Inter-domain Routing using Topology Information

Pedro Amaral Luís Bernardo Paulo Pinto

Abstract - This paper proposes an architecture for inter-domain routing, called DTIA – Dynamic Topological Information Architecture. It is designed to address current limitations of BGP like: scalability of the routing tables, multihoming problems, churn rate, range of routing events and policy coordination. Still it supports the most important functionalities of BGP, and maintains the current Internet routing distribution model based on Inter-AS business relationships. The architecture is based on the knowledge of a static network graph formed by the Autonomous Systems (AS) and an algorithm to manage link failures. We use the concept of a region as a mechanism to sustain scale. New possible features like multi-path routing and enhanced Traffic Engineering capabilities might benefit with DTIA’s usage.

I. INTRODUCTION

The Border Gateway Protocol (BGP) is the current inter-domain routing protocol. BGP’s decision process is based on attributes allowing very complex manipulations. The evolution of the Internet has led to a very complex system that is highly sensitive to coordination. Several weaknesses and inefficiencies [1] have been identified and will get worse with time. Some of most important are: the growth of the routing tables, convergence problems caused by lack of policy coordination [2], high churn rate, range of routing events and the lack of support for multihoming (which worsens all the above problems in BGP).

This paper proposes a new vision for inter-domain routing based on giving importance to three aspects: a) there is a separation between reachability and routing; b) the network is a static network and solely link failures produce dynamism; and c) attribute manipulation is replaced by a set of rules.

II. DESIGN CHOICES

Based on the current Internet structure DTIA is designed based in six choices:

1. The maintenance of the current business model based on Autonomous Systems and the commercial relationships between them [3].

2. The recognition of a hierarchical architecture based on customer-provider links forming a three-tier structure [3].

3. The stable nature of the links that form the Internet. The issue is whether the link failed or not, and not if it exists.

4. Reachability is based on AS connections and not on prefixes. This reduces the routing tables in two ways: first because there are fewer ASes than prefixes; second the use of multihoming, traffic engineering and load balancing schemes in BGP often impairs prefix aggregation. There are further advantages on using ASes: traffic engineering and load balancing can be performed based on a single graph of inter AS connections with no impact on scalability and multihoming is reduced to a choice of paths. The downside is that a mapping service between prefixes and ASes must exist.

5. Routers get a static map of the network and co-operate to learn about failures. A central entity (or various to provide reliability) delivers a static map of the network (or a region, see choice 6) to routers. The dissemination of failure information only “disturbs” the relevant routers with precise rules about its scope.

6. Maps and co-operations are limited to regions.

We propose to divide the Internet in regions and for each region an entity (e.g. RIPE [4] for Europe) builds the static graph delivered to routers. Packets going from one region to the other use either a direct link from one internal AS connected to the destination region (if valid), or have to climb up the hierarchy and go down in the destination region.

III. DTIA ARCHITECTURE

The static network map is modeled by a directed graph $G(V,A)$ with $V(G)$ vertices that model ASes and $A(G)$ arcs that model links between ASes and are labeled according to four types of inter AS relationships:

- Provider-Customer. One AS (the provider) accepts all traffic from the other AS (the client). Two labels apply: provider-customer ($p2c$) and customer-provider ($c2p$).
- Peer-to-peer ($p2p$). ASes provide connectivity for their direct or indirect customers. No transit traffic is allowed.
- Peer-to-peer allowing backup ($p2pbkup$). The same as before but allows transit traffic if no other path exist.
- Peer-to-peer allowing transit traffic ($p2patt$). Transit traffic is allowed in any situation.

The first two inter AS relationships deal with the so-called common policies [5] that cover 99% [3,6] of the policies used today in BGP. The two latter relationships increase our architecture expressiveness to include more complex relationships like siblings and the use of peers for backup.

The common policies define in the current architecture how routes are distributed and therefore how traffic flows in the network (i.e. they define the valid paths). In our approach policies are enforced using both the labels of the graph and a small set of rules that represent the current economic relationships. The rules are used to calculate a set of all valid paths in $G(V,A)$ from one AS X to any other AS in the region. This set is denoted as $P(X)$.

A. Searching and Pruning process

$P(X)$ is constructed in a hop-by-hop process. To control valley
paths a qualifier, named Direction (D), is added to each path. D indicates if a path is descending in the provider – customer direction or ascending in the customer-provider direction.

Table 1 and Table 2 contain the validity rules (valid (V) or invalid (X)) for an arriving arc in the row and a departing arc in the column. Table 1 is for descending paths (with D=0) in these paths there cannot be c2p arcs and p2p arcs are always invalid. Table 2 is for the D=1 case (ascending paths). In this case when the first p2c arc appears the Direction changes its value.

