Randomness for Reduced-State Inter-Domain Forwarding

Stephen D. Strowes
University of Glasgow
Department of Computing Science
sds@dcs.gla.ac.uk

Colin Perkins
University of Glasgow
Department of Computing Science
csp@csperkins.org

1. INTRODUCTION

The rate of growth of forwarding state at core Internet routers has prompted some concern about the scalability of the Internet in the future. This growth is affected by two factors: the number of prefixes advertised, and the density of the inter-domain graph of Autonomous Systems (ASes) [10, 5]. While it is unquestionable that the number of advertised prefixes will continue to rise, the increased adoption of multihoming by edge networks (for improved connection reliability) exacerbates the growth. Regardless of whether temporary core/edge separation solutions such as LISP [6] are deployed, the graph of networks providing the inter-domain service is increasingly densely connected, not just at the edge nodes, but also further into the network [8].

Theory tells us that in any packet forwarding system, there are two opposing constraints: routing table size and network path length [9]. The current routing system opts to retain all routing state injected into the inter-domain system to achieve policy routing with small stretch, but the density of the graph makes this increasingly burdensome. We are currently investigating the behaviour of random forwarding in the Internet graph, how it may reduce forwarding table sizes, the various trade-offs involved, and how these affect AS path stretch.

2. INTERNET GRAPH STRUCTURE

The graph of ASes demonstrates small-world properties [4], with the mean hop count between any pair of ASes in the network between 3 – 4 hops [8]. Given that the number of networks participating in inter-domain routing is constantly increasing, there must exist a corresponding increase in the level of connectivity between these networks, and hence the size of forwarding tables, to allow the average path length to essentially remain constant [8].

This connectivity is evident in the ranked graph of node degrees, Fig.1 (derived derived from [3]), which demonstrates a heavy tail of node connectivity. This heavy tail, however, features many reasonably well-connected nodes providing transit, bypassing the extremely well-connected ASes at the top. A well connected AS is likely to have no default route, and so forwarding information base (FIB) sizes will be large in these ASes to correctly handle longest-prefix matching. These networks form a dense mesh potentially offering many viable routes between pairs of ASes within a low number of AS hops (subject to policy constraints).

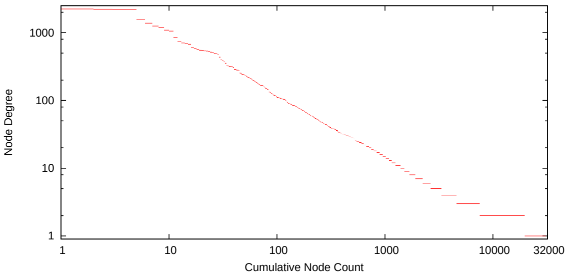

Figure 1: AS node degrees. On a log-log scale, we see many (˜20000) nodes of degree 2 or more in the AS graph.

Further, the mean length of the advertised prefixes is increasing, currently standing at around 22.24 (calculated from [1]), suggesting an increasing proportion of long prefixes advertised. There are good explanations for this increase, including endpoint multihoming, traffic engineering, and fragmentation of the address space. This expansion may be manageable in an IPv4 world; IPv6, however, with its larger address space, will exacerbate the problem.

3. RANDOM FORWARDING

We suggest that the AS graph may be dense enough, and the routing system may be holding enough redundant information, that careful abandonment of some forwarding state at heavily-connected ASes may allow us to greatly reduce the volume of this state without significantly affecting AS path stretch.

We are investigating this prospect for reducing forwarding state in the inter-domain routing system using an element of randomness. We assert that forwarding tables in some inter-domain routers can be shrunk substantially through the removal of some table entries for non-local destinations, and that the graph is dense enough to recover from these missing entries by randomly forwarding packets destined for destinations no longer covered by the FIB. Note that corresponding agreements between peering ASes would be required, and that that we are not suggesting border routers form route caches [7]. Specifically, when generating FIBs from BGP

Algorithm 1 FIB Generation.

```
On BGP update
if length(path to prefix) $<= 1$ then
    Add path to FIB
else
    Add path to FIB with probability $p$
end if
```

Algorithm 2 Random Forwarding Algorithm.

```
On packet, perform longest prefix-match
if match found then
    Forward out correct port
else if no default route then
    Forward to randomly chosen port (not incoming port)
else
    Use default route
end if
```

updates, a router may choose to leave a prefix out based on a given random factor, p (Alg.1); when forwarding packets, if the router does not know the next hop for a destination, it will choose a random outgoing port (Alg.2).

Provided *just enough* state remains in the FIB, and the network is sufficiently well-connected, we believe that if one AS cannot route to a particular destination, then, statistically, a neighbouring AS will and that the stretch due to this random walk will be small. If this is so, what sort of forwarding state reduction is achievable? What are the characteristics of the path stretch in this environment?

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 route drop probability (p), vs. the degree of connectivity, vs. the effect on path stretch across the network.

An interesting vector to explore is whether shorter prefixes could be weighted such that they are more likely to remain in the FIB than longer prefixes. This may not cause problems as one might expect: the largest volume of address space advertised is captured by the /16 prefixes (Fig.2); while many of the long prefixes may be provider independent addressing for multihomed networks, it may be that preferentially dropping long prefixes leaves most of the address space intact without relying on random forwarding, with a beneficial reduction in forwarding state. Further weighting may be introduced, such that advertised routes with longer paths are more likely to be dropped. (Though if strong clustering is common, this may deteriorate quickly, with packets bouncing within clusters until they expire.)

4. EVALUATION AND CONCLUSION

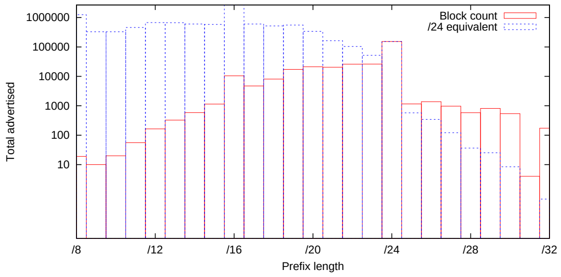

Figure 2: Prefix length distribution. Many entries exist to serve small portions of the address space.

The continued growth of the network may require some fundamental shift in how we perform packet forwarding in the future. This work explores one such shift.

Data from CAIDA [3, 2] offers one of the best representations of the current Internet graph available. Using this, we are constructing a highly distributed simulation environment to analyse the characteristics of random forwarding across this graph, and the feasibility of random forwarding as one mechanism for offering a scalable routing infrastructure as part of a future Internet, based on the questions above.

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. Random forwarding is a simple scheme which may be able to use the nature of the network graph to significantly reduce forwarding tables, but without adversely affecting performance levels.

5. REFERENCES

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

[2] Routeviews Prefix to AS mappings Dataset. http://caida.org/data/routing/routeviews-prefix2as.xml, 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] D.C. Feldmeier. Improving gateway performance with a routing-table cache. *INFOCOM*, Mar 1988.

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

[9] L. Kleinrock and F. Kamoun. Hierarchical Routing for Large Networks. *Computer Networks*, 1(3), 1977.

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