Solution Notes for

Analytical Network & System Admin.

These should be seen as a basis for class discussion. The aim is to get practice in modelling.

Solutions to exercises

Ex 1: Skim through course book ....

S1: e.g. i) how do I find the safest time to take a disk backup?

ii) which routers should I spend most money on?

iii) how should I give priority to the items in my budget, given a list of aims / policy that I need to fulfill.

(iv) how do I value / rank tasks to be carried out?

Ex 2: Comment on the difference between the scientific method and math.

S2: Science is about observing what we do not understand and guessing an explanation. There is no right / wrong answer.

Math starts with an assumption / axiom and determines a correct consequence of that assumption (assuming we do not make a mistake). - it is a modelling language for science.

Once we have formulated a hypothesis in science, we can determine logical consequences of that (e.g. cases of a formula), but a hypothesis is only a quasi-axiom.

We must perform experiments to see if it is <u>probably</u> true.

Popper said that, if we can show one example where the hypothesis is not true, then we have shown it is incorrect.

We cannot prove that something will never happen unless we know that our facts are 100% certain. But if we know 90% - we can prove with math the logical consequences, <u>iff true</u>

Ex 3: Comment on the ethics of science.

S3: Some thoughts.

- science has a 'perceived authority' which can be abused. (Many people believe that science is "the truth".)
- we can use scientific methods to find optimal answers, given an initial policy for what we want.
  It can be considered ethical to do the best job we can ✓
  But we have to take into account things (constraints) that might be difficult to model.
- It might be considered unethical to guess if we have a way of finding the best answer.
- We must be careful not to use words like "right" or "wrong" way to do things. We have to define these words. Even with science/math there might not be <u>one</u> <u>best</u> <u>answer</u>.
- Perhaps it is our ethical duty to be rational and critical of what we do. Science/math are tools that help us.
- Would it be as good to cheat? We could try to answer this later more formally using game theory — by inventing a numerical scale for <u>policy</u> <u>reward</u>. As always — it depends on priorities and circumstances — so it is complicated, but not impossible. By formulating something mathematically, it forces us to confront our assumptions and question them!

Ex 4: Happy users are well behaved users ... how do we test this?

S4: There are 2 concepts:

- happiness
- well-behavedness

How can we measure these things? (Blood test? Brainwaves?)
If we can invent a scale of measurement, we can collect statistics and see if there is a correlation / relationship between them. Ideally, we would like a numerical scale / something to count.

Happiness	Good Behaviour
How many times do users hit restrictions we have set?	Measure number of times a user breaks policy / or hits quotas.
(here if we choose these - they are the same!)	
Ask users with questionnaire	Ask users with questionnaire?
Is load / slowness related to stress → happiness?	Use unauthorised tools?
Must isolate other influences? or not?	How do we take account of personality?
Measure agressiveness of keystrokes / mouse clicks?	How many problems do we have to fix related to user behaviour?
Recognise bad words from the mouths ... how much people smile.	Conflicts with other users.

Can we make conclusions about individual users - or on average for the whole organization?

Question: do users have a 'right' to be happy? Should they have to follow any rules?

Idea: happiness makes us tolerant of others.
happiness makes us complacent about rules - inconvenience doesn't matter.

unhappiness - rules can be ignored if we are desperate or motivated.

This is an interesting example because it reveals the problem that one has in social sciences. It is easy to think of hypotheses, but it is very hard to test them rationally!

e.g. is laughter related to happiness.

Exercise 6

Consider the set / sequence:

$$\{x\} = \{1, 4, 3, 2, 6, 5, 8, 9, 7, 10\}$$

1.) What is $\sum_{i=1}^{3} x_i$?

$$= 1 + 4 + 3 = \underline{\underline{8}}$$

2) $\sum_{i=1}^{10} x_i = \sum_{i=1}^{10} i = \underline{\underline{55}}$

3) $\prod_{i=1}^{4} x_i = 1 \times 4 \times 3 \times 2 = \underline{\underline{24}}$

4) The mean value is sum of numbers / number.

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \underline{\underline{5.5}}$$

5) The expectation of this set, interpreted as a distribution is the most likely weighted value.

$$\langle x \rangle = \sum_{i=1}^{10} p(i)\, i$$

where $p(i)$ is the probability of finding the i-th number in the sequence. In this case, each number occurs only once in the sequence, so that probability, for all i, is

$$p(i) = \frac{1}{10} \quad , \forall i$$

Hence $\langle x \rangle = \sum_{i=1}^{10} \frac{1}{10} i = \frac{\sum_{i=1}^{10} x_i}{10} = \underline{\underline{5.5}}$

(as in (4) - i.e. just the same as the mean value).

(6) There is no difference between the mean value $(\overline{x})$ and expectation value of x_i: $(\langle x \rangle)$ except in the way we calculate them; provided we have access to all the data:

$$\overline{x} = \frac{\sum_{\text{numbers}} x_i}{\text{Number of } x_i}$$

$\boxed{\text{If we work out a mean based only on the sample classes, clearly we get a different answer, since we don't record how many times each occured.}}$

$$\langle x \rangle = \sum_{\text{types}} \text{probablity (type)} \times \text{type}$$

e.g. Consider $\{y\} = \{1, 1, 1, 1, 2, 3, 3\}$

$$\overline{y} = \frac{1+1+1+1+2+3+3}{7} = \frac{12}{7} \approx \underline{1 \cdot 9}$$ (average value)

$$\boxed{\neq \frac{1+2+3}{3} = \frac{1}{2}.}$$

$$\langle y \rangle = \sum_{i=\{1,2,3\}} p(i)\, i \, .$$

$$p(1) = \frac{4}{7}$$
$$p(2) = \frac{1}{7}$$
$$p(3) = \frac{2}{7}$$

$$\Rightarrow \quad \langle y \rangle = \left(\frac{4}{7} \times 1\right) + \left(\frac{1}{7} \times 2\right) + \left(\frac{2}{7} \times 3\right) = \underline{\underline{\frac{12}{7}}}$$ (probable value) $\approx 1 \cdot 7$

~~These differ when we consider functions of the variables~~.

Notice that there are two different interepretations – "middle value" and "likely value". One is based on data, the other is based on experience. It is a question of how we present the problem (how we approach it).

Suppose we only know the categories and the probability with which they occur; e.g. we know that network packed are either:

TCP	$p(TCP) = 5/8$
UDP	$p(UDP) = 2/8$
ICMP	$p(ICMP) = 1/8$

and we know the probabilities with which they arrive, i.e. TCP is 5 times more likely that ICMP.

Now, suppose we know a property that is related to this distribution, like the size of the header.

$$\text{size}(tcp) = \cancel{4}\ 6$$
$$\text{size}(udp) = \cancel{2}\ 3$$
$$\text{size}(icmp) = 1$$

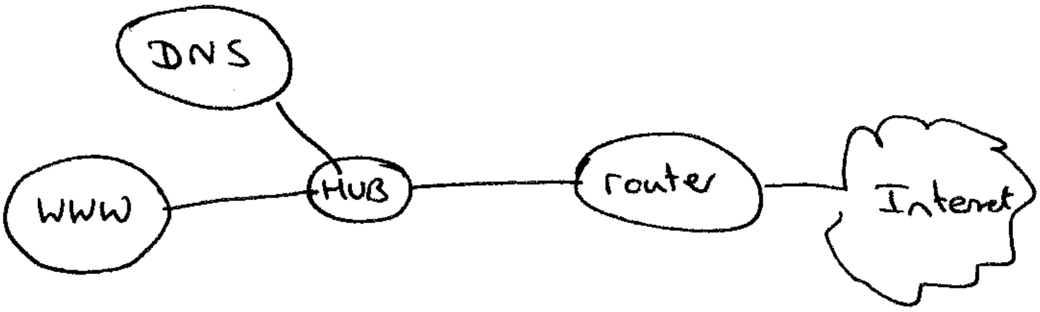
We can now work out the expected size of a packet

$$\langle \text{size} \rangle = \left(\frac{5}{8} \times 6\right) + \left(\frac{2}{8} \times 3\right) + \left(\frac{1}{8} \times 1\right)$$

$$= \frac{37}{8} = 4.6\ ?$$

This might be important for choosing how much buffer memory to include in a network device.

Exercise 7



Variables	Values	Variation	Perf. Arch or utilization
CPU % speed	MHz	Fixed value at location	P, A
net speed	bits per second	fixed values on links	P, A
collisions	% of packets?	random	P, U
DNS response time	milliseconds	varies in time	P
WWW response time	"	"	P
% idle CPU	1-100	variable at each host	U

etc. . .

Exercise 8

i) We have

$$W = kN + c$$

$$\text{kilogrammes} = \underset{(\text{computer})}{\underline{\text{kilogrammes}}} \cdot N + \text{kilogrammes}$$

i.e. both k and c are measured in kilogrammes, since N is a pure number.

Now $P = \alpha W$

$$\text{pounds} = \alpha \text{ kilos}$$

$$\Rightarrow [\alpha] = \left[\frac{\text{pounds}}{\text{kilo}}\right]$$

2. If A is measured in nanometres, and B is measured in kilogrammes, then A+B has <u>no</u> meaning!

3. Let A be arrivals in packets per second.
K is kilobytes per packet
T is seconds

Let $\mathbb{K}$ be kilobytes:

$$\underbrace{\text{kilobytes}}_{\mathbb{K}} = \underbrace{\left(\begin{array}{c}\text{kilobytes}\\\text{per packet}\end{array}\right)}_{K}\underbrace{\left(\frac{\text{packets}}{\text{seconds}}\right)}_{A}\underbrace{(\text{seconds})}_{T}$$

$$\therefore \underline{\mathbb{K} = KAT}$$

4. Let $\theta = kx - \omega t$

$$\left.\begin{array}{l}\theta \text{ is radians}\\ x \text{ is distance}\end{array}\right\} \Rightarrow k = \text{radians per metre.}$$

$$t \text{ is time} \quad \Rightarrow \omega = \text{radians per second.}$$

The wavelength λ is metres per ~~second~~ cycle. So to relate this to k, we try to match dimensions:

$$\underbrace{\left(\frac{\text{metres}}{\text{cycle}}\right)}_{\lambda} \overset{?}{\longleftrightarrow} \underbrace{\left(\frac{\text{radians}}{\text{metre}}\right)}_{k}$$

Try:

$$\underbrace{\left(\frac{\text{cycles}}{\text{metre}}\right)}_{1/\lambda} \overset{?}{\longleftrightarrow} \underbrace{\left(\frac{\text{radians}}{\text{metre}}\right)}_{k}$$

Now:

$$\underbrace{\left(\frac{\text{cycles}}{\text{metre}}\right)}_{1/\lambda}\underbrace{\left(\frac{\text{radians}}{\text{cycle}}\right)}_{2\pi} = \underbrace{\left(\frac{\text{radians}}{\text{metre}}\right)}_{k}$$

$$\therefore \boxed{k = \frac{2\pi}{\lambda}}$$

5. Frequency f is Hertz (cycles per second)
There are 2π radians per cycle
ω is radians per second.

$$\underbrace{\left(\frac{\text{radians}}{\text{cycle}}\right)}_{2\pi}\underbrace{\left(\frac{\text{cycles}}{\text{second}}\right)}_{f} = \underbrace{\left(\frac{\text{radians}}{\text{second}}\right)}_{\omega}$$

$$\therefore \boxed{\omega = 2\pi f}$$

6. The speed of a wave c is metres per second.
f is cycles per second
λ = metres per cycle

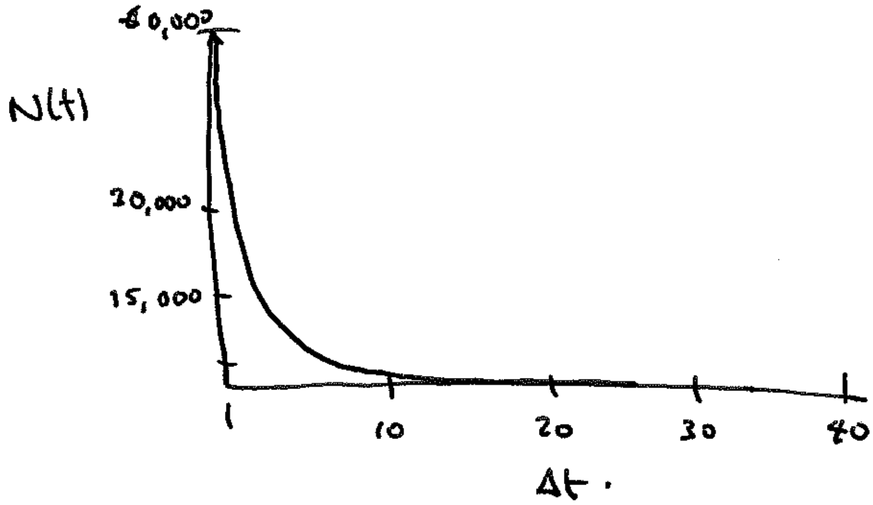
$$\underbrace{\left(\frac{\text{metres}}{\text{second}}\right)}_{c} = \underbrace{\left(\frac{\text{cycles}}{\text{second}}\right)}_{f}\underbrace{\left(\frac{\text{metres}}{\text{cycle}}\right)}_{\lambda}$$

$$\therefore \underline{c = f\lambda}.$$

(This is the wave "dispersion" relation).

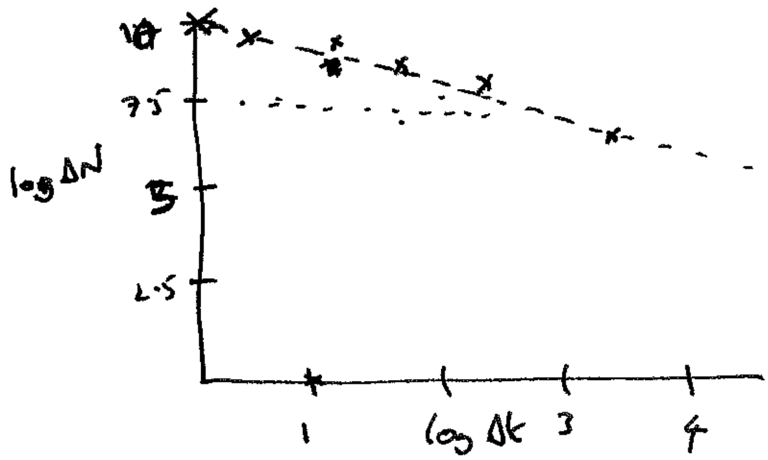
2) Observations etc. . . .

Ex 10.



1. These data tell us that packets arrive in short bursts (highly clustered events).
2.

Δt	$\log_e \Delta t$	N	$\log_e N$
1	0	57364	11
2 ~~0.69~~	0.69	12050	9.4
3 ~~1.09~~	1.09	7005	8.85
5 ~~1.6~~	1.6	4679	8.5
10	2.3	3050	8.0
30	3.4	795	6.7
37	3.6	531	6.3
38	3.6	509	6.2
40	3.7	552	6.3



$$\text{gradient} \simeq \frac{-2.5}{2.2}$$

$$\sim \underline{\underline{-1.2}}$$

3. Suppose we assume the empirical relation

$$N \propto \Delta t^{-\alpha}$$

i.e. $N = k \, \Delta t^{-\alpha}$

This is a straight-line relationship if we take logs.

$$\log N = \log k + \log(\Delta t^{-\alpha})$$

$$= \log k - \alpha \log \Delta t$$

Has the form:

$$y = + m x + c \qquad \text{i.e. a straight line}$$

where
$$\left.\begin{aligned} y &= \log N \\ x &= \log \Delta t \\ c &= \log k \\ m &= -\alpha \end{aligned}\right\} \quad \alpha = 1{\cdot}2 \text{ from the diagram.}$$

i.e. $\underline{N \sim \Delta t^{-1{\cdot}2}}$ fits quite well.

The exponential form does not give a straight line:

$$\log N = \log e^{-\beta t}$$

$$= -\beta t$$

i.e. if this were true, it would be a straight line between $\log N$ and t.

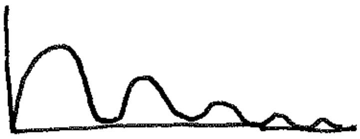
Why do we care? There are quite different processes that give rise to these different distributions. Power laws often come from 'social networks'.

Ex. 11

1. $$\alpha(t) = c_1 t + c_2 + c_3 t^3$$
$$\frac{d\alpha}{dt} = c_1 + 3c_3 t^2$$

2. $$\beta(t) = \alpha(t) e^{-\lambda t}$$
$$\frac{d\beta}{dt} = \frac{d\alpha}{dt} e^{-\lambda t} + \alpha(t) \cdot -\lambda e^{-\lambda t}$$
$$= \left(\frac{1}{\alpha}\frac{d\alpha}{dt} - \lambda\right) \underbrace{\alpha(t) e^{-\lambda t}}$$

4. $$q(t) = q_0\left(1 + \sin\left(\frac{2\pi t}{p}\right)\right) e^{-\lambda t}$$



$$\frac{dq}{dt} = q_0 \cdot \frac{2\pi}{p} \cos\left(\frac{2\pi t}{p}\right) e^{-\lambda t}$$
$$+ q_0\left(1 + \sin\left(\frac{2\pi t}{p}\right)\right) \cdot -\lambda e^{-\lambda t}$$

3 ~~4~~. $$q(t) = q_0\left(1 + \sin\left(\frac{2\pi t}{p}\right)\right)$$

$$\frac{dq}{dt} = q_0 \frac{2\pi}{p} \cos\left(\frac{2\pi t}{p}\right)$$

5. $$\rho(t) = \int_0^b q(t')\,(t'-t)\, dt'$$
$$\frac{\partial\rho}{\partial t} = \int_0^b \left[\frac{dq}{dt}(t'-t^*) + q(t) \cdot -1\right] dt'$$
$$= \int_0^b \left(\frac{dq}{dt}(t'-t^*) - q(t)\right) dt'$$
$$= -\int_0^b \left(\frac{dq}{dt}(t-t') + q(t)\right) dt'$$

6. $$p(t) = \int_0^b q(t')(t'-t)\,dt'$$

$$\frac{\partial p}{\partial t} = -\int_0^b q(t')\,dt'$$

7. $$\frac{d}{dt}\log(\alpha t) = \frac{\alpha}{\alpha t} = \frac{1}{t}$$ (independent of α! = scale free)

8. $$\int dt\, \sin(\omega t) = -\frac{1}{\omega}\cos(\omega t) + \text{const}$$

9. $$\int dt\, \cos(2\pi f t) = \frac{1}{2\pi f}\sin(\omega t) + \text{const}$$

10. $$\int dt\, e^{-\lambda t} = -\frac{1}{\lambda}e^{-\lambda t} + \text{const}$$

11. $$\int dt\, e^{-\lambda t}\sin(\omega t) \equiv I$$

Use partial integration, for instance.

$$I = \int u\frac{dv}{dx}\,dx = uv - \int v\frac{du}{dx}\,dx + \text{const}$$

Let: $$\frac{dv}{dt} = e^{-\lambda t} \qquad u = \sin(\omega t)$$

$$v = -\frac{1}{\lambda}e^{-\lambda t} \qquad \frac{du}{dt} = \omega\cos\omega t$$

$$I = -\frac{\sin(\omega t)e^{-\lambda t}}{\lambda} + \frac{\omega}{\lambda}\int e^{-\lambda t}\cos\omega t\,dt + \text{const}$$

This doesn't really help us – it's worse than the original expression! So there is no simple way to reduce this if we do not know the limits of integration.

12. $\int dt \left(c_1 t + c_2 t^2 + c_3 t^3\right)$

$$= \tfrac{1}{2} c_1 t^2 + \tfrac{1}{3} c_2 t^3 + \tfrac{1}{4} c_3 t^4 + \text{const}$$

13. $\int dt \cdot \sum_{n=1}^{m} c_n t^n \quad = \quad \int dt \sum_{n=1}^{m} \frac{1}{n+1} c_n t^{n+1}$

14. $\int dt \ \log \alpha t \ \equiv I$

Use partial integration

$$\frac{dV}{dt} = 1 \qquad u = \log \alpha t$$

$$V = t \qquad \frac{du}{dt} = \frac{\cancel{\alpha}\, 1}{t}$$

$$I = t \log \alpha t - \int \cancel{\alpha}\, dt + \text{const}$$

$$\underline{I = t \log \alpha t - \cancel{\alpha} t + \text{const}'}$$

15. $\int_0^1 \alpha t \, dt = \left[\tfrac{1}{2}\alpha t^2\right]_0^1 = \tfrac{1}{2}\alpha - 0 = \underline{\tfrac{1}{2}\alpha}.$

16. $\int_{-1}^{1} \alpha t \, dt = 0$! (odd integral).

17. $\int_0^{\pi} \sin(\omega t)\, dt = \left[\frac{-1}{\omega}\cos(\omega t)\right]_0^{\pi} = \frac{1}{\omega} - \frac{\cos(\omega\pi)}{\omega}$

$$= \underline{\frac{1}{\omega}\left(1 - \cos(\omega\pi)\right)}.$$

18. $\int_{-a}^{a} t \cos(\omega t)\, dt = 0$ (odd function).

19. $\int_{-\infty}^{\infty} e^{-at^2}\, dt \equiv \underline{I}.$

This integral is tricky - and we need to use a little creative trickery to solve it.

$$I = \int_{-\infty}^{\infty} e^{-at^2}\,dt$$

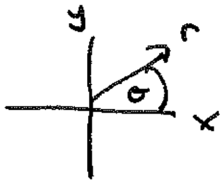
> the problem is t^2. Had it been t, it would be easy. To make this easy we need something like
>
> $$2\int x e^{-ax^2}\,dx \quad \rightarrow -\frac{1}{a} e^{-x^2}$$
>
> This is the deriv. of x^2, so we 'see' the answer but we don't have that .... yet!

What if we work out I^2?

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-at^2}\,dt\right)^2$$

$$= \int_{-\infty}^{\infty} e^{-ax^2}\,dx \int_{-\infty}^{\infty} e^{-ay^2}\,dy$$

$$I^2 = \iint_{-\infty}^{\infty} dx\,dy\; e^{-a(x^2+y^2)}$$



Now look! we have circular symmetry.

