Heavy-tailed distributions

What if service time S is not light-tailed, but instead has an infinite MGF, and hence can not be thought of as "approximately exponential"?

A good example is a Pareto distribution:

$$P(S > x) = x^{-k},\ x \geq 1,$$

where $k \geq 2$ is a constant. It is easily seen that the mean (expected value) of S is $E(S) = 1+(k-1)^{-1}$ and that $E(S^n) = \infty$ for any $n \geq k$. Thus a Pareto distribution has a very heavy or *fat* or *long* tail: it tends to 0 very slowly.

In particular

$$lim_{x\to\infty} \frac{P(S > x)}{e^{-\epsilon x}} = \infty, \tag{1}$$

for any $\epsilon > 0$.

Another interesting example (satisfying (1)) is provided by taking a random variable X with an exponential distribution (with rate α); $P(X > x) = e^{-\alpha x}$, and squaring it: $S = X^2$. This is an example of a *Weibull* distribution, and is not as heavy-tailed as a Pareto because (for example) it has finite moments of all orders: $E(S^n) < \infty$ for all $n \geq 1$. (A similar example is given by $S = e^X$ where X has a normal distribution; called a *lognormal* distribution; very important in the pricing of stock derivatives in financial engineering).)

(These kinds of distributions seem to appear in the data of many real systems such as the WWW; hence their importance in using them in modeling networks.)

Here we show how bad system behavior can get even for the FIFO single-server queue with Poisson arrivals, when S is heavy-tailed.

Let us suppose that $P(S > x) = x^{-2},\ x \geq 1$ so S is Pareto with mean $E(S) = 2$ (so $\mu = 0.5$). We will further assume that arrivals are Poisson at rate $\lambda = 0.25$, so that $\rho = \lambda/\mu = 0.5 < 1$.

Now it is well known that in steady-state, the probability that an arrival finds the server busy is ρ and that if an arrival finds the server busy, then the *remaining* service time S_r of the customer in service has the so-called *equilibrium* distribution with tail given by

$$P(S_r > x) = \mu \int_x^\infty P(S > y)dy,$$

which in this case yields $P(S_r > x) = 0.5x^{-1},\ x \geq 1$. In particular, the mean of S_r is infinite; $E(S_r) = \infty$. But $P(D > x) \geq \rho P(S_r > x)$, so $E(D) = \infty$ also.

Fluid Models

Deeply related to queueing models are *fluid models*. Instead of customers, fluid arrives continuously in time to a processor that processes the fluid at some given constant rate (say 1). This is meant to approximate packets of information travelling in a telecommunications network. The quantity of interest is then $V(t)$ = fluid level at time t (similar to workload).There are no service times.

In the simplest examples, the arrival of fluid has alternating "ON" and "OFF" periods. When ON, fluid arrives continuously at constant rate r; when OFF, no fluid arrives. The ON and OFF periods are random, and when the ON periods are light-tailed, we get the nice light-tailed results; when the ON periods are heavy-tailed we get the bad kind of behavior as in the queueing models. Allowing for the superposition of more than one independent such ON/OFF input leads to more complex models.

In our next lecture (Part II), we will go into heavy-tailed distributions in more detail, with more mathematics, more probability theory and more precise statements of results. See you then!