MINT: A Market for Internet Transit

Nick Feamster
Georgia Tech

Joint work with Vytautas Valancius, Ramesh Johari, Vijay Vazirani

Goals for Interdomain Routing

- Edge networks/users need **end-to-end paths with service level assurance**
  - Mechanism for verifying this level of service
- Service providers need **profitability**

- Requirements
  - Scalability (routing table size, churn, etc.)
  - Security
  - Manageability

Today: BGP has Market Inefficiencies

- Pair of ASes may decide to terminate connectivity arrangement
  - Even if end nodes would pay for the path to be there!

October 2005

31 Jul 2005: Level 3 Notifies Cogent of intent to disconnect.
16 Aug 2005: Cogent begins massive sales effort and mentions a 15 Sept. expected depeering date.
31 Aug 2005: Level 3 Notifies Cogent again of intent to disconnect (according to Level 3)
5 Oct 2005 9:50 UTC: Level 3 disconnects Cogent. Mass
hysteria ensues up to, and including policymakers in Washington, D.C.
7 Oct 2005: Level 3 reconnects Cogent

During the "outage", Level 3 and Cogent's singly homed customers could not reach each other. (~ 4% of the Internet's prefixes were isolated from each other)

April 2007

renesys | blog

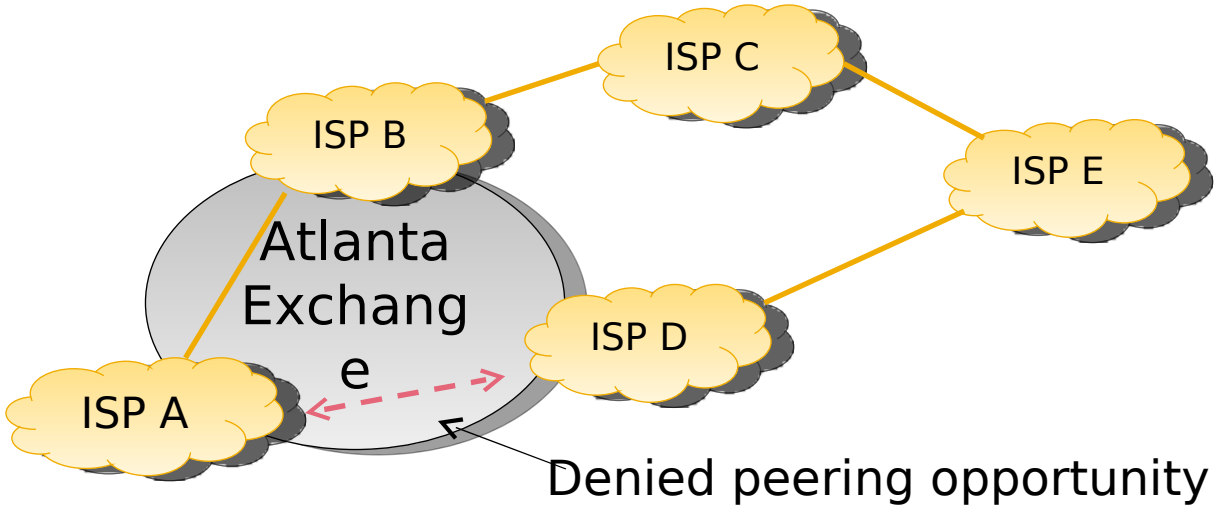
ROUGH DAY AT COGENT PART 1: COGENT DEPEERS SOMEONE?

These are the signs of the apocalypse: A worldwide earthquake, the sun ceasing to emit visible light, cats and dogs living together in harmony, and Cogent (AS174) depeering another AS. At least one of these happened earlier this week.

At about 10:00 UTC on Tuesday (6am EDT), Cogent depeered a couple of smaller, UK-based ISPs without notice. This was apparently intentional and due to a review of existing peers and whether they meet peering policies. Does this mean that Cogent is becoming more like its larger competitors that it so enjoys taunting? I'll take a look at who was depeered and speculate on why.