Let $r = \sqrt{x^2+y^2}$, $\theta = \tan^{-1} y/x$ $\quad dx\,dy \rightarrow r\,dr\,d\theta$

$$I^2 = \int_0^{2\pi} d\theta \int_0^{\infty} \underbrace{r\,dr\; e^{-ar^2}}$$

Now we have the form we want! (except for $\frac{1}{2}$)

$$= 2\pi \left[\frac{e^{-ar^2}}{2a}\right]_0^{\infty}$$

$$I = \frac{2\pi}{2a}$$

$$\Rightarrow I = \sqrt{\frac{\pi}{a}}$$

20. The fundamental theorem of calculus says that indefinite integration is anti-differentiation, i.e. differentiation and integration are mutual inverses.

Exercise 12

1. $V(t) = \sin(\omega t)$

"Vee of tee equals sine omega tee"

$V(t)$ is a function of t given by sine of omega t.

2. $V'(t) = \frac{d}{dt}\sin(\omega t) = \omega\cos(\omega t)$

"Vee primed of t equals dee by dee tee of sine omega tee – equals omega cos of omega tee"

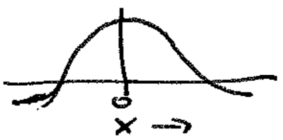
$V'(t)$ is the derivative with respect to t of sine (ωt).
$V'(t)$ is a function of t.

3. $V_0' = \frac{d}{dt}\sin(\omega t)\Big|_{t=0}$

"Vee zero primed equals dee by dee tee of ~~sign~~ sine omega tee evaluated at tee equals zero".

V_0' is a constant value equal to the value of the derivative of sine omega tee, evaluated at the point $t=0$.

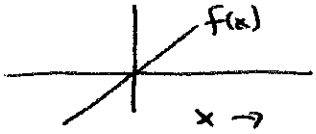
4. An even function is characterized by being symmetrical about zero in its control parameter. e.g.



is even.

i.e. $\boxed{f(x) = f(-x)}$.

5. An odd function is characterized by being anti-symmetric about zero in its control parameter. e.g.



$$\boxed{f(x) = -f(-x)}$$

6. What 'parity' does an even x odd function have.

Let $f(x)$ be <u>even</u>, i.e. $f(x) = f(-x)$.
Let $g(x)$ be <u>odd</u>, i.e. $g(x) = -g(-x)$

Now we want to know how $f(x)g(x)$ behaves when we reflect it about zero.

$$f(x)g(x) \rightarrow f(-x)g(-x) = f(x).-g(x)$$

$$\therefore \quad f(-x)g(-x) = -f(x)g(x)$$

$\Rightarrow f(x)g(x)$ is <u>odd</u>.

$\Rightarrow$ Since $f(x)$ and $g(x)$ are arbitrary, the product of any odd function with any even function is <u>odd</u>.

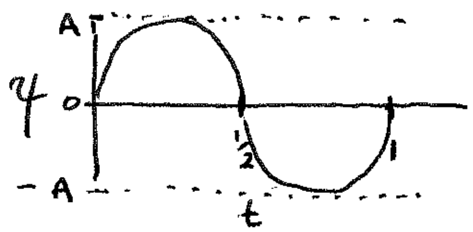
Exercise 13.

1. There are 2π radians in a single cycle.

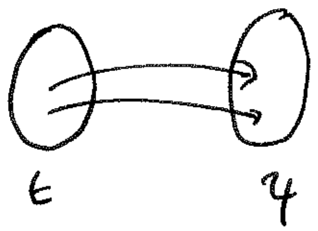
2. Sketch $\psi(t) = A \sin(2\pi t)$ for $t \in [0,1]$

   The max val of $\sin(2\pi t)$ is $+1$
   " min " $\sin(2\pi t)$ is -1

   $\Rightarrow$ Max val of ψ is A, min val is $-A$.



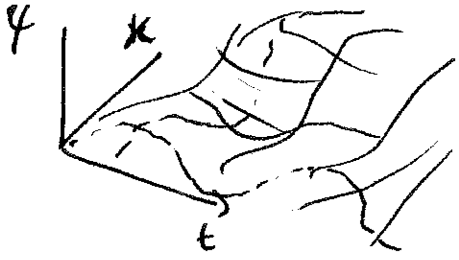
3. The function $\psi(t)$ can be considered a mapping of values t belonging to the domain $t \in [0,1]$ along the t-axis into the set $\psi \in [-A, A]$. The mapping is one to one.



4. The domain of the function is specified as $0 \to 1$ and the range is from $-A$ to A.

5. If we drop the restriction on t, then the function $\psi(t)$ has domain $-\infty$ to $+\infty$ and range $-A$ to A.

$$t \in [-\infty, \infty]$$
$$\psi \in [-A, A]$$

6. 

i.e. surface. Use mathematica or gnuplot.

7. The one-dimensional wave equation

$$\frac{1}{c^2}\frac{\partial^2 \psi}{\partial t^2} = \frac{\partial^2 \psi}{\partial x^2}$$

Show that $\psi(x,t) = A\sin(kx-\omega t)$ is a solution iff $k = \pm \omega/c$

Substitute in for $\psi(x,t)$.

$$\frac{\partial^2 \psi}{\partial x^2} = -k^2 A\sin(kx-\omega t) = -k^2\psi(x,t).$$

$$\frac{\partial^2 \psi}{\partial t^2} = -\omega^2 A\sin(kx-\omega t) = -\omega^2\psi(x,t)$$

$$\therefore \quad \frac{1}{c^2}\left(-\omega^2\psi\right) = -k^2\psi$$

$$\Rightarrow \quad \cancel{k^2 = \frac{\omega^2}{c^2}} \quad \text{~~since k,ω~~ are scalars.}$$

i.e. $$\left(k^2 - \frac{\omega^2}{c^2}\right)\psi(x,t) = 0$$

Thus either $\psi(x,t) = 0$, in which case the solution is trivial, or $(k^2 - \omega^2/c^2) = 0$. Note that $\psi(x,t) \neq 0$, for arbitrary x,t only special values, so $\psi(x,t)$ is a general solution for

$$k = \pm\omega/c$$

$\lambda f = c$, follows by substitution. See also earlier problem.

8| Show that a general combination

$$\Psi(x,t) = \int dk\, d\omega\; c(k,\omega)\, \sin(kx-\omega t)$$

is a solution.

$$\frac{\partial^2 \Psi}{\partial t^2} = -\int dk\, d\omega\; c(k,\omega)\, \omega^2 \sin(kx-\omega t).$$

$$\frac{\partial^2 \Psi}{\partial x^2} = -\int dk\, d\omega\, c(k,\omega)\, k^2 \sin(kx-\omega t)$$

Substitute into wave equation. (k,ω) have to remain under the integration sign; collect everything on LHS of equation.

$$\int dk\, d\omega\; \underbrace{\left(k^2 - \frac{\omega^2}{c^2}\right)}_{A} \underbrace{c(k,\omega)}_{B} \sin(kx-\omega t). = 0.$$

This means that either $A=0$ or $B=0$, for all x,t. If $c=0$, the solution is trivial (no solution!) hence $k^2 - \omega^2/c^2$ is zero and we have the same result as before. Thus general combinations of waves are also waves. e.g. Fourier superpositions, and all digital signals satisfy the wave equation.

13 Ex.

Calculus is good at problems of a continuous nature, like finding rates of change and adding together many special cases under integral signs. It is best dealing with slowly varying trends (smooth).

Calculus is the limit of discrete interval methods as the interval $\to 0$.

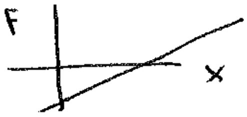
$$\frac{df}{dx} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

Thus it is the <u>embodiment</u> of the continuum approximation.

Exercise 15

Find maxima and minima.

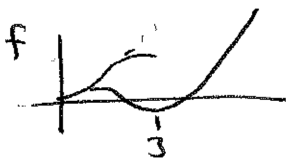
1. $f(x) = x - 3$.



Clearly x can have any value, so f(x) can have any value. There is no max or min. We can also see this by looking for stationary points where $\frac{df}{dx} = 0$.

$$\frac{df(x)}{dx} = 1 \neq 0 \Rightarrow \text{there are no stationary points.}$$

2. $f(x) = (x-3)^2 = x^2 - 6x + 9$.



Look for $\frac{df}{dx} = 0$

$$\frac{\partial f}{dx} = 2x - 6 = 0$$

$$\underline{\underline{x = 3}}$$

$$\frac{\partial^2 f}{\partial x^2} = 2 > 0 \Rightarrow \text{this is a } \underline{\text{minimum}}$$

3. $f(x,y) = \sqrt{x^2+y^2}$

This has spherical symmetry, so let $r^2 = x^2 + y^2$

$$f(r) = r$$

This is clearly a straight line relationship with no maxima or minima.

$$\frac{df}{dr} = 1 \neq 0 \quad . \quad (\text{same as } 1.)$$

4. $f(x,t) = A\sin(kx-\omega t)$ has two variables

$$\frac{\partial f}{dx} = -kA\cos(kx-\omega t)$$

This <u>does</u> have an infinite number of solutions for x

$$-kA\cos(kx-\omega t) = 0$$

$$\Rightarrow kx-\omega t = \frac{\pi}{2} \pm n\pi \qquad (n = 0, 1, 2 \ldots)$$

Similarly in t direction:

$$\frac{\partial f}{\partial t} = 0 = -\omega A\cos(kx-\omega t) = 0$$

also when

$$kx-\omega t = \frac{\pi}{2} \pm n\pi \qquad (n=0,1,2 \ldots)$$

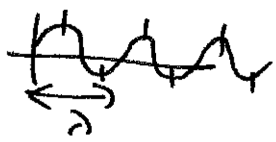
ie. x and t are not independent

Note: $$kx-\omega t = k\left(x - \frac{\omega}{k}t\right)$$
$$= k(x-ct) = \frac{\pi}{2} \pm n\pi$$

$$\therefore \frac{2\pi}{\lambda}(x-ct) = \frac{\pi}{2} \pm n\pi$$

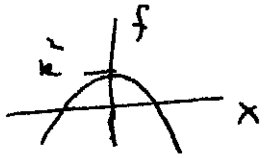
$$\underline{(x-ct) = \lambda\left(\tfrac{1}{4} \pm \tfrac{1}{2}n\right)} \qquad n = 0, 1, 2 \ldots$$

$x - ct$ is the distance travelled by the wave, in the wave's own frame of reference, ie. when the wave is stationary. At intervals of $\frac{1}{2}\lambda$ there are alternating maxima and minima. This is obvious if we <u>draw</u> it.



(*) This tells us the positions of 0, 1 bits in a digital signal.

5. $f(x) = k^2 - x^2$



$$\frac{\partial f}{\partial x} = -2x = 0 \quad \text{if } x = 0.$$

clearly this is a maximum.

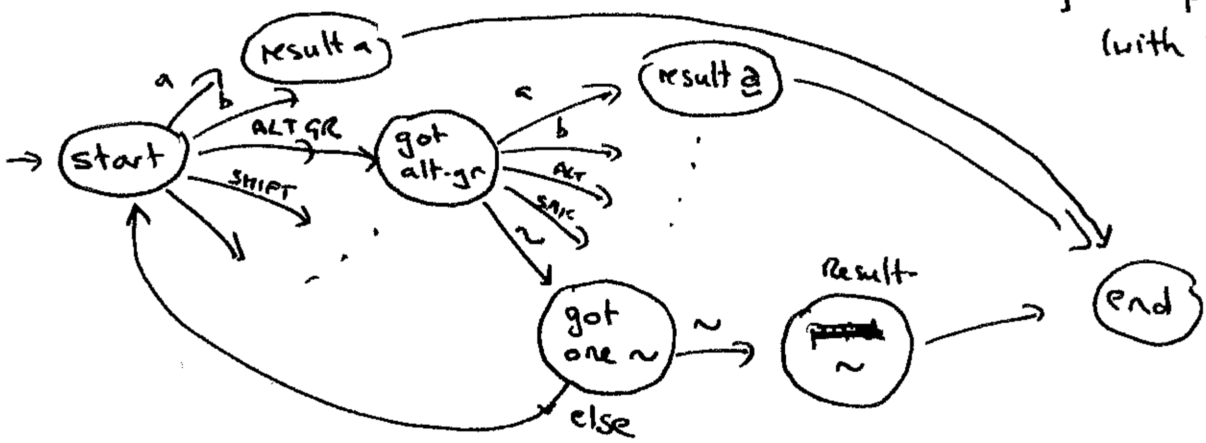
$$\frac{\partial^2 f}{dx^2} = -2 < 0 \quad \checkmark$$

WEEK 3: Discrete → CONTINUOUS.

Ex16. What is meant by a finite state machine?

There are many ways we can describe this. It is a 'system' or machine that can recognize certain regular patterns. We can think of it as a graph or as a tuple (Q, Σ, T, S, X) see 5.7.1. An FSM has a limited amount of memory in it's internal states.

Modelling a keyboard remember a ~ or ^ : e.g. simplistic: (with blanks).

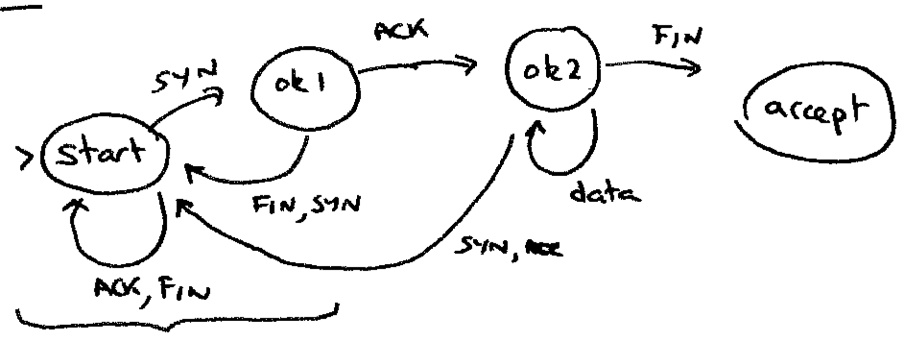


Internal states are like 'alert conditions' DEF CON 1 etc..

It would be much better to model this keyboard behaviour with a different kind of automaton.

Ex. 17

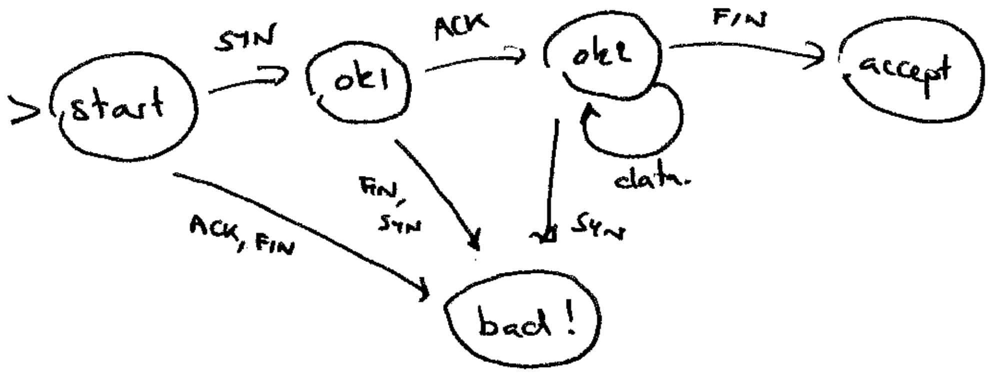
(i)



The 'back' arrows reset the connection if an incorrect sequence is found.

SYN → ACK → DATA → FIN (a simplified sketch of the TCP protocol)

(ii) To detect bad streams we can add a new state "bad" and instead of feeding the arrows back to "start", we feed to the bad state:



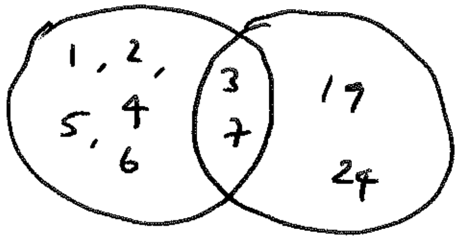
Once the bad state is reached, we can sound an alarm!

Ex 18

The question asks us to find the cases that agree with the intersection diagram:

For this to be the case, the sets must have both overlapping and non-overlapping parts.

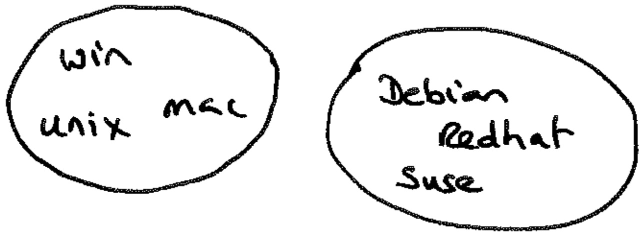
(i) Both sets A, B contain the elements 3 and 7, so this fits ✓
Both sets have more elements than these two



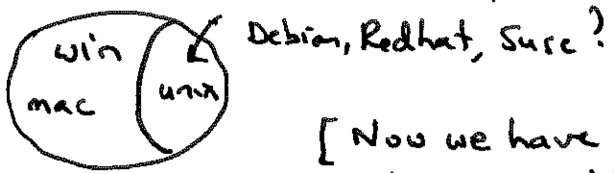
✓

(ii) Both sets contain file permission 555 ✓
Both sets are bigger ✓

(iii) There are no common elements in these sets X



or should we say that unix overlaps with Debian, RedHat, Suse?



[Now we have a modelling question to answer!]

Either way, this does not fit the picture X.

(iv) A consists of 2 continuous ranges $0 \leq t \leq 1$, $2 \leq t \leq 3$
B consists of 2 isolated values 0·5 and 0·7, both contained in A.

$$\Rightarrow \boxed{B \subset A}$$

Does not fit the picture X.

(v) Set A consists of all files called "passwd" in any subdirectory of "/" .

Set B consists of 3 named files, 3 of which match.

Set A contains all of B, so once again

$$\boxed{B \subset A}$$

Picture does not match X . (disjoint)

(6) These sets have no common elements.

This exercise mimics what cfengine does when checking file security attributes.

A = set of observed (mode, user, group ..)

B = set of policy (" .. ..)

Cfengine should not guess or make mistakes.

Ex. 19

1. AND = intersection  or 

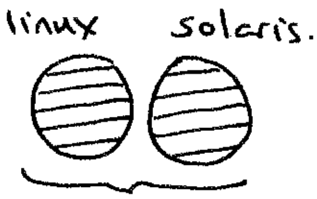
~~2.~~ OR = union 

These Venn diagrams are symbolic and seldom represent what the sets really look like or consist of.

2. (nothing to do!)

3. Venn diagrams:

(i)



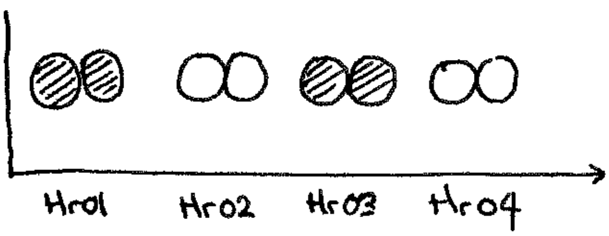
linux|solaris::

sets do not overlap, so the union is the sum of two disjointed sets.

(ii) Hr01.crayos::

or

(iii) (Hr01|Hr03).(linux|solaris)::



(see (i) and (ii))

(NB. In this classification of time, t is a finite set {Hr00, Hr01, ...} so we could do (E) linux.)

4. [redacted] linux.Hr01 is the set of all hosts that are ~~both~~ linux at ~~13:00 ≤ t ≤ 13:59~~ $13:00 \le t \le 13:59$. (cfengine language)

$\{linux\} \cap \{Hr01\}$ is the same thing in set language.

A hierarchical decomposition assumes that one class is a sub-class of another. Hierarchies don't cope well with multiplicity. There is no clear ordering of parent-child in these classes. However, the diagram shows that we can form an ad-hoc tree ....
the point is that the tree is not unique. (ontology)

Ex 20

1. Probability (windows) $= \frac{25}{25+16+3} = \frac{25}{44}$

   Prob (linux) $= \frac{16}{44} = \frac{4}{11}$

   Prob (solaris) $= \frac{3}{44}$

   Check $\sum_i$ prob (i) $= 1$ ✓

2. The numbers are based on very few data. Note that when we go to a probability picture, we lose the information about how many data are used to derive them.

   (a) It is reasonable to suppose that the current state of the company will be representative of its state in the future — assuming no environmental changes. So the distribution could be used predictively.

   (b) In Norway, we have no support for the hypothesis that our company statistics are representative of anywhere else.

   (c) This would be nonsense.

3. Assume that the probabilites are representative of growth and that the new total number of employees is N'

$$n(\text{windows}) = 3N' \cdot \frac{25}{44}$$

$$n(\text{linux}) = 3N' \cdot \frac{16}{44}$$

$$n(\text{solaris}) = 3N' \cdot \frac{3}{44}.$$

4. (a) if all employees work on similar tasks we can expect that they all generate 'typical' traffic. Then it is reasonable to postulate that

$$\text{packets (windows)} \propto p(\text{windows})$$
$$\text{packets (linux)} \propto p(\text{linux}) \quad (\text{Expectation value})$$
$$\text{packets (solaris)} \propto p(\text{solaris})$$

(b) if they have different tasks, we cannot say anything without more information. Then

packets (windows) $\propto$ p(windows) AND / OR other criteria.

Ex 21

1. $$A+B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} + \begin{pmatrix} 2 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1+2 & 2+1 & 3+1 \\ 4+2 & 5+1 & 6+2 \\ 7+1 & 8+1 & 9+1 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 3 & 4 \\ 6 & 6 & 8 \\ 8 & 9 & 10 \end{pmatrix}$$

2. $$AB = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\begin{pmatrix} 2 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} (2+4+3) & (1+2+3) & (1+4+3) \\ (8+10+6) & (4+5+6) & (4+10+6) \\ (14+16+9) & (7+8+9) & (7+16+9) \end{pmatrix} = \begin{pmatrix} 9 & 6 & 8 \\ 24 & 15 & 20 \\ 39 & 24 & 32 \end{pmatrix}$$

3. $BA = \begin{pmatrix} 2 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 13 & 17 & 21 \\ 20 & 25 & 30 \\ 12 & 15 & 18 \end{pmatrix}$

NOTE 1: octave

```
a = [1 2 3; 4 5 6; 7 8 9]
b = [2 1 1; 2 1 2; 1 1 1]

a+b
a*b
b*a
```

NOTE 2: Mathematica.

```
a = {{1,2,3}, {4,5,6}, {7,8,9}}   (SHIFT ENTER)
b = {{2,1,1}, {2,1,2}, {1,1,1}}   (SHIFT ENTER)

a+b   (SE)
a.b   (SE)
b.a   (SE)
```

Ex 22

This is about constraints.

