Randomness for Reduced-State Inter-Domain Forwarding

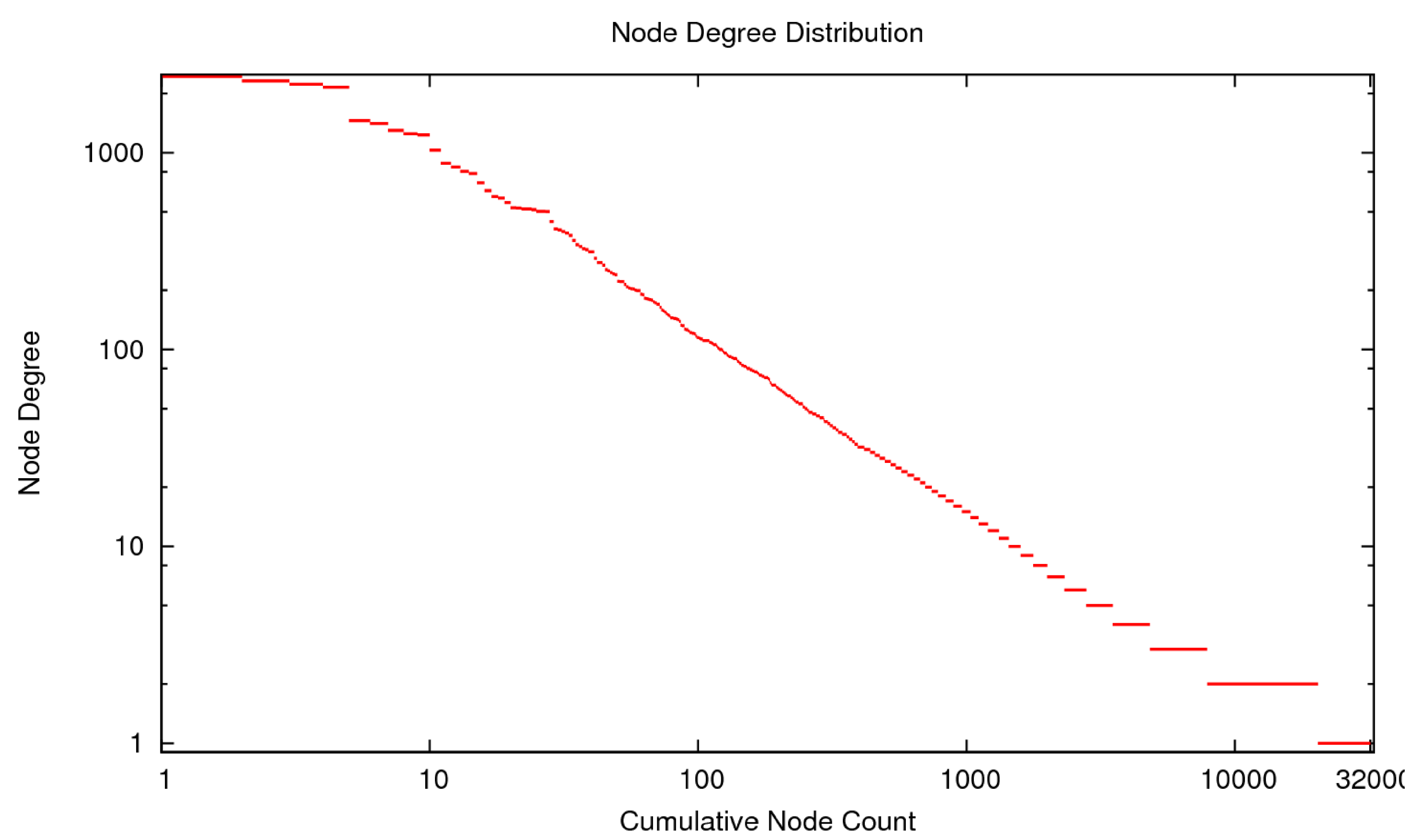
Stephen Strowes and Colin Perkins
sds@dcs.gla.ac.uk, csp@csperkins.org
Department of Computing Science
University of Glasgow

