Routing in the Future Internet

Marcelo Yannuzzi

Graduate Course (Slideset 7)
Institute of Computer Science
University of the Republic (UdelaR)

August 27th 2012, Montevideo, Uruguay

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

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

Review of the Readings

- First Reading
- Second Reading

Review of the Readings

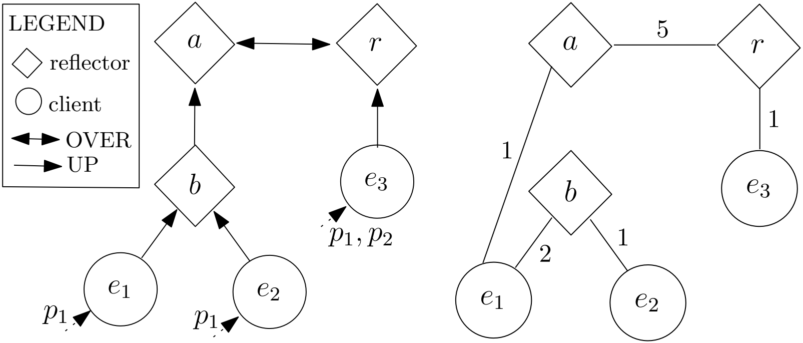
- **First Reading**
- Second Reading

Route Reflectors....

“Each iBGP router propagates its best route according to the following rules:

if the route is learned from a peer or from a route-reflector, then it is relayed only to clients, otherwise it is reflected to all iBGP neighbors."

iBGP Deceptions

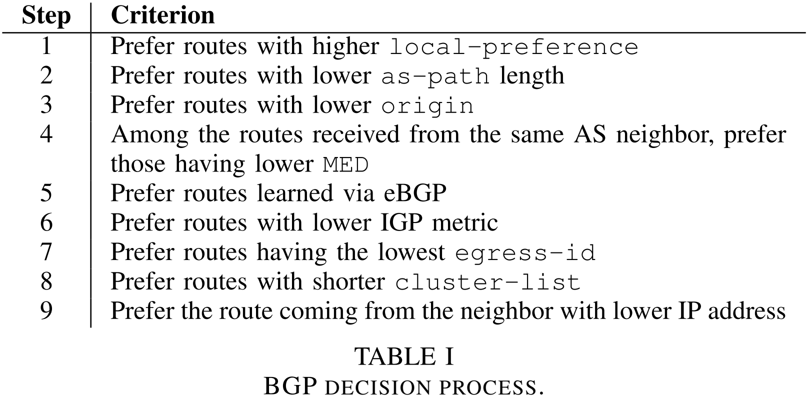


(a) iBGP topology (b) IGP topology

Fig. 1. A simple network that exhibits visibility issues.

- Source: S. Vissicchio et al., “iBGP Deceptions: More Sessions, Fewer Routes,” IEEE INFOCOM 2012.

iBGP Deceptions (cont.)

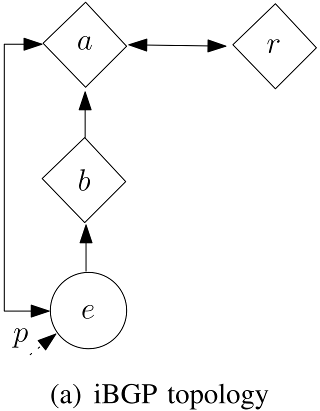
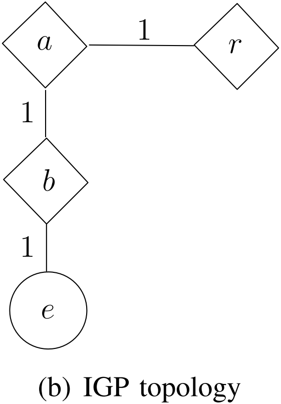


Step	Criterion
1	Prefer routes with higher `local-preference`
2	Prefer routes with lower `as-path` length
3	Prefer routes with lower `origin`
4	Among the routes received from the same AS neighbor, prefer those having lower `MED`
5	Prefer routes learned via eBGP
6	Prefer routes with lower IGP metric
7	Prefer routes having the lowest `egress-id`
8	Prefer routes with shorter `cluster-list`
9	Prefer the route coming from the neighbor with lower IP address

TABLE I
BGP DECISION PROCESS.