1. Consider the matrix equation

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

(a) This is 2 equations:

$$\left.\begin{aligned} x + 2y &= 0 \\ 3x + 4y &= 0 \end{aligned}\right\}$$

(b) Substitute in $x=y=0$, both equations are satisfied trivially: $0=0$.

(c) To determine if any other solution exists we check $\det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 0$

$\det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = (1\cdot 4 - 3\cdot 2) = -2 \neq 0 \Rightarrow$ no other solutions.

2. Consider

$$\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

(a) Same as before $(x,y) = 0$ is always a solution.

(b) check $\det \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} = 4 - 4 = 0$ ✓

$\Rightarrow$ other solutions exist. There is really only one equation:

$$2x + y = 0$$

i.e. any value of x, y satisfying this constraint are solutions.

3. ~~(a) as before~~

~~(b) $\det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = -1$~~

First write the equation in the same form:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

(a) $\begin{pmatrix} x \\ y \end{pmatrix} = 0$ is a trivial solution.

(b) $\det \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} = 0 \Rightarrow$ other solutions exist.

Write out:

$$\left.\begin{aligned} x &= y \\ y &= x \end{aligned}\right\}.$$

i.e. any value of (x,y) where $x=y$ is a solution.

WEEK 4 CONFIGURATION MANAGEMENT

Ex 23

1. Which operation is represented by the following matrix operation

$$\vec{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ r \\ w \\ x \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ r \\ 1 \\ x \end{pmatrix}$$

The operation sets $w = 1$, regardless of its previous value, i.e. it is equivalent to

$\boxed{\text{chmod u+w f.}}$

Ex 24

1. $\vec{\psi}_0$ is the policy conformant state, and ψ is any other state. Let $\hat{O}$ be any operator such that $\hat{O}^2 = \hat{O}$, if this satisfies the immunity property then

$$\hat{O}^n \vec{\psi} = \vec{\psi}_0$$
$$\hat{O}\,\vec{\psi}_0 = \vec{\psi}_0$$

Now since $\hat{O}^2 = \hat{O}$, $\Rightarrow$ $\hat{O}^n = \hat{O}$.

$$\therefore \left.\begin{aligned} \hat{O}\vec{\psi} &= \vec{\psi}_0 \\ \hat{O}\vec{\psi}_0 &= \vec{\psi}_0 \end{aligned}\right\}$$ if $\hat{O}$ is convergent provided

Now the second line is:

$$\hat{O}\vec{\psi}_0 = \hat{O}^2\vec{\psi} = \vec{\psi}_0 \text{, i.e.}$$

$$\left.\begin{aligned} \hat{O}\vec{\psi} &= \vec{\psi}_0 \\ \hat{O}^2\vec{\psi} &= \psi_0 \end{aligned}\right\}$$ provided that the first line is true. (there is nothing in $\hat{O}^2 = \hat{O}$ that demands this).

Ex 24 (improved)

Let $\vec{\psi}_{any}$ be an arbitrary state, and $\hat{O}$ be an arbitrary operator. Then we have:

$$\hat{O}\,\vec{\psi}_{any} = \vec{\psi}_{new\cdot any} \qquad (1)$$

In general $\vec{\psi}_{any} \neq \vec{\psi}_{new\cdot any}$ (though this could be true by chance/luck). Now consider a special $\hat{O}$ that is idempotent. Call it $\hat{I}$, and we know that $\hat{I}^2 = \hat{I}$. Now,

$$\left.\begin{aligned} \hat{I}\,\vec{\psi}_{any} &= \vec{\psi}_{new\text{-}any} \\ \hat{I}^2\,\psi_{any} &= \vec{\psi}_{new\text{-}any} \end{aligned}\right\} \qquad (2)$$

Suppose we apply $\hat{I}$ again:

$$\hat{I}^3\,\vec{\psi}_{any} = \hat{I}^2\psi_{any} = \underbrace{\hat{I}\,\vec{\psi}_{any} = \hat{I}\,\vec{\psi}_{new\text{-}any}}_{*\text{ see below}} = \vec{\psi}_{newer}.$$

This means we have

$$\left.\begin{aligned} \hat{I}\,\vec{\psi}_{any} &= \vec{\psi}_{new} \\ \hat{I}\,\vec{\psi}_{new} &= \vec{\psi}_{newer} \end{aligned}\right\} \qquad (3)$$

This is not convergent, since we cannot say that $\vec{\psi}_{newer} = \vec{\psi}_{new}$. (Again, it <u>could be</u>, but this is not guaranteed).

> * (footnote) It looks as though
> $$\hat{I}\,\vec{\psi}_{any} = \hat{I}\,\psi_{new\text{-}any}$$
> $$\Rightarrow \psi_{any} = \psi_{new\text{-}any}$$
> but $\hat{I}^2 = \hat{I}$ means that
> either $$\vec{\psi}_{new\text{-}any} = \begin{cases} \vec{\psi}_{any} \\ \hat{I}\,\vec{\psi}_{any} \end{cases}$$ <u>or</u>

Now consider a special op. $\hat{C}$ which is convergent

$$\hat{C}\,\vec{\psi}_{any} = \vec{\psi}_0 \qquad \text{(constant, fixed).}$$
$$\hat{C}\,\vec{\psi}_{new\text{-}any} = \vec{\psi}_0$$

and since $\vec{\psi}_0$ is in $\vec{\psi}_{any}$

$$\left.\begin{aligned}\hat{C}\,\vec{\psi}_{any} &= \vec{\psi}_0\\ \hat{C}\,\vec{\psi}_0 &= \vec{\psi}_0\end{aligned}\right\} \qquad (4)$$

Apply $\hat{C}$, or combine these

$$\hat{C}^2\,\vec{\psi}_{any} = \hat{C}\,\vec{\psi}_0 = \hat{C}^2\vec{\psi}_0 = \vec{\psi}_0 \qquad (5)$$

i.e. $\hat{C}^2 = \hat{C}$, so $\hat{C}$ is idempotent, but <u>more</u> (stronger constraint).

<u>Examples (text editing in cfengine).</u>

$\hat{O}$ = "insert at current position"

$\hat{O}^2 \neq \hat{O}$ (inserts twice at random/unknown location)

$\hat{D}$ = "append"

$\hat{D}^2 \neq \hat{D}$ (appends twice) at unpredictable location.

$\hat{C}$ = "append if no such line"

$\hat{C}^2 = \hat{C}$ (convergent / idempotent)

$\hat{I}$ = "insert at current position if no such line"

$\hat{I}^2 = \hat{I}$ (but we don't know where, so if start state changes we could insert many different places).

<u>Exercise 25</u>

Extend the matrix model

σ = string $= 0$ if $\nexists$.

ε = empty file

1. A state now has the form:

$$\vec{\psi}_f = \begin{pmatrix}1\\ \sigma\\ r\\ \omega\\ x\end{pmatrix}$$ where σ, r, ω, x can take on any value.

The creat (create empty file) operator in unix with start permissions is

$$\hat{C}_f\,(\varepsilon, R, W, X)$$

Apply this to an arbitrary state

$$\psi = \begin{pmatrix}1\\ \alpha\\ \beta\\ \gamma\\ \delta\end{pmatrix}$$

$$\hat{C}_f\psi = \begin{pmatrix}1 & 0 & 0 & 0 & 0\\ \varepsilon & 0 & 0 & 0 & 0\\ R & 0 & 0 & 0 & 0\\ W & 0 & 0 & 0 & 0\\ X & 0 & 0 & 0 & 0\end{pmatrix}\begin{pmatrix}1\\ \alpha\\ \beta\\ \gamma\\ \delta\end{pmatrix} = \begin{pmatrix}1\\ \varepsilon\\ R\\ W\\ X\end{pmatrix} \quad ✓$$

i.e. if the file exists previously it is emptied and its permissions are set. If it doesn't exist ($\alpha=0$) it is created.

2. Show that $\hat{C}_f^2 = C_f$:

$$\begin{pmatrix}1 & 0 & 0 & 0 & 0\\ \sigma & 0 & 0 & 0 & 0\\ r & 0 & 0 & 0 & 0\\ w & 0 & 0 & 0 & 0\\ x & 0 & 0 & 0 & 0\end{pmatrix}\begin{pmatrix}1 & 0 & 0 & 0 & 0\\ \sigma & 0 & 0 & 0 & 0\\ r & 0 & 0 & 0 & 0\\ w & 0 & 0 & 0 & 0\\ x & 0 & 0 & 0 & 0\end{pmatrix} = \begin{pmatrix}1 & 0 & 0 & 0 & 0\\ \sigma & 0 & 0 & 0 & 0\\ r & 0 & 0 & 0 & 0\\ w & 0 & 0 & 0 & 0\\ x & 0 & 0 & 0 & 0\end{pmatrix} \quad ✓$$

3. Show that $\hat{P}(r,w,x)$ is idempotent, i.e.

$$\hat{P}^2(r,w,x) = \hat{P}(r,w,x)$$

$$\hat{P}^2 = \begin{pmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ r&0&0&0&0 \\ w&0&0&0&0 \\ x&0&0&0&0 \end{pmatrix}\begin{pmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ r&0&0&0&0 \\ w&0&0&0&0 \\ x&0&0&0&0 \end{pmatrix} = \begin{pmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ r&0&0&0&0 \\ w&0&0&0&0 \\ x&0&0&0&0 \end{pmatrix} = \hat{P} \quad \checkmark$$

4. Show that $\hat{C}_f \hat{P}_f = \hat{C}_f$:

$$\hat{C}_f \tilde{P}_f' = \begin{pmatrix} 1&0&0&0&0 \\ \sigma&0&0&0&0 \\ r&0&0&0&0 \\ w&0&0&0&0 \\ x&0&0&0&0 \end{pmatrix}\begin{pmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ r'&0&0&0&0 \\ w'&0&0&0&0 \\ x'&0&0&0&0 \end{pmatrix} = \begin{pmatrix} 1&0&0&0&0 \\ \sigma&0&0&0&0 \\ r&0&0&0&0 \\ w&0&0&0&0 \\ x&0&0&0&0 \end{pmatrix} = \hat{C}_f(\sigma, r, w, x)$$

$$\tilde{P}_f' \hat{C}_f = \begin{pmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ r'&0&0&0&0 \\ w'&0&0&0&0 \\ x'&0&0&0&0 \end{pmatrix}\begin{pmatrix} 1&0&0&0&0 \\ \sigma&0&0&0&0 \\ r&0&0&0&0 \\ w&0&0&0&0 \\ x&0&0&0&0 \end{pmatrix} = \begin{pmatrix} 1&0&0&0&0 \\ \sigma&0&0&0&0 \\ r'&0&0&0&0 \\ w'&0&0&0&0 \\ x'&0&0&0&0 \end{pmatrix} \neq \hat{C}_f(\sigma, r, w, x)$$

$$= \hat{C}_f(\sigma, r', w', x')$$

5. If the order of operations is incorrect, the permissions of files could be left in a non-policy state.

6. What modification could we make to $\hat{C}$ to allow $\hat{C}$ and $\hat{P}$ to commute. – We can make sure that $\hat{C}$ does nothing to overlap with $\hat{P}$ – i.e. we can make it orthogonal.

$$\hat{C}_f(\sigma, r, w, x) \rightarrow \hat{C}_f(\sigma, \underset{\uparrow \text{identity}}{I})$$

$$\hat{C}_f \rightarrow \left(\begin{array}{cc|ccc} 1&0&0&0&0 \\ \sigma&0&0&0&0 \\ \hline 0&0&1&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&1 \end{array}\right) = \hat{C}_f'$$

i.e. we replace the overlapping part with the identity matrix.

$\hat{C}_f'$ and $\hat{P}_f$ are <u>othogonal</u>.

Ex 26

1. Show that $\hat{M}_f(\delta\sigma)$ adds (appends) a string to the end of a file. Start with file

$$\psi_f = \begin{pmatrix} 1 \\ \sigma \\ r \\ w \\ x \end{pmatrix}$$

$$\hat{M}_f(\delta\sigma)\psi_f = \begin{pmatrix} 1&0&0&0&0 \\ \delta\sigma&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&1 \end{pmatrix}\begin{pmatrix} 1 \\ \sigma \\ r \\ w \\ x \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ \sigma+\delta\sigma \\ r \\ w \\ x \end{pmatrix} \quad \checkmark$$

This transforms $\sigma \rightarrow \sigma + \delta\sigma$ and leaves $\begin{pmatrix} r \\ w \\ x \end{pmatrix} \rightarrow \begin{pmatrix} r \\ w \\ x \end{pmatrix}$ unchanged.

We can write $\hat{M}_f$ as the identity

$$I = \begin{pmatrix} 1&&& \\ &1&& \\ &&1& \\ &&&1 \end{pmatrix}$$

plus a generator

$$\hat{M}_f = I + \begin{pmatrix} 0&0&0&0&0 \\ \delta\sigma&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \end{pmatrix}$$

2. If we apply this operation when the file does not exist: $\vec{\psi}_n = \begin{pmatrix} 1 \\ 0 \\ r \\ w \\ x \end{pmatrix}$

$$\hat{M}_f \vec{\psi}_n = \begin{pmatrix} 1&0&0&0&0 \\ \delta\sigma&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ r \\ w \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ \delta\sigma \\ r \\ w \\ x \end{pmatrix}$$

3. Clearly $\hat{M}_f$ is not idempotent $\hat{M}_f^2 \neq \hat{M}_f$

$$\hat{M}_f^2 \vec{\psi}_n = \begin{pmatrix} 1 \\ \delta\sigma+\delta\sigma \\ r \\ w \\ x \end{pmatrix}$$

4. Does the operation commute with $\hat{C}$ and $\hat{P}$?

$$\hat{P}\hat{M}_f = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\hat{M}_f\hat{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix}$$

i.e. $\boxed{[\hat{P}, \hat{M}_f] = \hat{P}\hat{M}_f - \hat{M}_f\hat{P}_f = 0}$ i.e. we can change the permissions independently of editing its contents.

$$\hat{C}_f\hat{M}_f = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\hat{M}_f\hat{C}_f = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma+\delta\sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\boxed{[\hat{C}_f, \hat{M}_f] = \hat{C}_f\hat{M}_f - \hat{M}_f\hat{C}_f = \hat{\alpha}}$$

i.e. we cannot create a file independently of editing it — the result depends on the order.

5. Do modifications of the same file commute?

$$\hat{M}_f(\alpha)\hat{M}_f(\beta) = (I+\hat{\alpha})(I+\hat{\beta}) = I + \hat{\alpha}\hat{\beta} + \hat{\alpha} + \hat{\beta}$$

$$\hat{M}_f(\beta)\hat{M}_f(\alpha) = (I+\hat{\beta})(I+\hat{\alpha}) = I + \hat{\beta}\hat{\alpha} + \hat{\beta} + \hat{\alpha}$$

$$\therefore \quad [\hat{M}_f(\alpha), \hat{M}_f(\beta)] = [\hat{\alpha}, \hat{\beta}] + (\widehat{\alpha+\beta}) - (\widehat{\beta+\alpha})$$

Now, what is special here is that "+" means concatenation of strings and thus + is not associative

$$\hat{\alpha} + \hat{\beta} \neq \hat{\beta} + \alpha$$

$$\therefore \quad [\hat{M}_f(\alpha), \hat{M}_f(\beta)] \neq 0 .$$

6. Different modifications to different files must commute, since the operations have no common attributes.

7. In cfengine, we have Append If No Such Line is

$$\hat{m}_f = \begin{cases} \hat{I} + \delta\hat{\sigma} & \text{iff } \delta\sigma \text{ is not in } \sigma \\ \hat{I} & \text{if } \delta\sigma \text{ is in } \sigma \end{cases}$$

Is this idempotent?

$$\hat{m}_f^2 = \begin{cases} \hat{I} & \text{if acting on a state containing } \delta\sigma. \\ \hat{I} + \delta\hat{\sigma} & \text{if acting on a state not containing } \delta\sigma \end{cases}$$

So ~~$\hat{m}_f^2 \neq \hat{m}_f$ in general.~~

However, the result of $\hat{m}_f^2$ on any state is the same as the result of $\hat{m}_f$ on any state. So, since the operator needs to observe the state it is acting on for its definition, we can define it to be idempotent.

Is the operator convergent? If we assume that policy means a state in which $\delta\sigma$ is contained, (once only) then:

$$\hat{m}_f \vec{q} = \vec{q}_{policy}$$
$$\hat{m}_f \vec{q}_{policy} = \vec{q}_{policy} \quad \checkmark$$

$\Rightarrow$ $\hat{m}_f$ is convergent.

Ex 27

1. If writing a tool for automatic configuration, I would want two properties
   - (i) orthogonality (no overlap) in operators
   - (ii) convergence towards a policy state.

2. The difference between idempotence and convergence depends on context. Idempotence is a property of an operator. Convergence requires that an operator understands the state it acts upon.

3. (a) Assuming that $\hat{T}$ includes resetting the system to the base state $\hat{T}^2 = \hat{T}$ by definition, so it is idempotent.

(little $\hat{m}$) $\rightarrow$ $\hat{C}, \hat{P}$ and $\hat{m}$ are idempotent provided we consider their action on states. So in both cases, the same result is achieved, provided the states are taken into account.

(b) Both methods are as reliable as each other. They both do what they claim.

(c) Both methods are equally correct – they both do what they claim.

(d) The convergence method can be used 'on the fly' while devices are in operation.

The congruence method requires the system to undergo a catastrophic reset; hence it cannot be used 'on the fly'.

(e) Given that maintenance is necessary often, convergence is the least 'violent' method of change – since it does not require a system to be halted.

(f) SNMP's operations can only read and write values, not modify values.

$$PUT^2 = PUT$$
$$GET^2 = GET$$

$\Rightarrow$ these are idempotent.

The problem of using this for workstations is that these operations cannot model the complexity of configuration without memorizing the <u>entire</u> configuration at a remote location.

NOTES ABOUT THIS WEEK:

These exercises show that it is possible to implement a 'dumb' mechanical algorithm for implementing convergence. Instead of complex logic in the change process, we enumerate the set of allowed operators, satisfying the constraint of orthog. and convergence.

Config. management is still widely discussed.

Ex 24 (improved)

Let $\vec{\psi}_{any}$ be an arbitrary state, and $\hat{O}$ be an arbitrary operator. Then we have:

$$\hat{O}\,\vec{\psi}_{any} = \vec{\psi}_{new\cdot any} \qquad (1)$$

In general $\vec{\psi}_{any} \neq \vec{\psi}_{new\cdot any}$ (though this could be true by chance/luck).

Now consider a special $\hat{O}$ that is idempotent. Call it $\hat{I}$, and we know that $\hat{I}^2 = \hat{I}$. Now,

$$\left.\begin{aligned} \hat{I}\,\vec{\psi}_{any} &= \vec{\psi}_{new\text{-}any} \\ \hat{I}^2 \psi_{any} &= \vec{\psi}_{new\text{-}any} \end{aligned}\right\} \qquad (2)$$

Suppose we apply $\hat{I}$ again:

$$\hat{I}^3\,\vec{\psi}_{any} = \hat{I}^2 \psi_{any} = \underbrace{\hat{I}\,\vec{\psi}_{any} = \hat{I}\,\vec{\psi}_{new\text{-}any}}_{*\text{ see below}} = \vec{\psi}_{newer}.$$

This means we have

$$\left.\begin{aligned} \hat{I}\,\vec{\psi}_{any} &= \vec{\psi}_{new} \\ \hat{I}\,\vec{\psi}_{new} &= \vec{\psi}_{newer} \end{aligned}\right\} \qquad (3)$$

This is not convergent, since we cannot say that $\vec{\psi}_{newe} = \vec{\psi}_{new}$.
(Again, it <u>could be</u>, but this is not guaranteed).

> * (footnote) It looks as though
> $$\hat{I}\,\vec{\psi}_{any} = \hat{I}\,\psi_{new\text{-}any}$$
> $$\Rightarrow \quad \psi_{any} = \psi_{new\text{-}any}$$
> but $\hat{I}^2 = \hat{I}$ means that
> $$\text{either} \quad \vec{\psi}_{new\text{-}any} = \begin{cases} \vec{\psi}_{any} & \underline{\text{or}} \\ \hat{I}\,\vec{\psi}_{any} \end{cases}$$

WEEK 4 CONFIGURATION MANAGEMENT

Ex 23

1. Which operation is represented by the following matrix operation

$$\vec{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ r \\ w \\ x \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ r \\ 1 \\ x \end{pmatrix}$$

The operation sets $w=1$, regardless of its previous value, i.e. it is equivalent to

$$\boxed{\text{chmod u+w f.}}$$

Ex 24

1. ~~$\vec{\psi}_0$ is the policy conformant state, and ψ is any other state.~~
~~Let $\hat{o}$ be any operator such that $\hat{o}^2 = \hat{o}$, if this satisfies the immunity property then~~

$$\cancel{\hat{o}^n \vec{\psi} = \vec{\psi}_0}$$
$$\cancel{\hat{o}\, \vec{\psi}_0 = \vec{\psi}_0}$$

~~Now since $\hat{o}^2 = \hat{o}$, $\Rightarrow$ $\hat{o}^n = \hat{o}$.~~

$$\cancel{\therefore \left.\begin{aligned} \hat{o}\,\vec{\psi} &= \vec{\psi}_0 \\ \hat{o}\,\vec{\psi}_0 &= \vec{\psi}_0 \end{aligned}\right\}} \text{ if } \hat{o} \text{ is convergent provided}$$

~~Now the second line is:~~

$$\cancel{\hat{o}\,\psi_0 = \hat{o}^2 \vec{\psi} = \vec{\psi}_0}, \text{ i.e.}$$

$$\cancel{\left.\begin{aligned} \hat{o}\,\vec{\psi} &= \vec{\psi}_0 \\ \hat{o}^2 \vec{\psi} &= \psi_0 \end{aligned}\right\}} \text{ provided that the first line is true. (there is nothing in } \hat{o}^2 = \hat{o} \text{ that demands this).}$$

Now conside a special op. $\hat{c}$ which is convergent

$$\hat{c}\,\vec{\psi}_{any} = \vec{\psi}_0 \qquad \text{(constant, fixed).}$$

$$\hat{c}\,\vec{\psi}_{new\text{-}any} = \vec{\psi}_0$$

and since $\vec{\psi}_0$ is in $\vec{\psi}_{any}$

$$\left.\begin{aligned} \hat{c}\,\vec{\psi}_{any} &= \vec{\psi}_0 \\ \hat{c}\,\vec{\psi}_0 &= \vec{\psi}_0 \end{aligned}\right\} \qquad (4)$$

Apply $\hat{c}$, or combine these

$$\hat{c}^2\,\vec{\psi}_{any} = \hat{c}\,\vec{\psi}_0 = \hat{c}^2\vec{\psi}_0 = \vec{\psi}_0 \qquad (5)$$

i.e. $\hat{c}^2 = \hat{c}$, so $\hat{c}$ is idempotent, but more (stronger constraint).

