Fluid Queues and Their Applications in Telecommunications

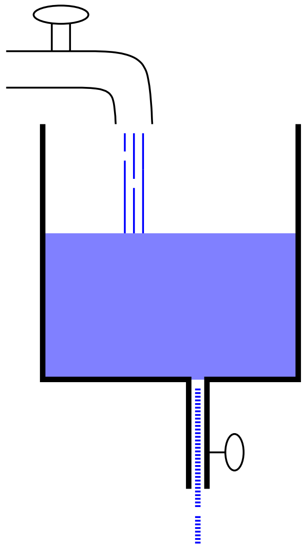
S-38.215 Special Course in Networking Technology, Spring 2002

Prof. Jacques Resing

Date: 22nd March 2002

Contents

1 Introduction



Notation: $Z(t)$ = content of the buffer at time t.

From the figures one sees that in fluid queue the inflow and outflow are "gradual", unlike with the ordinary M/M/1-queue.

Characteristics:

- Finite or infinite fluid buffer
- Inflow/outflow is regulated by some underlying stochastic process

Measures of interest are

- buffer content distribution
- average buffer content
- overflow probability (finite buffer)
- output process

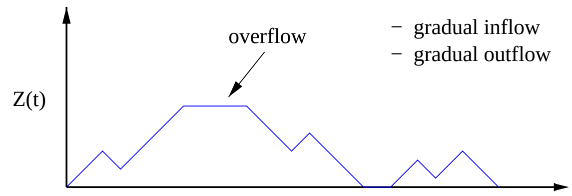


Figure 1: Typical fluid queue realization.

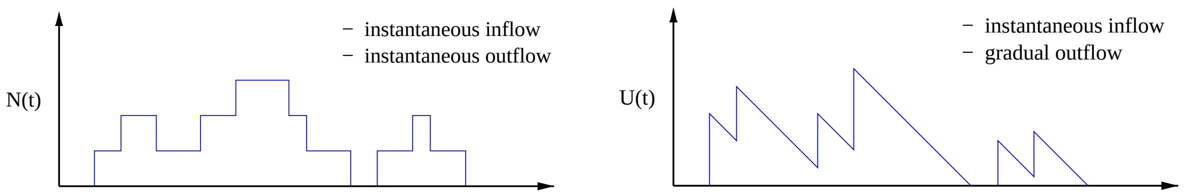


Figure 2: Typical realizations of occupancy and unfinished work in M/M/1 queue.

Applications:

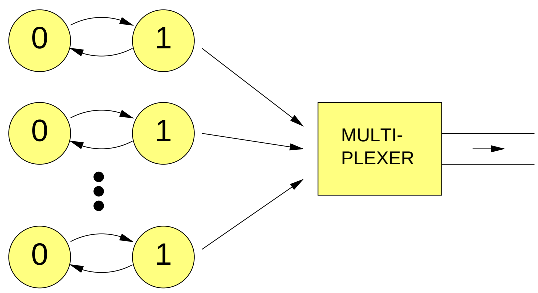
- Real fluid systems
- Systems with small entities, e.g. packets in communication systems, having (almost) deterministic processing times

History:

- Water reservoirs / dams (1960's)
- Motorway traffic (Newell, 1970's)
- Production systems
- Communication systems in burst level (Anick-Mitra-Sondhi, 1982)

Anick-Mitra-Sondhi Model:

- N users which are ON/OFF
- Inflow is proportional to the number of active users



2 Mathematical Background

2.1 Differential Equations

Linear differential equation system is

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t), \tag{1}$$

where $\mathbf{A}_{k\times k}$ is a constant matrix and $\mathbf{x}(t)$ are the unknown functions.

