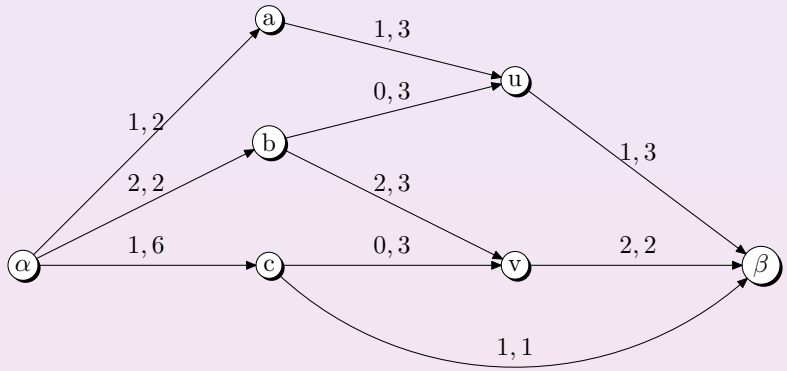
Modelling data networks – research summary and modelling tools



Richard G. Clegg (richard@richardclegg.org)—
Available online at `http://www.richardclegg.org/lectures` accompanying printed notes provide full bibliography.
(Prepared using LaTeX and beamer.)

Research topics and modelling demonstration

- This final lecture describes some important topics and using this work in network research.
- Firstly a word about the applicability of some of the approximations we made.
- These are a personal view of some interesting research topics.
- The mathematical techniques from previous lectures today are important in many of these research areas.
- Finally a demonstration of some well-known modelling tools ns-3 and mininet.

Real life network measurements

- We learned how to completely model a queue with Poisson process input and Poisson service time. Is modelling the internet solved?
- Think back to the start of the lecture we need to think about the input process, the server process and the network topology.
- In fact there is a rather beautiful queuing theory result about networks with all Poisson input (and some other conditions) – the Jackson Network.
- How realistic is it to model our network as M/M/*?
- What is the topology like anyway?

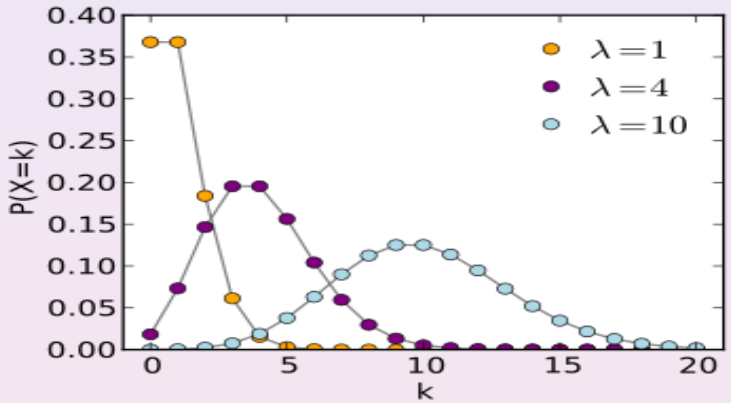
What is a Heavy-Tailed distribution?

Heavy-Tailed distribution

A variable X has a heavy-tailed distribution with $\beta \in (0, 2)$ if

$$\mathbb{P}[X > x] \sim x^{-\beta},$$

where $\sim$ means asymptotically proportional to as $x \to \infty$.



- Obviously an example of a power law *extreme values* are still quite common.
- Examples: Heights of trees, frequency of words, populations of towns.

Heavy tails and the internet

- The following internet distributions have (approximate) heavy tails:
  1. Files on any particular computer.
  2. Files transferred via ftp.
  3. Bytes transferred by single TCP connections.
  4. Files downloaded by the WWW.
- This is more than just a statistical curiosity.
- Consider what this distribution would do to queuing performance (no longer Poisson).
- Non mathematicians are starting to take an interest in heavy tails (reference to "the long tail".

Why do we care? Drastically impacts performance of queues, loading of servers, etc etc.

AS level topology

Power law networks

The node-degree distribution in AS networks is particularly well-studied. Let $P(k)$ be the proportion of nodes with degree k (having k neighbours). To a good approximation

$$P(k) \sim k^{-\alpha},$$

where α is a constant and $\sim$ is asymptotically proportional to.

- Power law topology of the AS graph shown by [Faloutsos x3].
- This graph has some interesting properties — some extremely highly connected nodes, what happens if they fail?
- Same type of graph as:
  1. Links on websites, Wikipedia and many other similar online systems.
  2. Academic citations in papers.
  3. Human frientship, sexual contacts.

