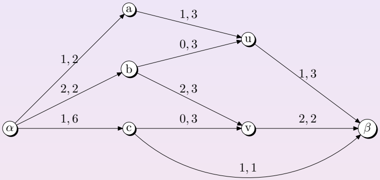
Modelling data networks – introduction



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.)

Difficulties in modelling the Internet

- See "Difficulties in simulating the internet" Floyd & Paxson 2001.
- The internet is big (and growing).
- The internet is heterogenous to a large degree.
- No central maps exist of the internet.
- The internet is not always easy to measure.
- The internet is rapidly changing.
- It is extremely important to be able to model the internet.

The internet cannot possibly be modelled, yet we must model the internet. How can this be resolved?

Aim of these of lectures

Learn about basic mathematics necessary to model the internet

Stochastic processes

Modelling how things arrive on a network (in particular the Poisson process).

Markov chains

A useful way of modelling the state of networks and necessary for queuing theory.

Queu(e)ing theory

The mathematics of how things queue – packets in routers, processes in a CPU...

Steps to modelling

- How you model the network depends critically on the problem you are solving.
- What are you trying to show with your model?
- Metrics: what are we trying to measure?
  1. Throughput?
  2. Goodput?
  3. System efficiency?
- Validation: what real data can be used to check the model?
- Sensitivity: what happens if your assumptions change?
  1. What if the demand on the system is slightly different?
  2. What happens if delays and bandwidths are changed?
  3. What happens if users stay longer or download more?

Important questsions for modelling

1. How much of the network do we model?
   - Whole internet (then we can't even model every computer – every AS?)
   - A few typical nodes?
   - A sub net?
   - A single queue and buffer?
2. What level of modelling is appropriate?
   - Mathematical – solution "instant" (or quick)
   - Detailed simulation
   - Combined approach (equations abstract away some details with approximations)
3. How far down the network stack need we go?

Model example one – peer-to-peer network

Modelling Task

Test the possible improvements expected if we try a locality aware peer selection policy on a global bittorrent network.

What must our model include?

1. The distribution of nodes (peers) on the overlay network (not the whole network).
2. The delay and throughput between these peers (must depend on distance to some extent).
3. How users arrive and depart.
4. What users choose to download.

Note that this might already be a vast modelling task with hundreds of thousands or even millions of nodes.

Approach to model one – peer-to-peer network

- Research existing P2P models, do any fit? Don't reinvent the wheel.
- Real data: What real-life measurements exist to validate against?
- If we are modelling a new peer selection we must be sure our model covers existing peer selection well.
- Metrics: what must we measure in our model?
  1. Overall throughput/goodput?
  2. Distribution of time taken for peers to make their download?
  3. Total resources used in system?
- Validation: Instrumented P2P clients exist – how do they compare to our simulation.
- Sensitivity: Different distribution of users? Different delays and throughputs?

Model example two – Buffer overflow model

Modelling task

Given a router with a buffer, how does the buffer size in packets affect the probability of packet loss?

What must our model include?

1. A model of the incoming packets to the buffer.
2. The rate at which packets leave the buffer.
3. Possibly distribution of packet lengths in bytes.
4. Possibly the feedback (TCP) between packet loss and arrival rate.

Approach to model two – Buffer overflow model

- Research: what is known about the statistics of internet traffic?
- What is the distribution of inter-arrival times and packet lengths?
- Metrics:
  1. Packet loss.
  2. Packet delay.
- Sensitivity: What if we change the following parameters:
  1. The total arrival rate.
  2. The bandwidth of the outgoing link.
- Validation: Real traffic traces (CAIDA has a collection).

Model example three – TCP protocol model

Modelling Task

Test a possible improvement to the TCP model which aims to improve fairness and throughput when flows share a link.

What must our model include?

1. Individual packet model (existing TCP protocol well modelled).
2. A reasonable estimate of how long each connection lasts and the rate at which new connections.
3. A model of the probability of round trip time for the parts of the connection not on the link being modelled.
4. A model of the probability of packet loss on the link (due to buffer overflow?)

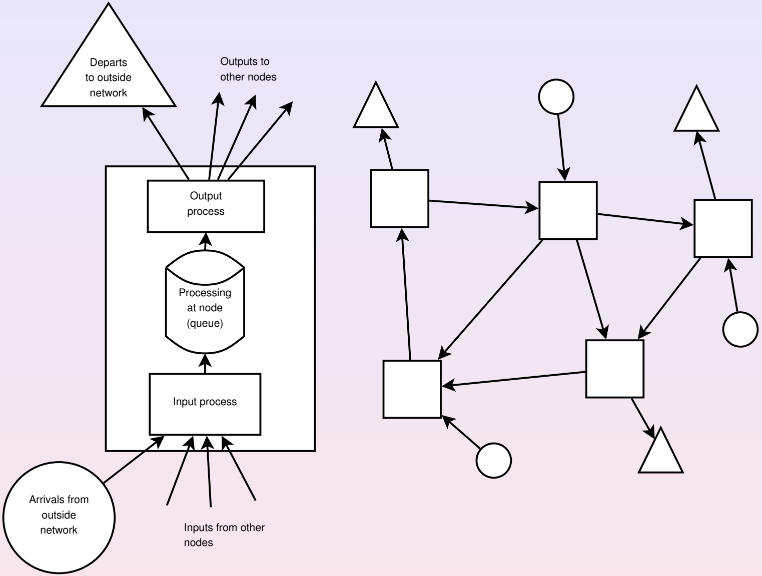
Approach to model three – TCP protocol model

- Can existing network models help (ns-3/mininet could be an obvious choice)?
- What if the existing protocol shares a link with flows using the old protocol.
- Metrics:
  1. Throughput and goodput.
  2. Fairness between flows.
- Sensitivity, what if we change these parameters:
  1. Number of flows using existing and new protocol.
  2. Bandwidth of link.
  3. Round trip time of flows.
  4. Probability of packet loss.
- Validation: Does our model agree with real measurements?

Other things to model

- Of course depending on the nature of your modelling, there may well be other aspects of the network to be modelled.
- Some examples might be:
    1. Reliability of nodes and links.
    2. Possible hostile attacks to the network.
- In all cases, an important starting point is to find out what research already exists in the area.
- Are any real-life data sets available which could inform your modelling? Could you gather such data?

The basics of a network model



The basics of a network model

A network model could be viewed as these components.

Arrival process

A statistical process describing how objects (packets) arrive in the network – statistical process modelling.

Queueing process

A model which describes how objects (packets) are processed by a network node – queuing theory.

Topology

A "network topology" – the wiring diagram which shows how these things connect together.

Check your understanding

What are the concepts being described here?

- The response of a model to changes in its input data.
- Checking a model's performance against real world data.
- The process describing how and when new objects arrive in a network.

Check your understanding

What are the concepts being described here?

- The response of a model to changes in its input data. Sensitivity
- Checking a model's performance against real world data. Validation
- The process describing how and when new objects arrive in a network. Arrival process

Rest of lectures

The rest of these lectures will, therefore, cover:

- Statistical processes – in particular the Poisson process.
- Markov chains – a useful modelling tool in themselves and a prerequisite for...
- Queuing theory – the mathematical study of how things join and leave queues.

Finally a short lecture will cover:

- A summary of basic research in the internet.
- A brief demo of a well-known network modelling tool ns-3 and a virtual network controller.