Theorem 1 *Let $\lambda_1,\ldots,\lambda_k$ be all the eigenvalues of $\mathbf{A}$ and let $\mathbf{v}_1,\ldots,\mathbf{v}_n$ be the corresponding eigenvectors. If $\lambda_i \neq \lambda_j$ for all $i \neq j$, then the general solution to* (1) *is,*

$$\mathbf{x}(t) = c_1 \cdot \mathbf{v}_1 e^{\lambda_1 t} + \ldots + c_k \cdot \mathbf{v}_k e^{\lambda_k t}.$$

2.2 Phase Type Distributions

distribution	$F(x)$	$f(x)$	$\mathrm{E}[X]$	$\mathrm{V}[X]$	$c_X^2 = \frac{\mathrm{V}[X]}{\mathrm{E}[X]^2}$
Exponential	$1-e^{-\mu x}$	$\mu e^{-\mu x}$	$1/\mu$	$1/\mu^2$	1
Erlang-r (or Gamma)	-	$\frac{\mu(\mu x)^{r-1}}{(r-1)!}e^{-\mu x}$	r/μ	r/μ^2	$1/r < 1$
Hyperexponential	$\sum_i p_i F_i(x)$	$\sum_i p_i \mu_i e^{-\mu_i x}$	-	-	> 1

All three distributions above are examples of so called *phase type distributions*. Phase type distributions are characterized by:

- initial distribution $(p_1, p_2, \ldots, p_N)$
- exponential residence times in states with parameters $\mu_1, \mu_2, \ldots, \mu_N$
- transient[1] transition probability matrix $\mathbf{P}$

Def 1 (phase type) *Random variable X is phase type if it is the residence time of the Markov process described above.*

2.3 Renewal Theory

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with pdf $f_X(x)$, and let t be a random point of time. In a typical example X_i corresponds to the lifetime of a light bulb.

We have random variables:

- *Curr*: total lifetime,
- *Res*: residual lifetime,
- *Elaps*: elapsed lifetime.

Q: What are pdf's of *Curr*, *Res* and *Elaps*?

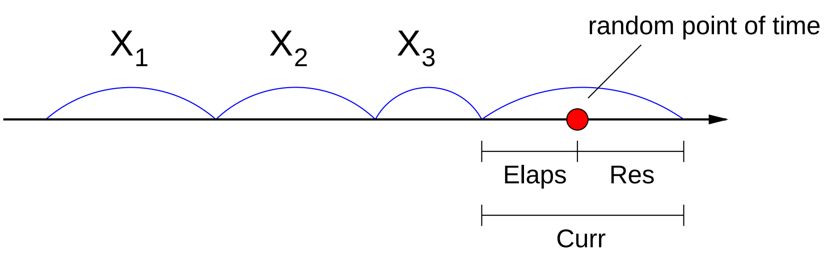


Figure 3: Sequence of i.i.d. random variables.

Results from renewal theory:

$$f_{\mathrm{Curr}}(x) = \frac{x \cdot f_X(x)}{\mathrm{E}[X]} \qquad \Rightarrow \qquad \mathrm{E}[\mathrm{Curr}] = \frac{\mathrm{E}[X^2]}{\mathrm{E}[X]}. \tag{2}$$

[1] $\lim_{n\to\infty} \mathbf{P}^n = 0$

$$f_{\text{Elaps}}(x) = f_{\text{Res}}(x) = \int_x^\infty \frac{f_{\text{Curr}}(y)}{y} dy = \frac{1 - F_X(x)}{\mathrm{E}[X]} \qquad \Rightarrow \qquad \mathrm{E}[\text{Elaps}] = \mathrm{E}[\text{Res}] = \frac{\mathrm{E}\left[X^2\right]}{2\mathrm{E}[X]}. \tag{3}$$

2.4 Queueing Theory

M/G/1 queue:

- Interarrival times are exponentially distributed with λ, and $A_i \sim \text{Exp}(\lambda)$
- Service times S_i are i.i.d. with some mean $\mathrm{E}[S]$ and the second moment $\mathrm{E}\left[S^2\right]$
- Single server capable of doing one unit of work per unit time
- Infinite number of waiting places
- Stability: $\rho := \lambda \mathrm{E}[S] < 1$

(Little's theorem: $\overline{N} = \lambda \overline{T}$. average occupancy = arrival intensity times average sojourn time)

Laplace-Stieltjes transform: $\tilde{S}(s) = \mathrm{E}\left[e^{-sS}\right]$.

