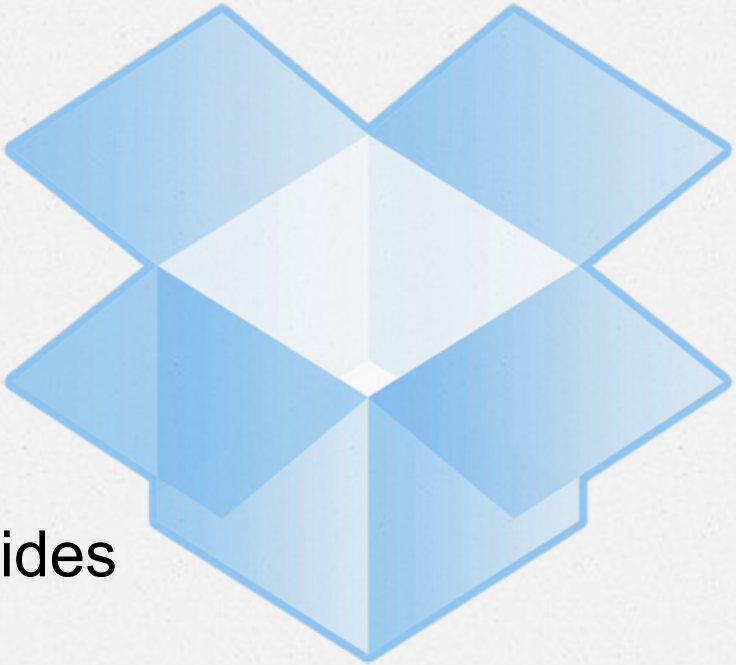
Private Dropbox

Calum O'Hare

Supervisor: David Eyers

Dropbox

- What is dropbox?



- What are the downsides to such a service?

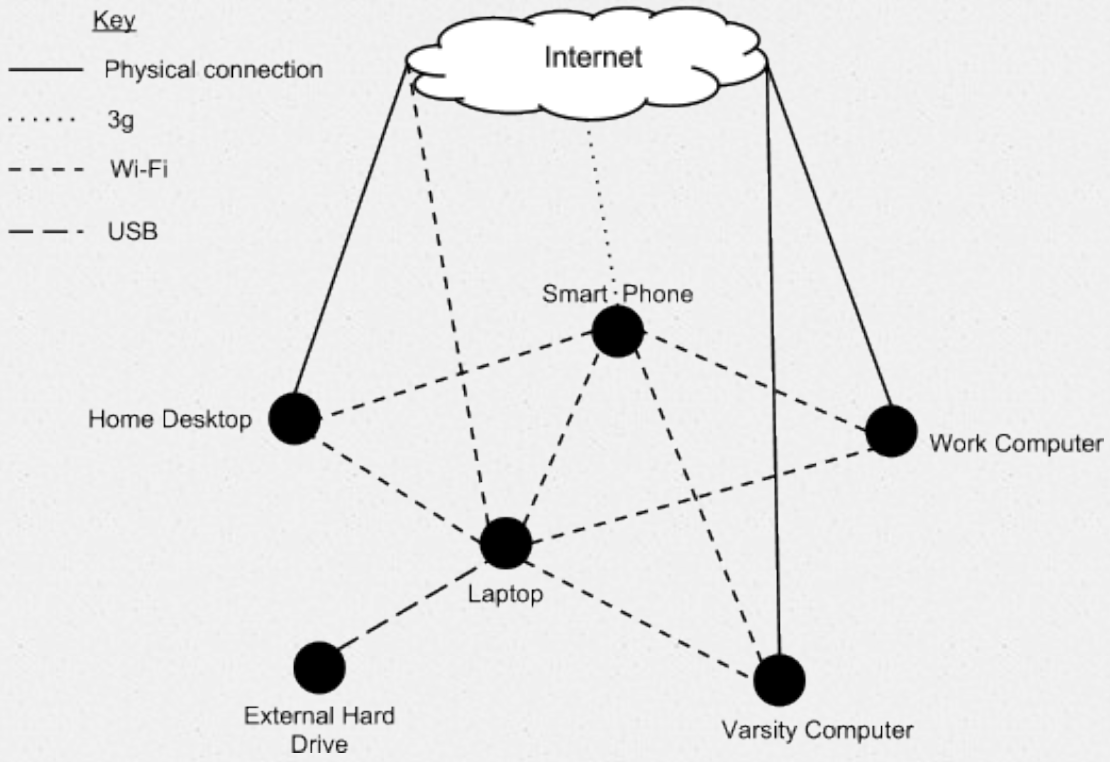
Private Dropbox

- A file synchronisation tool.
- Its main function should be to keep data synchronised between multiple devices.
- What makes ‘private’ dropbox?