<u>Examples (text editing in cfengine).</u>

$\hat{O}$ = "insert at current position"

$\hat{O}^2 \neq \hat{O}$ (inserts twice at random/unknown location)

$\hat{O}$ = "append"

$\hat{O}^2 \neq \hat{O}$ (appends twice) at unpredictable location.

$\hat{C}$ = "append if no such line"

$\hat{C}^2 = \hat{C}$ (convergent / idempotent)

$\hat{I}$ = "insert at current position if no such line"

$\hat{I}^2 = \hat{I}$ (but we don't know where, so if start state changes we could insert many different places).

Exercise 25

Extend the matrix model

$$\sigma = \text{string} \quad = 0 \text{ if } \nexists .$$
$$\varepsilon = \text{empty file}$$

1. A state now has the form:

$$\vec{\psi}_f = \begin{pmatrix} 1 \\ \sigma \\ r \\ \omega \\ x \end{pmatrix}$$

where σ, r, ω, x can take on any value.

The creat (create empty file) operator in unix with start permissions is

$$\hat{C}_f(\varepsilon, R, W, X)$$

Apply this to an arbitrary state

$$\psi = \begin{pmatrix} 1 \\ \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}$$

$$\hat{C}_f \psi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \varepsilon & 0 & 0 & 0 & 0 \\ R & 0 & 0 & 0 & 0 \\ W & 0 & 0 & 0 & 0 \\ X & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix} = \begin{pmatrix} 1 \\ \varepsilon \\ R \\ W \\ X \end{pmatrix} \quad \checkmark$$

i.e. if the file exists previously it is emptied and its permissions are set. If it doesn't exist $(\alpha = 0)$ it is created.

2. Show that $\hat{C}_f^2 = C_f$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} \quad \checkmark$$

3. Show that $\hat{P}(r,w,x)$ is idempotent, i.e.

$$\hat{P}^2(r,w,x) = \hat{P}(r,w,x)$$

$$\hat{P}^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} = \hat{P} \quad \checkmark$$

4. Show that $\hat{C}_f \hat{P}_f = \hat{C}_f$:

$$\hat{C}_f \tilde{P}'_f = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ r' & 0 & 0 & 0 & 0 \\ w' & 0 & 0 & 0 & 0 \\ x' & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} = \hat{C}_f(\sigma, \cancel{r}, w, x)$$

$$\tilde{P}'_f \hat{C}_f = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ r' & 0 & 0 & 0 & 0 \\ w' & 0 & 0 & 0 & 0 \\ x' & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r' & 0 & 0 & 0 & 0 \\ w' & 0 & 0 & 0 & 0 \\ x' & 0 & 0 & 0 & 0 \end{pmatrix} \neq \hat{C}_f(\sigma, r, w, x)$$

$$= \hat{C}_f(\sigma, r', w', x')$$

5. If the order of operations is incorrect, the permissions of files could be left in a non-policy state.

6. What modification could we make to $\hat{C}$ to allow $\hat{C}$ and $\hat{P}$ to commute. – We can make sure that $\hat{C}$ does nothing to overlap with $\hat{P}$ – i.e. we can make it orthogonal.

$$\hat{C}_f(\sigma, r, w, x) \rightarrow \hat{C}_f(\sigma, \underset{\uparrow \text{ identity.}}{I})$$

$$\hat{C}_f \rightarrow \left(\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right) \equiv \hat{C}'_f$$

i.e. we replace the overlapping part with the identity matrix.

$\hat{C}_f'$ and $\hat{P}_f$ are <u>othogonal</u>.

Ex 26

1. Show that $\hat{M}_f(\delta\sigma)$ adds (appends) a string to the end of a file. Start with file

$$\psi_f = \begin{pmatrix} 1 \\ \sigma \\ r \\ w \\ x \end{pmatrix}$$

$$\hat{M}_f(\delta\sigma)\psi_f = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ \sigma \\ r \\ w \\ x \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ \sigma+\delta\sigma \\ r \\ w \\ x \end{pmatrix} \quad \checkmark$$

This transforms $\sigma \to \sigma+\delta\sigma$ and leaves $\begin{pmatrix} r \\ w \\ x \end{pmatrix} \to \begin{pmatrix} r \\ w \\ x \end{pmatrix}$ unchanged.
We can write $\hat{M}_f$ as the identity

$$\underline{I} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

plus a generator

$$\hat{M}_f = \underline{I} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ \delta\sigma & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

2. If we apply this operation when the file does not exist: $\vec{\psi}_n = \begin{pmatrix} 1 \\ 0 \\ r \\ w \\ x \end{pmatrix}$

$$\hat{M}_f \vec{\psi}_n = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ r \\ w \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ \delta\sigma \\ r \\ w \\ x \end{pmatrix}$$

3. Clearly $\hat{M}_f$ is not idempotent $\hat{M}_f^2 \neq \hat{M}_f$

$$\hat{M}_f^2 \vec{\psi}_n = \begin{pmatrix} 1 \\ \delta\sigma+\delta\sigma \\ r \\ w \\ x \end{pmatrix}$$

4. Does the operation commute with $\hat{C}$ and $\hat{P}$?

$$\hat{P}\hat{M}_f = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\hat{M}_f\hat{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix}$$

i.e. $\boxed{[\hat{P}, \hat{M}_f] = \hat{P}\hat{M}_f - \hat{M}_f\hat{P}_f = 0}$

i.e. we can change the permissions independently of editing its contents.

$$\hat{C}_f\hat{M}_f = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\hat{M}_f\hat{C}_f = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \delta\sigma & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \sigma+\delta\sigma & 0 & 0 & 0 & 0 \\ r & 0 & 0 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \\ x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\boxed{[\hat{C}_f, \hat{M}_f] = \hat{C}_f\hat{M}_f - \hat{M}_f\hat{C}_f = \hat{\alpha}}$$

i.e. we cannot create a file independently of editing it — the result depends on the order.

5. Do modifications of the same file commute?

$$\hat{M}_f(\alpha)\hat{M}_f(\beta) = (I+\hat{\alpha})(I+\hat{\beta}) = I + \hat{\alpha}\hat{\beta} + \hat{\alpha} + \hat{\beta}$$

$$\hat{M}_f(\beta)\hat{M}_f(\alpha) = (I+\hat{\beta})(I+\hat{\alpha}) = I + \hat{\beta}\hat{\alpha} + \hat{\beta} + \hat{\alpha}$$

$$\therefore \quad [\hat{M}_f(\alpha), \hat{M}_f(\beta)] = [\hat{\alpha}, \hat{\beta}] + (\hat{\alpha}+\hat{\beta}) - (\hat{\beta}+\hat{\alpha})$$

Now, what is special here is that "+" means concatenation of strings and thus + is not associative

$$\hat{\alpha}+\hat{\beta} \neq \hat{\beta}+\alpha$$

$$\therefore \quad [\hat{M}_f(\alpha), \hat{M}_f(\beta)] \neq 0 .$$

6. Different modifications to different files must commute, since the operations have no common attributes.

7. In cfengine, we have Append If No Such Line is

$$\hat{m}_f = \begin{cases} \hat{I} + \delta\hat{\sigma} & \text{iff } \delta\sigma \text{ is not in } \sigma \\ \hat{I} & \text{if } \delta\sigma \text{ is in } \sigma \end{cases}$$

Is this idempotent?

$$\hat{m}_f^{\,2} = \begin{cases} \hat{I} & \text{if acting on a state containing } \delta\sigma. \\ \hat{I} + \delta\hat{\sigma} & \text{if acting on a state not containing } \delta\sigma \end{cases}$$

~~So $\hat{m}_f^{\,2} \neq \hat{m}_f$ in general.~~

However, the result of $\hat{m}_f^{\,2}$ on any state is the same as the result of $\hat{m}_f$ on any state. So, since the operator needs to observe the state it is acting on for its definition, we can define it to be idempotent.

Is the operator convergent? If we assume that policy means a state in which δ_0 is contained, (once only) then:

$$\hat{m}_f \vec{q} = \vec{q}_{policy}$$
$$\hat{m}_f \vec{q}_{policy} = \vec{q}_{policy} \quad \checkmark$$

$\Rightarrow$ $\hat{m}_f$ is convergent.

Ex 27

1. If writing a tool for automatic configuration, I would want two properties
   (i) orthogonality (no overlap) in operators
   (ii) convergence towards a policy state.

2. The difference between idempotence and convergence depends on context. Idempotence is a property of an operator. Convergence requires that an operator understands the state it acts upon.

3. (a) Assuming that $\hat{T}$ includes resetting the system to the base state $\hat{T}^2 = \hat{T}$ by definition, so it is idempotent.

(little $\hat{m}$) $\rightarrow$ $\hat{C}$, $\hat{P}$ and $\hat{M}$ are idempotent provided we consider their action on states. So in both cases, the same result is achieved, provided the states are taken into account.

(b) Both methods are as reliable as each other. They both do what they claim.

(c) Both methods are equally correct – they both do what they claim.

(d) The convergence method can be used 'on the fly' while devices are in operation.

The congruence method requires the system to undergo a catastrophic reset; hence it cannot be used 'on the fly'.

(e) Given that maintenance is necessary often, convergence is the least 'violent' method of change – since it does not require a system to be halted.

(f) SNMP's operations can only read and write values, not modify values.

$$PUT^2 = PUT$$
$$GET^2 = GET$$

$\Rightarrow$ these are idempotent.

The problem of using this for workstations is that these operations cannot model the complexity of configuration without memorizing the <u>entire</u> configuration at a remote location.

NOTES ABOUT THIS WEEK:

These exercises show that it is possible to implement a 'dumb' mechanical algorithm for implementing convergence. Instead of complex logic in the change process, we enumerate the set of allowed operators, satisfying the constraint of orthog. and convergence.

Config. management is still widely discussed.

Week 5: Simple Systems

Exercise 28

1. The difference between a static and a dynamic system is whether there is perceptible activity in the system.
   - A static system is an organisational (data) structure.
   - A dynamic system contains at least one process.

   There is the question of time-scale here. If a system changes only very slowly compared to other changes around it we can consider it to be static. e.g. the categorization of a library is static compared to the rate at which books are borrowed. So it depends on our perspective.

2. The components in a dynamic system:

   Freedoms, constraints, processes etc..

3. Freedom = capacity to change.
   Constraint = restriction to a subset of allowed freedoms.

   (a) pendulum: freedom = movement in x-y plane, time.
   constraint = the pivot, the rod, gravitational force.

   (b) Web-server: freedoms = read data, transmit data, receive data
   Constraints = data rates limited by protocols, read access limited by access controls.

   (c) A help-desk: Freedoms = receive information, process information, transmit information.
   constraints = number of processors, queuing.

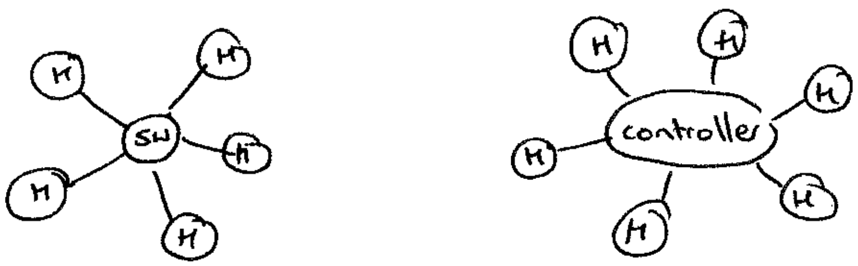
   (d) Config engine: freedoms = to change disk, start-stop processes ...
   constraints = policy, (access rights)

(e) SNMP monitor: freedoms = network read access
constraints = network communication.

4. Explain relationship between an algorithm and a protocol.
An algorithm is a constraint on a process that causes it to operate within its specification. A protocol is a specification of structure in a message that causes it to operate within its specification.
An algorithm is used to verify a protocol. The freedom is the freedom to send/process information. Constraints are the specs of the algorithm or protocol themselves (rules).

Exercise 29

1. Compare StarLAN network with Beowulf architecture.



LAN - hosts + switch.

Beowulf hosts + controller host.

Topology is similar. Both have centralized "switch" which constraints the flow of work. In LAN hosts talk to hosts through switch. In cluster, hosts talk mainly to controller. Freedoms are similar: each host is free to work as it likes, start/stop processes etc. Constraints are different: in Beowulf controller determines task schedule for hosts. In StarLAN, hosts determine their own tasks.

These architectures are good for coordinating small groups of nodes. Not so good for large numbers, since the centralization ⇒ serialization i.e. bottleneck which throttles performance.

2. Top-down means organization by increasing detail. i.e. high-level purpose motivates low level technicalities.

Bottom-up means organization of existing resources into functional entities that can be applied to a high-level attack.

(a) system design: top-down is good if you can start with a fresh slate (tabula rasa), but involves disruption or waste if you have an existing organisation. Bottom up allows us to adapt existing resources to new tasks.

(b) system maintenance: top down useful for planning strategies, but not for fixing things. e.g. take backup hierarchically is inefficient (tar), whereas bottom-up (dump) requires less overhead.

3. Hierarchy means 'tree-like' structure of levels. If we have to search deep trees, it can impact performance. Every level requires an additional level of processing.

4. Hierarchy uses decisions at each branch point to divide the set of all items into subsets. i.e. it is about subdivisions. We can also use overlapping set models (like cfengine), or relational database model.

5. System normalization is about finding an efficient characterization of subdivision, i.e. making sure that the decisions made at each level of subdivision are not repeated inefficiently.

Exercise ~~29~~ 30

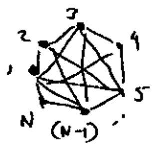
More of the same.

WEEK 7: DIAGRAMS AND GRAPHS

Exercise 31

Encryption keys: (i) shared-keys - 128 bytes
(ii) public-private keys - 1032 bytes

Using shared keys, N persons can point to (N-1) others. Each pair of N(N-1) persons needs one key, so the number of keys to ensure private communication is $N(N-1)/2$.



If each person has their own key pair, clearly 2N keys.

When is: $2N << N(N-1)/2$? Let's say that "<<" means that there is an order of magnitude (a factor of 10) difference.

i.e. $\frac{1}{2}(N-1)N = 20N$

$N^2 - N - 40N = 0$

$N(N-41) = 0$ i.e. $N=0$ or $N=41$.

So when the number of persons is about 40 it is worthwhile (by a factor of 10) to introduce PP-keys. The threshold, when they are equally numbered.

$\frac{1}{2}N(N-1) = 2N$

i.e. $N(N-5) = 0$ i.e. $\underline{N=5}$!

Suppose we compare memory instead. The break-even point is

$\frac{1}{2}N(N-1) \times 128 = 2N \times 1032$

$64N^2 - 64N - 2064N = 0$

$N(N - \frac{2128}{64}) = 0$

$N = \frac{2128}{64} = 33\frac{1}{4}$

So, in terms of storage space, the methods are equally efficient up to about 30 persons.

In order to be efficient by a factor of 10, we have

$\frac{1}{2}N(N-1) \times 128 = 20N \times 1032$

i.e. $N(N - \frac{20714}{64}) = 0$

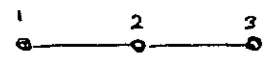
$N = \cancel{324}\ 324$

So, in terms of storage space, PP keys are efficient when we get to <u>hundreds</u> of users.

NOTE: normally space is not an issue, and we are more interested in managing the keys - i.e. how many we have to handle.

Exercise 32

1.



Adjacency matrix is 1 for link, 0 for no-link. Graph is undirected, so matrix is symmetrical.

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

2. Find the eigenvalues of A;

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \vec{\psi} = \lambda \vec{\psi} \qquad (*)$$

$$= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \vec{\psi} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \vec{\psi}$$

$$\begin{pmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{pmatrix} \vec{\psi} = 0$$

$\therefore$ $\det \begin{pmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{pmatrix} = 0$ for non-trivial solutions.

$$\det() = -\lambda\left(+\lambda^2 - 1\right) - 1\left(-\lambda . 1\right)$$

$$= -\lambda^3 + \lambda + \lambda = 0$$

$$\Rightarrow \lambda(\lambda^2 - 2) = 0$$

$\Rightarrow \lambda = 0$ or $\lambda = \pm\sqrt{2}$ (3 solutions).

Now find the eigenvalues:

$\lambda = 0$: substitute in $(*)$ and solve for $\vec{\psi} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$$

$$\Rightarrow \begin{aligned} a &= 0 \\ a + c &= 0 \\ b &= 0 \end{aligned}$$

$$\Rightarrow \psi_{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

$\lambda = \pm\sqrt{2}$

$$\begin{pmatrix} \mp\sqrt{2} & 1 & 0 \\ 1 & \mp\sqrt{2} & 1 \\ 0 & 1 & \mp\sqrt{2} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$$

$\Rightarrow$ $\mp\sqrt{2}a + b = 0$ $\Rightarrow b/a = \pm\sqrt{2}$ (1)

$a \mp\sqrt{2}b + c = 0$ $\Rightarrow a + c = \pm\sqrt{2}\,b$ (2)

$b \mp \sqrt{2}c = 0$ $\Rightarrow b/c = \pm\sqrt{2}$ (3)

(1) and (3) $\Rightarrow$

$a = c$

(2) $\Rightarrow$ $\pm\sqrt{2}\,a = b$

Take $c = 1$.

$\therefore$ $\vec{\psi}_{\pm} \propto \begin{pmatrix} 1 \\ \pm\sqrt{2} \\ 1 \end{pmatrix}$ (up to a scale factor).

The principal eigenvector belongs to $\lambda = +\sqrt{2}$ is $\vec{\psi}_{+}$ which reflects the symmetry of the graph's connectivity.

3. Consider now the metropolitan area network with 10 nodes.

$$
\begin{array}{c|cccccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
4 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
5 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
6 & 0 & 0 & 0 & 0 & 1 & 0 & \phi & 1 & 0 & 0 \\
7 & 0 & 0 & 0 & 0 & 0 & \phi & 0 & 0 & 0 & 0 \\
8 & 0 & 0 & 0 & 0 & 0 & \phi & 0 & 0 & 0 & 0 \\
9 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
10 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}
$$

Use "octave"

$$A = [\cdots ; \cdots]$$

$$[v, d] = \text{eig}(A)$$

gives, $\lambda_{max} = 2.5$

$$\boxed{\vec{\psi}^T_{max} = k\big(\underbrace{\underset{(1)}{1.8}\ \ \underset{(2)}{1.8}}\ \ \underbrace{\underset{(3)}{3.9}\ \ \underset{(4)}{3.9}}\ \ \underset{(5)}{5.1}\ \ \underset{(6)}{3.0}\ \ \underbrace{\underset{(7)}{1.2}\ \ \underset{(8)}{1.2}}\ \ \underset{(9)}{2.0}\ \ \underset{(10)}{4.6}\big)}$$

Notice which nodes are symmetrical and which nodes is biggest.

∴ Nodes are ranked by the principal eigenvector above. Node ~~10~~ 5 is the best connected node. This assumes that each of the routers in the diagram contributes as much as any other.

5. We rank the importance by the number of neighbours.

Let I_j be the j-th component of the importance vector $\vec{I}$. The (local) local importance is just the number of links.

$$I_j = \#\text{links to } j$$

$$= \sum_{i = \text{neighbours of } j}$$

$$I_j = \sum_i A_{ij} = \underline{\underline{A}}_j \vec{1}$$

If we we weight the sum

$$I_j = \sum_i \alpha_i A_{ij} = \underline{\underline{A}}\vec{\alpha} = \vec{I}$$

Suppose we now let $\vec{\alpha} \propto I = \beta\vec{I}$ then,

$$\vec{I} = \underline{\underline{A}}\vec{\alpha}$$

$$= \underline{\underline{A}}\,\beta\vec{I}$$

Let $\lambda = \frac{1}{\beta}$ $\quad \boxed{\therefore\ \underline{\underline{A}}\vec{I} = \lambda\vec{I}}$ $\quad$ ie $\lambda^{-1} = \beta$ is the constant of proportionality.

6. The ranking is limited by the assumption that connectivity is everything and that all links are equally important.

7. The principal eigenvector is the one in which the solution is formed from the strictly positive sum of the contributions.

8) Modelling human-computer Systems

Related exercises.

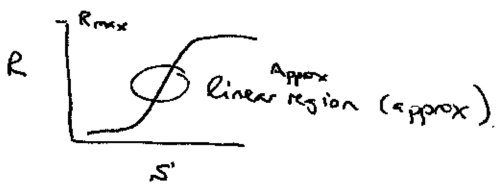
Ex 33 Antarctic research station measures signal strength from satellite changes periodically due to orbital effects. Rough approximation:

$$S' = S'_0 \left(1 + \sin\left(\frac{2\pi t}{24}\right)\right); \quad t = \text{hrs}.$$

(1) Should signal strength be $\propto$ data rate?

