (2007)

Problem: 2 Link Disjoint Paths

Given a source node s and a destination node t, find two link-disjoint (s, t)-paths p_1 and $_2$ of minimum total weight $W(p_1) + W(p_2)$.

- The path with minimum weight can be used as the primary path and the second one as the backup path.
- A relevant problem is to find two paths p_1 and p_2 that minimize $\max\{W(p_1), W(p_2)\}$. The solution to this problem can achieve a better balance between the delay of the primary and backup path, but this problem is *NP*-hard.
- The standard algorithm used for solving this problem is the one provided by Suurballe and Tarjan (full topology must be known to every node in the network).

Security Issues...

Security vulnerabilities

Add-on instead of built-in

- BGP lacks both path and origin authentication
- A BGP router can be perfectly used to advertise any possible (prefix, path vector) pair to the Internet
- This makes the inter-domain routing system extremely vulnerable to certain attacks, since both IP prefixes and routes can be hijacked

Adding security

Proposals	Origin authentication				Path authentication			
	Design	Security	Overhead		Design	Security	Overhead	
			Time	Space			Time	Space
S-BGP	Hierarchical PKI local memory	Strong	Low	High	Signatures in message	Strong	High	High
soBGP	Hierarchical PKI separate database	Strong	Low	Low	Topology map	Medium	Low	Low
psBGP	Distribute PALs local memory	Medium	Low	High	Signatures bit vector	Strong	Low	Very high
IRV	Separate IRV servers	Strong	Low	Low	Distributed database	Medium	High	Low
OA	Delegation OATs in message	Strong	Low	High	–	–	–	–
S-A	–	–	–	–	Signature bit vector hash tree	Strong	Low	Very high
APA	–	–	–	–	Aggregate signature, bit vector, hash tree	Strong	Low	Medium
SPV	–	–	–	–	Hash chain hash tree one-time signature	Medium	Low	Very high
Listen Whisper	–	–	–	–	Consistency check TCP flow	Low	Low	Low

- Source: M. Zhao, S. W. Smith, and D. M. Nicol, “The performance impact of BGP security,” IEEE Network, vol. 19, no. 6, Nov./Dec. 2005.

Adding security (cont.)

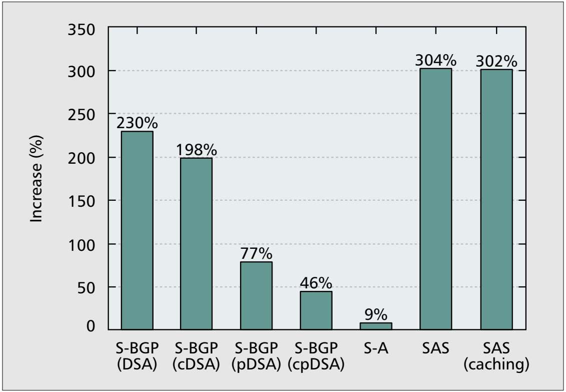


Figure 1. *Relative increase in convergence time of path authentication schemes relative to ordinary BGP.*

- Source: M. Zhao, S. W. Smith, and D. M. Nicol, "The performance impact of BGP security," IEEE Network, vol. 19, no. 6, Nov./Dec. 2005.

Adding security (cont.)

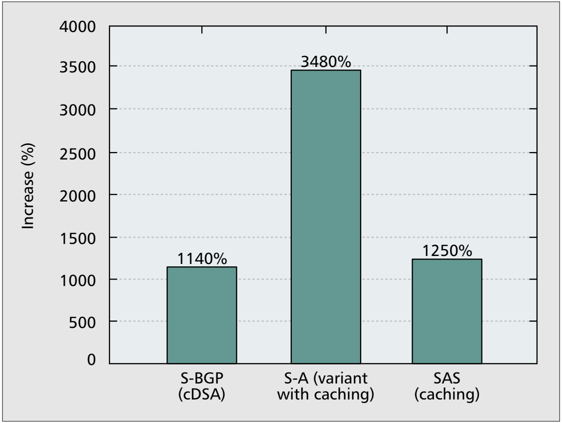
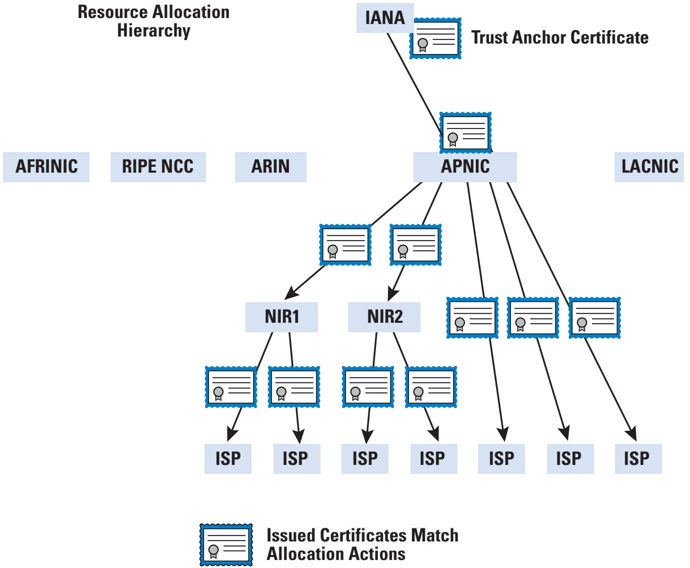


