Advanced Computer Architecture: Google 1

Jeremy Bradley

1 March 2004

1. Assume you are planning to layout a computing cluster in a cohosting site. Each rack unit PC outputs 40W of heat. Each rack requires a single switch which can output a maximum of 50W. You can fit up to 32 PCs and one switch in a rack. The air-conditioning in a hosting site can typically expel 1000W per square metre. If each rack takes 1m^2 of floor space and your proposed site is 100m^2, what is the maximum number of PCs you can locate in your site?

   - **32 PCs at 40W + 1 switch at 50W = 1330W per rack**
   - **Air-con can expel 100,000W over 100m^2**
   - $\Rightarrow$ **Site can support 100000/1330 racks = 75.188**
   - **Allowing a partially populated 76th rack (for which there would still be physical room) then you could have another 5 PCs + a switch.**
   - **i.e. total number of machines is: 2405 PCs**

2. You have a cryptographic key that needs to be broken. It is 128 bits in length i.e. there are 2^{128} possible keys. You have a budget of £500,000. A standard PC costs £1000 and can perform 100,000 key tests per second today. Assuming that the average PC doubles in speed every 18 months (indefinitely), how long must you wait before you can afford to buy sufficient computing power to test all 2^{128} keys within 6 months – ignoring inflation. (If you have time, have a go at solving the harder problem of finding the shortest time to search a 128 bit key given Moore's Law)

   - **number of seconds in 6 months,** $m = 0.5 \times 365 \times 24 \times 3600$
   - **number of computers is 500 (whenever they are bought)**
   - **required "keys per second per computer" rate,** $r = 2^{128}/500m$
   - **"doubles every 18 months"** $\Rightarrow r = 100,000 \times 2^n$
   - **number of 18 months epochs,** $n = 128 - (\log_2 50,000,000 + log_2 m) = 78.51$
   - **number of years** $= 1.5 \times n = 117.771$ **(or 177 years 9 months)**

3. You are running a disk array in a computing facility which receives 90,000 file requests every minute. In order to fulfil your SLA (service level agreement) your disk array must process each request (from arrival to service completion) within 0.01 seconds Assuming you are able to software configure the disk IO buffer, how large should you set it in order to deal with the average request load? How fast must the overall array service requests in order to meet the mean response time requirement? Answer the latter question for the different models of the disk array:

   (a) M/M/1

   (b) M/D/1

   - **Little's law:** $\mathbb{E}(N) = \lambda\, \mathbb{E}(R)$
   - **i.e. mean buffer length** = **arrival rate** $\times$ **mean response time also,** μ = **service rate,** $\rho = \lambda/\mu$ = **server utilisation**
   - **NB care with units: 90,000 per minute = 1500 per second**
   - **Buffer size requirement is independent of model used:**
   - $\mathbb{E}(R) = \mathbb{E}(N)/\lambda < 0.01$
   - $\Rightarrow \mathbb{E}(N) < 0.01\ \lambda = 15$
   - **i.e.** $\mathbb{E}(N) < 15$**, buffer must be at least capable of holding 15 files**

   M/M/1 model

   $\Rightarrow \mathbb{E}(N) = \rho/(1-\rho) < 15$
   $\Rightarrow 15 - 16\rho > 0$
   $\Rightarrow \rho < 15/16$
   $\Rightarrow \mu > 1600$**, i.e. must exceed 1600 requests per second**

   M/D/1 model

   $\Rightarrow \mathbb{E}(N) = \rho + \rho^2/(2(1-\rho)) < 15$
   $\Rightarrow \rho < 16 - \sqrt{226}$
   $\Rightarrow \mu > 1551.7$ **requests per second**

4. You are given a new model for the expected buffer length in an intranet search engine:
   $$\mathbb{E}(N) = \rho + \frac{3\rho}{1-\rho}$$
   Given that the arrival rate for requests, λ, is 2 per second and the average response time is to be kept below t seconds. Find an upper bound expression for ρ in terms of t.

- Little's Law: $\mathbb{E}(N)/2 = \mathbb{E}(R) < t \Rightarrow \mathbb{E}(N) < 2t$

$\Rightarrow$ $\rho + 3\rho/(1-\rho) < 2t$

$\Rightarrow$ $\rho^2 - 2(t + 2\rho) + 2t > 0$

$\Rightarrow$ $\rho < 2 + t - \sqrt{t^2 + 2t + 4}$