This is certainly not true for individual packets; we know that transmission is either successful or unsuccessful. But an unsuccessful transmission would lead to retransmission which would reduce the efficiency of transmission. We also know that, over a certain threshold, data-rate is not dependent on signal strength — as the likelihood of errors becomes very small. So this approximation might be roughly ok over long times (statistically), but the real relationship between signal strength and data-rate is much more non-linear and complicated

Would guess the real form looks more like the logistic function:



(2) $R(t) = R_0\left(1 + \sin\left(\frac{2\pi t}{24}\right)\right)$. If the maximum rate is 1 Gbps, then this must occur at the maximum value of $\sin(x)$, i.e.

$$R_{max} = \frac{10^9}{8} = R_0\,(1+1)$$

$$\Rightarrow R_0 = \underline{6.25 \times 10^7} \text{ bps}.$$

(3) To find the time for data transfer we have to be careful of units. $R(t)$ is measured in G bytes per second, but we are integrating over time in hours, so let

$$\frac{1}{\text{sec}} \quad \frac{\text{sec}}{\text{hr}} \quad \text{hr}$$

$$d(t) = \int dt \,(3600)\, R(t).$$

(4) So, we have:

$$d(t) = 3600 \int_{t_s}^{12.5} (6.25\times10^7)\left(1 + \sin\left(\frac{2\pi t'}{24}\right)\right) dt'$$

We want the movie downloaded by 18:00 hrs GMT, and $t=0$ corresponds to 5:30 a.m., then 18:00 hrs is $18 - 5.5 = 12.5 = t_f$.

so:

$$4\times10^9 \text{ bytes} = 3600 \int_{t_s}^{12.5} (6.25)\times10^7 \left(1 + \sin\left(\tfrac{\pi}{12}t'\right)\right) dt'$$

$$\frac{4}{36\times6.25} = \int_{t_s}^{12.5} \left(1 + \sin\left(\tfrac{\pi}{12}t'\right)\right) dt'$$

$$0.017 = \left[t - \frac{12}{\pi}\cos\left(\frac{\pi}{12}t\right)\right]_{t_s}^{12.5}$$

$$0.017 = (12.5 - t_s) - \frac{12}{\pi}\left(\cos\left(\pi\frac{12.5}{12}\right) - \cos\left(\frac{\pi}{12}t_s\right)\right)$$

This must be solved graphically or numerically. Use Mathematica:

```
FindRoot[12.5-0.017-t - 12/Pi*(Cos[12.5*Pi/12]- Cos[Pi*t/12])==0,
    {t,10}]
```

Gives $t_s = 12.48$ i.e. it takes just 1.2 minutes for the film to be downloaded.

Share by 10 penguins:

(5) Suppose $R_o \to R_o/10$, recalculate: $(0.017 \to 0.17)$

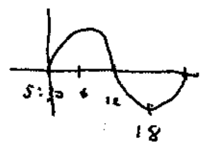
This solves to $t_s = 12.31$, so this take $12.5 - 12.31 = 0.19$ hrs $= 11.4$ mins.

Share by 100

Suppose $R_o \to R_o/100$, $0.017 \to 1.7$

$t_s \to 10.91$, ie. $(12.5 - 10.91) * 60 = \underline{94.8\text{mins}}$.

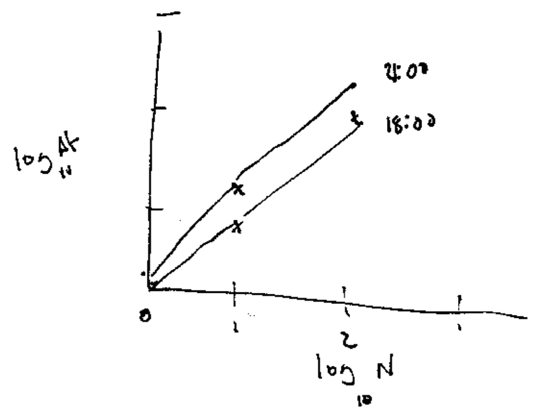
The behavior is surprisingly linear here.



Suppose we change the download time to 21:00 hrs GMT. ie $t_f = 15.5$

	t_s	Δt (mins)
R_o	15.42	4.8
$R_o/10$	14.85	39
$R_o/100$	12.31	190.8

At this time perfomance is much worse, and less linear.
So actually, it pays to choose the time of day appropriately.



$t_s = 12.48$ fits ~~This corresponds to 1 hour 1-2 mins before the film is shown, ie. 17:30 GMT. So we should start downloading before this.~~

MATHEMATICA:
~~FindRoot[12.5 - 0.017 - t - 12/Pi * (Cos[12.5 * Pi / 12] - Cos[Pi * t / 12]) == 0, {t, 10}]~~

Exercise 34

1. The number of connections, at any one time, is an integer, but the average number is divided by the total number of observations. That means it can be any rational number.

2. Maximum download activity is at $t = 12$ (hours). The maxima are found when

$$\frac{dN}{dt} = 0 = -10^4 \cdot \frac{12}{\pi} \cos\left(\frac{\pi}{12}t + \phi\right),$$

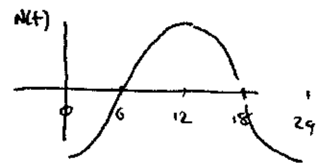
This is zero when $\cos(x) = \cos\left(\frac{\pi}{2} \pm n\pi\right)$ for $n = 0, 1, 2 \ldots$

$$\therefore \quad \frac{\pi}{12}t + \phi = \frac{\pi}{2} \pm n\pi$$

If we want to make $t = 12$ a maximum then

$$\frac{\pi}{12} \times 12 + \phi = \frac{\pi}{2} \pm 2n\pi \qquad \text{(some of these are minima)}$$

$$\therefore \quad \phi = -\frac{\pi}{2} \pm 2n\pi$$



3. Times of minimum activity are, according to the formula, $t = 0 \pmod{24}$. ie. when

$$\frac{\pi}{12}t - \frac{\pi}{2} = \frac{\pi}{2} \pm (2n+1)\pi \qquad n = 0, 1, \ldots$$

Ex 3~~4~~5

1. "Time is money" means that we value time spent by humans or computers and can relate this to a quantity of money, by some unspecified relationship. e.g. $t = m\alpha + \beta$.

2. Knowledge is stored in a binary tree – when grows in size like 2^n, for answering n questions. How can we relate CPU time or network resources to an amount of cache.

e.g. dimensionally:

memory (bytes) = (bytes per second) x seconds = bytes per query x queries

= (bytes per query) x (queries per second) x seconds.
(β) (dn/dt) (t)

So if n is the number of queries:

cache size = $2^n = \beta \frac{dn}{dt} \cdot t$. (*)

$$\int \beta \, dn \, 2^{-n} = \int \frac{dt}{t}$$

$$\beta \int dn \, e^{-n \log_e 2} = \log_e t + \text{const}$$

$$-\frac{\beta}{\log_e 2} e^{-n \log_e 2} = \log_e t/t_0$$

$$-\frac{\beta}{\log_e 2} 2^{-n} = \log_e t/t_0$$

$$\boxed{\begin{aligned} 2^{-n} &= e^{-x} \\ -n \log_e 2 &= -x \\ e^{-n \log_e 2} &= 2^{-n}. \end{aligned}}$$

Thus t is equivalent to a number of bytes in the cache by (*) and, in terms of queries n,

$$t = t_0 \exp\left(-\frac{\beta}{\log_e 2} 2^{-n}\right) = \text{time ~~saved~~ for query as cache grows}$$

3. Security is the opposite of convenience.

How do we measure security? (invent scale σ)
How do we measure convenience. (invent scale K)

What about $\sigma = \frac{\alpha}{K}$ α = constant ?

If security = 0 ⇒ convenience is ∞ ! (This is not true)
If convenience = 0 ⇒ security is ∞ ! (not true).

So introduce some constants σ_0 and K_0.

$$(\sigma + \sigma_0) = \frac{\alpha}{(K + K_0)}$$

Now if security $\sigma = 0$, $K = \frac{\alpha}{\sigma_0} - K_0$

i.e. $\left(\frac{\alpha}{\sigma_0} > K_0\right)$.

If convenience $K = 0$, $\sigma = \frac{\alpha}{K_0} - \sigma_0 > 0$

i.e. $\boxed{\alpha > \sigma_0 K_0}$ is okay ✓

So there is a consistent formulation we can use to define policy exactly.

4) Convenience of having all hosts identical must be inversely proportional somehow to the amount of work required to make them all different.

Define convenience of being same $\propto$ work saved
$\propto$ number of hosts x average work

$$K = \beta N \bar{W} \quad (\text{const } \beta).$$

(When $W = 0$, $K = 0$ ✓)

What happens if we substitute for κ in the answer to (3)

$$(\sigma + \sigma_0) = \frac{\alpha}{(\beta N W + \kappa_0)}$$

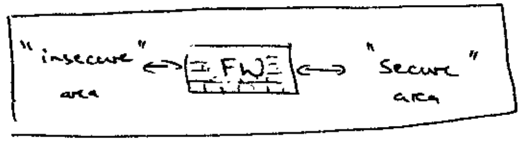
ie. security is inversely proportional to the amount of work saved.

or – security would be improved if we did more work!

(This is not necessarily true – it depends on so many hidden assumptions! Clearly formalizing policy is not easy.)

Ex 36

Biba model;



(a) If hosts in the insecure area read data from the secure area, the information is exposed to a low security area $\Rightarrow$ downgraded in the insecure area, but not in secure area.

(b) If hosts in the insecure area can write info to the secure area, the information is not downgraded, but the security level of the secure area might be, if the data are used in trust.

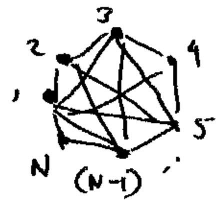
(c) Dubious.

WEEK 7: DIAGRAMS AND GRAPHS

Exercise 31

Encryption keys: (i) shared-keys - 128 bytes
(ii) public-private keys - 1032 bytes

Using shared keys, N persons can point to (N-1) others. Each pair of N(N-1) persons needs one key, so the number of keys to ensure private communication is N(N-1)/2.



If each person has their own key pair, clearly 2N keys.

When is: $2N << N(N-1)/2$? Let's say that "<<" means that there is an order of magnitude (a factor of 10) difference.

i.e. $\frac{1}{2}(N-1)N = 20N$

$$N^2 - N - 40N = 0$$

$$\underline{\underline{N(N-41) = 0}}$$ i.e. $N=0$ or $N=41$.

So when the number of persons is about 40 it is worthwhile (by a factor of 10) to introduce PP-keys. The threshold, when they are equally numbered.

$$\frac{1}{2}N(N-1) = 2N$$

i.e. $N(N-5) = 0$ i.e. $\underline{N=5}$!

Suppose we compare memory instead. The break-even point is

$$\tfrac{1}{2}N(N-1) \times 128 = N(1024 + 8)$$

(2 items)

$$64N^2 - 64N - 1032N = 0$$

$$N = \frac{1096}{64} = \underline{17}.$$

So, in terms of storage, the methods are just as good up to about 20 persons. To be efficient by, say, a factor of 10,

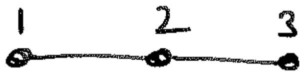
$$\tfrac{1}{2}N(N-1) = 10N \times 1032$$

$$N \simeq 161$$

So in terms of storage, PP-keys are efficient when we get to hundreds of users.

NOTE: memory is not usually an issue today (especially where security is involved) We are more concerned with managing the keys – i.e. how many we have to handle.

Ex. 32

1. 

The adjacency matrix is 1 for a link, 0 for no-link.

Graph is undirected, so matrix is symmetrical.

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

2. Find the eigenvalues of A;

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \vec{\psi} = \lambda \vec{\psi} \qquad (*)$$

$$= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \vec{\psi} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \vec{\psi}$$

$$\begin{pmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{pmatrix} \vec{\psi} = 0$$

$\therefore \det \begin{pmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{pmatrix} = 0$ for non-trivial solutions.

$$\det() = -\lambda\left(+\lambda^2 - 1\right) - 1\left(-\lambda \cdot 1\right)$$

$$= -\lambda^3 + \lambda + \lambda = 0$$

$$\Rightarrow \underline{\lambda\left(\lambda^2 - 2\right) = 0}$$

$\Rightarrow \lambda = 0$ or $\lambda = \pm\sqrt{2}$ (3 solutions).

<u>Now find the eigenvalues:</u>

$\underline{\lambda = 0}$: substitute in $(*)$ and solve for $\vec{\psi} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$$

$$\Rightarrow \quad a = 0$$
$$a + c = 0$$
$$b = 0$$

$$\Rightarrow \underline{\psi_{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}}.$$

$\lambda = \pm\sqrt{2}$

$$\begin{pmatrix} \mp\sqrt{2} & 1 & 0 \\ 1 & \mp\sqrt{2} & 1 \\ 0 & 1 & \mp\sqrt{2} \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$$

$$\Rightarrow \quad \mp\sqrt{2}a + b = 0 \qquad \Rightarrow \quad b/a = \pm\sqrt{2} \qquad (1)$$

$$a \mp\sqrt{2}b + c = 0 \qquad \Rightarrow \quad a + c = \pm\sqrt{2}b \qquad (2)$$

$$b \mp \sqrt{2}c = 0 \qquad \Rightarrow \quad b/c = \pm\sqrt{2} \qquad (3)$$

(1) and (3) $\rightarrow$

$$a = c$$

$$(2) \Rightarrow \pm\sqrt{2}\,a = b$$

Take $c = 1$.

$$\therefore \quad \vec{q}_{\pm} \propto \begin{pmatrix} 1 \\ \pm\sqrt{2} \\ 1 \end{pmatrix} \qquad \text{(up to a scale factor)}.$$

The principal eigenvector belongs to $\lambda = +\sqrt{2}$ is $\vec{q}_{+}$ which reflects the symmetry of the graph's connectivity.

(547)

3. Consider now the metropolitan area network with 10 nodes.

$$
\begin{array}{c|cccccccccc|}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
4 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
5 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
6 & 0 & 0 & 0 & 0 & 1 & 0 & \phi & 1 & 0 & 0 \\
7 & 0 & 0 & 0 & 0 & 0 & \phi & 0 & 0 & 0 & 0 \\
8 & 0 & 0 & 0 & 0 & 0 & \phi & 0 & 0 & 0 & 0 \\
9 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
10 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}
$$

Use "octave"

$$A = [\cdots ; \cdots]$$

$$[v, d] = \mathrm{eig}(A)$$

gives, $\lambda_{max} = 2.5$

$$\boxed{\vec{\psi}^{\,T}_{max} = k(\underbrace{\underset{(1)}{1.8}\ \ \underset{(2)}{1.8}}\ \ \underbrace{\underset{(3)}{3.9}\ \ \underset{(4)}{3.9}}\ \ \underset{(5)}{5.1}\ \ \underset{(6)}{3.0}\ \ \underbrace{\underset{(7)}{1.2}\ \ \underset{(8)}{1.2}}\ \ \underset{(9)}{2.0}\ \ \underset{(10)}{4.6})}$$

Notice which nodes are symmetrical and which nodes is biggest.

4. Nodes are ranked by the principal eigenvector above. Node ~~10~~ 5 is the best connected node. This assumes that each of the routers in the diagram contributes as much as any other.

S48

5. We rank the importance by the number of (local) neighbours.

Let I_j be the j-th component of the importance vector $\vec{I}$. The local importance is just the number of links.

$$I_j = \#\text{links to } j$$

$$= \sum_{i = \text{neighbours of } j}$$

$$I_j = \sum_i A_{ij} = \underline{\underline{A}} \cdot \vec{1}$$

If we we weight the sum

$$I_j = \sum_i \alpha_i A_{ij} = \underline{\underline{A}}\,\vec{\alpha} = \vec{I}$$

Suppose we now let $\vec{\alpha} \propto I = \beta \vec{I}$ then,

$$\vec{I} = \underline{\underline{A}}\,\vec{\alpha}$$

$$= \underline{\underline{A}}\,\beta \vec{I}$$

Let $\lambda = \frac{1}{\beta}$ $\boxed{\therefore \ \underline{\underline{A}}\,\vec{I} = \lambda \vec{I}}$ i.e. $\lambda^{-1} = \beta$ is the constant of proportionality.

6. The ranking is limited by the assumption that connectivity is everything and that all links are equally important.

7. The principal eigenvector is the one in which the solution is formed from the strictly positive sum of the contributions.

8) Modelling human-computer Systems

Related exercises.

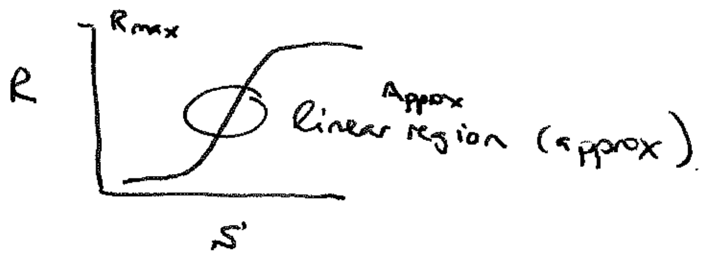
Ex 33 Antarctic research station measures signal strength from satellite changes periodically due to orbital effects. Rough approximation:

$$S' = S'_0 \left(1 + \sin\left(\frac{2\pi t}{24}\right)\right) ; \; t = \text{hrs}.$$

(1) Should signal strength be $\propto$ data rate?

This is certainly not true for individual packets; we know that transmission is either successful or unsuccessful. But an unsuccessful transmission would lead to retransmission which would reduce the efficiency of transmission. We also know that, over a certain threshold, data-rate is not dependent on signal strength – as the likelihood of errors becomes very small. So this approximation might be roughly ok over long times (statistically), but the real relationship between signal strength and data-rate is much more non-linear and complicated.

Would guess the real form looks more like the logistic function:



(2) $R(t) = R_0 \left(1 + \sin\left(\frac{2\pi t}{24}\right)\right)$. If the maximum rate is 1 Gbps, then this must occur at the maximum value of $\sin(x)$, i.e.

$$R_{max} = \frac{10^9}{8} = R_0 (1+1)$$

$$\Rightarrow R_0 = \underline{6.25 \times 10^7} \text{ bps}.$$

(3) To find the time for data transfer we have to be careful of units. $R(t)$ is measured in G bytes per second, but we are integrating over time in hours, so let

$$\frac{1}{\text{sec}} \quad \frac{\text{sec}}{\text{hr}} \quad \text{hr}$$

$$d(t) = \int dt \; (3600) \; R(t).$$

(4) So, we have:

$$d(t) = 3600 \int_{t_s}^{12.5} (6.25 \times 10^7)\left(1 + \sin\left(\frac{2\pi t'}{24}\right)\right) dt'$$

We want the movie downloaded by 18:00 hrs GMT, and $t=0$ corresponds to 5:30 a.m., then 18:00 hrs is $18 - 5.5 = 12.5 = t_f$.

So: $$4 \times 10^9 \text{ bytes} = 3600 \int_{t_s}^{12.5} (6.25) \times 10^7 \left(1 + \sin\left(\frac{\pi}{12} t'\right)\right) dt'$$

$$\frac{4}{36 \times 6.25} = \int_{t_s}^{12.5} \left(1 + \sin\left(\pi/_{12}\, t'\right)\right) dt'$$

$$0.017 = \left[t - \frac{12}{\pi} \cos\left(\frac{\pi}{12} t\right) \right]_{t_s}^{12.5}$$

$$0.017 = (12.5 - t_s) - \frac{12}{\pi}\left(\cos\left(\pi \frac{12.5}{12}\right) - \cos\left(\frac{\pi}{12} t_s\right)\right)$$

This must be solved graphically or numerically. Use Mathematica:

```
FindRoot [12.5-0.017-t - 12/Pi*(Cos[12.5*Pi/12]- Cos[Pi*t/12])==0,
                                                      {t,10}]
```

Gives $t_s = 12.48$ i.e. It takes just 1.2 minutes for the film to be downloaded.

```
for (t = 12.0 ; t ≤ 12.8 ; t++)
    {
    z = "(formula = 0)"
    print t, z
    }
```

(now look for the answer t when z is zero).

share by 10 penguins :

(15) Suppose $R_o \to R_o/10$, recalculate : $(0.017 \to 0.17)$ (55/a)

This solves to $t_s = 12.31$, so this take $12.5 - 12.31 = 0.19$ hrs

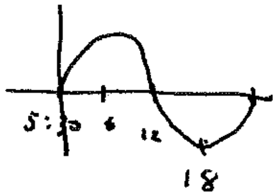
$= 11.4$ mins.

Share by 100

Suppose $R_o \to R_o/100$, $0.017 \to 1.7$

$t_s \to 10.91$, ie. $(12.5 - 10.91) * 60 = \underline{94.8 \text{mins}}$.

The behavior is surprisingly linear here.

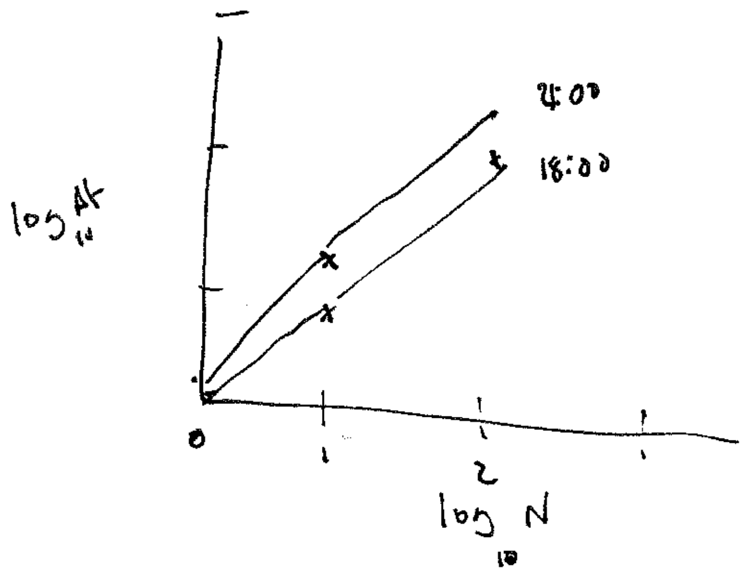


Suppose we change the download time to 21:00 hrs GMT . ie $t_f = 15.5$

	t_s	Δt (mins)
R_o	15.42	4.8
$R_o/10$	14.85	39
$R_o/100$	12.31	190.8

At this time perfomance is much worse, and less linear.

So actually, it pays to choose the time of day appropriately.



$t_s = 12.98$ fits. This corresponds to ~~$\frac{1}{2}$ hour~~ 12 mins before the film is shown, i.e. 17:30 GMT. So we should start downloading before this.

MATHEMATICA:

```
FindRoot[12.5 - 0.017*t - 12/Pi*(Cos[12.5*Pi/12] - Cos[Pi*t/12]) == 0, {t,10}]
```

Exercise 34

1. The number of connections, at any one time, is an integer, but the average number is divided by the total number of observations. That means it can be any rational number.