BGP has Connectivity Inefficiencies

- Denied peering opportunities exist in every exchange
  - Disagreements over payment direction
  - Bilateral nature of contracts introduces information asymmetry



Reconsidering the Abstraction

- **Path Segments:** Unit of connectivity
  - Each independently operated network advertises connectivity between two intermediate points (ingress, egress, price, <path properties>)

- **Intermediaries:** Mechanism for “stitching” paths together to match segments to end-to-end paths

Benefits

- Independent innovation and evolution
  - Each independently operated network can deploy its own protocols within the network
- Improved isolation
  - As long as independently operated network can maintain the abstraction, no need for readvertisement
- Direct value transfer
  - Possibility of directly paying for and end-to-end path

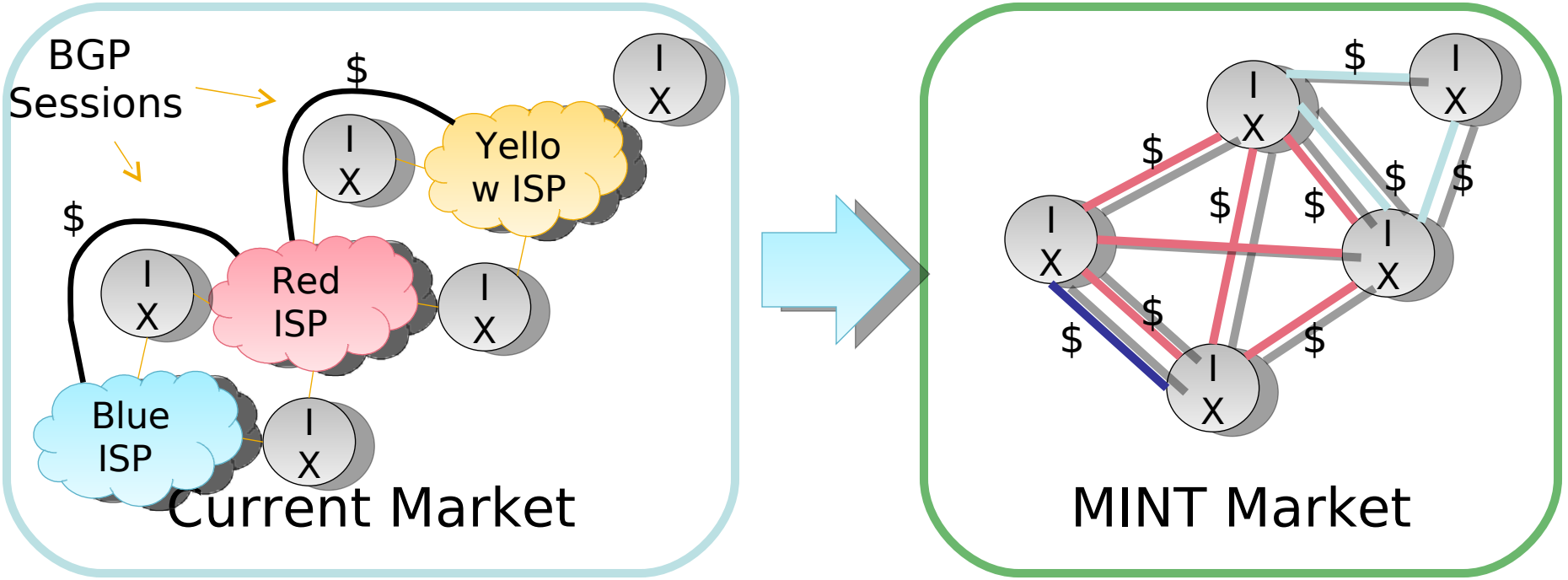
MINT in a Nutshell

- Replace bilateral contracts with **path auctions**
- **Sellers**
  - Sell **segments** from exchange to exchange
- **Buyers**
  - Buy connections of segments that form **paths**
  - Use legacy BGP for best-effort flows

- Market and connectivity efficiency
  - End networks can directly express their valuation of network-to-network paths
  - No incentive to de-peer as long as end-networks are valuing the paths