Figure 2. *Relative increase in memory costs of path authentication schemes relative to ordinary BGP.*

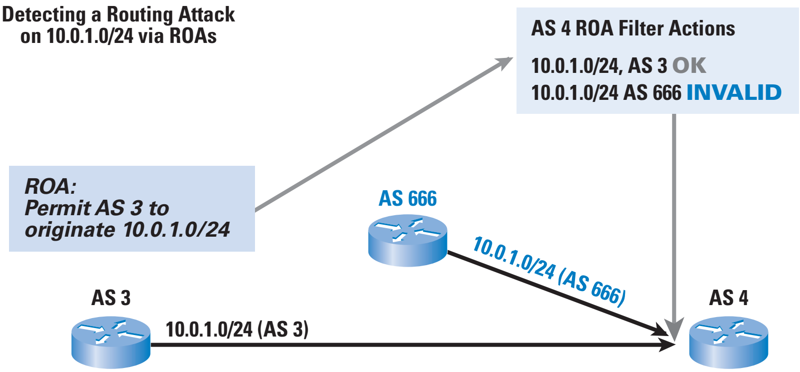
- Source: M. Zhao, S. W. Smith, and D. M. Nicol, “The performance impact of BGP security,” IEEE Network, vol. 19, no. 6, Nov./Dec. 2005.

Current Activities at the IETF...



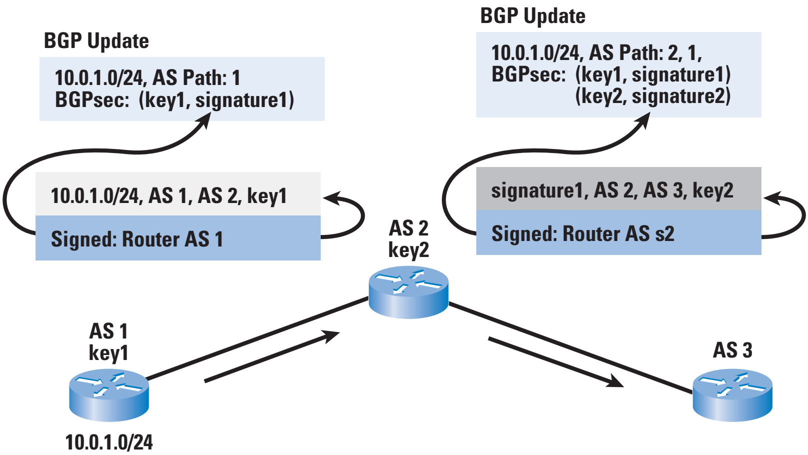
- Source: G. Huston and R. Bush, “Securing BGP with BGPsec,” Internet Protocol Journal, June 2011.

Current Activities at the IETF...



- Source: G. Huston and R. Bush, "Securing BGP with BGPsec," Internet Protocol Journal, June 2011.

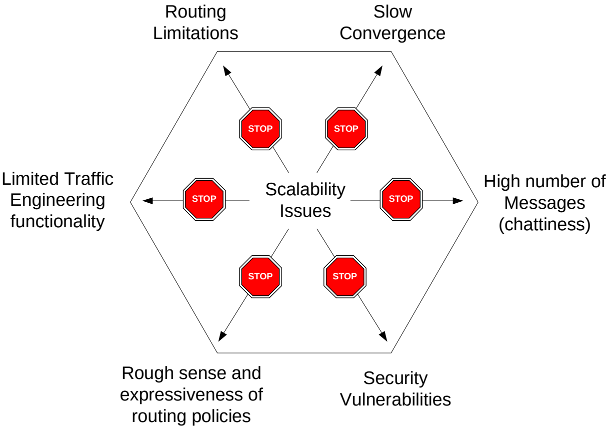
Current Activities at the IETF...



- Source: G. Huston and R. Bush, “Securing BGP with BGPsec,” Internet Protocol Journal, June 2011.

The atomic approach ... a big mistake ...

- Cross dependencies are strong, issues cannot be addressed isolatedly ...



Questions?