2.4.1 Pollaczek-Khinchine Formulas

- Number of customers in the system, N

$$p_N(z) = \sum_n \mathrm{P}\{N = n\} \cdot z^n = \frac{(1-\rho)\tilde{S}(\lambda(1-z))(1-z)}{\tilde{S}(\lambda(1-z)) - z}.$$

- Waiting time, W

$$\tilde{W}(s) = \mathrm{E}\left[e^{-sW}\right] = \frac{(1-\rho)s}{\lambda\tilde{S}(s) + s - \lambda}.$$

- Sojourn time, $T = W + S$

$$\tilde{T}(s) = \mathrm{E}\left[e^{-sT}\right] = \frac{(1-\rho)s\tilde{S}(s)}{\lambda\tilde{S}(s) + s - \lambda}.$$

2.4.2 Mean Value Analysis for M/G/1

Often the mean values are enough. The mean waiting time becomes

$$\mathrm{E}[W] = \mathrm{E}\left[N_q\right] \cdot \mathrm{E}[S] + \rho \mathrm{E}[\text{Res}] = \frac{\rho \mathrm{E}[\text{Res}]}{1-\rho} = \frac{\rho}{1-\rho} \frac{\mathrm{E}\left[S^2\right]}{2\mathrm{E}[S]} = \frac{\lambda \mathrm{E}\left[S^2\right]}{2(1-\rho)}, \tag{4}$$

and the average sojourn time is

$$\mathrm{E}[T] = \mathrm{E}[W] + \mathrm{E}[S], \tag{5}$$

and finally by using Little's theorem for the queue length N_q gives,

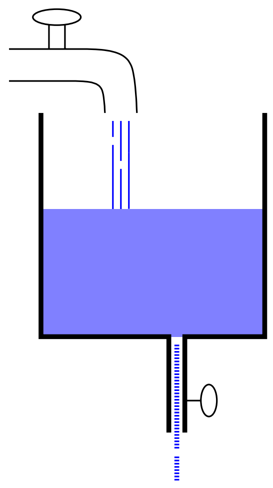
$$\mathrm{E}\left[N_q\right] = \lambda \cdot \mathrm{E}[W]. \tag{6}$$

3 Basic Fluid Model

Assumptions:

- Infinite buffer
- Constant outflow with rate equal to 1
- Inflow regulated by an ON/OFF–source with Exp(μ)-distributed ON-times, and Exp(λ)-distributed OFF-times ($\mu > \lambda$)
- During ON times inflow with rate equal to 2, and no inflow during OFF times

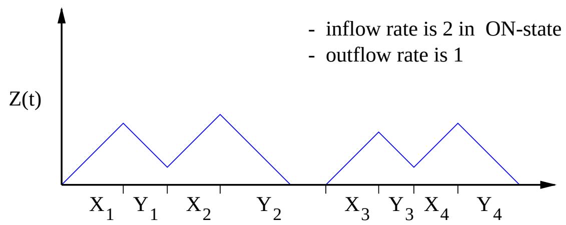
Performance measure: $\mathrm{P}\{Z(t) \leq x\}$ "in steady-state" or "in the long run".



3.1 Continuous Time Approach

First note that $Z(t)$ alone is not a Markov process ("direction" matters). However, two-dimensional process $(Z(t), I(t))$, where $I(t)$ is the state of the source, is a Markov process with state space $[0,\infty) \times \{0,1\}$.
Denote,

$$F_i(t,x) := \mathrm{P}\{Z(t) \leq x,\, I(t) = i\}.$$



Then one interesting measure is the steady state distribution, i.e. what happens when $t \to \infty$. From Fig. one gets for small interval Δ, that

$$F_0(t+\Delta,x) = F_0(t,x+\Delta)\cdot(1-\lambda\Delta+O(\Delta^2)) + F_1(t,x+O(\Delta))(\mu\Delta+O(\Delta^2)), \tag{7}$$
$$F_1(t+\Delta,x) = F_1(t,x-\Delta)\cdot(1-\mu\Delta+O(\Delta^2)) + F_0(t,x+O(\Delta))(\lambda\Delta+O(\Delta^2)). \tag{8}$$