- Source: S. Vissicchio et al., “iBGP Deceptions: More Sessions, Fewer Routes,” IEEE INFOCOM 2012.

iBGP Deceptions (eBGP NH remains unchanged)



(a) iBGP topology

(b) IGP topology

Fig. 2. OVER-RIDE GADGET

- Source: S. Vissicchio et al., “iBGP Deceptions: More Sessions, Fewer Routes,” IEEE INFOCOM 2012.

Model...

“We model an iBGP topology as a directed labeled multigraph $B = (V, E)$ where nodes in V represent routers and edges in E represent iBGP sessions."

“We define a valid signaling path as a path $(u, \ldots, v)$ on B that can be used to advertise routes from u to v (or vice versa). A **valid signaling path** consists of zero or more UP sessions, followed by zero or one OVER session, followed by zero or more DOWN sessions. This means that a valid signaling path matches regular expression:

UP*OVER?DOWN*"

Signaling correctness: The BGP configuration is free from routing anomalies, i.e., BGP is guaranteed to always converge to a single predictable stable state.

Forwarding correctness: Guarantees the absence of packet deflections along the forwarding path.

Sufficient Conditions for Correctness

- Set of sufficient conditions that guarantee that an iBGP topology *B* is both signaling and forwarding correct:

  1. *B* has no cycles consisting of UP sessions only
  2. Any route-reflector prefers paths propagated by its clients over paths propagated by non-clients
  3. All shortest paths must also be valid signaling paths.

- Note that Conditions 1 and 2 ensure that the iBGP configuration is signaling correct, while Condition 3 guarantees forwarding correctness.

  ...in Condition 3 there is an issue with the graphs (remember that valid signaling paths are defined on *B*) ...

Sufficient Conditions for Correctness (cont.)

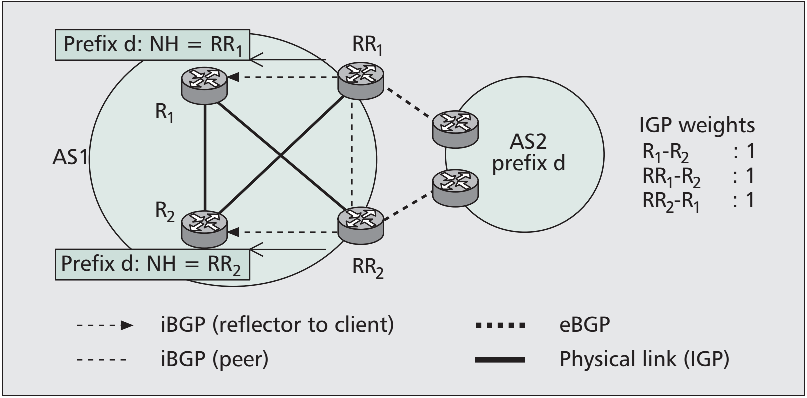
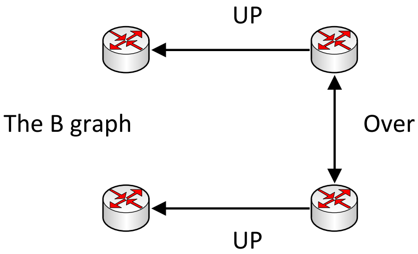


Figure 2. *Route reflection with data forwarding loop.*

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

Sufficient Conditions for Correctness (cont.)



1) B has no cycles consisting of UP sessions only

2) Any route-reflector prefers paths propagated by its clients over paths propagated by non-clients

3) All shortest paths must also be valid signaling paths (on B?).

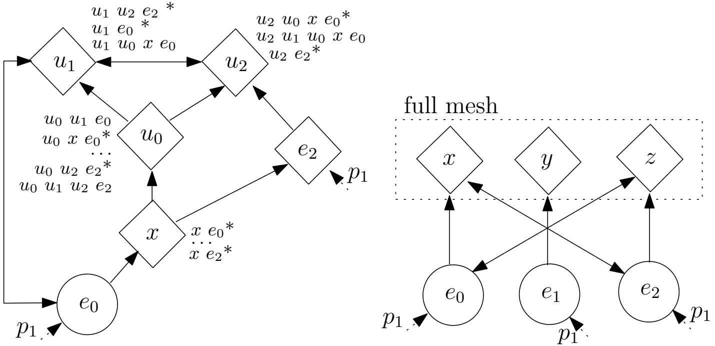
Dissemination Correctness...

“Let *B* be a signaling correct iBGP topology. Then, *B* is dissemination correct if all the routers in *B* are guaranteed to receive at least one route to prefix *p* in the stable state, for any non-empty set of egress points for *p*".

