Performance Modelling — Lecture 1

Jane Hillston

School of Informatics
The University of Edinburgh
Scotland

Key notions

A model can be constructed to represent some aspect of the dynamic behaviour of a system.

Key notions

A model can be constructed to represent some aspect of the dynamic behaviour of a system.

Once constructed, such a model becomes a tool with which we can investigate the behaviour of the system.

Modelling computer systems: the challenges

- Physical distance
    - Network latency

Modelling computer systems: the challenges

- Physical distance
  - Network latency
- Partial failures
  - Server may be down
  - Routers may be down

Modelling computer systems: the challenges

- Physical distance
    - Network latency
- Partial failures
    - Server may be down
    - Routers may be down
- Scale
    - Workload characterisation

Modelling computer systems: the challenges

- Physical distance
  - Network latency
- Partial failures
  - Server may be down
  - Routers may be down
- Scale
  - Workload characterisation
- Resource sharing
  - Network contention
  - CPU load

Modelling computer systems: the challenges

- Physical distance — need to represent time
  - Network latency
- Partial failures
  - Server may be down
  - Routers may be down
- Scale
  - Workload characterisation
- Resource sharing
  - Network contention
  - CPU load

Modelling computer systems: the challenges

- Physical distance — need to represent time
  - Network latency
- Partial failures — randomness and probability
  - Server may be down
  - Routers may be down
- Scale
  - Workload characterisation
- Resource sharing
  - Network contention
  - CPU load

Modelling computer systems: the challenges

- Physical distance — need to represent time
  - Network latency
- Partial failures — randomness and probability
  - Server may be down
  - Routers may be down
- Scale — need to quantify population sizes
  - Workload characterisation
- Resource sharing
  - Network contention
  - CPU load

Modelling computer systems: the challenges

- Physical distance — need to represent time
  - Network latency
- Partial failures — randomness and probability
  - Server may be down
  - Routers may be down
- Scale — need to quantify population sizes
  - Workload characterisation
- Resource sharing — need to express percentages
  - Network contention
  - CPU load

Modelling computer systems: the challenges

Time	What representation of time will we use?
Randomness	What kind of random number distributions will we use?
Probability	How can we have probabilities in the model without uncertainty in the results?
Scale	How can we escape the state-space explosion problem?
Percentages	What can it mean to have a fraction of a process?

The discrete event view

In this course we will consider discrete event systems.

The discrete event view

In this course we will consider discrete event systems.

The state of the system is characterised by variables which take distinct values and which change by discrete events, i.e. at a distinct time something happens within the system which results in a change in one or more of the state variables.

The discrete event view: example

We might be interested in the number of nodes in a communication network which are currently waiting to send a message N.

- If a node, which was not previously waiting, generates a message and is now waiting to send then $N \rightarrow N + 1$, or
- If a node, which was previously waiting, successfully transmits its message then $N \rightarrow N - 1$.

Discrete time vs. Continuous time

Within discrete event systems there is a distinction between a discrete time representation and a continuous time representation:

Discrete time vs. Continuous time

Within discrete event systems there is a distinction between a discrete time representation and a continuous time representation:

Discrete time: such models only consider the system at predetermined moments in time, which are typically evenly spaced, eg. at each clock "tick".

Continuous time: such models consider the system at the time of each event so the time parameter in such models is conceptually continuous.

Discrete time vs. Continuous time

Within discrete event systems there is a distinction between a discrete time representation and a continuous time representation:

Discrete time: such models only consider the system at predetermined moments in time, which are typically evenly spaced, eg. at each clock "tick".

Continuous time: such models consider the system at the time of each event so the time parameter in such models is conceptually continuous.

At levels of abstraction above the hardware clock continuous time models are generally appropriate for computer and communication systems.

Performance Modelling

Performance modelling is concerned with the dynamic behaviour of systems and quantified assessment of that behaviour.

Performance Modelling

Performance modelling is concerned with the dynamic behaviour of systems and quantified assessment of that behaviour.

There are often conflicting interests at play:

- Users typically want to optimise external measurements of the dynamics such as response time (as small as possible), throughput (as high as possible) or blocking probability (preferably zero);

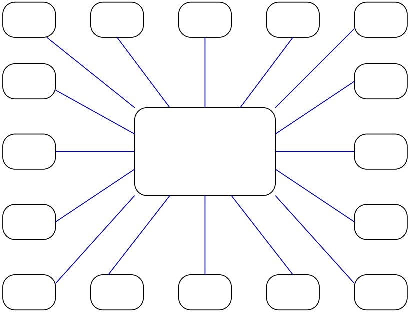
Performance Modelling

Performance modelling is concerned with the dynamic behaviour of systems and quantified assessment of that behaviour.