From (7) one gets

$$\frac{F_0(t+\Delta,x)-F_0(t,x)}{\Delta} + \frac{F_0(t,x)-F_0(t,x+\Delta)}{\Delta} = -\lambda F_0(t,x+\Delta) + \mu F_1(t,x+O(\Delta)) + O(\Delta)),$$

and as $\Delta \to 0$,

$$\frac{\partial}{\partial t}F_0(t,x) - \frac{\partial}{\partial x}F_0(t,x) = -\lambda F_0(t,x) + \mu F_1(t,x).$$

By doing a similar derivation for (8) one finally gets,

$$\frac{\partial}{\partial t}F_0(t,x) - \frac{\partial}{\partial x}F_0(t,x) = -\lambda F_0(t,x) + \mu F_1(t,x), \tag{9}$$
$$\frac{\partial}{\partial t}F_1(t,x) + \frac{\partial}{\partial x}F_1(t,x) = \lambda F_0(t,x) - \mu F_1(t,x). \tag{10}$$

3.1.1 Steady-state Distribution

It is obvious from the situation that steady-state distribution exists, i.e. there is $F_i(x)$ such that,

$$\lim_{t\to\infty} F_i(t,x) \to F_i(x).$$

In the steady-state the time derivative is zero, i.e.

$$\frac{\partial}{\partial t}F_i(t,x) \to 0, \quad \text{as} \quad t\to\infty.$$

Similarly,

$$\frac{\partial}{\partial x}F_i(t,x) \to F_i'(x), \quad \text{as} \quad t\to\infty.$$

Thus at the limit $t\to\infty$ one gets

$$\begin{aligned} -F_0'(x) &= -\lambda F_0(x) + \mu F_1(x), \\ F_1'(x) &= \lambda F_0(x) - \mu F_1(x), \end{aligned}$$

which in the matrix form becomes,

$$\mathbf{F}'(x) = \mathbf{A}\mathbf{F}(x), \qquad \text{where } \mathbf{F}(x) = \begin{pmatrix} F_0(x) \\ F_1(x) \end{pmatrix} \text{ and } \mathbf{A} = \begin{pmatrix} \lambda & -\mu \\ \lambda & -\mu \end{pmatrix}. \tag{11}$$

Theorem 1 gives the general solution for (11). Eigenvalues and corresponding eigenvectors of $\mathbf{A}$ are

$$\lambda_1 = 0,\ \lambda_2 = \lambda - \mu \quad \text{and} \quad \mathbf{v}_1 = \begin{pmatrix} \mu \\ \lambda \end{pmatrix},\ \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and thus the general solution for (11) is

$$\mathbf{F}(x) = c_1 \begin{pmatrix} \mu \\ \lambda \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-(\mu-\lambda)x}.$$

Constants c_1 and c_2 can be obtained from the boundary conditions:

- When $x\to\infty$ one gets

$$\lim_{x\to\infty} \mathbf{F}(x) = \begin{pmatrix} \mathrm{P}\{I=0\} \\ \mathrm{P}\{I=1\} \end{pmatrix} = \frac{1}{\lambda+\mu}\begin{pmatrix} \mu \\ \lambda \end{pmatrix}, \quad \text{and thus} \quad c_1 = 1/(\lambda+\mu).$$

- When $x=0$ the second component of $\mathbf{F}(0)$ must be zero, i.e. $F_1(0)=0$, because when $I(t)=1$ the buffer content immediately grows bigger than zero. Hence,

$$\mathbf{F}(0) = \begin{pmatrix} F_0(0) \\ 0 \end{pmatrix} = \frac{1}{\lambda+\mu}\begin{pmatrix} \mu \\ \lambda \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and hence} \quad c_2 = -\frac{\lambda}{\lambda+\mu}.$$