The authors claim that a signaling correct topology is not guaranteed to be dissemination correct. Moreover, a dissemination correct topology is not guaranteed to be forwarding correct.

Another Pathological Example...

- Why are those the preferences in the first place??



(a) A spurious OVER can create routing oscillations.

(b) A spurious OVER can cause forwarding loops.

Fig. 3. Two cases in which adding a spurious OVER creates signaling and forwarding anomalies.

- Source: S. Vissicchio et al., “iBGP Deceptions: More Sessions, Fewer Routes,” IEEE INFOCOM 2012.

Dissemination Correctness...intractability

Dissemination Correctness Problem (DCP): Given a signaling correct iBGP topology B and the underlying IGP topology I, decide if B is dissemination correct. One More Session Problem (OMSP): Given a dissemination correct iBGP topology $B = (V, E)$, the underlying IGP topology I, and a spurious OVER session $o = (x, y), x, y \in V$, decide if $B' = (V, E \cup (x, y))$ is dissemination correct.

Dissemination Correctness...intractability (cont.)

In practice, intractability is not necessarily an issue:

- NP-completeness only refers to the run-time of the **worst case instances**.....note that many of the instances that occur in practical applications can be solved in polinomial time!
- We also need to distinguish between online computations and offline computations.....
- F.A. Kuipers, “Quality of Service Routing in the Internet: Theory, Complexity and Algorithms", Ph.D. thesis, Delft University Press, The Netherlands, ISBN 90-407-2523-3, September 2004.

Guidelines...

- “In redundant iBGP configurations, redundant route-reflectors must belong to the same cluster in order to enforce the prefer-client condition."
- “Whenever an additional session is needed to solve visibility issues, an UP session should be deployed, in order to enforce the no-spurious-OVER condition."

Review of the Readings

- First Reading
- **Second Reading**

Highlights...

“Each network advertises pathlets—fragments of paths represented as sequences of virtual nodes (vnodes) along which the network is willing to route. A sender concatenates its selection of pathlets into a full end-to-end source route."

“Pathlet routing can be seen as source routing over a virtual topology whose nodes are vnodes and whose edges are pathlets."

- It enables an exponentially large number of path choices.
- It offers flexibility, routing scalability (smaller FIBs), and source-controlled routing.
- It supports complex routing policies.

Pathlet Routing (Forwarding Identifiers (FIDs))

A, C, and D have policies that are local, i.e., they depend only on their neighbors. B has a BGP-like policy which depends on the destination: it allows transit from B to C only when the ultimate destination is E.
Sort of MPLS push/pop forwarding style...based on one-hop/multi-hop pathlets

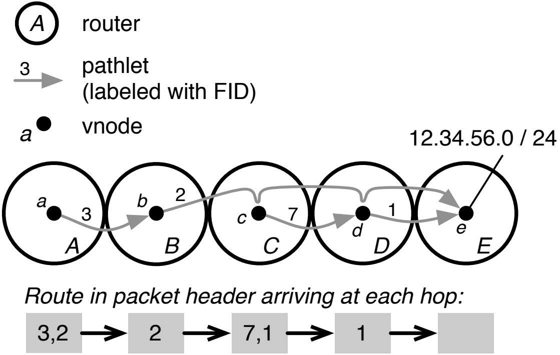


Figure 1: A pathlet routing example.

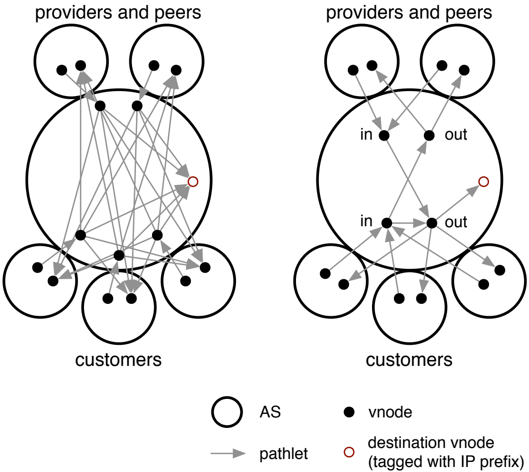
- Source: P. Brighten Godfrey et al. “Pathlet Routing,” ACM SIGCOMM 2009.

The Basics...

