Stor: An Anonymous Data Storage Framework

Daniel Playle

The University of Southampton

djap1g11@soton.ac.uk

Based on report available at https://github.com/dp0/IP-Final-Report

Why?

Wanted to create an anonymous collaboration platform

There wasn’t a framework that could assist me

What are the requirements?

Secure – CIA Triad

- Confidentiality
- Integrity
- Availability

Distributed

Anonymous

Asynchronous

Performance

Anonymous Layer

Use Tor’s hidden services

Each node should run a hidden service

How do we find other nodes?

Network Topology

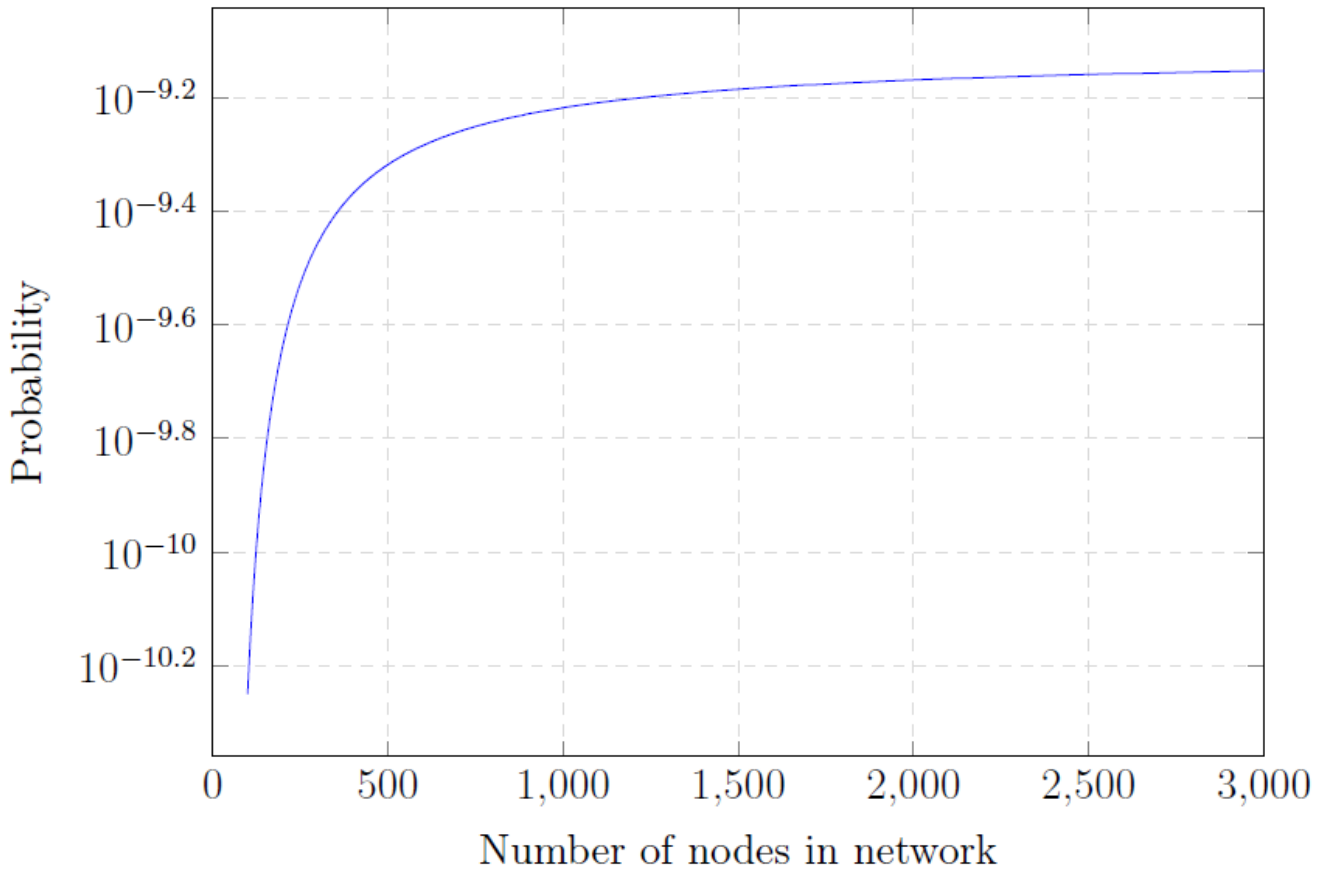
How should nodes connect? Are all nodes created equal?