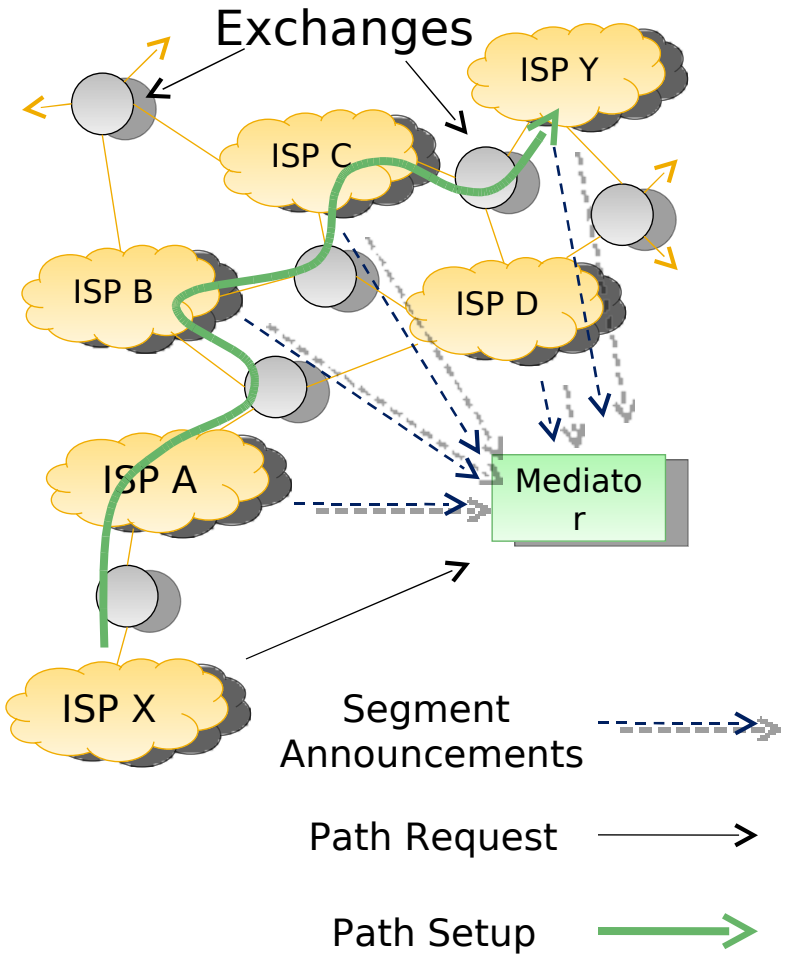
What Good Should be Priced?

- Current market: pricing connections
  - No control to end-networks, coarse granularity
- MINT market: pricing segments
  - High granularity, possibility to value/construct entire paths
  - Pricing bandwidth, delay, loss or any combination



Market

- Modeling Internet is an Auction
  - **Sellers** advertise prices (or “asks”) for each segment
  - **Buyers** issue “bids” for paths
- Auction properties:
  - **Reverse** and **Continuous:** ISPs are setting the prices to attract or repel traffic
  - **Combinatorial**: bids are for set of goods
  - **First-price:** the lowest cost path is chosen



The Mediator

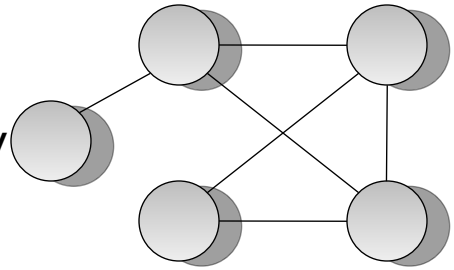
- Mediator runs the auction, matches bids and asks
  - Bidding for price and Boolean constraints (i.e. bandwidth, diversity)
    - First-price: O(E+Vlog(V))
    - VCG: O(E(E+Vlog(V))) – truthful, but slow and expensive
  - Bidding for price and additive/multiplicative constraints: NP-Complete, approximations available

- Solution:
  - Charge for path requests
  - Allow multiple mediators to compete

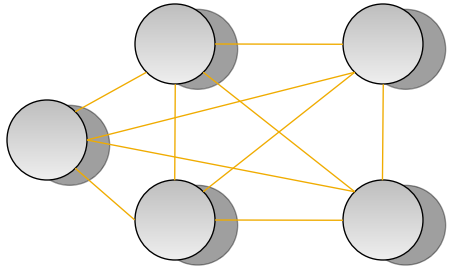
Control Plane: Path Segments

- Link-state protocols to dimension available segments
  - Reuse OSPF-TE, ISIS-TE
- Export the segment prices to mediator to attract or reduce traffic
  - Charge for each update
  - Number of states can be reduced using virtual nodes (Godfrey et. al.)

