P2N: Cloud Control

David Tarrant
davetaz@ecs.soton.ac.uk

Ben O’Steen
benjamin.osteen@ouls.ox.ac.uk
Problem

• Everyone loves the cloud

• No one in this room would use it as their primary storage.

• Would anyone use it as a long term preservation storage solution?
More Questions

• Does the cloud do backup/replication/multi-site replication?

• Where are my files stored (geographically)?

• What is the long term pricing strategy of the cloud?
Influences

- Simplistic Cloud API
- High resilience and distribution of resources
- Transparent Expansion
- Low Barrier to Entry
The API

- Amazon S3
- PUT, GET, POST, HEAD, DELETE
- HTTP has all the tools we need!
High Resilience & Distribution

- Erasing coding (Honeycomb & RAID)
  - More efficient than replication

- Resilience of Bit Torrent

- Nodes in the network are geographic aware
Transparent Expansion

- Nodes can be added to the network arbitrarily
- Network re-distributes data for even spread
Low Barrier to Entry

- Provide a node
- Full machine
- Spare space on an existing machine
The P2N

N_1  N_2  N_3  N_4  N_5  N_6
The P2N

N

N_1  N_2  N_3  N_4  N_5  N_6
The P2N
The P2N

[Diagram showing a network with nodes N1 to N6 and N, with a question 'Single Point of Failure?']
The P2N

Single Point of Failure?
Institutional Distribution
Flexability

- Object level granularity
- Basic metadata support (through POST, HEAD)
- Object reporting, available via HEAD (single object) or GET (network report)
- Extensions to S3 API without breaking core functionality.
Progress so far

- Feasibility study has been done
- Now re-modularising the core
- P2N1 - Localised Network (Spare space)
- P2N2 - Thumper Network (200Tb+)
Thank-You

P2N: Cloud Control

David Tarrant
davetaz@ecs.soton.ac.uk

Ben O’Steen
benjamin.osteen@ouls.ox.ac.uk