There are often conflicting interests at play:

- Users typically want to optimise external measurements of the dynamics such as response time (as small as possible), throughput (as high as possible) or blocking probability (preferably zero);
- In contrast, system managers may seek to optimize internal measurements of the dynamics such as utilisation (reasonably high, but not too high), idle time (as small as possible) or failure rates (as low as possible).

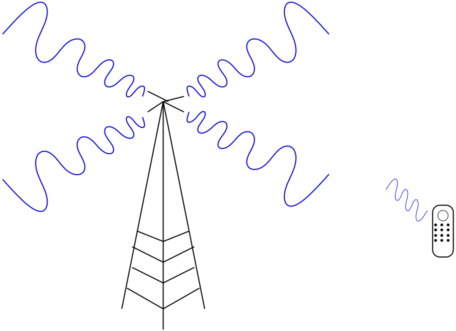
Performance Modelling: Motivation



Capacity planning

- How many clients can the existing server support and maintain reasonable response times?

Performance Modelling: Motivation

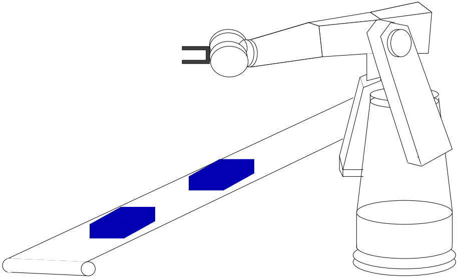


Mobile Telephone Antenna

System Configuration

- How many frequencies do you need to keep blocking probabilities low?

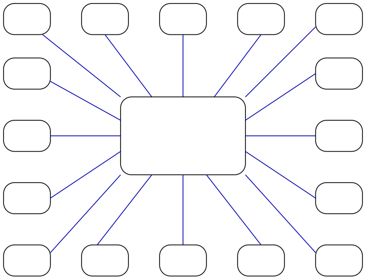
Performance Modelling: Motivation



System Tuning

- What speed of conveyor belt will minimize robot idle time and maximize throughput whilst avoiding lost widgets?

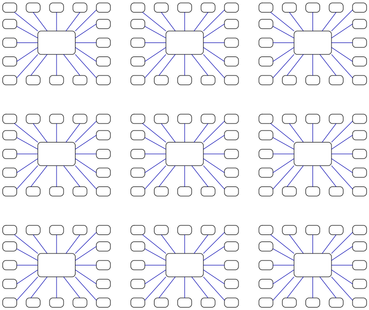
Performance Modelling: Response time analysis



Quality of Service issues

- Can the server maintain reasonable response times?

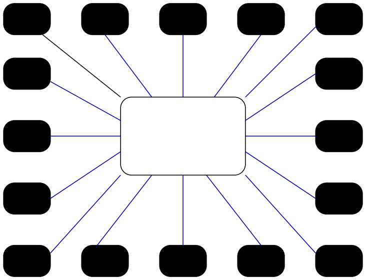
Performance Modelling: Capacity planning



Scalability and capacity planning issues

- How many times do we have to replicate this service to support all of the subscribers?

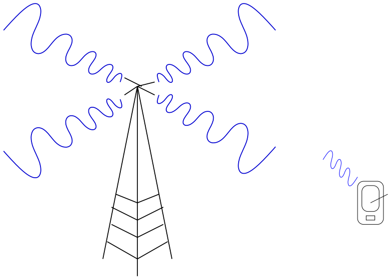
Performance Modelling: Scalability analysis



Robustness and scalability issues

- Will the server withstand a distributed denial of service attack?

Performance Modelling: Service Level Agreements



Service-level agreements

- What percentage of downloads do complete within the time we advertised?

Probability Theory

1. There is a sample space, Ω, which encompasses all possible observations or outcomes.
2. There is a collection of subsets of Ω, denoted E, and termed events; these subsets are usually identified as sample points which satisfy some condition.
3. There is a probability mapping, Pr, from E to $\mathbb{R}$. Pr must satisfy three simple conditions:
   1. For any event A, $A \in E$, the mapping Pr is defined and satisfies $0 \leq \Pr(A) \leq 1$.
   2. $\Pr(\Omega) = 1$.
   3. If A and B are mutually exclusive, that is, they contain no sample points in common, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

Simple properties of probabilities

Various properties of probabilities can be derived from the axioms and simple set theory.