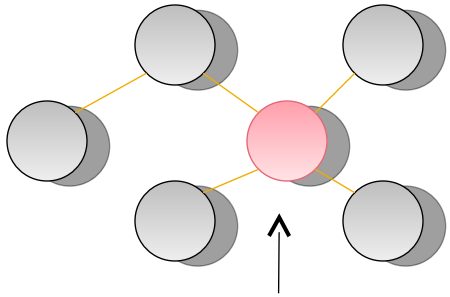
Physical Topology



Full Segment Announcement



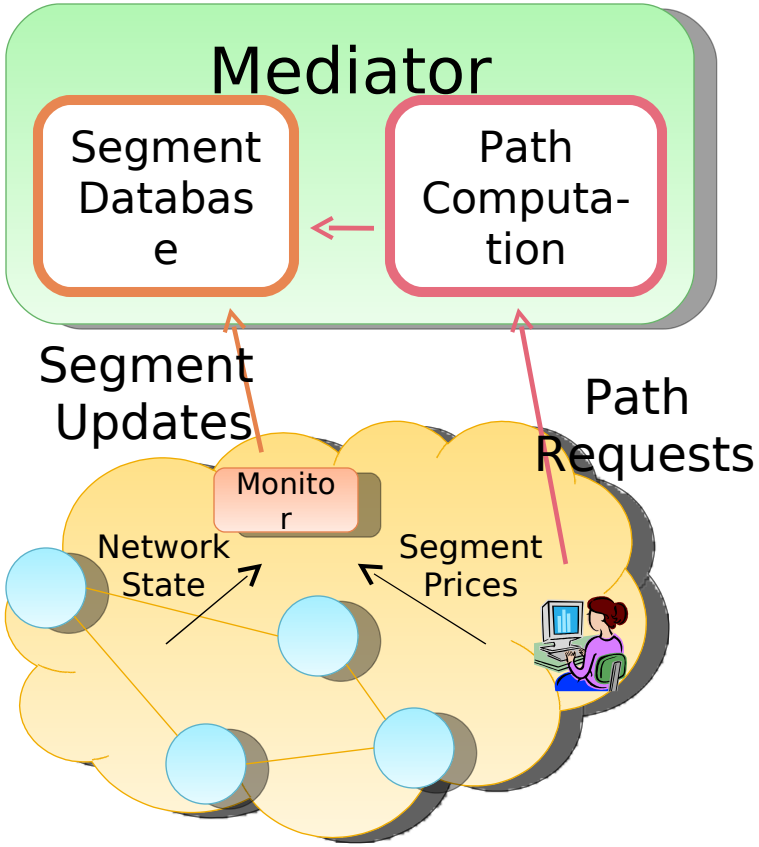
Reduce state with Virtual Nodes



Virtual Node

Control Plane: Computing Paths

- Separate segment database and path computation elements
  - Computation triggered only by path requests
  - Path computation works only on a snapshot of the database



Data Plane

- Scaling segment to node mapping
  - Path Computation Element (PCE)
- Scaling number of paths
  - 1,000,000's of tunnel tags per Interface (Cisco, Juniper)
  - 100,000's new path per second (Wang et. al. 2004)
  - PCE architecture to load paths between the ingress points
  - 50,000 networks, 1% significant destinations = only 25M paths for the whole Internet

Data Plane

- End-networks are responsible for establishing and maintaining the paths
  - Mediator provides only segment level information that needs to be mapped to link-level paths
  - Any tunneling technology would do
    - MPLS scales to millions of labels
- Segment to link mapping
  - Ingress node computation, or
  - PCE architecture (as in RFC4655)
    - Dedicated computing element
    - Ingress/egress selection, node load balancing
- Reservation protocol depends on node-to-node connections
  - As in BGP, transitive trust on inter-domain boundaries, “path-stitching” (as in RFC 5150)
  - 100,000’s new paths per second (Wang et. al. 2004)
- Failure detection and recovery
  - In-band detection, sub-second switchover to back-up path
  - Without a backup, initiate a new path bid or revert to legacy best-effort

Scalability

- How many segments MINT can distribute?
- How does path computation scale?
- How many paths MINT can support?
- How MINT integrates with current infrastructure?

Conclusion

- BGP has market and connectivity inefficiencies. Does not achieve basic goals for interdomain routing
- Consider segment as an abstraction
- MINT: alternative way of structuring inter-domain bandwidth market
  - Rather trading connectivity, trade transit segments
- Benefits
  - No notion of customer-provider or peer-peer
  - Policy expression though price
- Evaluation
  - Scalable control and data plane
  - Market stability and incentive compatibility

http://www.cc.gatech.edu/research/reports/GT-CS-09-01.pdf