Finally the cumulative steady-state distribution becomes,

$$\mathbf{F}(x) = \frac{1}{\lambda+\mu}\begin{pmatrix} \mu \\ \lambda \end{pmatrix} - \frac{\lambda}{\lambda+\mu}\begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-(\mu-\lambda)x}. \tag{12}$$

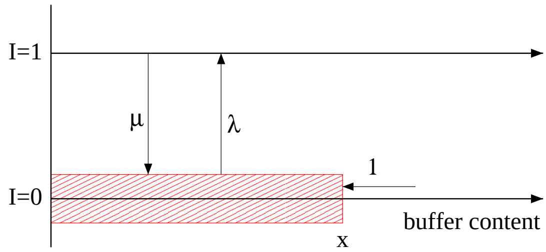
Especially,

$$\mathrm{P}\{Z\le x\} = F_0(x) + F_1(x) = \frac{\lambda+\mu-2\lambda e^{-(\mu-\lambda)x}}{\lambda+\mu}.$$

The average outflow from the system is, $1\cdot\left(1-\frac{\mu-\lambda}{\lambda+\mu}\right) = \frac{2\lambda}{\mu+\lambda}$, i.e. equal to the inflow as it should be.

3.1.2 Probability Flows

The set of differential equations for the steady-state can be also determined from the Fig. on the right. Consider the "probability flows" leaving and entering the red rectangle, i.e. $S = [0,x) \times \{0\}$, depicted in the figure. Clearly the flow leaving the subset S during a short time interval Δ is equal to,



$$F_0(x) \cdot \lambda\Delta.$$

Similarly, the probability flow entering the subset S must be

$$F_1(x) \cdot \mu\Delta + F_0(x+\Delta) - F_0(x) \approx F_1(x) \cdot \mu\Delta + F_0'(x) \cdot \Delta.$$

In the steady-state the flow leaving and entering S must be equal, so combining the above gives,

$$\lambda F_0(x) = \mu F_1(x) + F_0'(x),$$

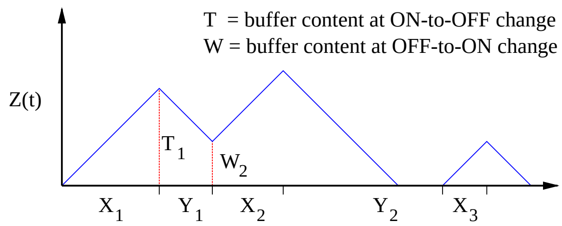
which is equivalent to (11). Similar reasoning can be conducted for the subset $[0,x) \times \{1\}$.

3.2 Discrete Time Approach

Definitions:

W_n: the buffer content $Z(t)$ at the time of the nth OFF $\rightarrow$ ON switch

T_n: the buffer content $Z(t)$ at the time of the nth ON $\rightarrow$ OFF switch



Considering W_n first gives,

$$\begin{aligned} W_1 &= 0 \\ W_{n+1} &= \max\{W_n + X_n - Y_n,\ 0\} \end{aligned}$$

which is *Lindley's equation* for the waiting time in an ordinary queueing system with service time X_i and interarrival times Y_i. As both X_i's and Y_i's are exponentially distributed this corresponds to the waiting time of M/M/1-queue, and

$$\mathrm{P}\{W \le x\} = (1 - \frac{\lambda}{\mu}) + \frac{\lambda}{\mu}(1 - e^{-(\mu-\lambda)x}).$$

For T_n:s one notices that $T_{n+1} = W_n + X_n =$ sojourn time for the same M/M/1-queue, and thus

$$\mathrm{P}\{T \le x\} = 1 - e^{-(\mu-\lambda)x}.$$

Consider now an arbitrary point of time, t_0:

- With probability of $\lambda/(\lambda+\mu)$ the source is in ON-state, and with probability of $\mu/(\lambda+\mu)$ the source is in OFF-state.