For example, the probability of the union of two events A and B which are **not** mutually exclusive is

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

Similarly, the probability of the complement of event A, denoted by $\neg A$, is

$$\Pr(\neg A) = 1 - \Pr(A).$$

Conditional Probability

The conditional probability of A occurring, given that B has occurred, is

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Conditional Probability

The conditional probability of A occurring, given that B has occurred, is

$$\boxed{\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}}$$

- If A and B are mutually exclusive $\Pr(A \mid B) = 0$.
- If B is a precondition for A, then $\Pr(A \cap B) = \Pr(A)$.
- Two events are independent if knowledge of the occurrence of one of them tells us nothing about the probability of the other, i.e. $\Pr(A \mid B) = \Pr(A)$, or

$$\boxed{\Pr(A \cap B) = \Pr(A) \times \Pr(B).}$$

Random experiments and events

- To apply probability theory to the process under study, we view it as a random experiment.

Random experiments and events

- To apply probability theory to the process under study, we view it as a random experiment.
- The sample space of a random experiment is the set of all individual outcomes of the experiment.

Random experiments and events

- To apply probability theory to the process under study, we view it as a random experiment.
- The sample space of a random experiment is the set of all individual outcomes of the experiment.
- These individual outcomes are also called sample points or elementary events.

Random experiments and events

- To apply probability theory to the process under study, we view it as a random experiment.
- The sample space of a random experiment is the set of all individual outcomes of the experiment.
- These individual outcomes are also called sample points or elementary events.
- An event is a subset of a sample space.

Random variables

We are interested in the dynamics of a system as events happen over time.

A function which associates a (real-valued) number with the outcome of an experiment is known as a random variable.

Formally, a random variable X is a real-valued function defined on a sample space Ω.

Measurable functions

If X is a random variable, and x is a real number, we write $X \leq x$ for the event

$$\{\, \omega : \omega \in \Omega \text{ and } X(\omega) \leq x \,\}$$

Measurable functions

If X is a random variable, and x is a real number, we write $X \leq x$ for the event

$$\{\, \omega : \omega \in \Omega \text{ and } X(\omega) \leq x \,\}$$

and we write $X = x$ for the event

$$\{\, \omega : \omega \in \Omega \text{ and } X(\omega) = x \,\}$$

Measurable functions

If X is a random variable, and x is a real number, we write $X \leq x$ for the event

$$\{\, \omega : \omega \in \Omega \text{ and } X(\omega) \leq x \,\}$$

and we write $X = x$ for the event

$$\{\, \omega : \omega \in \Omega \text{ and } X(\omega) = x \,\}$$

Another property required of a random variable is that the set $X \leq x$ is an event for each real x. This is necessary so that probability calculations can be made. A function having this property is said to be a measurable function or measurable in the Borel sense.

Distribution function

For each random variable X we define its distribution function F for each real x by

$$F(x) = \Pr[X \leq x]$$

Distribution function

For each random variable X we define its distribution function F for each real x by

$$F(x) = \Pr[X \leq x]$$

We associate another function $p(\cdot)$, called the probability mass function, with X (pmf), for each x:

$$p(x) = \Pr[X = x]$$

Distribution function

For each random variable X we define its distribution function F for each real x by

$$F(x) = \Pr[X \leq x]$$

We associate another function $p(\cdot)$, called the probability mass function, with X (pmf), for each x:

$$p(x) = \Pr[X = x]$$

A random variable X is continuous if $p(x) = 0$ for all real x.

Distribution function

For each random variable X we define its distribution function F for each real x by

$$F(x) = \Pr[X \leq x]$$

We associate another function $p(\cdot)$, called the probability mass function, with X (pmf), for each x:

$$p(x) = \Pr[X = x]$$

A random variable X is continuous if $p(x) = 0$ for all real x.

(If X is a continuous random variable, then X can assume infinitely many values, and so it is reasonable that the probability of its assuming any specific value we choose beforehand is zero.)

