Deterministic, Reduced-Visibility Inter-Domain Forwarding

Stephen D. Strowes and Colin Perkins
Department of Computing Science
University of Glasgow, U.K.
sds@dcs.gla.ac.uk, csp@csperkins.org

ABSTRACT

Inter-domain forwarding state is growing at a super-linear rate, rendering older routers obsolete and increasing the cost of replacement. A reduction of state will alleviate this problem. In this paper, we outline a new reduced-state inter-domain forwarding mechanism. We carefully drop portions of the advertised forwarding state using a utility measure for prefixes based on the length of the prefix and the path length to its origin. A deterministic forwarding algorithm uses the resulting partial view. The graph of connections between autonomous systems is shallow, offering many viable paths for data flows, a property we aim to use to achieve minimal detrimental effect on delay and AS path stretch.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Packet-switching networks*

General Terms

Algorithms

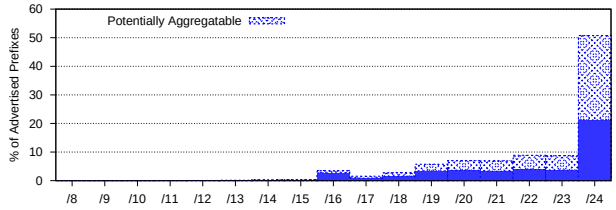
1. INTRODUCTION

The rate of growth of forwarding state in the Internet's default-free zone (DFZ) is encouraging the routing community to investigate alternative routing protocols and operational procedures [10]. There are currently over 300,000 entries in the full BGP table, and current trends suggest a year-on-year increase of around 17% [5, 7]. While we expect the number of advertised prefixes to naturally rise, increased adoption of multihoming by edge-sites (for improved connection reliability) exacerbates the growth. The larger address space of IPv6 will not solve this problem.

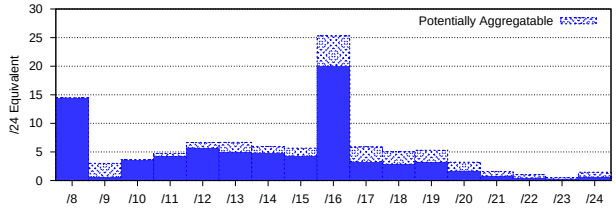
The current system opts to retain all state advertised to achieve policy routing with small stretch, but the volume of advertisements makes this increasingly burdensome. Regardless of core/edge separation solutions such as LISP [6], the network will continue to grow; the transit networks providing the inter-domain service are already well-connected [7]. There is scope for leveraging this property to achieve inter-domain packet forwarding with reduced state.

We are currently investigating the potential reduction in forwarding state if we drop portions of this state according to a measure of each advertisement's utility. Our work achieves deterministic packet forwarding with carefully limited forwarding state (reducing visibility), and aims to achieve minimal detrimental effect on delay and AS path stretch.

CoNEXT'09 Student Workshop, December 1, 2009, Rome, Italy.
ACM 978-1-60558-751-6/09/12.



(a) Distribution of advertised prefixes.



(b) The /24 equivalent distribution of advertised prefixes.

Figure 1: Prefix distributions.

2. ADVERTISED IP ADDRESS SPACE

Over time, the proportion of long prefixes in the full BGP table has increased; the mean prefix length is now approximately 22.24 (calculated from [1]). Long prefixes dominate the full BGP table, but cover a disproportionately small volume of the advertised address space. Edge-site multihoming, traffic engineering, and fragmentation of the address space explain this increase.

Fig. 1(a) shows the distribution of prefix lengths. The /24's dominate the table with 51.9% of the prefixes, yet cover only 1.52% of the advertised address space (Fig 1(b)). The /8 to /16 range, however, occupies only 4.3% of the table, and yet covers 75.3% of the advertised address space.

Further, more than half (52.3%) of the prefixes in the table are de-aggregates from other advertised address blocks; 58% of the /24's are contained within shorter prefixes advertised elsewhere. Specific prefixes advertised to support edge-site multihoming seem to create much of the burden on routing state in the Internet today, though the utility of this state may diminish as the distance from the origin increases.

3. INTERNET GRAPH STRUCTURE

The graph of ASes consists around 32,000 networks, and demonstrates small-world properties [4], with the mean hop count between any pair of ASes between 3 – 4 hops [7]. Theory tells us that in any packet forwarding system, there are two opposing constraints: routing table size and network path length [8].