Mathematics to generate AS topology

Albert–Barabasi [Barabasi 99] "Preferential attachment" model

Constructive Start with a small "core" network. When a new node arrives, attach it to an old node with the following probability

$$\mathbb{P}\left[\text{Attaching to node } i\right] = \frac{d(i)}{\sum_{j \in \text{all nodes}} d(j)},$$

where $d(i)$ is the degree of node i.

- This model "grows" a network with a powerlaw.
- Many similar models have been created which are more general.

Why do we care? Structure of network tells us about resilience, number of hops to connect and how we should model large scale internet.

Albert-László Barabási (1967 –)

More recently though

- More recently researchers have questioned how exact these power law fits are.
- All of this is still a rapidly evolving reseearch area.
- New results are changing our ideas of what is the most accurate model.
- This is a very active research area – new publications and new results every month.
- What is the "true" model for internet traffic and topology? Nobody knows – I and others are trying to find out.
- We didn't even think about the server process which is mostly deterministic but depends on packet size.

Software Defined Networks (SDN)

Software Defined Networks

Allows more flexible use of the control plane by separating it from the data plane. Programmable control of your network.

OpenFlow

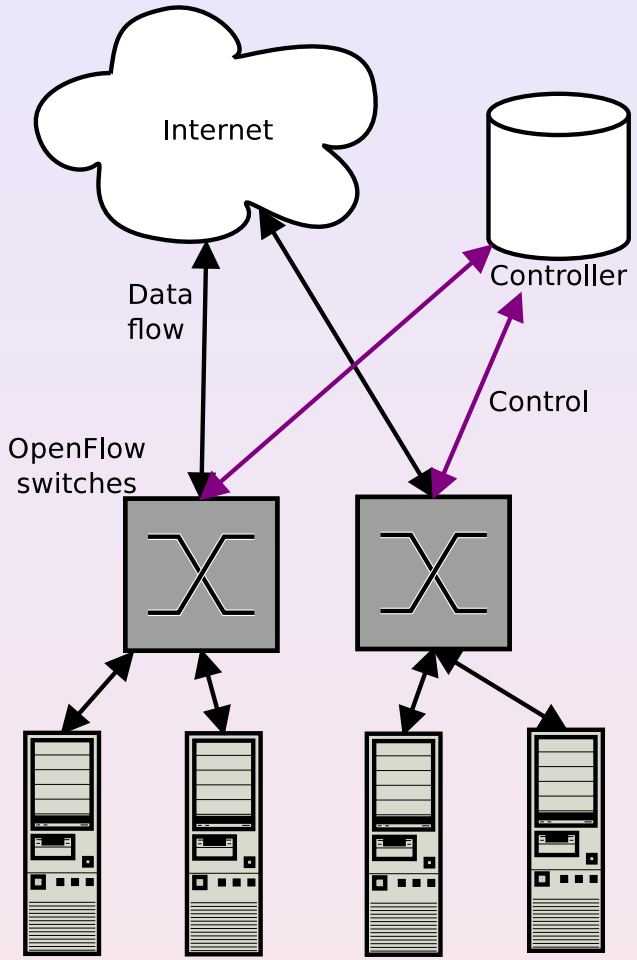
A limited form of SDN which allows programmatic direction of data based on layer-2, layer-3 and layer-4 headers.

Controller contains list of matching rules and actions The first packet of every non matched flow is sent to the controller. Controller uses standard coding languages (python, Java, C++) to decide how to deal with packet and to insert rules.