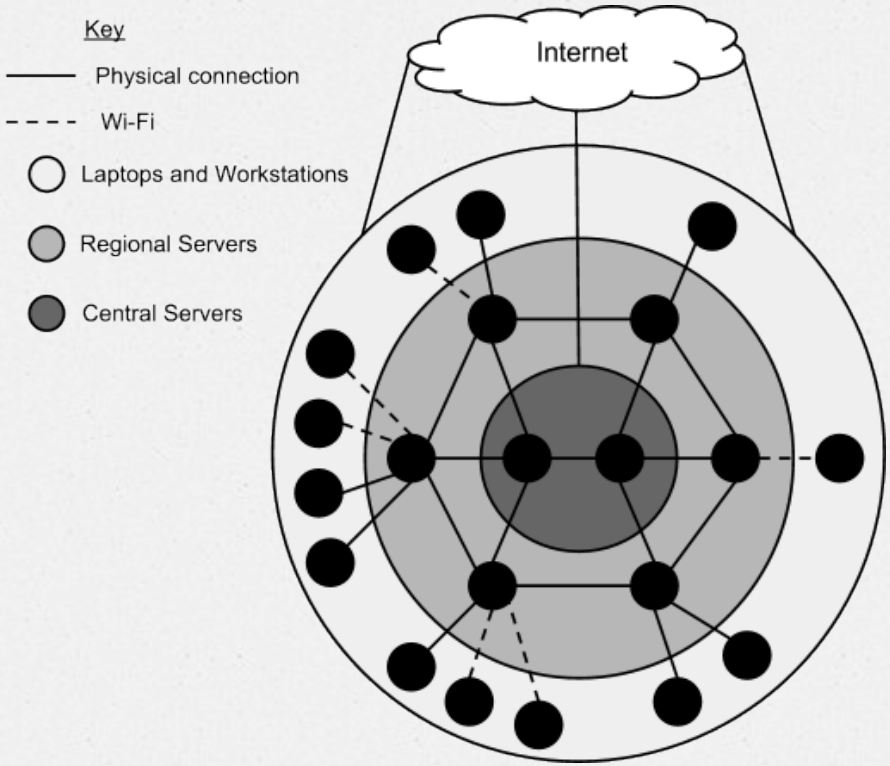
Important features

- **Decentralised**
- **File synchronisation**
- **Fine-grained user control**
- **Statistics**

Use case 1


Key
Physical connection
3g
Wi-Fi
USB
Internet
Smart Phone
Home Desktop
Work Computer
Laptop
External Hard Drive
Varsity Computer