Within the AS graph, there is a subgraph of ASes consisting almost 5,000 networks offering transit services. According to the AS relationships dataset [3], the mean out-degree in this (discounting links to edge-sites) region is 9.8. The node degree distribution for this transit subgraph demonstrates a heavy tail: some networks are extremely well connected while many others are reasonably well connected. A router in a well connected AS is likely to have no default route, and so will carry a full BGP table to correctly handle longest-prefix matching. These networks potentially offer many viable routes between pairs of ASes with few AS hops (subject to policy constraints).

4. REDUCED STATE FORWARDING

We assert that forwarding tables in some inter-domain routers can be shrunk substantially through the selective removal of state. We suggest that the graph is dense enough to recover from missing entries by deterministically forwarding packets for destinations no longer covered by the FIB to neighbours (Alg.1). We believe that if one AS cannot route to a particular destination, then, statistically, a neighbouring AS will and that the additional stretch will be small. Determinism is retained by hashing over the IP header to choose an output port, thus maintaining packet ordering. Note that corresponding transit agreements between peering ASes would be required.

Given that the majority of the advertised address space is captured by shorter prefixes, we suggest that FIB construction may drop certain prefixes based on their utility (Alg.2). We define the utility of a prefix as a function of its length and the number of AS hops it has travelled, on the observation that the actual usefulness of a prefix may diminish rapidly as the path length and prefix length increase. We peg certain advertisements at utility 1: all /8 to /16 prefixes, and all advertisements originating from within one hop. Traffic matrices may also be used to peg popular prefixes at utility 1. The utility of the remaining advertisements tends rapidly to 0 as path length and prefix length increase.

We are investigating this prospect for reducing forwarding state in the inter-domain routing system. Key to understanding the feasibility of this technique and its potential gains is an investigation on the probabilities at play and, statistically, just how much of the table an AS can throw away before stretch degrades substantially. The analysis will investigate the relationship between the size of the forwarding tables generated at a given node, vs. the advertisement utility (u), vs. the degree of connectivity, vs. the effect on path stretch across the network.

5. EVALUATION AND CONCLUSION

The continued growth of the network may require a fundamental shift in how we perform packet forwarding in the future. This work explores one such shift, which would tend the network toward a form of limited visibility, but retains determinism important for current transport protocols. If strong clustering is common in the AS graph, alternative measures of utility may be required to avoid packets bouncing within clusters, or altogether different routing schemes [9].

Data from CAIDA [2, 3] offers some of the best representations of the current Internet graph available. Using these data, we are constructing a highly distributed simulation environment to analyse the characteristics of this forwarding scheme on the AS graph, and its feasibility as a mechanism for offering a scalable routing infrastructure as part of a future Internet.

Shortest-path routing led to the longest-prefix matching scheme we have today, with as much state retained as possible. This state, however, represents a heavy burden to achieve best possible path stretch. The scheme presented here may be able to use the nature of the network graph to significantly reduce forwarding state, but without adversely affecting performance levels.

Algorithm 1 Deterministic Packet Forwarding

```
On packet, perform longest prefix-match
if match found then
  Forward on correct port
else if no default route then
  Forward on port[hash(header) mod num_ports]
else
  Use default route
end if
```

Algorithm 2 Partially-Randomised FIB Generation

```
On BGP update, consult random number generator to yield r
if r > u then
  Add entry to FIB
end if
```

6. REFERENCES

[1] http://archive.routeviews.org/oix-route-views/2009.03/.

[2] Routeviews Prefix to AS mappings Dataset. http://caida.org/data/routing/, May 2009.

[3] The CAIDA AS Relationships Dataset. http://caida.org/data/active/as-relationships/, May 2009.

[4] M. Boguñá and D. Krioukov. Navigating ultrasmall worlds in ultrashort time. *Physical Review Letters*, 102(058701), 2009.

[5] T. Bu, L. Gao, and D. Towsley. On Characterizing BGP Routing Table Growth. *Computer and Telecomm Networking*, 45(1), 2004.

[6] D. Farinacci et al. Locator/ID Separation Protocol (LISP), Mar 2009. Work in Progress.

[7] G. Huston. BGP in 2008. http://potaroo.net/ispcol/2009-03/bgp2008.html.

[8] L. Kleinrock and F. Kamoun. Hierarchical Routing for Large Networks. *Computer Networks*, 1977.

[9] D. Krioukov et al. Compact Routing on Internet-Like Graphs. In *INFOCOM*, Mar. 2004.

[10] D. Meyer et al. Report from the IAB Workshop on Routing and Addressing (RFC 4984), Sept. 2007.