Mean, or expected value

If X is a continuous random variable with density function $f(\cdot)$, we define the mean or expected value of X, $\mu = E[X]$ by

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Mean, or expected value

If X is a continuous random variable with density function $f(\cdot)$, we define the mean or expected value of X, $\mu = E[X]$ by

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

If X is a discrete random variable with probability mass function $p(\cdot)$, we define the mean or expected value of $X \in S$, $\mu = E[X]$ by

$$E(X) = \sum_{x \in S} x p(x)$$

Variance

The expectation only gives us an idea of the average value assumed by a random variable, not how much individual values may differ from this average.

The variance, $Var(X)$, gives us an indication of the "spread" of values:

$$\boxed{Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2.}$$

Exponential random variables, distribution function

The random variable X is said to be an *exponential random variable with parameter* λ ($\lambda > 0$) or to have an exponential distribution with parameter λ if it has the distribution function

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

Exponential random variables, distribution function

The random variable X is said to be an *exponential random variable with parameter* λ ($\lambda > 0$) or to have an exponential distribution with parameter λ if it has the distribution function

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

Some authors call this distribution the negative exponential distribution.

Exponential random variables, density function

The density function $f = dF/dx$ is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Mean, or expected value, of the exponential distribution

Suppose X has an exponential distribution with parameter $\lambda > 0$. Then

$$\mu = E[X] = \int_{-\infty}^{\infty} x\lambda e^{-\lambda x} dx$$

Mean, or expected value, of the exponential distribution

Suppose X has an exponential distribution with parameter $\lambda > 0$. Then

$$\mu = E[X] = \int_{-\infty}^{\infty} x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

Exponential inter-event time distribution

The time interval between successive events can also be deduced.

Let $F(t)$ be the distribution function of T, the time between events. Consider $\Pr(T > t) = 1 - F(t)$:

$$\Pr(T > t) = \Pr(\text{No events in an interval of length } t)$$

Exponential inter-event time distribution

The time interval between successive events can also be deduced.

Let $F(t)$ be the distribution function of T, the time between events. Consider $\Pr(T > t) = 1 - F(t)$:

$$\begin{aligned} \Pr(T > t) &= \Pr(\text{No events in an interval of length } t) \\ &= 1 - F(t) \end{aligned}$$

Exponential inter-event time distribution

The time interval between successive events can also be deduced.

Let $F(t)$ be the distribution function of T, the time between events. Consider $\Pr(T > t) = 1 - F(t)$:

$$\begin{aligned}\Pr(T > t) &= \Pr(\text{No events in an interval of length } t) \\ &= 1 - F(t) \\ &= 1 - (1 - e^{-\lambda t})\end{aligned}$$

Exponential inter-event time distribution

The time interval between successive events can also be deduced.

Let $F(t)$ be the distribution function of T, the time between events. Consider $\Pr(T > t) = 1 - F(t)$:

$$\begin{aligned} \Pr(T > t) &= \Pr(\text{No events in an interval of length } t) \\ &= 1 - F(t) \\ &= 1 - (1 - e^{-\lambda t}) \\ &= e^{-\lambda t} \end{aligned}$$

Memoryless property of the exponential distribution

The memoryless property of the exponential distribution is so called because the time to the next event is independent of when the last event occurred.

Memoryless property of the exponential distribution

Suppose that the last event was at time 0. What is the probability that the next event will be after $t + s$, given that time t has elapsed since the last event, and no events have occurred?

Memoryless property of the exponential distribution

Suppose that the last event was at time 0. What is the probability that the next event will be after $t + s$, given that time t has elapsed since the last event, and no events have occurred?

$$\Pr(T > t + s \mid T > t) = \frac{\Pr(T > t + s \text{ and } T > t)}{\Pr(T > t)}$$

Memoryless property of the exponential distribution

Suppose that the last event was at time 0. What is the probability that the next event will be after $t+s$, given that time t has elapsed since the last event, and no events have occurred?

$$\begin{aligned}\Pr(T > t+s \mid T > t) &= \frac{\Pr(T > t+s \text{ and } T > t)}{\Pr(T > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}}\end{aligned}$$

Memoryless property of the exponential distribution

Suppose that the last event was at time 0. What is the probability that the next event will be after $t + s$, given that time t has elapsed since the last event, and no events have occurred?

$$\begin{aligned} \Pr(T > t + s \mid T > t) &= \frac{\Pr(T > t + s \text{ and } T > t)}{\Pr(T > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \end{aligned}$$

Memoryless property of the exponential distribution

Suppose that the last event was at time 0. What is the probability that the next event will be after $t+s$, given that time t has elapsed since the last event, and no events have occurred?

$$\begin{aligned}\Pr(T > t+s \mid T > t) &= \frac{\Pr(T > t+s \text{ and } T > t)}{\Pr(T > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s}\end{aligned}$$

This value is independent of t (and so the time already spent has not been remembered).