2. Maximum download activity is at $t = 12$ (hours). The maxima are found when.

$$\frac{dN}{dt} = 0 = -10^4 \cdot \frac{12}{\pi} \cos\left(\frac{\pi}{12} t + \phi\right)$$

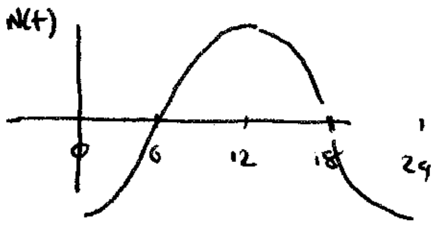
This is zero when $\cos(x) = \cos\left(\frac{\pi}{2} \pm n\pi\right)$ for $n = 0, 1, 2 \ldots$,

$$\therefore \quad \frac{\pi}{12} t + \phi = \frac{\pi}{2} \pm n\pi$$

If we want to make $t = 12$ a maximum then

$$\frac{\pi}{12} \times 12 + \phi = \frac{\pi}{2} \pm 2n\pi \qquad \text{(some of these are minima)}$$

$$\therefore \quad \phi = -\frac{\pi}{2} \pm 2n\pi$$



3. Times of minimum activity are, according to the formula. $t = 0 \pmod{24}$. i.e. when

$$\frac{\pi}{12} t - \frac{\pi}{2} = \frac{\pi}{2} \pm (2n+1)\pi \qquad n = 0, 1, \ldots$$

Ex 345

1. "Time is money" means that we value time spent by humans or computers and can relate this to a quantity of money, by some unspecified relationship. e.g. $t = m\alpha + \beta$.

2. Knowledge is stored in a binary tree – when grows in size like 2^n, for answering n questions. How can we relate CPU time or network resources to an amount of cache.

e.g. dimensionally:

$$\text{memory (bytes)} = (\text{bytes per second}) \times \text{seconds} = \text{bytes per query} \times \text{queries}$$

$$= \underbrace{(\text{bytes per query})}_{(\beta)} \times \underbrace{(\text{queries per second})}_{(dn/dt)} \times \underbrace{\text{seconds.}}_{(\Delta t)}$$

So if n is the number of queries:

$$\text{cache size} = 2^n = \beta \frac{dn}{dt} \cdot \Delta t \cdot \qquad (*)$$

$$\int \beta \, dn \, 2^{-n} = \int \frac{dt}{\Delta t}$$

$$\beta \int dn \, e^{-n \log_e 2} = \log_e t + \text{const}$$

$$-\frac{\beta}{\log_e 2} e^{-n \log_e 2} = \log_e t/t_0$$

$$-\frac{\beta}{\log_e 2} 2^{-n} = \log_e t/t_0$$

$$\boxed{\begin{aligned} 2^{-n} &= e^{-x} \\ -n \log_e 2 &= -x \\ e^{-n \log_e 2} &= 2^{-n}. \end{aligned}}$$

Thus t is equivalent to a number of bytes in the cache by $(*)$ and, in terms of queries n,

$$t = t_0 \exp\left(-\frac{\beta}{\log_e 2} 2^{-n}\right) = \text{time ~~spent~~ for query as cache grows}$$

3. Security is the opposite of convenience.

How do we measure security? (invent scale σ)

How do we measure convenience. (invent scale K)

What about $\sigma = \dfrac{\alpha}{K}$ α = constant ?

If security $= 0$ $\Rightarrow$ convenience is ∞ ! (This is not true)

If convenience $= 0$ $\Rightarrow$ security is ∞ ! (not true).

So introduce some constants σ_0 and K_0.

$$(\sigma + \sigma_0) = \frac{\alpha}{(K + K_0)}$$

or $\sigma = 100\% - K$.

Now if security $\sigma = 0$, $K = \dfrac{\alpha}{\sigma_0} - K_0$

i.e. $\left(\dfrac{\alpha}{\sigma_0} > K_0\right)$.

If convenience $K = 0$, $\sigma = \dfrac{\alpha}{K_0} - \sigma_0 > 0$

i.e. $\boxed{\alpha > \sigma_0 K_0}$ is okay ✓

So there is a consistent formulation we can use to define policy exactly.

4) Convenience of having all hosts identical must be inversely proportional somehow to the amount of work required to make them all different.

Define convenience of being same $\propto$ work saved

$\propto$ number of hosts $\times$ average work

$$K = \beta N \bar{W} \quad (\text{const } \beta).$$

(When $W = 0$, $K = 0$ ✓)

What happens if we substitute for K in the answer to (3)

$$(\sigma + \sigma_0) = \frac{\alpha}{(\beta N W + K_0)}$$

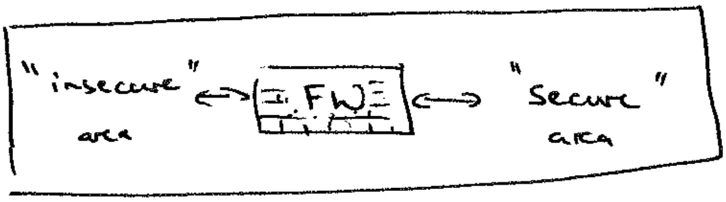
i.e. security is inversely proportional to the amount of work saved.

or – security would be improved if we did more work!

(This is not necessarily true – it depends on so many hidden assumptions! Clearly formalizing policy is not easy.)

Ex 36

Biba model;



(a) If hosts in the insecure area read data from the secure area, the information is exposed to a low security area ⇒ downgraded in the insecure area, but not in secure area.

(b) If hosts in the insecure area can write info to the secure area, the information is not downgraded, but the security level of the secure area might be, if the data are used in trust.

(c) Dubious.

WEEK 9 : Integrity, Information & Noise

Ex 37

1. The uncertainty of an observation / measurement is how much scatter there is (or might be) in the value on repeated measurement. e.g. the std-deviation of a set of measurements, or the std-error of the mean.

2. The Shannon entropy of a data stream with C 'symbols' is H

$$H = -\sum_{i=1}^{C} p_i \log p_i$$

3. The classes represent coded digits in a message.
   (Note that the uncertainty in each symbol is the width of the class in terms of the resolution of a measurement.)

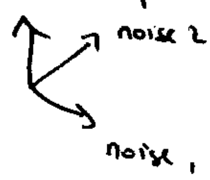
4. A high entropy message has nearly equal numbers of all symbols.
   A low entropy message is dominated by only a few (e.g. one) symbols.

5. Two ways of coding symbols on fibre:
   - Using wavelength / colour range (frequency) FM
   - Using intensity of light (amplitude) AM.

6. The error formula says.

$$Err^2 = \sum_i \Delta q_i^2 \quad \text{(Pythagorean sum)}.$$

   i.e. each source of uncertainty is independent, and affects the symbols independently.



(orthogonal ⇒ independent)

7/

$$L = -\sum_{i=1}^{c} p_i \ln p_i - \alpha\left(\sum_{i=1}^{c} p_i - 1\right) - \beta\left(\sum_{i=1}^{c} p_i \frac{(q_i-\bar{q}_i)^2}{\sigma^2} - 1\right)$$

Maximize wrt p_i, α, β:

$$\frac{\partial L}{\partial p_i} = -\ln p_i - \frac{p_i}{p_i} - \alpha - \beta\,(q_i-\bar{q}_i)^2/\sigma^2 = 0$$

i.e. $\ln p_i = -1 - \alpha - \beta(q_i-\bar{q}_i)^2/\sigma^2$

$$p_i = \exp\left(-1-\alpha-\beta(q_i-\bar{q}_i)^2/\sigma^2\right)$$

$\frac{\partial L}{\partial \alpha} = \sum_{i=1} p_i - 1 = 0$, i.e. sum of probabilities $= 1$

$$\sum_i \exp(-1-\alpha)\exp\left(-\beta(q_i-\bar{q}_i)^2/\sigma^2\right) = 1$$

i.e. $\exp(-1-\alpha) = \left(\sum_i \exp\left(-\beta(q_i-\bar{q}_i)^2/\sigma^2\right)\right)^{-1}$

$$\Rightarrow \boxed{p_i = \frac{e^{-\beta(q_i-\bar{q}_i)^2/\sigma^2}}{\sum e^{-\beta(q_i-\bar{q}_i)^2/\sigma^2}}}$$

This is the Gaussian distribution!

8/. (see book, section 15.6)

part 8 supplement to (15.6) in book (leading to eqn.(15.19))

The Gaussian distribution is

$$p(q) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-q^2}{2\sigma^2}\right)$$

$$\Rightarrow -\ln p(q) = \ln\sqrt{2\pi\sigma^2} + \frac{q^2}{2\sigma^2}$$

The Shannon entropy is (up to a conversion of log-base)

$$H(q) = -\int p(q)\ln p(q)$$

$$= \ln(\sqrt{2\pi\sigma^2})\int p(q)\,dq + \int \frac{q^2}{2\sigma^2} p(q)\,dq$$

$$H \equiv H_1 + H_2$$

$$H_1 = \ln(\sqrt{2\pi\sigma^2}) \underbrace{\int p(q)\,dq}_{\equiv 1} = \ln(\sqrt{2\pi\sigma^2}) = \underline{\tfrac{1}{2}\ln(2\pi\sigma^2)}.$$

$$H_2 = \frac{1}{2\sqrt{2\pi}}\frac{1}{|\sigma|^3}\int_{-\infty}^{+\infty} q^2 e^{-q^2/2\sigma^2}\,dq.$$

Now, we know $\int_{-\infty}^{+\infty} e^{-x^2\alpha}\,dx = \sqrt{\frac{\pi}{\alpha}}$, so $\frac{d}{d\alpha}\int e^{-x^2\alpha}\,dx = -\int x^2 e^{-x^2\alpha}\,dx = -\frac{1}{2}\sqrt{\frac{\pi}{\alpha^3}}$

Let $\alpha = 1/2\sigma^2$

$$H_2 = \frac{1}{2\sqrt{2\pi}}\frac{1}{|\sigma|^3} \times +\frac{1}{2}\sqrt{\frac{\pi}{\left(\frac{1}{2\sigma^2}\right)^3}}$$

$$= \frac{\cancel{\sqrt{\pi}}\,(\sqrt{2}\,|\sigma|)^3}{4\sqrt{2}\,\cancel{\sqrt{\pi}}\,|\sigma|^3} = \underline{\underline{\frac{1}{2}}} = \frac{1}{2}\ln e$$

$$\Rightarrow H = \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2}\ln e$$

$$= \underline{\frac{1}{2}\ln(2\pi e\sigma^2)}$$

Ex 38 /

1. Show that the Shamon entropy multiplied by N is the average length of a string with fixed alphabet

(See section 9.8)

2. The entropy of an encryption key refers to the amount of symbol randomness in the stream of random numbers used to generate a key.

e.g. password. AAAAB has low entropy (2 symbols / 26)

passwd AXCDFM has higher entropy (6 symbols / 26)

If over all passwds of all users, we use only 2 symbols, it would be low entropy ⇒ easy to guess.

If all passwds are taken randomly from a stream of 26 symbols, ⇒ high entropy ⇒ hard to guess.

3. A system policy can be thought of as a stream of assertions, or operators that describe state.

4. A simple compression of

BIGCATBIGCAT .... XYXYXY....

X ⇔ BIG

Y ⇔ CAT

H = Average information per symbol.

5. Length of message = 18, over alphabet $\{A, B, C, I, G, T\}$; $c = 6$

$p(A) = 3/18$

$p(B) = 3/18$

$p_i = 3/18$.

$i = 1..c$

$$H = -\sum_{i=1}^{6} \frac{3}{18} \log_6 \frac{3}{18}$$

$$= -\log_6 \frac{1}{6}$$

$$= 1$$ (uncertainty per symbol is 100%)

⇒ $18H = 18$ symbols are uncertain.

The uncertainty per symbol is 100% because our statistics only tell us that the probability of finding each symbol is equal – i.e. there is no statistical way to choose based on past experience.

6. The message has $HN = 18$ – i.e. the compressed length is the same as the full length. This is true, even though we know that the message can be redigitized from "xml" $\widehat{BIG}\,\widehat{CAT}$ to binary $\hat{X}\,\hat{Y}$. The entropy cannot take into account a redigitization, because the expression assumes that the symbols are immutable.

The entropy tells us about the length of a string that can be reduced on statistical grounds – by virtue of the frequency of symbols in the string. The point here is that the statistical distribution of symbols has maximum entropy, i.e. $p_i = p_j$ for all i, j. All symbols are equally likely, so there is no way to reduce the message by inference. In the 'big cat' example, we are using pattern recognition, not statistics to compress.

7. The entropy distinguishes one path in the lattice of all possible messages.  i.e. it counts uncertainty based on distinguishability of ABC from ACB.
If we form an alphabet $\Sigma = \{\hat{O}_1, \hat{O}_2 \ldots\}$ of operators that commute and are convergent / idempotent.

This means that there is no difference between different orderings or multiple operations,

$$\text{i.e.} \quad \hat{O}_1\hat{O}_1\hat{O}_2\hat{O}_3 = \hat{O}_3\hat{O}_2\hat{O}_1 = \hat{O}_1\hat{O}_3\hat{O}_2$$

If there are C operators in the alphabet, then no message can be longer than C symbols.

The total number of different messages one can create from C symbols, if

(i) order matters

$$= C^N$$

(ii) if order does not matter.

$$= C. \qquad (\text{equiv to } \hat{O}_1 \hat{O}_2 \hat{O}_3 \ldots \hat{O}_C)$$

8. If faults Δ_i $(i = 1..8)$ occur with probability $p_i = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}$

they form strings

$$\hat{\Delta}_1 \hat{\Delta}_3 \hat{\Delta}_1 \hat{\Delta}_2 \hat{\Delta}_1 \hat{\Delta}_1 \hat{\Delta}_3 \ldots$$

In N symbols, we learn the probabilities above. This tells us the uncertainty per symbol is:

$$H = \sum_i p_i \log_2 p_i$$

$$= \underline{2 \text{ bits}} \quad \text{per symbol}$$

To correct these faults we need the exact antidote:

$$\hat{\Delta}_3^{-1} \hat{\Delta}_1^{-1} \ldots \text{ etc.}$$

which contains the same information as the faults.

(or, if the faults are idempotent, we need only to preserve the order)

If we do not respond to faults immediately but wait, then we ~~If the corrective operators~~ only need a compressed summary of the faults to correct. $T_m/\langle T\rangle$ = units of fault

$$\Rightarrow \text{Entropy} \times \left(\frac{\langle T\rangle}{T_m}\right)^{-1} = \text{fault number}$$

This avoids double counting, so reduces number of operations required. But we can never get shorter than a set of commuting idempotent / convergent operations.

WEEK 9 : Integrity, Information & Noise

Ex 37

1. The uncertainty of an observation / measurement is how much scatter there is (or might be) in the value on repeated measurement. e.g. the std-deviation of a set of measurements, or the std-error of the mean.

2. The Shannon entropy of a data stream with C 'symbols' is H

$$H = -\sum_{i=1}^{C} p_i \log p_i$$

3. The classes represent coded digits in a message.
(Note that the uncertainty in each symbol is the width of the class in terms of the resolution of a measurement.)

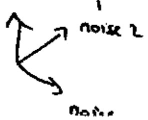
4. A high entropy message has nearly equal numbers of all symbols.
A low entropy message is dominated by only a few (e.g. one) symbols.

5. Two ways of coding symbols on fibre:
   - Using wavelength / colour range (frequency) FM
   - Using intensity of light (amplitude) AM.

6. The error formula says.

$$Err^2 = \sum_i \Delta q_i^2 \quad \text{(Pythagorean sum).}$$

i.e. each source of uncertainty is independent, and affects the symbols independently.



(orthogonal ⇒ independent)

7/

$$L = -\sum_{i=1}^{C} p_i \ln p_i - \alpha\left(\sum_{i=1}^{C} p_i - 1\right) - \beta\left(\sum_{i=1}^{C} p_i \frac{(q_i-\bar{q}_i)^2}{\sigma^2} - 1\right)$$

Maximize wrt p_i, α, β:

$$\frac{\partial L}{\partial p_i} = -\ln p_i - \frac{p_i}{p_i} - \alpha - \beta\,(q_i-\bar{q}_i)^2/\sigma^2 = 0$$

i.e. $\ln p_i = -1 - \alpha - \beta(q_i-\bar{q}_i)^2/\sigma^2$

$$p_i = \exp\left(-1-\alpha-\beta(q_i-\bar{q}_i)^2/\sigma^2\right)$$

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1} p_i - 1 = 0$$, i.e. sum of probabilities $= 1$

$$\sum_i \exp(-1-\alpha)\exp\left(-\beta(q_i-\bar{q}_i)^2/\sigma^2\right) = 1$$

i.e. $\exp(-1-\alpha) = \left(\sum_i \exp\left(-\beta(q_i-\bar{q}_i)^2/\sigma^2\right)\right)^{-1}$

$$\Rightarrow \boxed{p_i = \frac{e^{-\beta(q_i-\bar{q}_i)^2/\sigma^2}}{\sum e^{-\beta(q_i-\bar{q}_i)^2/\sigma^2}}}$$

This is the Gaussian distribution!

8/. (see book, section 15.6)

part 8 supplement to (15·6) in book (leading to eqn (15·19))

The Gaussian distribution is

$$p(q) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-q^2}{2\sigma^2}\right)$$

$$\Rightarrow -\ln p(q) = \ln\sqrt{2\pi\sigma^2} + \frac{q^2}{2\sigma^2}$$

The Shannon entropy is (up to a conversion of log-base)

$$H(q) = -\int p(q)\ln p(q)$$

$$= \ln(\sqrt{2\pi\sigma^2})\int p(q)\,dq + \int \frac{q^2}{2\sigma^2} p(q)\,dq$$

$$H \equiv H_1 + H_2$$

$$H_1 = \ln(\sqrt{2\pi\sigma^2}) \underbrace{\int p(q)\,dq}_{\equiv 1} = \ln(\sqrt{2\pi\sigma^2}) = \underline{\tfrac{1}{2}\ln(2\pi\sigma^2)}.$$

$$H_2 = \frac{1}{2\sqrt{2\pi}}\frac{1}{|\sigma|^3}\int_{-\infty}^{+\infty} q^2 e^{-q^2/2\sigma^2}\,dq.$$

Now, we know $\int_{-\infty}^{+\infty} e^{-x^2\alpha}dx = \sqrt{\frac{\pi}{\alpha}}$, so $\frac{d}{d\alpha}\int e^{-x^2\alpha}dx = -\int x^2 e^{-x^2\alpha}dx = -\frac{1}{2}\sqrt{\frac{\pi}{\alpha^3}}$

Let $\alpha = 1/2\sigma^2$

$$H_2 = \frac{1}{2\sqrt{2\pi}}\frac{1}{|\sigma|^3} \times +\frac{1}{2}\sqrt{\frac{\pi}{(\frac{1}{2\sigma^2})^3}}$$

$$= \frac{\sqrt{\pi}\,(\sqrt{2}\,|\sigma|)^3}{4\sqrt{2}\sqrt{\pi}\;|\sigma|^3} = \underline{\underline{\frac{1}{2}}} = \frac{1}{2}\ln e$$

$$\Rightarrow H = \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2}\ln e$$

$$= \underline{\frac{1}{2}\ln(2\pi e\sigma^2)}$$

Ex38 /

1. Show that the Shannon entropy multiplied by N is the average length of a string with fixed alphabet

   (See section 9.8) 

2. The entropy of an encryption key refers to the amount of symbol randomness in the stream of random numbers used to generate a key.

   e.g. password. AAAAB has low entropy (2 symbols / 26)

   passwd AXCDFM has higher entropy (6 symbols / 26)

   If over all passwds of all users, we use only 2 symbols, it would be low entropy ⇒ easy to guess.

   If all passwds are taken randomly from a stream of 26 symbols, ⇒ high entropy ⇒ hard to guess.

3. A system policy can be thought of as a stream of assertions, or operators that describe state.

4. A simple compression of

   BIGCATBIGCAT .... XYXYXY. . . .

   X ⇔ BIG

   Y ⇔ CAT

   H = Average information per symbol.

5. Length of message = 18, over alphabet $\{A, B, C, I, G, T\}$; $c = 6$

   $p(A) = 3/18$

   $p(B) = 3/18$

   $p_i = 3/18$.

   $i = 1..c$

$$H = -\sum_{i=1}^{6}\frac{3}{18}\log_6\frac{3}{18}$$

$$= -\log_6\frac{1}{6}$$

$= 1$ (uncertainty per symbol is 100%)

$\Rightarrow 18H = 18$ symbols are uncertain.

The uncertainty per symbol is 100% because our statistics only tell us that the probability of finding each symbol is equal – i.e. there is no statistical way to choose based on past experience.

6. The message has $HN = 18$ – i.e. the compressed length is the same as the full length. This is true, even though we know that the message can be redigitized from "xml" $\widehat{BIG}\,\widehat{CAT}$ to binary $\hat{X}\hat{Y}$. The entropy cannot take into account a redigitization, because the expression <u>assumes</u> that the symbols are immutable.

The entropy tells us about the length of a string that can be reduced on <u>statistical grounds</u> – by virtue of the frequency of symbols in the string. The point here is that the statistical distribution of symbols has <u>maximum</u> entropy, i.e. $p_i = p_j$ for all i,j. All symbols are equally likely, so there is no way to reduce the message by inference. In the 'big cat' example, we are using pattern recognition, not statistics to compress.

7. The entropy distinguishes <u>one</u> path in the lattice of all possible messages.  i.e. It counts uncertainty based on distinguishability of ABC from ACB.
If we form an alphabet $\Sigma = \{\hat{O}_1, \hat{O}_2 \dots\}$ of operators that <u>commute</u> and are convergent / idempotent.

This means that there is no difference between different orderings or multiple operations,

i.e. $\hat{O}_1\hat{O}_1\hat{O}_2\hat{O}_3 = \hat{O}_3\hat{O}_2\hat{O}_1 = \hat{O}_1\hat{O}_3\hat{O}_2$

If there are C operators in the alphabet, then no message can be longer than C symbols.

The total number of different messages one can create from C symbols, if

(i) <u>order matters</u>

$$= C^N$$

(ii) <u>if order does not matter.</u>

$$= C. \qquad (\text{equiv to } \hat{O}_1\hat{O}_2\hat{O}_3 \dots \hat{O}_C)$$

8. If faults Δ_i $(i = 1..8)$ occur with probability $p_i = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64})$
they form strings

$$\hat{\Delta}_1\hat{\Delta}_3\hat{\Delta}_1\hat{\Delta}_2\hat{\Delta}_1\hat{\Delta}_1\hat{\Delta}_3 \dots$$

In N symbols, we learn the probabilities above. This tells us the uncertainty per-symbol is:

$$H = \sum_i p_i \log_2 p_i$$

$$= \underline{2 \text{ bits}} \text{ per symbol}$$

To correct these faults we need the exact antidote:

$$\hat{\Delta}_3^{-1}\hat{\Delta}_1^{-1} \dots \text{ etc.}$$

which contains the same information as the faults.