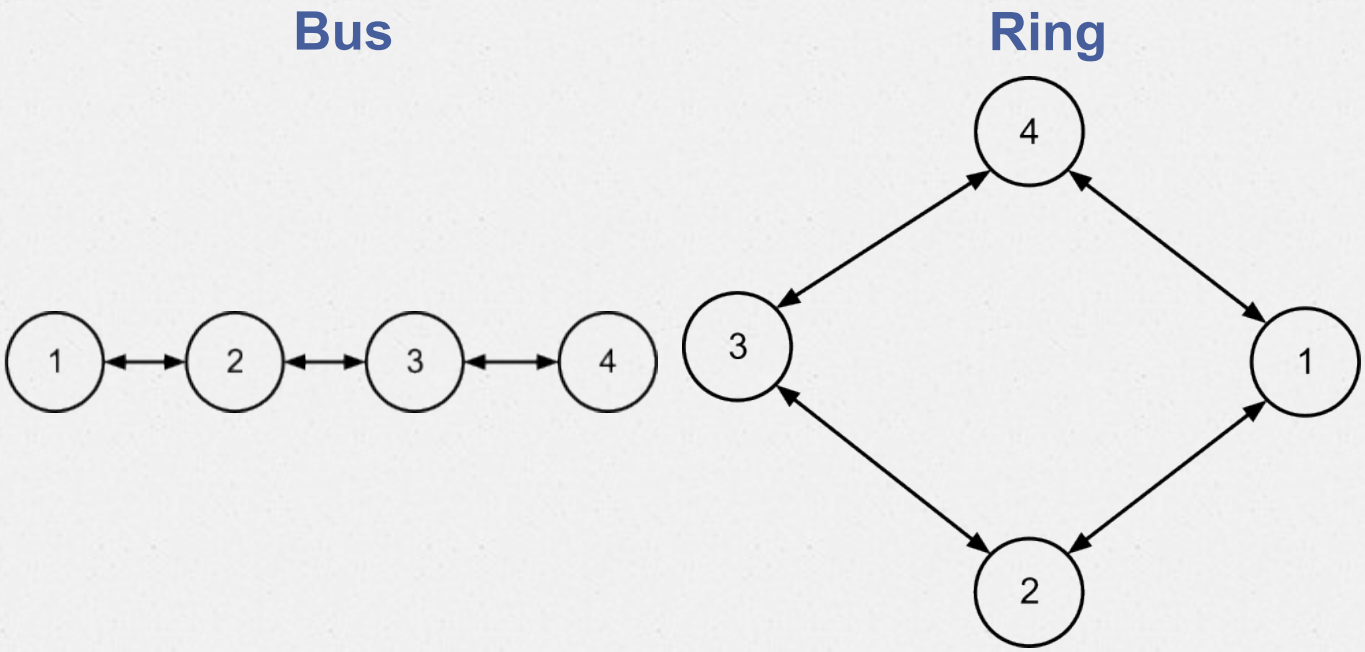
Use case 2


Key
Physical connection
Wi-Fi
Laptops and Workstations
Regional Servers
Central Servers
Internet

Progress so far

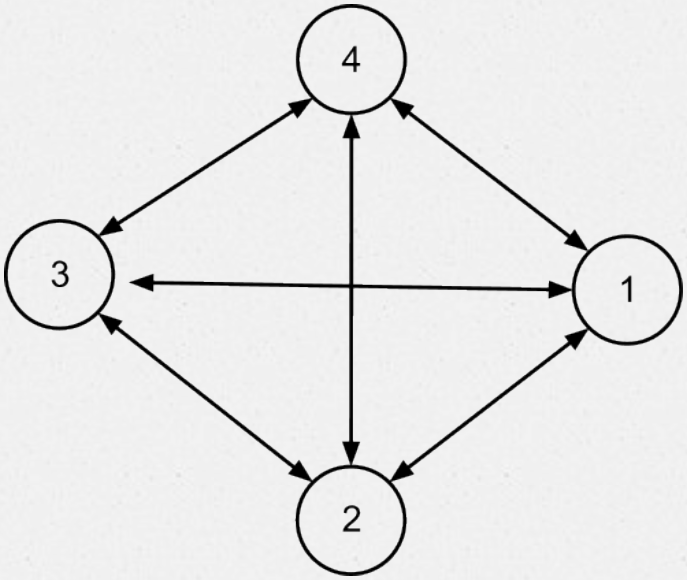
- Virtual machines
- Network topologys
- Performance measures
- Point to point synchronisation
- Detecting filesystem changes

Network topologys

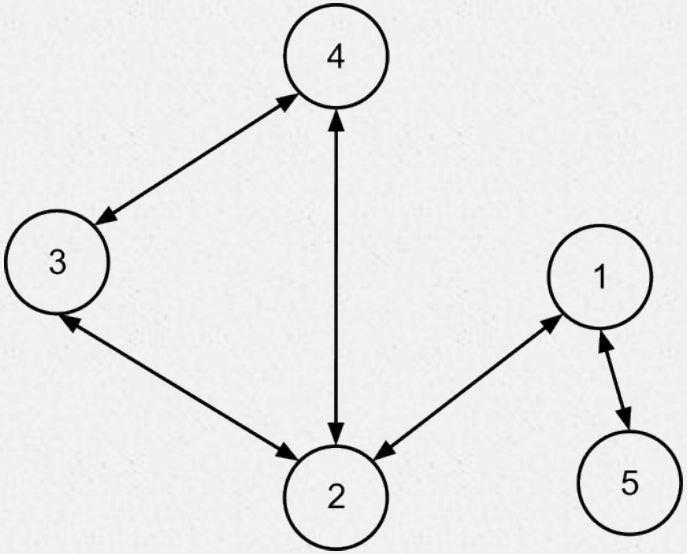


Network topologys

Mesh



Unstructured



Performance measures

- Throughput
- Latency

- Monitoring network interfaces

Point to point synchronisation

- Copy
- Rsync
- Unison
  - Two way replication tool
  - Free open source
  - Runs on many platforms (windows,os x, linux)
  - Keeps itself in a tidy state incase of failure

