Implications of Source Routing

Onur Ascigil
Laboratory for Advanced Networking
University of Kentucky
onur@netlab.uky.edu

Ken Calvert
Laboratory for Advanced Networking
University of Kentucky
calvert@netlab.uky.edu

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Internet

Keywords

Routing, Topology, Economics

1. REFACTORING ROUTING AND FORWARDING

Over the years, a number of source routing architectures have been put forward [2, 1, 5]. A source routing architecture is one where packets indicate the paths that they follow. Because such architectures seem to offer a number of advantages, especially with respect to competition and evolvability, we believe the case for a source-routed network layer is worth examining in detail.

Our model of source routing consists of packets carrying (loose) provider-level source routes (i.e., interdomain paths) and distributed *path providing servers* that collect routing updates disseminated by transit providers, compute (inter-domain level) paths, and provide those paths to end systems for placement in packets. Thus routing is provided as a service with distributed *path providers*, similar to ROSE [6]. We also assume a mechanism along the lines of Platypus [7] so that providers can verify packets for routing-policy-compliance and proof-of-payments.

While it offers significant benefits, source routing also introduces some new technical and economic challenges. With such a different "factoring" of the routing process, we envision such an architecture to have a different business model in selling transit services to users. Because source routing empowers users to select indirect providers from whom users obtain transit service, users would be required to compensate their indirect providers in addition to the direct providers for the transit services. In order to scale the money transactions between users and all the direct and indirect transit providers, we envision a set of third party resellers, *path brokers*, who buy transit (relaying) service from providers, and then sell paths to users. Path brokers and path providers can be the same entity or can be separate.

A major barrier towards deploying source routing is the "conventional wisdom" that source routing architectures do not scale. This wisdom is mainly due to an assumption that performing route computation at the edges of the network (as opposed to within the network) requires awareness to very short-term changes in the routing state of the network. However, through (possibly in-band) end-to-end measurements, cooperating endpoints can obtain up-to-date routing state of the relevant parts of the topology and can respond to short-term changes in routing state. Therefore, the routing computation can use long-term measurements of routing state (e.g. statistical information) that can be disseminated at coarse time-scales.

With significant changes in routing and economics, we expect differences in the structure of the AS-level topology to arise. In order to understand the possible effects of the above-mentioned architectural changes in the structure of the inter-domain topology, we use a game-theoretic model to simulate the behaviors of transit providers and users. The next section outlines our model.

2. TOPOLOGY EVOLUTION MODEL

We consider a population of N Autonomous Systems (ASes), or just "nodes". Each node in the topology acts as an agent and makes individual decisions that trade-off the costs of actions against the potential pay-off (revenue) from doing so. Possible actions are peering/depeering, link capacity upgrade and so on. Nodes deterministically select a strategy depending on their current costs and pay-off. Similar to our work, ITER [4] use an agent-based model to predict changes in the connections between ASs; however we incorporate price and quality competition between ASs and deploy a different economic model and routing mechanism.

Our model of peering consists of ASs leasing dark fiber to an Internet Exchange Point (IXP), and an Ethernet connection to a public peering fabric to send and receive traffic. Because the traffic demands of stub nodes remain bounded during the time-scales of interest in our model, stub nodes do not upgrade their hardware and simply keep their total peering capacity constant during the evolution of the topology. Transit nodes, on the other hand, are subject to potential changes in transit traffic volumes, therefore they can change their hardware, lease more or less capacity over time. Although they are unable to add new capacity, multi-homed stub nodes are capable of moving (transferring) capacity from one peer to another (simple rewiring on the switching fabric). We model geographical constraints in peering by assigning each node to a set of locations where each location roughly corresponds to an IXP. Two nodes must have a common location if they are to establish a link between each other. In our model, the world consists of a fixed number of locations and the geographical expanse of nodes are selected according to a skewed distribution.

We assume that the traffic flowing between two nodes is the aggregate of many "small" users and we represent the total amount

CoNEXT Student'12, December 10, 2012, Nice, France.

of inter-domain traffic flowing between pairs of stub nodes with a traffic matrix T. However, the user traffic demand is elastic, which means that the demand may decrease (increase) with the increasing (decreasing) cost of sending traffic. Therefore, each element T_{xy} denotes an upper bound on the average volume of traffic demand that is originated by a stub node x and consumed by another stub node y.

In addition to peering and capacity allocation activities, nodes take part in *a price competition* where each node takes turn to set prices for their *relaying services*. A relaying service is the forwarding of packets from an ingress channel to an egress channel. Relay prices indicate the amount charged to *users* for relaying a unit volume (e.g. mbps) of traffic. Relay prices are used by nodes as a knob to control the amount of traffic volumes on the relays. Because we do not model intra-domain channel connectivities and capacities, nodes individually assign prices to the ingress channel and egress channel portions of the relays. However, users are only exposed to relay prices which are the sums of corresponding ingress and egress (inter-domain) channels' prices. With these changes in money flow, peering incentives are different from current Internet, where the direction of the traffic determines the direction of pay-off. In our money flow model, both endpoints of a link gain from traffic flowing between each other regardless of the direction of the traffic as they directly charge users.