(or, if the faults are idempotent, we need only to preserve the order.)

If we do not respond to faults immediately but wait, then we ~~If the corrective operators~~ only need a compressed summary of the faults to correct. $T_m/\langle T\rangle$ = units of fault

$$\Rightarrow \text{Entropy} \times \left(\frac{\langle T\rangle}{T_m}\right)^{-1} = \text{fault number}$$

This avoids double counting, so reduces number of operations required. But we can never get shorter than a set of commuting idempotent / convergent operations.

WEEK 10: ARRIVALS & QUEUES

Ex 39 A time series is a set of sampled observations measured at regular time intervals. i.e. a set of pairs $(t, q(t))$

The Hurst exponent of a time series is a measure of how self-similar a time-series is in terms of its fluctuations. It is defined in terms of a scaling relation.

$$s^{-H} \langle q(st) \rangle = \langle q(t) \rangle$$

This says that if we stretch a time series sample to s times its length, the average result is the same up to some factor.

$\underset{\Delta t}{\vdash\!\!\dashv} \rightarrow \underset{s\Delta t}{\vdash\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!\dashv}$ (s x magnification)

i.e. the two regions have the same statistical properties, no matter if we zoom-in and look at a smaller region, apart from a number s^H. H indicates whether a time-series has any long-range correlations, i.e. whether there are any long-memory processes.

Ex ~~38~~ 40

1. An arrival process is a random variable that occurs at random times. i.e. it is a pair $\{(t, q(t))\}$, where both t and $q(t)$ are random variables. The common model for an ~~poisson~~ arrival process is the Poisson process.

2. Arrival of jobs at a help-desk. How would we determine the type of arrival process? If there are both long and short jobs, we need to decide whether we care about distinguishing them. We could i) ignore the difference ii) treat a long job as the arrival of several small "units" at the same time.

Then, we measure the inter-arrival times and see what kind of distribution we get (e.g. see exercise 10). We can use log plots to see if the distribution is an exponential process or a power-law.

The best way to model a queue is "probably" to treat is as a Poisson processe – or Markov process, just because it is so much easier to deal with — and the benefits of the extra complexity using other models is not worth it, for most applications.

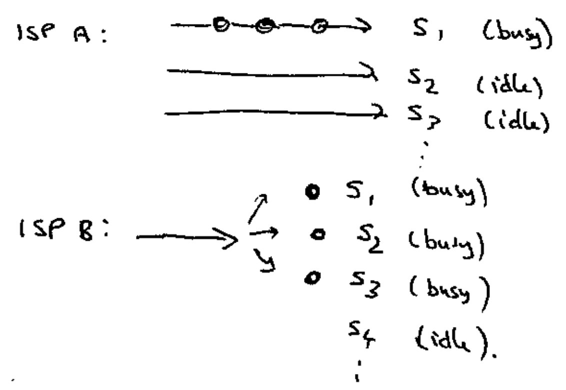
3/. Two ISPs compete for efficiency using different queue models. Both have six staff on telephones.

ISP A : separates queries into 6 different, pre-sorted queues

ISP B : keeps all queries in 1 long queue, with six 'servers'.

Intuitively, 1 long queue with 6 servers seems more efficient because every request receives service from the first available server. In separate queues, a long job will prevent a queue from being emptied, while other queues are idle.

If there are 3 calls on the same topic:



(4) ISP A and B both get 0·9 calls per minute.

The average time per call is 5 minutes $\Rightarrow$ service rate $= \frac{1}{5}\ \text{min}^{-1}$

(compare this to example 109) Chapter 12.

<u>ISPA uses 6 × M/M/1 queues</u>

Assume that the call types are evenly distributed

$$\lambda_A^i = \frac{0.9}{6} = 0.15$$

$$\mu_A^i = 0.2 \quad \text{for all queues.}$$

Traffic intensity $\rho^i = \frac{\lambda_A^i}{\mu_A^i} = \frac{0.15}{0.2} = \underline{0.75}$

This is less than 1, so the queue is stable. The average response time is:

$$R = \frac{\langle n \rangle}{\lambda} = \frac{1}{\mu - \lambda} = \frac{1}{0.05} = \underline{20 \text{ mins}}.$$

<u>ISPB uses M/M/6 queue</u>

We have the same traffic intensity, since now $\rho = \lambda/\mu k$

$\lambda = 0.9$, $\mu = 0.2$, $k = 6$. To find R for M/M/6, we need

$$P_0 = \left(1 + \sum_{n=1}^{5} \frac{(6\times0.75)^n}{n!} + \frac{(6\times0.75)^6}{6!\,(1-0.75)}\right)^{-1}$$

$$= \left(1 + 6\times0.75 + \frac{(6\times0.75)^2}{2} + \right.$$

$$= \left(1 + 4.5 + 1495 + \dots \frac{(6\times0.75)^k}{6!\,(1-0.75)}\right)^{-1}$$

$$K = \frac{(6\times0.75)^6}{6!\,(1-0.75)}\left(1 + 4.5 + 1495 + \dots\right)^{-1}$$

$$K = \frac{(4.5)^6}{6!\,0.25}\left(1 + 4.5 + \frac{(4.5)^2}{2} + \frac{(4.5)^3}{6} + \frac{(4.5)^4}{24} + \frac{(4.5)^5}{120} + \frac{(4.5)^6}{6!\,0.25}\right)^{-1}$$

$$= 0.422$$

Hence

$$R = \frac{1}{\mu}\left(1 + \frac{K}{k(1-\rho)}\right)$$

$$= \frac{1}{0.2}\left(1 + \frac{0.422}{6\times0.25}\right)$$

$$= \underline{6.4 \text{ mins.}}$$

The waiting time is significantly less at high load for the single queue. (It is close to the average call time.). The much longer time for the single queues. arises because there is a significant chance that arrivals will be delayed by build ups in <u>one</u> of the queues. (ie. it depends on the order).

5. The utilization is ~~75% in both cases~~. $\cancel{= \frac{\text{hit rate}}{\text{time to reply}}}$

$$U = \frac{\text{Busy time}}{\text{Total time}} = \frac{\text{busy time}}{\text{No. completions}} \times \frac{\text{completions}}{\text{time}}$$

$$= S\mu.$$

If hit rate $= 0.15x$ hits per minute, $\lambda/6$ for each queue.

time to reply $= R$ = average busy time.

Utilization $\approx 0.15R$.

ISP A: $U = \frac{0.15}{\cancel{6}} \times 20 = 0.5 = \cancel{50\%}\ \ 300\%$

ISP B: $U = \frac{0.15}{\cancel{6}} \times 6.4 = 0.16 = \cancel{16\%}\ \ 96\%$.

Week 11: Workflow Model.

Ex 41/

1. If the failure rate for total network is I, failure rate per device is I/N.

2. Work capacity C must be: $(C \geq I)$. i.e. the work that can be done on average by each node is C/N.

3. Scaling refers to the behaviour of workflow as a function of N. "Good scaling" means Workflow (N) is constant or grows only slowly. i.e.

$$W(N) \sim N^{\alpha} \quad \text{where } \alpha \leq 0 .$$

"Bad scaling" means

$$W(N) \sim N^{\alpha} \text{ for } \alpha > 0 .$$

NB. $W(N)$ is the amount of work that must be carried out at a single point, or that which must flow over a fixed link (bottleneck).

4. (18.55).

$$I_{fail} = \left(I_{err} - \frac{C_s}{N}\right) \Theta\left(I_{err} - \frac{C_s}{N}\right)$$

The $\Theta(x)$ function is non-zero when $x > 0$, so this means that the failure rate $I_{fail} = 0$ as long as the error rate $I_{err} \leq C_s/N$, i.e. as long as we can cope with the workload. Then, when it exceeds this limit, it grows like $(I_{err} - C_s/N)$.

5. If the probability of communication < 1, the scalability of the model is not affected (by the argument in (18.56)).

6. 'Each machine for itself' scales as a constant, i.e. $W(N) = N^0 = $ const

Ex 42/ Consider LDAP.

1. LDAP is a directory service, answering queries belonging to yellow and white page lookups.

2. LDAP has a hierarchical data model, and hierarchical distribution, with replication. Systems performing lookups are dependent on a remote server.

3. Principles at work: redundancy, hierarchy, scalability, partial centralization.....

4. Testing effectiveness by measuring: latency, scalability per number of hosts or request/second, numbers of users....

Ex 43

The connectivity:

$$\chi = \frac{1}{N(N-1)} h^T A h$$

Use condition $h^T h = H$, i.e. the number of active hosts is constant, and maximize the connectivity:

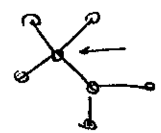
$$\chi = \frac{1}{N(N-1)} h^T A h - \alpha \left(h^T h - H\right)$$

$$\frac{\partial \chi}{\partial h^T} = \frac{1}{N(N-1)} A\vec{h} - \alpha\vec{h} = 0$$

i.e. $\boxed{A\vec{h} = \lambda\vec{h}}$, where $\lambda = N(N-1)\alpha$

i.e. we have an eigenvalue equation and $\vec{h}$ is the principal eigenvector of A.

In the graph:



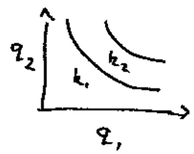
we find the principal eigenvector.

$$\vec{h}^T_{max} = (0.28\ \ 0.28\ \ 0.63\ \ 0.28\ \ 0.5\ \ \underbrace{0.22\ \ 0.22})$$

Node (3), the heart of the stick-man is ~~the~~ the central node.

Ex 44

1. If $q_1 q_2 = \text{const}$ $\quad q_1 = \frac{k}{q_2}$, i.e.



2. Define productivity

$$P = \vec{L}^T A \vec{L}$$

$$P = (L_1\ L_2)\begin{pmatrix} \alpha_1 & \beta \\ \beta & \alpha_2 \end{pmatrix}\begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = (L_1 L_2)\begin{pmatrix} \alpha_1 L_1 + \beta L_2 \\ \beta L_1 + \alpha_2 L_2 \end{pmatrix}$$

$$= \alpha_1 L_1^2 + 2\beta L_1 L_2 + \alpha_2 L_2^2 \quad \text{(scalar)}$$

WEEK 10: ARRIVALS & QUEUES

Ex 39 A time series is a set of sampled observations measured at regular time intervals. i.e. a set of pairs $(t, q(t))$

The Hurst exponent of a time series is a measure of how self-similar a time-series is in terms of its fluctuations. It is defined in terms of a scaling relation.

$$s^{-H} \langle q(st) \rangle = \langle q(t) \rangle$$

This says that if we stretch a time series sample to s times its length, the average result is the same up to some factor.

$$\vdash\!\!\dashv_{\Delta t} \rightarrow \vdash\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!\dashv_{s\Delta t} \qquad (s \times \text{magnification})$$

i.e. the two regions have the same statistical properties, no matter if we zoom-in and look at a smaller region, apart from a number s^H. H indicates whether a time-series has any long-range correlations, i.e. whether there are any long-memory processes.

Ex ~~38~~ 40

1. An arrival process is a random variable that occurs at random times. i.e. it is a pair $\{(t, q(t))\}$, where both t and $q(t)$ are random variables. The common model for an ~~poisson~~ arrival process is the Poisson process.

2. Arrival of jobs at a help-desk. How would we determine the type of arrival process? If there are both long and short jobs, we need to decide whether we care about distinguishing them. We could i) ignore the difference ii) treat a long job as the arrival of several small "units" at the same time.

Then, we measure the inter-arrival times and see what kind of distribution we get (e.g. see exercise 10). We can use log plots to see if the distribution is an exponential process or a power-law.

The best way to model a queue is "probably" to treat is as a Poisson processe – or Markov process, just because it is so much easier to deal with – and the benefits of the extra complexity using other models is not worth it, for most applications.

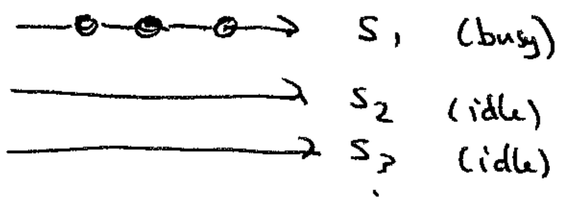
3/. Two ISPs compete for efficiency using different queue models. Both have six staff on telephones.

ISP A : separates queries into 6 different, pre-sorted queues.
ISP B : keeps all queries in 1 long queue, with six 'servers'.

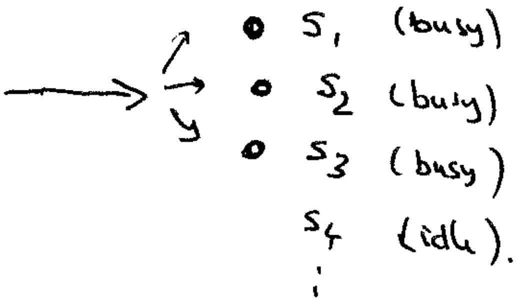
Intuitively, 1 long queue with 6 servers seems more efficient because every request receives service from the first available server. In separate queues, a long job will prevent a queue from being emptied, while other queues are idle.

If there are 3 calls on the same topic:

ISP A :



ISP B :



(4) ISP A and B both get 0·9 calls per minute.

The average time per call is 5 minutes $\Rightarrow$ service rate $= \frac{1}{5}\ \text{min}^{-1}$

(compare this to example 109.) Chapter 12.

ISPA uses $6 \times M/M/1$ queues

Assume that the call types are evenly distributed

$$\lambda_A^i = \frac{0.9}{6} = 0.15$$

$$\mu_A^i = 0.2 \quad \text{for all queues.}$$

Traffic intensity $\rho^i = \dfrac{\lambda_n^i}{\mu_A^i} = \dfrac{0.15}{0.2} = \underline{0.75}$

This is less than 1, so the queue is stable. The average response time is:

$$R = \frac{\langle n \rangle}{\lambda} = \frac{1}{\mu - \lambda} = \frac{1}{0.05} = \underline{20 \text{ mins}}.$$

ISPB uses $M/M/6$ queue

We have the same traffic intensity, since now $\rho = \lambda/\mu k$

$\lambda = 0.9$, $\mu = 0.2$, $k = 6$. To find R for $M/M/6$, we need

$$p_0 = \left(1 + \sum_{n=1}^{5} \frac{(6\times 0.75)^n}{n!} + \frac{(6\times 0.75)^6}{6!(1-0.75)}\right)^{-1}$$

$$= \left(1 + 6\times 0.75 + \frac{(6\times 0.75)^2}{2} + \right.$$

$$= \left(1 + 4.5 + 1495 + \cdots \frac{(6\times 0.75)^k}{6!(1-0.75)}\right)^{-1}$$

$$K = \frac{(6\times 0.75)^6}{6!(1-0.75)} \left(1 + 4.5 + 1495 + \cdots \right)^{-1}$$

$$K = \frac{(4.5)^6}{6!\,0.25}\left(1 + 4.5 + \frac{(4.5)^2}{2} + \frac{(4.5)^3}{6} + \frac{(4.5)^4}{24} + \frac{(4.5)^5}{120} + \frac{(4.5)^6}{6!\,0.25}\right)^{-1}$$

$$= 0.422$$

Hence

$$R = \frac{1}{\mu}\left(1 + \frac{K}{k(1-\rho)}\right)$$

$$= \frac{1}{0.2}\left(1 + \frac{0.422}{6 \times 0.25}\right)$$

$$= \underline{6.4 \text{ mins}}.$$

The waiting time is significantly less at high load for the single queue. (It is close to the average call time.). The much longer time for the single queues. arises because there is a significant chance that arrivals will be delayed by build ups in <u>one</u> of the queues. (i.e. it depends on the order).

5. The utilization is ~~75% in both cases~~. ~~= hit rate / time to reply~~

$$U = \frac{\text{Busy time}}{\text{Total time}} = \frac{\text{busy time}}{\text{No. completions}} \times \frac{\text{completions}}{\text{time}}$$

$$= S\mu.$$

If hit rate = 0.15λ hits per minute, $\lambda/6$ for each queue.

time to reply = R = average busy time.

Utilization $\approx 0.15R$.

ISP A: $U = \frac{0.15}{\cancel{\lambda}} \times 20 = 0.5 = \cancel{50\%}$ 300%

ISP B: $U = \frac{0.15}{\cancel{\lambda}} \times 6.4 = 0.16 = \cancel{16\%}$ 0.96%

Utilization (dimensional analysis).

Defined as $\frac{\text{time busy}}{\text{total time}} = \frac{\text{number busy}}{1} = \langle n \rangle$.

i.e. the length of the queue (like Unix load average).

By Little's law

$$\boxed{\langle n \rangle = \lambda R.}$$

Also we know:

$$S' = \frac{\text{busy time}}{\text{no. of completions}} = \text{busy number} \times \frac{\text{time unit}}{\text{completions}}$$

$$= \frac{\langle n \rangle}{\mu}$$

Just like $R = \frac{\langle n \rangle}{\lambda}$ = response time.

$$\Rightarrow \quad \langle n \rangle = \lambda R = \mu S'.$$

$$\frac{S'}{R} = \frac{\text{service time}}{\text{response time}} = \rho \quad \text{(traffic intensity)}$$

in problem.

$U_A = 0.15 \times 20 = \lambda R = 3$ (300% util.).
(losing this battle).

$U_B = 0.15 \times 6.4 = \lambda R = 0.96$ (96% util).
(coping).

WEEK 11: WORKFLOW MODEL.

Ex 41/

1. If the failure rate for total network is I, failure rate per device is I/N.

2. Work capacity C must be: $(C \geq I)$. ie. the work that can be done on average by each node is C/N.

3. Scaling refers ~~on~~ to the behaviour of workflow as a function of N. "Good scaling" means Workflow (N) is constant or grows only slowly. ie.

$$W(N) \sim N^{\alpha} \quad \text{where } \alpha \leq 0 .$$

"Bad scaling" means

$$W(N) \sim N^{\alpha} \text{ for } \alpha > 0 .$$

NB. $W(N)$ is the amount of work that must be carried out at a single point, or that which must flow over a fixed link (bottleneck).

4. (18.55).

$$I_{fail} = \left(I_{err} - \frac{C_s}{N} \right) \Theta \left(I_{err} - \frac{C_s}{N} \right)$$

The $\Theta(x)$ function is non-zero when $x > 0$, so this means that the failure rate $I_{fail} = 0$ as long as the error rate $I_{err} \leq C_s/N$, ie. as long as we can cope with the workload. Then, when it exceeds this limit, it grows like $(I_{err} - C_s/N)$.

5. If the probability of communication < 1, the scalability of the model is not affected (by the argument in (18.56)).

6. 'Each machine for itself' scales as a constant, i.e. $\omega(N) = N^0 = \text{const}$

Ex 42/ Consider LDAP.

1. LDAP is a directory service, answering queries belonging to yellow and white page lookups.

2. LDAP has a hierarchical data model, and hierarchical distribution, with replication. Systems performing lookups are dependent on a remote server.

3. Principles at work: redundancy, hierarchy, scalability, partial centralization.....

4. Testing effectiveness by measuring: latency, scalability per number of hosts or request/second, numbers of users....

Ex 43

The connectivity:

$$\chi = \frac{1}{N(N-1)} h^T A h$$

Use condition $h^T h = H$, i.e. the number of active hosts is constant, and maximize the connectivity:

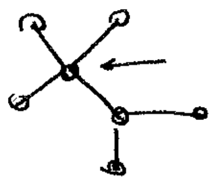
$$\chi = \frac{1}{N(N-1)} h^T A h - \alpha \left(h^T h - H \right)$$

$$\frac{\partial \chi}{\partial h^T} = \frac{1}{N(N-1)} A\vec{h} - \alpha \vec{h} = 0$$

i.e. $\boxed{A\vec{h} = \lambda \vec{h}}$, where $\lambda = N(N-1)\alpha$

i.e. we have an eigenvalue equation and $\vec{h}$ is the principal eigenvector of A.

In the graph:



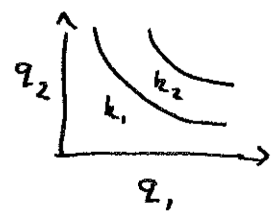
we find the principal eigenvector.

$$\vec{h}^T_{max} = (0.28 \quad 0.28 \quad \underset{\uparrow}{0.63} \quad 0.28 \quad 0.5 \quad \underbrace{0.22 \quad 0.22})$$

Node (3), the heart of the stick-man is the central node.

Ex 44

1. If $q_1 q_2 = \text{const}$ $\quad q_1 = \frac{k}{q_2}$, i.e.



2. Define productivity

$$P = \vec{L}^T A \vec{L}$$

$$P = (L_1 \; L_2)\begin{pmatrix} \alpha_1 & \beta \\ \beta & \alpha_2 \end{pmatrix}\begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = (L_1 L_2)\begin{pmatrix} \alpha_1 L_1 + \beta L_2 \\ \beta L_1 + \alpha_2 L_2 \end{pmatrix}$$

$$= \underline{\alpha_1 L_1^2 + 2\beta L_1 L_2 + \alpha_2 L_2^2} \quad \text{(scalar)}$$

$$= \alpha_1 L_1^2 + 2\beta k + \alpha_2 k / L_1^2$$

3) The minimum of P, using $L_1 L_2 = k$.

$$P = \alpha_1 L_1^2 + 2\beta k + \alpha_2 k / L_1^2$$

is found from

$$\frac{dP}{\partial L_1} = 2\alpha_1 L_1 - 2\alpha_2 k / L_1^3 = 0$$

$$L_1^4 = \frac{\alpha_2}{\alpha_1} k \,, \quad L_2^4 = \frac{\alpha_1}{\alpha_2} k.$$

$$\frac{d^2 P}{dL^2} = 2\alpha_1 + 6\alpha_2 k / L_1^4 \geq 0$$ $\Rightarrow$ minimum.

4) There is no maximum value (∞).

5) Stationary values of P

$$\frac{\delta P}{\delta L^T} = \begin{pmatrix} \alpha_1 & \beta \\ \beta & \alpha_2 \end{pmatrix} \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = 0$$

$$\left.\begin{array}{l} \alpha_1 L_1 + \beta L_2 = 0 \\ \beta L_1 + \alpha_2 L_2 = 0 \end{array}\right\}$$

$$\Rightarrow \quad \beta L_1 + \alpha_2 \left(- \frac{\alpha_1 L_1}{\beta} \right) = 0$$

$$\therefore \quad \underline{\beta^2 = \alpha_1 \alpha_2}$$

The significance of the solution is that β constrains the flow between the producers of α_1, α_2, so you cannot communicate more work over the channel than the link can cope with. β turns out to be the geometric mean of of the α_1, α_2 work rates.