- If the source is in ON-state at t_0, then applying the renewal theory gives,

$$\text{buffer content} \stackrel{d}{=} W + \text{Elaps}_X \stackrel{d}{=} W + X \stackrel{d}{=} T.$$

- Similarly, if the source is in OFF-state at t_0, then

$$\text{buffer content} \stackrel{d}{=} \max\{T - \text{Elaps}_Y, 0\} \stackrel{d}{=} \max\{T - Y, 0\} \stackrel{d}{=} W.$$

Thus one gets,

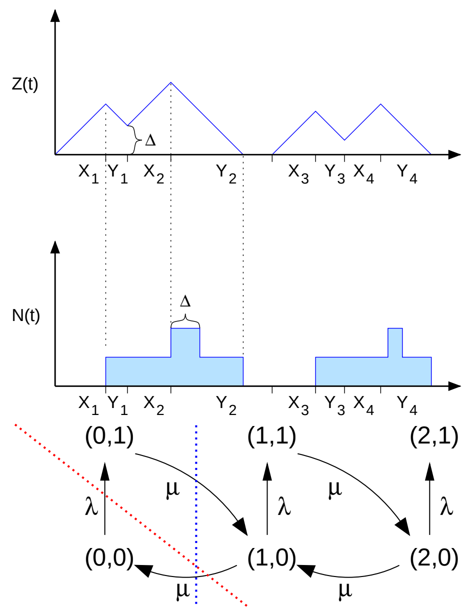
$$F_0(x) = \frac{\mu}{\lambda+\mu} \cdot \mathrm{P}\{W \le x\},$$
$$F_1(x) = \frac{\lambda}{\lambda+\mu} \cdot \mathrm{P}\{T \le x\},$$

which are equivalent to (12).

3.3 Stochastic Discretization Approach

At the end of ON period add extra portion (bag) of fluid (exp(μ)).
During OFF period remove bags if they become empty.
Study $(N(t), I(t))$ instead of $(Z(t), I(t))$, and do as if it is a Markov Process (which is not the case).
Flow diagram of $(N(t), I(t))$ is depicted on the right.
Let $p(n,i)$ be limiting probabilities of the above. Then,

(i) $\lambda \cdot p(n,0) = \mu p(n+1,0), \quad \forall\, n \ge 0,$
(ii) $\mu \cdot p(n,1) = \mu p(n+1,1), \quad \forall\, n \ge 0,$



$$\Rightarrow \quad p(n,0) = \left(\frac{\lambda}{\mu}\right)^n p(0,0), \quad p(n,1) = \left(\frac{\lambda}{\mu}\right)^{n+1} p(0,0).$$

$$\Rightarrow \qquad p(0,0) \cdot \left(\sum_n a^n + \sum_n a^{n+1}\right) = 1, \text{ and after some manipulation,} \quad p(0,0) = \frac{\mu-\lambda}{\mu+\lambda}.$$

Then,

$$\left\{ \begin{array}{r} (N(t), I(t)), \\ p(n,i) \end{array} \right. \qquad \Rightarrow \qquad (Z(t), I(t)).$$

$$\begin{array}{ll} F_0(x) & = \mathrm{P}\{Z(t) \le x,\ I(t) = 0\} \\ F_1(x) & = \mathrm{P}\{Z(t) \le x,\ I(t) = 1\} \end{array} \quad \Rightarrow \quad \begin{array}{ll} F_0(x) & = \sum_{n=0}^{\infty} \overbrace{p(n,0)}^{I=0} \cdot \mathrm{P}\{\overbrace{\text{ sum of } n \text{ Exp}(\mu)}^{\text{Erlang}_n(\mu)} \le x\} \\ F_1(x) & = \sum_{n=0}^{\infty} p(n,1) \cdot \mathrm{P}\{\text{Erlang}_{n+1}(\mu) \le x\} \end{array}$$

where $n+1$ in the lower Eq. comes from the added current "bag".