Users route their traffic *selfishly* on minimum "cost" paths. We assume for the purposes of this study that quality and price attributes of relays are fully exposed to the users. Similar to the existing literature on selfish routing, we quantify the quality of a channel by its average latency computed as follows: $\frac{1}{C-v} + prop$, where C is the capacity of the channel, v is the average volume of traffic on the channel, and $prop$ is the propagation delay constant. We model the variety in path selection policies of different users with a number of *user traffic classes*. Users belonging to the same traffic class are homogenous in that they have exactly the same path selection policy. We model a path selection policy as a trade-off between quality (i.e. latency) and price.

We approximate the traffic flow in the internetwork with selfish routing with a *traffic flow equilibrium*. A traffic flow equilibrium is a state where no user can improve the cost of its traffic by unilaterally changing the amount of traffic it sends along different network paths. We compute the traffic equilibrium using a multi-threaded variant of Frank-Wolfe algorithm.

Transit nodes incur various costs such as hardware costs (e.g. routers, switches, interface cards), operational costs (e.g. hardware maintenance, power consumption), co-location costs at an IXP and so on. Because these individual costs are very hard to come by, we only estimate the total cost of a node as the cost of its peering links. We approximate the various costs associated with a single peering link p of node x, with a concave step function of p's capacity as $C(p)^{\alpha}$ where $C(p)$ is the capacity of the link p (in mbps). In a previous work [3], costs associated with settlement-free connections are estimated using the same function where α is estimated to be between 0.4 and 0.75. We do not model stub node costs and pay-off and assume that they are profitable.

The topology evolution process proceeds as follows: We first build an initial topology that resembles the current Internet AS graph in terms of AS degree distributions and geographical expanse. Starting from the initial topology, the model goes through an iterative process consisting of i) price competition game and then ii) capacity allocation game. During the price competition game, nodes attempt to make small modifications to their channel prices (one at a time) as long as the modification leads to an increase in pay-off. The price competition continues until no price changes are made on any link, a state which we call *a price equilibrium*. We limit nodes to only increase the prices of over-utilized (utilization $> U_H$) links, decrease the prices of under-utilized (utilization $< U_L$) links and make no changes to the price of channels whose utility is in $[U_L, U_H]$. We assume that keeping channel utilizations within a certain interval is a default strategy for all nodes.

Once the price game reaches an equilibrium, nodes take part in a capacity allocation game where each makes a single "move" in random order. A move consists of capacity allocation/deallocation and peering/depeering decisions according to a strategy. The exact strategy of an AS during its move depends on its current profitability (total pay-off minus total costs). The strategy of a profitable node is to perform actions to further increase its pay-off while the strategy of an unprofitable (lossy) node is to reduce costs by shrinking capacities of links or terminating links (depeering) until it is profitable.

The iterative process, consisting of a series of price and capacity allocation games, terminate when no node makes any changes to its connectivity and link capacities during the capacity competition game, at which point the system reaches a *topology equilibrium*. A traffic equilibrium indicates that the evolution reaches a stable point when certain external parameters such as number of ASs, IXPs, average traffic demands and so on are constant. Our assumption is that the topology convergence happens at shorter time-scales compared to the changes in the above-mentioned external parameters.

3. RESULTS

We plan to study the structure of the resulting equilibrium topologies and properties of the nodes such as profitability. At this point, we have observed topology equilibrium in our tests on multiple occasions; however we do not have sufficient amount of results to draw any conclusions. We observed that the price game reaches equilibrium and does not oscillate under the specified strategies. A challenge with the model is the substantial amount of computation because of repeating traffic equilibrium computations upon changes in the prices, peering and link capacities. Therefore we were only able to run tests with topologies up to 500 ASs.

4. ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under grant CNS-1111040.

5. REFERENCES

[1] Bobby Bhattacharjee, Ken Calvert, Jim Griffioen, Neil Spring, and James Sterbenz. Postmodern Internetwork Architecture. Technical report, 2006.

[2] Isidro Casteneyra, Noel Chiappa, and Martha Steenstrup. The Nimrod Routing Architecture. RFC 1992, August 1996.

[3] Hyunseok Chang, Sugih Jamin, and Walter Willinger. To Peer or Not to Peer: Modeling the evolution of the internet's as-level topology. In *Proceedings of INFOCOM*, 2006.

[4] Amogh Dhamdhere and Constantine Dovrolis. The Internet is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh. In *Proceedings of ACM CoNEXT*, 2010.

[5] Philip Brighten Godfrey, Igor Ganichev, Scott Shenker, and Ion Stoica. Pathlet Routing. In *Proceedings of SIGCOMM*, 2009.

[6] Karthik Lakshminarayanan, Ion Stoica, and Scott Shenker. Routing as a Service. Technical report, 2004.

[7] B. Raghavan and A. Snoeren. A System for Authenticated Policy-Compliant Routing. In *Proceedings of ACM SIGCOMM 2004, Portland, Oregon*, August 2004.