6) The coupled equations are now

$$\frac{\delta}{\delta L^T} P = A\vec{L} - \lambda \vec{L}^* = 0 \quad , \text{ where } L^* = (L_2, L_1)$$

$$= A\begin{pmatrix} L_1 \\ L_2 \end{pmatrix} - \lambda \begin{pmatrix} L_2 \\ L_1 \end{pmatrix}$$

Note, this is not an eigenvalue equation:

$$\begin{pmatrix} \alpha_1 & \beta \\ \beta & \alpha_2 \end{pmatrix}\begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = \lambda \begin{pmatrix} L_2 \\ L_1 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \alpha_1 & \beta-\lambda \\ \beta-\lambda & \alpha_2 \end{pmatrix}\begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = 0 \qquad (*)$$

$$\alpha_1 L_1 + (\beta-\lambda) L_2 = 0 \quad \Rightarrow L_1 = \frac{\lambda-\beta}{\alpha_1} L_2 \qquad (A)$$

$$(\beta-\lambda) L_1 + \alpha_2 L_2 = 0 \quad \Rightarrow L_1 = \frac{\alpha_2}{\lambda-\beta} L_2 \qquad (B)$$

Combining (A) and (B)

$$(\lambda-\beta)^2 = \alpha_1\alpha_2$$

$$\lambda^2 - 2\beta\lambda + (\beta^2 - \alpha_1\alpha_2) = 0$$

$$\therefore \underline{\lambda = \beta \pm \sqrt{\alpha_1\alpha_2}} \qquad (C)$$

Sub'ing for $\beta - \lambda = \mp\sqrt{\alpha_1\alpha_2}$ in (*)

$$\left.\begin{aligned} \alpha_1 L_1 \mp \sqrt{\alpha_1\alpha_2}\, L_2 &= 0 \\ \mp\sqrt{\alpha_1\alpha_2}\, L_1 + \alpha_2 L_2 &= 0 \end{aligned}\right\} \Rightarrow \underline{L_1 = \pm\sqrt{\frac{\alpha_2}{\alpha_1}}\, L_2}$$

i.e. $\begin{pmatrix} L_1 \\ L_2 \end{pmatrix}_{\pm} \propto \begin{pmatrix} \pm\sqrt{\alpha_2} \\ \sqrt{\alpha_1} \end{pmatrix}$

but $L_1 L_2 = k \quad \Rightarrow \quad L_1^2 = \pm\sqrt{\frac{\alpha_2}{\alpha_1}}\, k. \quad L_2^2 = \pm\sqrt{\frac{\alpha_1}{\alpha_2}}\, k$

7) Sub in for $\begin{pmatrix} L_1 \\ L_2 \end{pmatrix}_{\pm}$ in the productivity

$$P = \alpha_1 L_1^2 + 2\beta L_1 L_2 + \alpha_2 L_2^2$$

$$= \alpha_1 \left(\pm \sqrt{\frac{\alpha_2}{\alpha_1}}\, k \right) + 2\beta k + \alpha_2 \left(\pm \sqrt{\frac{\alpha_1}{\alpha_2}}\, k \right)$$

$$= k \left(\pm \sqrt{\alpha_1 \alpha_2} + 2\beta \right)$$

$$= k \left(\pm \beta + 2\beta \right)$$

$$\underline{P = \left. \begin{matrix} \beta L_1 L_2 \\ 3\beta L_1 L_2 \end{matrix} \right\}}$$

Expect this proportional to β, because this is the bottleneck which throttles the flow.

(8) Repeat calculations with $A = \begin{pmatrix} \alpha_1 & \beta \\ 0 & \alpha_2 \end{pmatrix}$ $\quad \alpha_1 \circ \xrightarrow{\beta} \circ\, \alpha_2$

Now: $\underline{P = \alpha_1 L_1^2 + 2\beta L_1 L_2}$

$$\begin{pmatrix} \alpha_1 & \beta \\ 0 & \alpha_2 \end{pmatrix} \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = 0$$

$$\alpha_1 L_1 + \beta L_2 = 0$$

$$\alpha_2 L_2 = 0 \quad \Rightarrow L_2 = 0$$

$$\Rightarrow L_1 = 0 \qquad \text{(unconstrained)}$$

$$\text{or} \quad \alpha_1 \alpha_2 = 0$$

Add constraint:

$$\begin{pmatrix} \alpha_1 & \beta - \lambda \\ \beta - \lambda & \alpha_2 \end{pmatrix} \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = 0$$

$$\alpha_1 L_1 + (\beta - \lambda) L_2 = 0$$

$$\cancel{\alpha_2 L_2 + (\beta - \lambda) L_1 = 0 \quad \Rightarrow \lambda = \beta}$$

$$\cancel{\Rightarrow \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = \begin{pmatrix} 0 \\ L \end{pmatrix}}$$

Non-trivial solutions for L if

$$\det\begin{pmatrix}\alpha_1 & \beta-\lambda\\ -\lambda & \alpha_2\end{pmatrix} = 0$$

$$\alpha_1\alpha_2 + \lambda(\beta-\lambda) = 0$$

$$\lambda^2 - \beta\lambda - \alpha_1\alpha_2 = 0$$

$$\lambda = \frac{\beta \pm \sqrt{\beta^2 + 4\alpha_1\alpha_2}}{2} = \underline{\tfrac{1}{2}\beta \pm \sqrt{\tfrac{1}{4}\beta^2 + \alpha_1\alpha_2}}$$

The vector

$$\begin{pmatrix}\alpha_1 & \frac{1}{2}\beta \mp \sqrt{\beta^2/4 + \alpha_1\alpha_2}\\ -\frac{1}{2}\beta \mp \sqrt{\ } & \alpha_2\end{pmatrix}\begin{pmatrix}L_1\\ L_2\end{pmatrix} = 0$$

$$L_1/L_2 = \frac{-\frac{1}{2}\beta \pm \sqrt{\beta^2/4 + \alpha_1\alpha_2}}{\alpha_1}$$

$$L_1/L_2 = \frac{\alpha_2}{+\frac{1}{2}\beta \pm \sqrt{\beta^2/4 + \alpha_1\alpha_2}}$$

check

$$\frac{-\frac{1}{2}\beta \pm \sqrt{\ }}{\alpha_1} = \frac{\alpha_2}{\frac{1}{2}\beta \pm \sqrt{\ }}$$

$$\cancel{\tfrac{1}{4}\beta^2 \mp (\beta^2/4 + \alpha_1\alpha_2) = \alpha_1\alpha_2}$$

$$-\tfrac{1}{4}\beta^2 + \sqrt{\ }^2 = \alpha_1\alpha^2 \quad \checkmark \qquad \text{ok} \checkmark$$

$$\Rightarrow \quad L_1^2 = \underbrace{L_1L_2}_{k}\left(-\tfrac{1}{2}\beta \pm \sqrt{\beta^2/4 + \alpha_1\alpha_2}\right)$$

$$L_2^2 = \underbrace{L_1L_2}_{k.}\left(\tfrac{1}{2}\beta \pm \sqrt{\beta^2/4 + \alpha_1\alpha_2}\right)$$

$$L_1^2L_2^2 = k^2 = k^2\left(-\tfrac{1}{4}\beta^2 + \beta^2/4 + \alpha_1\alpha_2\right) = \alpha_1\alpha_2 = 1.$$

Now $P = \alpha_1 L_1^2 + \beta L_1L_2 + \alpha_2^2 L_2^2$

$$P = \alpha_1 k\left(-\tfrac{1}{2}\beta \pm \sqrt{\beta^2/4 + \alpha_1\alpha_2}\right) + 2\beta k + \alpha_2 k\left(\tfrac{1}{2}\beta \pm \sqrt{\ }\right)$$

$$\underline{P = (\alpha_2 - \alpha_1)k \pm (\alpha_1 + \alpha_2)\sqrt{\beta^2/4 + \alpha_1\alpha_2} + 2\beta k.}$$

To make productivity large we want $\alpha_2 > \alpha_1$, i.e. make host 2 big. (throttle).

New exercises added

(Marked "new" in problems

– numbers need to be adjusted)

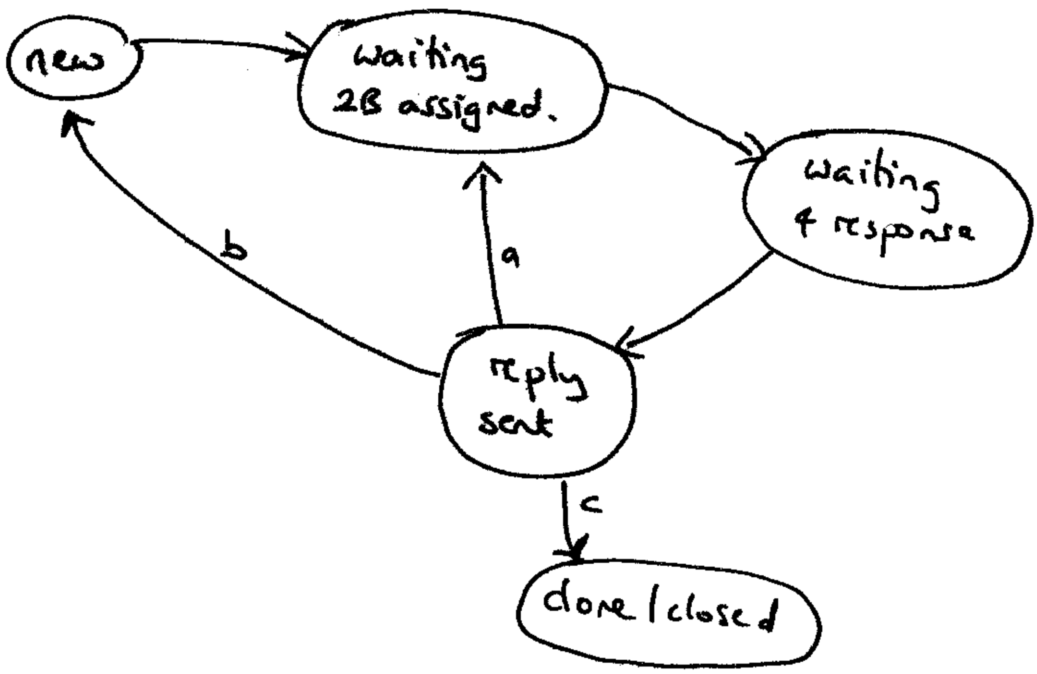
New exercises

1) Ticket handling system. (DISCRETE + CONTINUOUS)

This classifies states into

1) New problem arrives
2) Waiting to be assigned
3) Waiting for response
4) Reply sent
5) Ticket closed.

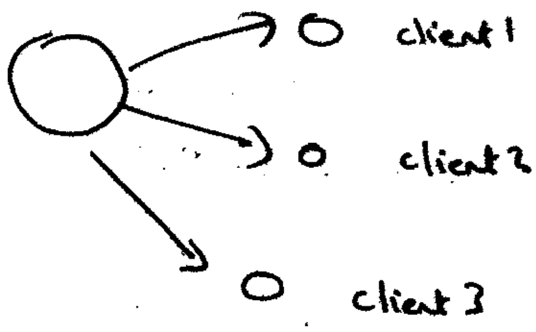
Transition diagram:



The top arrows are obvious, but what happens when the user replies to the "solution" sent by the help-desk? Either it fixes the problem (c) or it does not (a or b). If not, the problem would normally be reassigned (a). In the worst case — or at least in some cases the problem might fork into a complete new problem (b).

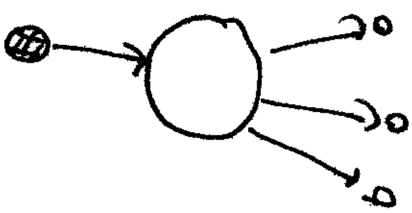
(2) DIAGRAMS + GRAPHS.

1. Company providing a service to 3 clients:



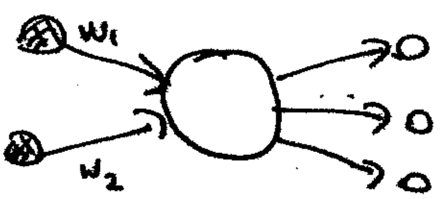
The arrows represent "provide service". Or we could reverse them and think of the dual picture "depends on".

2. The same company outsources part of its work:



i.e. now it depends on a provider itself for a service.

3. Redundant outsourcing:

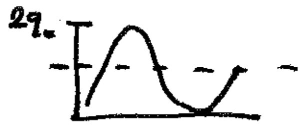


4. Load balancing means dividing the flow of service amongst a number of servers. We can show this using weights w_1, w_2 above

Snow Clearing Model.

(1) To make a probability function, we must be sure that it is bounded in $[0,1]$ and single-valued, and $\int dt\, p(t) = 1$

$$q(t) = q_0\left(1 + \sin\left(\frac{2\pi t}{P}\right)\right)$$



$q(t)$ is bounded in $[0, 2q_0]$. So let $p(t) = q(t)/N$

where N is a normalization constant.

$$\int_0^P dt\; p(t) = 1$$

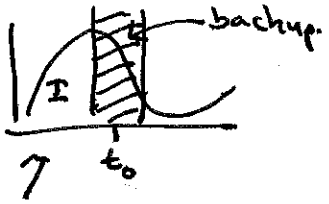
$$\Rightarrow \int_0^P dt\; \frac{q_0}{N}\left(1 + \sin\left(\frac{2\pi t}{P}\right)\right) = 1.$$

$$\Rightarrow \quad \underline{N = \left(q_0 \int_0^P dt\left(1 + \sin\left(\frac{2\pi t}{P}\right)\right)\right)^{-1}}$$

(We don't care what this value is. The important thing is that it exists and is well defined. – i.e. we <u>can</u> make probabilites just by letting $q(t) \to q/N$).

(2) Expectation: $\langle T \rangle = \int_0^P T(t)\,p(t)\,dt$

(3) Time to repair function. Regions I and III are straightforward.



A change that occurs here has to be dealt with in ▨, i.e. the backup region II. So if it occurs at time t, the average time before backup will be approximately $t_0 - t$.

why?

- We don't know the exact time at which the work will be done inside region II, so the answer can only be approximate.
- It could occur anywhere between

$$t_0 - \tfrac{1}{2} t_b$$
and $$t_0 + \tfrac{1}{2} t_b$$

Changes are arriving, modulated by $q(t)$ which is not flat.
If t_b is short compared to P $(t_b \ll P)$ then we can ignore the error.

II : $(t_0 - t)$ or $(P-t)+t_0$ $\rightarrow$ $\frac{1}{2}(t_0 - t) + \frac{1}{2}(P - t + t_0) = \frac{1}{2}P + t_0 - t$.

In region III, $t > t_0$ so we have to wait until the next pass.

The time to wait is then $\underbrace{P - t}_{\text{complete this cycle}} + \underbrace{t_0}_{\text{next cycle}}$.

(4) In region II, we have, on average, half $t < t_0$ and half $t > t_0$; by the same argument $t_b \ll P$. So take the average

$$T_{II} \approx \tfrac{1}{2}(T_I + T_{III}) = \tfrac{1}{2}(P + t_0 - t).$$

(5) Now we combine the result from (2) with the piecewise

$$T = \begin{cases} t_0 - t & , \quad t \text{ in region I}, \text{ i.e. } 0 < t < t_0 - \frac{1}{2}b \\ \frac{1}{2}(P + t_0 - t) & , \quad t \text{ in region II}, \ t_0 - \frac{1}{2}b < t < t_0 + \frac{1}{2}b \\ P + t_0 - t & , \quad t \text{ in region III}, \ t_0 + \frac{1}{2}t_b < t < P \end{cases}$$

$$\boxed{\langle T \rangle = \int_0^{t_0 - \frac{1}{2}t_b} (t_0 - t)\, p(t)\, dt + \int_{t - \frac{1}{2}t_b}^{t + \frac{1}{2}t_b} \tfrac{1}{2}(P + t_0 - t)\, p(t)\, dt \; \ldots + \int_{t + \frac{1}{2}t_b}^{P} (P + t_0 - t)\, p(t)\, dt}$$

$\rho = \rho_0 \langle T \rangle$. QED.

Business Activity Diagram.

Transition matrix.

$$\begin{array}{ccc}
 & \begin{array}{cccccccc} e & h & s & v & g & c & b & x \end{array} & \\
\begin{array}{c} e \\ h \\ s \\ v \\ g \\ c \\ b \\ x \end{array} &
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.7 & 0 & 0.1 & 0 & 0 & 0.2 \\
0 & 0 & 0.4 & 0.2 & 0.15 & 0 & 0 & 0.25 \\
0 & 0 & 0 & 0 & 0.65 & 0 & 0 & 0.35 \\
0 & 0 & 0 & 0 & 0 & 0.3 & 0.6 & 0.1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} &
\begin{array}{c} \Sigma = 1 \\ \Sigma = 1 \\ \Sigma = 1 \\ " \\ " \\ " \\ " \\ " \end{array}
\end{array}$$

This for formality.
(We don't model reentry)

(i) Directed graph $\Rightarrow$ non-symmetrical.

(ii) Each row should sum to 1, since these are probabilities.

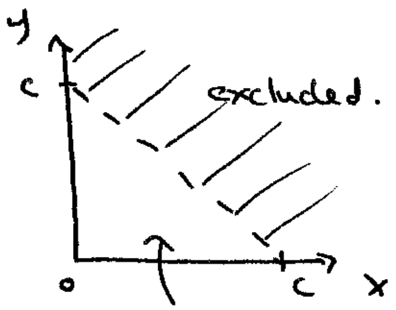
Linear Programming

Two services X and Y, with rates x and y Gigabits per second each. We don't know x and y yet.

We know that the maximum capacity available to the company over their leased line is C.

$$\Rightarrow \quad x + y \leq C \qquad (1)$$

1)

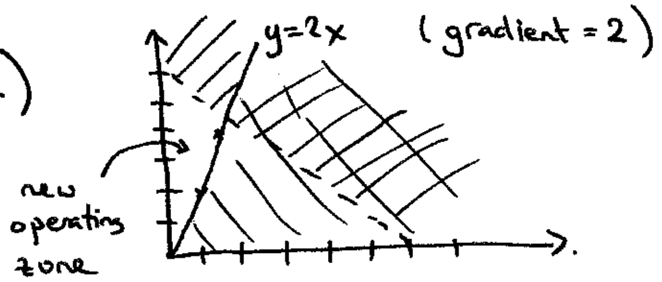


This is our operating region, in the absence of any other constraints.

i.e. any value of (x, y) inside this region is okay by (1).

So now we look at the other constraints.

2)



↑ Notice this prevents x from getting big. (bad news since that's where the money's going to be).

Our "business plan" is thus to exploit this region of 'achievable states'. It is a continuum (at least on paper)

(3) Show $\frac{P_x x}{D_x}$ is a rate of earning.

$$[P_x] = \frac{\text{euros}}{\text{transactions}}$$

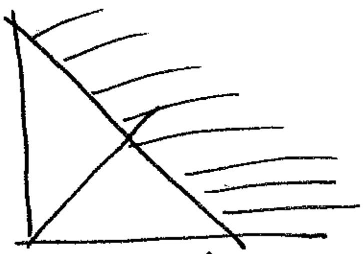
$$[x] = \frac{\text{Gbytes}}{\text{seconds}}$$

$$[D_x] = \frac{\text{Gbytes}}{\text{transactions}}$$

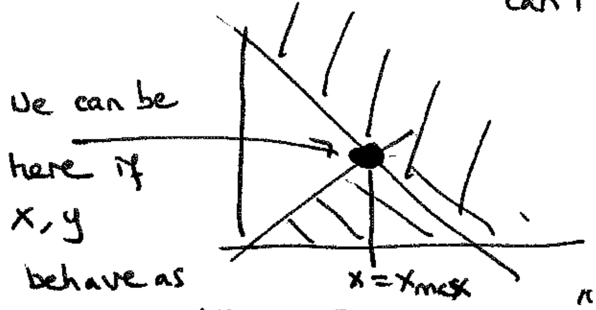
$$\left[\frac{P_x x}{D_x}\right] = \frac{\text{euros}}{\cancel{\text{trans}}.} \times \frac{\cancel{\text{Gbytes}}}{\text{secs}} \times \frac{\cancel{\text{trans}}.}{\cancel{\text{Gbytes}}} = \frac{\text{euros}}{\text{secs}}.$$

Thus this ratio is an earning rate.

4. If $P_x \leq 3P_y$ we should make x as big as possible to maximize earnings.



↑ We would like to sell just x for profit, but we can't.



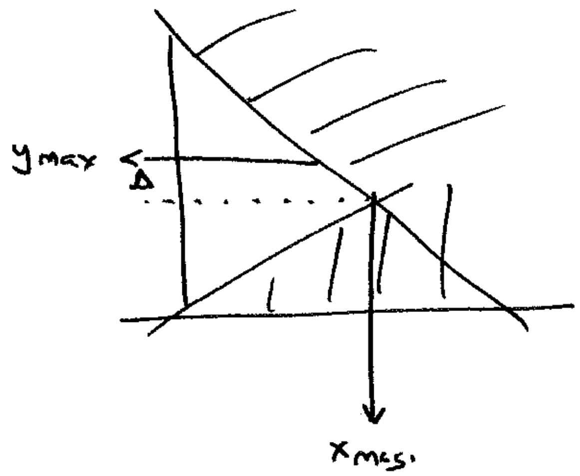
we would like. But x, y are "random variables" (averaged) – can't predict the actual rates, clients do what they like.

But – we can set this value $x = x_{max}$ as the throttle rate in the router, so we don't prevent any 'legal' possibilites, according to

Our constraints.

Now, if they constrain y to the y-coordinate of this blob, that will ~~xx~~ prevent y from growing bigger. That would preclude an opportunistically large value of y from making up for the lower than 'best' value of x. So we might want to set the ceiling value of y somewhere above this.

If, however, we set $y_{max} = C$, then it could prevent any x from occurring, in the worst case.



So we could leave a margin for error Δ.

There is no way to derive the 'right thing to do' here. It requires a <u>POLICY</u> choice/decision to be made.

5. Condition for the company to make a profit:

$$\boxed{\frac{P_x x}{D_x} + \frac{P_y y}{D_y} > R.}$$