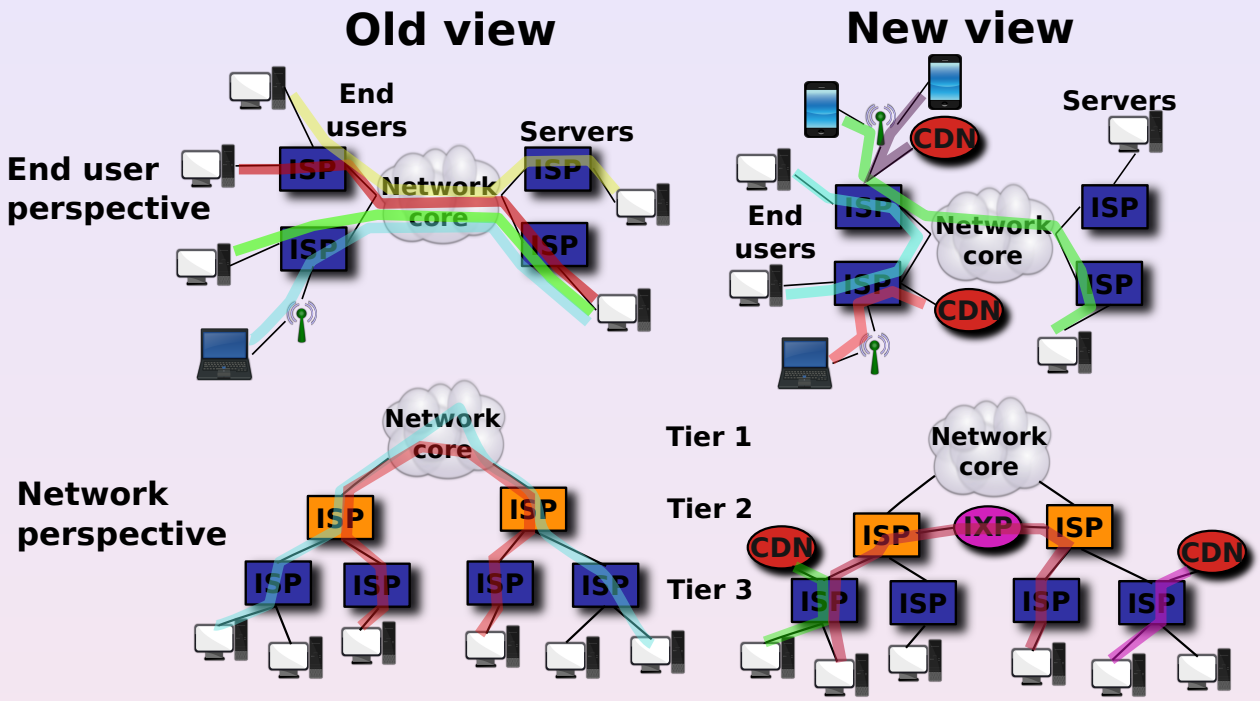
OpenFlow – Big idea

Why do we care?

Can test new ideas on our own computers as if it were a real network. Can deploy new ideas on a testbed with relative ease. Don't need to learn how to wrangle each new router. Flexible and interchangeable routers.



Flattening of the network



Old: A hierarchical network with content far from the user

New: A flatter network with content nearer to the user.

Flattening of the network (2)

Content pushed towards the edges

Old: Client gets content from server

New: Content pushed to content distribution networks (CDNs)

Network topology flattens

Old: Hierarchical network, eyeball ISPs connect to transit providers then tier ones.

New: Internet eXchange Points (IXPs) connect ISPs directly

Why do we care? Structure of network is changing. Content is closer to the user.

Information Centric Networking

How do we connect to a web address?

Put URL in browser `http://www.richardclegg.org/erdos`, translate first part to IP address `89.16.173.124`, connect to that machine, ask it for that data labelled `erdos`.

What's wrong with this? We don't want to connect to that machine, we want to retrieve that content (web page).
How else could we do it? Address maps to content, not machine and does so in a cryptographically signed way. Provably get the data you want from a nearby host.
Why do we care? Interesting new trend. If deployed, will radically change how networks operate.

The ns-3 simulation

- ns-3 is a freely available event-driven simulator which simulates packet-level traffic.
- It is available from `http://nsnam.isi.edu/nsnam`
- It does allow you to connect real traffic sources into your simulation.
- It has many prebuilt modelling scenarios which may be useful for the modelling you wish to do.
- The scripts used for these examples are available at `http://www.richardclegg.org/modelling_networks`
- It is available from `https://www.nsnam.org/`

OpenFlow and mininet

- Mininet is a tool which can be used for modelling network topologies.
- Real network connections are made on a single machine.
- We can use this in combination with the Floodlight OpenFlow controller to demonstrate OpenFlow networking.
- More like emulation than simulation – actual packets are being sent.

Use cases

ns-3 use case

Larger simulation, packet level simulation with real applications injecting packets into the simulation.

mininet use case

Smaller simulation, actual programs run on VMs sending data between them on simple topologies.

Summary of today's lectures

Stochastic processes

A building block for many mathematical models of the internet. Characterising some part of the network in terms of a stochastic process allows you to understand it better.

Markov chains

A flexible mathematical tool allowing you to model how a system evolves in time and a necessary prerequisite for queuing theory.

Simulation modelling

Sometimes simulation is the only way to proceed but study of the existing system is necessary to make the simulation match reality.

Final thoughts

- Select an appropriate level of modelling — if you need to model the whole internet you cannot do packet level modelling. If you need to model intricate protocol details for packets you cannot model the whole internet.
- Check against real data where possible that your modelling assumptions are justified.
- Is your experiment repeatable? Do you get similar results if you try slightly different starting scenarios?
- Remember sensitivity analysis: What happens if the bandwidth is a little less? What if the demand is a little more?
- Can statistical analysis of your results help?
- Don't reinvent the wheel – what is out there to help you already?