Fully connected?

Hybrid, with some nodes more important than others?

All nodes should be equal!

Lonely Nodes



Universal Information Distribution

How to distribute information to all nodes in the network when no one is to be trusted?

Assume each node has some packets

Let’s also assume we can identify a packet with a short reference – this could be implemented with a checksum

Universal Polling

Very simple – just ask a node for the references of the packets it has

Wasteful in that there isn't much new information between requests for this data

Need to poll often in order to reduce packet propagation latency in the network

Active Forwarding

A more elegant solution – instead of polling other nodes, have the nodes forward new packets to you

Could easily send the packet reference first so we can reject any packets we already have

Doesn’t work if some nodes go offline, but it’s a good start for most of the transfers that will happen

Selective Polling

Rather than ask for everything a node has, why not just ask for the new packets?

Need to store the time of receipt for each packet and also the last time we queried each node

Potential for de-anonymzation vector? As long as the original inserter has to “acquire” the packet like other nodes. Such an attack should be hard

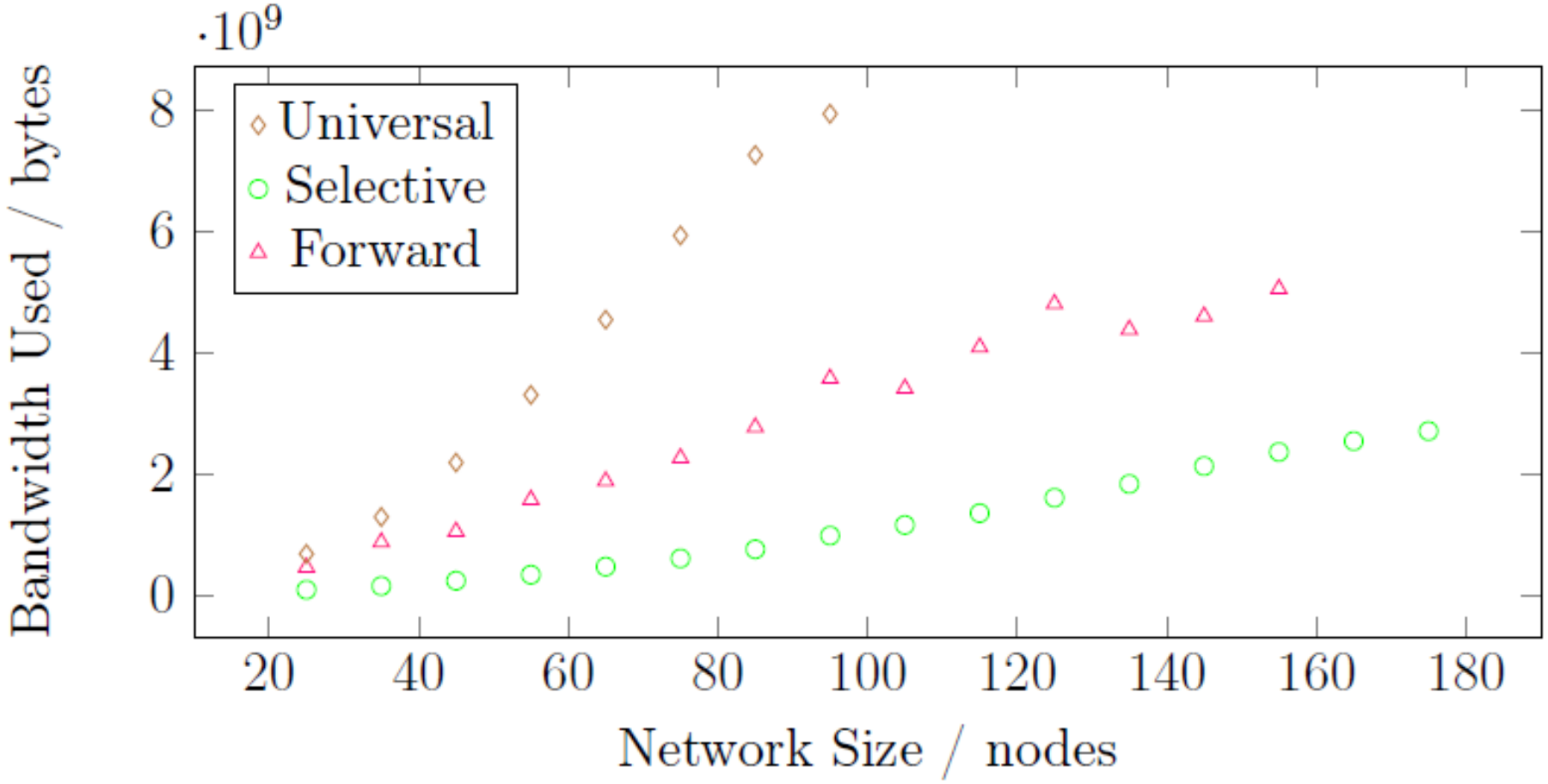
Universal Information Distribution

What is the best?

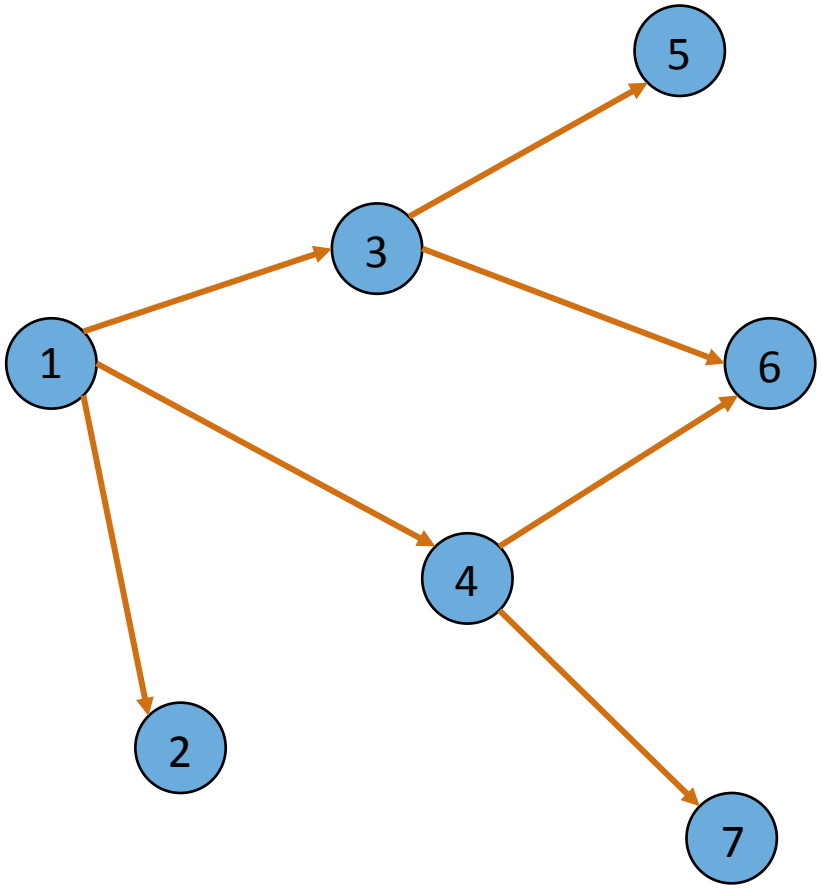
I modelled each to determine their scalability and performance

The basic overheads associated with Selective Polling are greater than the overheads for Active Forwarding. It’s obvious who the winner will be, right?

Performance of Distribution Methods



Active Forwarding Race Condition



Node 1 generates a packet

Node 1 forwards the packet to Nodes 2, 3 and 4

Node 2, 3 and 4 now own the packet

Node 3 forwards the packet to Nodes 5 and 6

Node 4 forwards the packet to Nodes 6 and 7

Anonymous Abuse

What stops a single party spamming the network?

Need to rate limit users

Introduce a trusted authority that issues “pseudo-anonymous identities”?

Proof-of-work?

Daniel Playle

djap1g11@soton.ac.uk

https://github.com/dp0/IP-Final-Report

See report for references