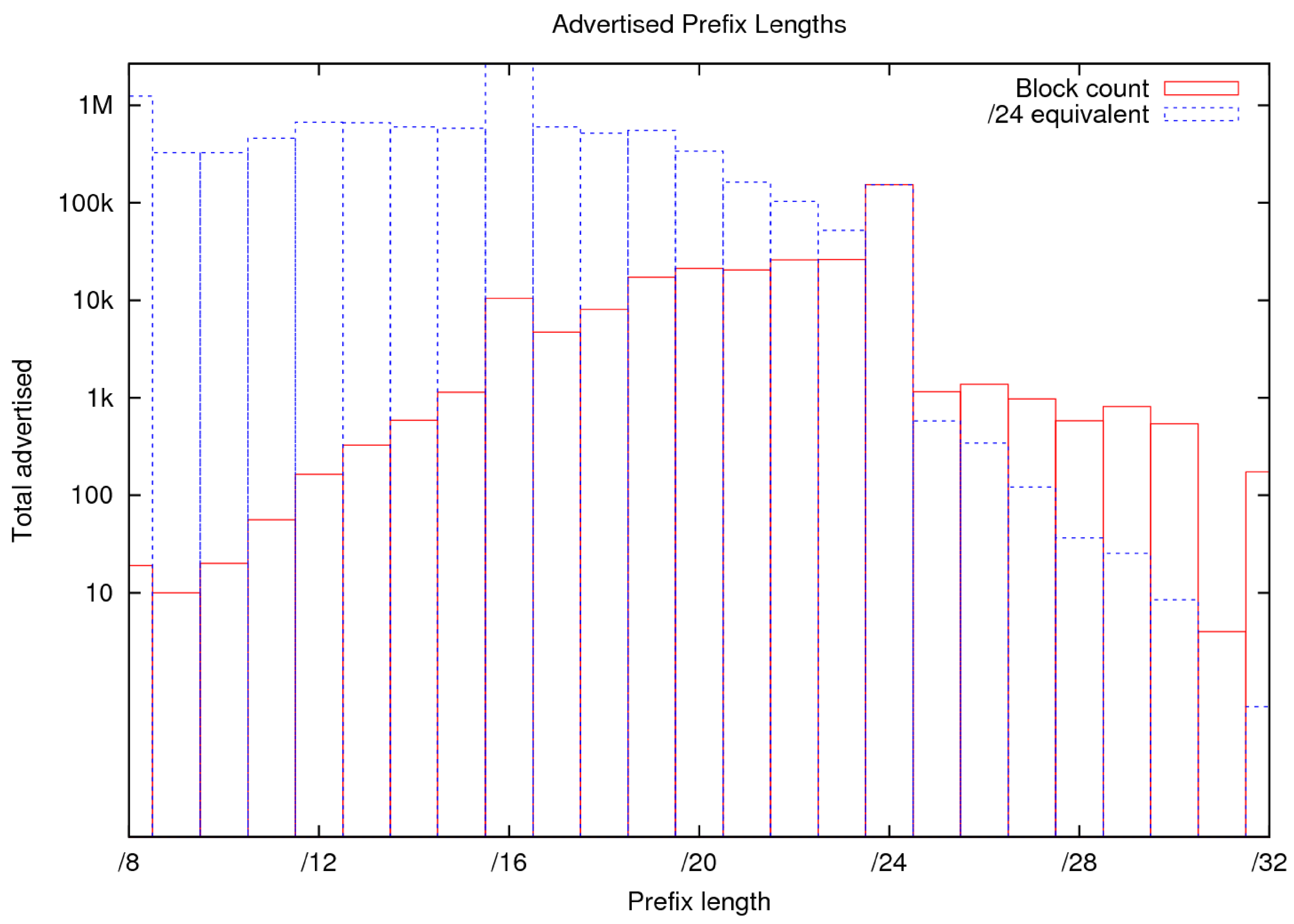
1.Problem: Increased Forwarding State

The rate of growth of forwarding state at core Internet routers has prompted concern about the scalability of the Internet in the future. This growth is exacerbated by increased adoption of end-site multi-homing, over and above the natural growth of the number of advertised prefixes [3, 4, 6].

Although this rapid expansion may be manageable with IPv4, this will not be the case given a transition to the larger address space of IPv6. Larger forwarding tables incur more time-consuming lookups, threatening a router's ability to forward packets at line-rate.

At the same time, the graph of Autonomous Systems (ASes) is becoming increasingly densely connected [4], a property we may be able to harness to reduce forwarding state.





2.Network Structure

Table size and network path length are inversely proportional [5]; thus, path length must increase rapidly to allow forwarding tables to scale.

The mean distance between any pair of ASes has been nearly constant at ~3 -- 4 AS hops [4], despite continuous growth of the number of networks. Thus, there must be a corresponding increase in the level of connectivity [4].

Today's connectivity is evident in the node-degree distribution (above, derived from [2]). The transit region of the graph forms a dense mesh of links offering many potential routes between pairs of ASes.

Further, the mean length of the advertised prefixes is increasing, currently at ~22.24 (calculated from [1]), reflecting the increasing proportion of long prefixes advertised from site multi homing, traffic engineering, and fragmentation of the address space.

3.Random Forwarding

Careful abandonment of some forwarding state at transit ASes may allow us to greatly reduce the volume of this state without significantly affecting AS path stretch.

Forwarding tables in some inter-domain routers may be shrunk substantially through the removal of some table entries, using the two algorithms opposite. We assert that the graph is dense enough to recover from these missing entries by randomly forwarding packets to neighbouring Ases.

Provided *just enough* state remains in the FIB, then, statistically, a neighbouring AS will be able to continue forwarding along a correct path.

FIB Generation:

On BGP update:
If (length(path to prefix) <= 1):
    Add path to FIB
Else:
    Add path to FIB with probability *p*

Random Forwarding:

An data packet:
    perform longest prefix match
    If (match found):
        Forward out correct port
    Else if (no default route):
        Forward out randomly chosen port (not incoming port)
    Else:
        Use default route

References

[1] http://archive.routeviews.org/oix-routeviews/2009.03/.
[2] The CAIDA AS Relationships Dataset. http://caida.org/data/active/as-relationships/, May 2009.
[3] T. Bu, L. Gao, and D. Towsley. On Characterizing BGP Routing Table Growth. Computer and Telecomm Networking, 45(1), 2004.
[4] G. Huston. BGP in 2008.http://potaroo.net/ispcol/2009-03/bgp2008.html.
[5] L. Kleinrock and F. Kamoun. Hierarchical Routing for Large Networks. Computer Networks, 1(3), 1977.
[6] D. Meyer et al. Report from the IAB Workshop on Routing and Addressing (RFC 4984), Sept. 2007.

4.Ongoing Work

Our 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.

We are building a simulation environment to test these ideas. In particular, we will investigate dropping FIB entries intelligently. For example the algorithm should be more likely to retain short prefixes, which cover most of the address space (see left-most figure), and should be more likely to retain prefixes advertised by nearby Ases, so advertised routes with longer paths are more likely to be dropped. It should retain a small set of distant ASes.