File system changes

- How does it work?
  - Python script built on pyinotify
  - Reads directories and hosts from config file

  - Inotify (Linux)
    - Linux kernel subsystem to notices changes to the filesystem
  - FSEvents, kqueue (OS X)
  - ReadDirectoryChangesW (Windows)

Next steps

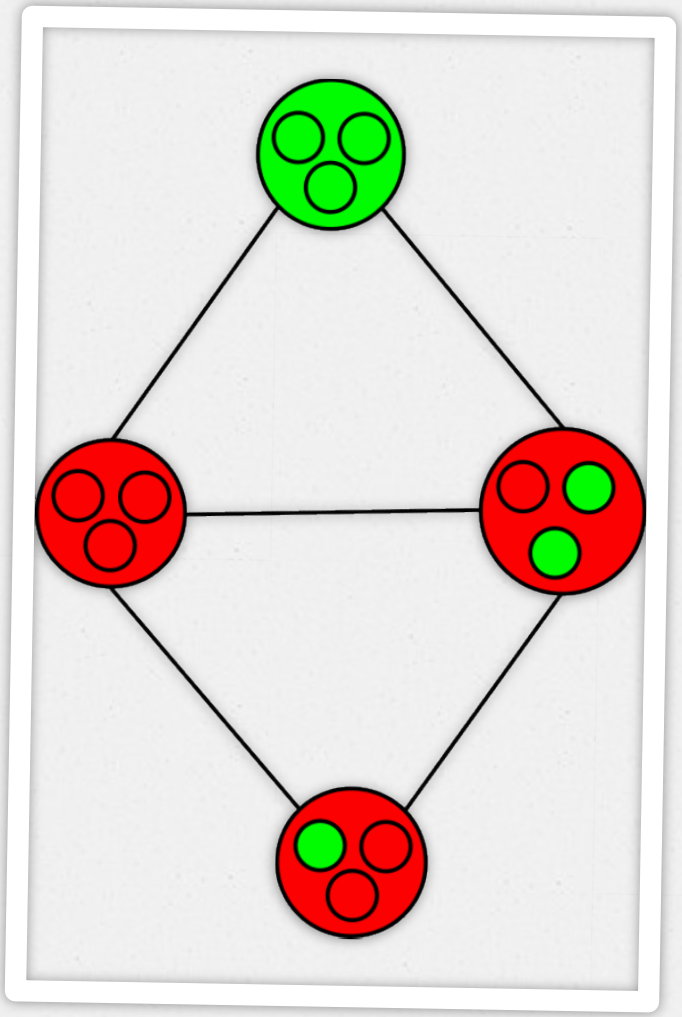
- Sub-nodes
- Mobile nodes
- Statistics
- More user control

Sub-nodes

How out of date is the graph at any given time?

How long do we expect it to take to become completely up to date?

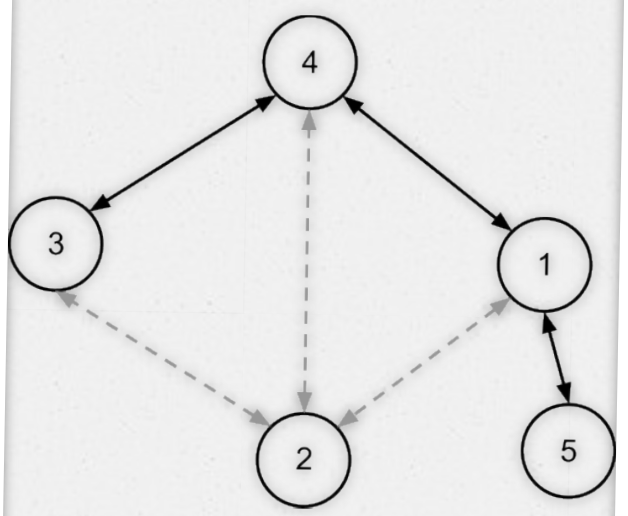
Can we improve this?



Mobile nodes

Say node two intermittently has a connection to a subset of nodes 1,3 and 4

How does this affect the graph over time?



User control

- More user control
- How often to replicate a directory
- Special conditions i.e. only over wi-fi
- Feedback on how changing settings may impact performance

Statistics

- What meaningful data can we gather from the program over time?
- How is each edge in the graph being used?
- How up to date is the graph?
- Can we improve the settings from this knowledge?

The end

- Thanks for listening