- It uses path vectors to disseminate pathlets....much as BGP notifies the Internet of the existence of IP prefixes....
  1. Announce pathlets which form a **shortest path tree from *v* to all destination vnodes reachable from *v***.
  2. Announce any additional pathlets that are reachable from *v*, up to limit(δ) pathlets originating at each AS with δ AS-level neighbors (the authors use: limit(δ) = 10 + δ).....**recall the abstract...“It enables an exponentially large number of path choices."....**
- Pathlet advertisements contain: the pathlet’s FID and its sequence of vnode identifiers.

The Basics (cont.)

“Two ways to implement a local transit policy are to connect the appropriate ingress-egress pairs (left), or to group neighbors into classes and connect the appropriate classes (right). Here we show the vnodes and pathlets in one AS to permit valley-free routes."...**note that BGP can do the same! (right)**



- Source: P. Brighten Godfrey et al. “Pathlet Routing,” ACM SIGCOMM 2009.

Weaknesses...tons of...

- 1) Disruptive....it requires drastic changes in hardware (new processors, new equipment,.....) ...note that now routers do not route based on IP prefix destinations.......since Pathlet advertisements contain: the pathlet's FID and its sequence of vnode identifiers....
- 2) ...magically...IP prefixes are out of the picture...so requirements such as mobility are someone else's problem... ☺
- **(Page 2)** In Fig. 1....how are packets delivered once they arrive at e? Note that they ingress with an empty route...
- **(Page 3)** What if we want to send part of the traffic from Y to Z and part to Z' and Z"? The problem is that we lost control based on the destination (prefixes)...or once we put prefixes into the picture we might scale even worse than we do today...

Weaknesses...tons of...(cont.)

- **(Page 4)** “A simple way to do this is to build a graph in which each vnode is a node, and each pathlet $v1 \rightarrow \cdots \rightarrow vn$ is a single edge $v1 \rightarrow vn$ (perhaps given a cost equal to the number of ASes through which the pathlet travels). Then, similar to link state routing, run a shortest path algorithm on this graph to produce a sequence of edges (i.e., pathlets) to each destination. After the router has made its path selection, it places the sequence of FIDs associated with the chosen pathlets into the packet header, and sends it."

Recall that shortest paths based on the AS length are often not the ones that show the best performance...

Weaknesses...tons of...(cont.)

- **(Page 4)** Not clear how loops are detected and handled in practice (note that a pathlet is a sequence of “virtual nodes"....)
- **(Page 4)** The simplest optimization is that we never need to switch to a more preferred dissemination path, since they are all equally acceptable...wow!
- **(Page 7)** QoS....
- **(Pages 9 and 10)** **Results are biased!**...consider multi-connectivity between domains and IP prefix reachability with TE objectives...especially when the address space is break down into more specific prefixes and scattered inside the ASs (e.g., hundreds or even thousands of nets for a class B)...for sure the authors used single node abstraction in their results...

Weaknesses...tons of...(cont.)

- **(Page 5)** Local Transit Policies....**too weak, too vague**...no clue about how it can be implemented considering IGPs and pathlets....no clue about how IP prefix destinations may affect the overall routing decisions and scalability...
- **(Page 6)** Indeed...“If it owns an IP prefix, then it has a second vnode *w* tagged with the prefix, from which no pathlets depart." .... so we might need lots of *w* vnodes due to prefixes
- **(Page 5)** The authors claim that the primary disadvantage is that policies cannot depend on a route’s destination....whereas an Internet routing system “must have that".

In Summary...

A new interdomain routing scheme.....

Does Pathlet Routing...

- ...solve the churn issue? **No**. Even with biased experiments the plots show that the results for pathlet are worse.
- ...solve the convergence issue? **No**. Destinations (IP prefixes) are basically out of scope (loseely treated)....so no clue about it.
- ...solve the security vulnerabilities of the Internet's routing system? **No**. Actually, it will suffer from the same issues as BGP (route attestations, route origination, and route dissemination and propagation problems)
- ...improve intra/inter TE objectives? **No**. TE is out of the scope of this paper.
- ...improve internal routing apects (compared to iBGP....RR, iBGP/IGP interactions, route deflections, oscillations, etc.). **No**. Actually this is not even described as it deserves....
- ...support partial deployments? **No**. The paper provides no clue on how to transition from BGP-4 to pathlets.

Then, why does the community need this paper?? This paper looks like the X-files .... yields more questions than answers .... SIGCOMM is highly overrated....

Questions?