<table>
<thead>
<tr>
<th>Result</th>
<th>p2c</th>
<th>c2p</th>
<th>p2pbkup</th>
<th>p2p</th>
<th>p2patt</th>
</tr>
</thead>
<tbody>
<tr>
<td>p2c</td>
<td>V</td>
<td>X</td>
<td>V</td>
<td>X</td>
<td>V</td>
</tr>
<tr>
<td>c2p</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>p2pbkup</td>
<td>V</td>
<td>X</td>
<td>If (AS in set)X else V</td>
<td>X</td>
<td>If (AS in set)X else V</td>
</tr>
<tr>
<td>p2p</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>p2patt</td>
<td>V</td>
<td>X</td>
<td>If (AS in set)X else V</td>
<td>X</td>
<td>If (AS in set)X else V</td>
</tr>
</tbody>
</table>

Table 1 – Rules to validate paths for D=0.

<table>
<thead>
<tr>
<th>Result</th>
<th>p2c</th>
<th>c2p</th>
<th>p2pbkup</th>
<th>p2p</th>
<th>p2patt</th>
</tr>
</thead>
<tbody>
<tr>
<td>p2c</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c2p</td>
<td>V;D=0</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>p2pbkup</td>
<td>V;D=0</td>
<td>V</td>
<td>If (AS in set)X else V</td>
<td>X</td>
<td>If (AS in set)X else V</td>
</tr>
<tr>
<td>p2p</td>
<td>V;D=0</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>p2patt</td>
<td>V;D=0</td>
<td>V</td>
<td>If (AS in set)X else V</td>
<td>X</td>
<td>If (AS in set)X else V</td>
</tr>
</tbody>
</table>

Table 2 – Rules to validate paths for D=1.

A table, FH(X) is generated from P(X), containing the different first hop exits for each AS in the region. It is up to higher-level algorithms to perform routing, traffic engineering, load balancing, etc. using this FH(X) table. Note that it is up to the routing protocol to solve loops due to the existence of multi-paths to a destination. The policy rules however guarantee that each path in P(X) is loop free providing that there are no cycles in the provider customer relationships [7].

Regions are connected at tier-1 or with direct links between lower level ASes. In the latter case these links can be either p2p or p2patt. p2p links are a private business of the ASes involved and are not used by other ASes. p2patt can be used by any other AS in the region only if the path towards the AS in the border is in ascending direction. ASes connected by p2patt exchange the set of ASes reachable through them. When a packet arrives for a destination in the other region that is not in the set, it is sent to tier-1 ASes.

B. Failure Management

The dynamic part of the architecture is used to create awareness on link failures and mark paths in P(X) as not available during the failure.

Only links fail (a failing AS means all its links failed). Routers at the endpoints of the failed link disseminate a control packet with the link identification. The dissemination is stopped if an AS checks that it can still reach all reachable ASes. If, at least one reachable AS becomes unreachable, the dissemination continues following the rules of the previous tables. The scope of the dissemination decreases with the degree of multihoming in the region reducing churn and the range of the event. The mechanism guarantees that P(X) at each node contains only valid paths and that no packets are lost if alternative Paths exist [7]. Note that the routing protocol in each router is warned about the link failure and there might be consequences at routing level to prevent loops.

C. Experiments

Our first set of experiments used a topology of the RIPE region with 11,335 ASes and about 21,000 links obtained from the CAIDA AS Relationships Data research project [8]. P(X) had a mean value of 22,563 paths with a standard deviation of 7,275 (see [7] for details). The calculation to obtain the largest P(X) took 0.2s with a 2,4GHz microprocessor, 4G of memory and using JAVA. The FH(X) table has a maximum of 169,293 different paths for one AS, taking 0.7s to calculate. The control packet dissemination mechanism was also tested and the time needed for an AS to check if it lost reachability to any AS was 0.2s in the worst case.

IV. CONCLUSIONS AND FURTHER WORK

This paper proposed a possible architecture for Internet interdomain reachability offering some advantages over BGP in issues such as multi-path routes, multihoming, routing table scalability and policy coordination. There is a strong containment on churn and route events which is greater with the degree of multihoming. Packets could be forwarded quicker based solely on AS numbers that could exist in an optional IP header.

However there are no secrets. Compromises have to be made in some issues and certain current BGP features have to be considered secondary. Nevertheless, we think it preserves the main characteristics of the Internet, and specially its business model. Above all its deployment is also credible[7].

This work opens new directions of research. The design of a new routing protocol that works with multi-path routes over a constrained graph (P(X)) and featuring load balancing and traffic engineering (or having economic considerations) can be very challenging.

REFERENCES