The Poisson distribution

The exponential distribution function is closely related to a discrete random variable, the Poisson distribution.

This random variable takes values in the set $\{0, 1, 2, \ldots\}$ and has the mass function

$$p_i = e^{-\mu} \frac{\mu^i}{i!} \qquad i \geq 0.$$

The Poisson distribution

The exponential distribution function is closely related to a discrete random variable, the Poisson distribution.

This random variable takes values in the set $\{0, 1, 2, \ldots\}$ and has the mass function

$$p_i = e^{-\mu} \frac{\mu^i}{i!} \qquad i \geq 0.$$

The expectation of a Poisson random variable with parameter μ is also μ.

The Poisson random variable

The Poisson random variable is typically used as a counting variable, recording the number of events that occur in a given period of time.

The Poisson random variable

The Poisson random variable is typically used as a counting variable, recording the number of events that occur in a given period of time.

If we observe a Poisson process with parameter μ for some short time period of length h then:

- the probability that one event occurs is $\mu h + o(h)$.
- the probability that two or more events occur is $o(h)$.
- the probability that no events occur is $1 - \mu h + o(h)$.

Relationship between the Poisson and Exponential distributions

If we observe a Poisson process for a infinitesimal time period dt the probability that an event occurs is μdt.

Relationship between the Poisson and Exponential distributions

If we observe a Poisson process for a infinitesimal time period *dt* the probability that an event occurs is μdt.

If the occurrence of events is governed by a Poisson distribution then the inter-event times are governed by an exponential distribution with the same parameter, and vice versa.

Relationship between the Poisson and Exponential distributions

If we observe a Poisson process for a infinitesimal time period dt the probability that an event occurs is μdt.

If the occurrence of events is governed by a Poisson distribution then the inter-event times are governed by an exponential distribution with the same parameter, and vice versa.

Therefore, if we know that the delay until an event is exponentially distributed then the probability that it will occur in the infinitesimal time interval of length dt, is μdt.

Topics Covered in the Course

- Operational Laws
- Modelling with Continuous Time Markov Chains
- High-level modelling formalisms:
    - Queueing Networks
    - Stochastic Petri Nets
    - Stochastic Process Algebra
- Stochastic Simulation
- Model Parameterisation
- Model Verification and Validation

Practicalities

- There will be eighteen 50 minute lectures;
  Tuesdays in F21 in 7 George Square and Fridays in AT2.12.

Practicalities

- There will be eighteen 50 minute lectures;
  Tuesdays in F21 in 7 George Square and Fridays in AT2.12.
- There are no formal tutorials for the course but individual help will be available from the lecturer by arrangement.

Practicalities

- There will be eighteen 50 minute lectures;
  Tuesdays in F21 in 7 George Square and Fridays in AT2.12.
- There are no formal tutorials for the course but individual help will be available from the lecturer by arrangement.
- Learn by experimentation: example models and instructions on how to use appropriate software will be given in lectures and on the web page.

Practicalities

- There will be eighteen 50 minute lectures; Tuesdays in F21 in 7 George Square and Fridays in AT2.12.
- There are no formal tutorials for the course but individual help will be available from the lecturer by arrangement.
- Learn by experimentation: example models and instructions on how to use appropriate software will be given in lectures and on the web page.
- There is a set of lecture notes. These will appear on the web page at least 24 hours prior to the lecture. You are encouraged to read them prior to the lecture.

Practicalities

- There will be eighteen 50 minute lectures;
  Tuesdays in F21 in 7 George Square and Fridays in AT2.12.
- There are no formal tutorials for the course but individual help will be available from the lecturer by arrangement.
- Learn by experimentation: example models and instructions on how to use appropriate software will be given in lectures and on the web page.
- There is a set of lecture notes. These will appear on the web page at least 24 hours prior to the lecture. You are encouraged to read them prior to the lecture.
- Thee are two practicals which together account for 25% of the marks for the course. The first will be due in the early November; the second in the second week of December.