$$F_1(x) = \sum_{n=0}^{\infty} \frac{\mu-\lambda}{\mu+\lambda} \left(\frac{\lambda}{\mu}\right)^{n+1} \cdot \mathrm{P}\{\text{Erlang}_{n+1}(\mu) \le x\}$$
$$= \frac{\lambda}{\mu+\lambda} \underbrace{\sum_{n=0}^{\infty} \frac{\mu-\lambda}{\mu} \left(\frac{\lambda}{\mu}\right)^{n} \cdot \mathrm{P}\{\text{Erlang}_{n+1}(\mu) \le x\}}_{\text{sojourn time in M/M/1-queue}} = \frac{\lambda}{\mu+\lambda}\left(1 - e^{-(\mu-\lambda)x}\right).$$

Similarly,

$$F_0(x) = \sum_{n=0}^{\infty} \frac{\mu-\lambda}{\mu+\lambda}\left(\frac{\lambda}{\mu}\right)^n \cdot \mathrm{P}\{\mathrm{Erlang}_n(\mu) \le x\}$$
$$= \frac{\mu}{\mu+\lambda} \underbrace{\sum_{n=0}^{\infty} \frac{\mu-\lambda}{\mu}\left(\frac{\lambda}{\mu}\right)^n \cdot \mathrm{P}\{\mathrm{Erlang}_n(\mu) \le x\}}_{\text{waiting time in M/M/1-queue}} = \frac{\mu}{\mu+\lambda}\left(1 - \frac{\lambda}{\mu} + \frac{\lambda}{\mu}(1-e^{-(\mu-\lambda)x})\right).$$

3.3.1 Summary so far

Basic Model:

- infinite buffer size
- constant outflow 1
- inflow alternates between 0 and 2, distributed with $\mathrm{Exp}(\lambda)$ and $\mathrm{Exp}(\mu)$ respectively

What directions can the model be extended? Different Approaches and their Usage:

I. – finite/infinite buffer model
   – process regulating the in/out flow can be any finite state Markov process

II. – useful for non-Markovian processes regulating the inflow of fluid buffer (e.g. $X_1, X_2, \ldots$ are arbitrary r.v.:s, $Y_1, Y_2, \ldots$ exponentially distributed r.v.:s), (M/G/1 formula)

III. – useful with more than one fluid buffer

3.4 General Model for Approach I

Let:

- $u(t)$ be a finite state Markov process regulating both inflow and outflow:
- **Q**: infinitesimal generator of $u(t)$
- If $u(t) = i$, then the net flow (inflow-outflow) equals to d_i
- π: stationary distribution of $u(t)$
- $\sum_i \pi_i d_i < 0$ for stability
- Technical assumption 1: $d_i \neq 0$ for all i
- $\mathbf{D} = \begin{pmatrix} d_0 & & & 0 \\ & d_1 & & \\ & & \ddots & \\ 0 & & & d_n \end{pmatrix}$ and $F_i(x) = \lim_{t\to\infty} \mathrm{P}\{Z(t) \le x,\ u(t) = i\}$

These give,

$$\mathbf{D} \cdot \mathbf{F}'(x) = \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{F}(x), \text{ where } \mathbf{F}(x) = \begin{pmatrix} F_0(x) \\ F_1(x) \\ \vdots \\ F_n(x) \end{pmatrix}.$$

For the basic model,
$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} F_0'(x) \\ F_1'(x) \end{pmatrix} = \begin{pmatrix} -\lambda & \mu \\ \lambda & -\mu \end{pmatrix} \cdot \begin{pmatrix} F_0(x) \\ F_1(x) \end{pmatrix}.$$

Technical Assumption 2: *All eigenvalues* $\lambda_1, \ldots, \lambda_{n+1}$ *are different.*

With these assumptions the general solution is,

$$\mathbf{F}(x) = \sum_{i=1}^{n+1} c_i \cdot \mathbf{v}_i \cdot e^{\lambda_i x}.$$

Boundary conditions (here we choose $\lambda_1 = 0$):

- Infinite buffer case:
  - From $\lim_{x\to\infty} F_i(x) = \pi_i$ we can find $c_i = 0$, if $\mathrm{Re}(\lambda_i) > 0$, and c_1 follows from $\lim_{x\to\infty} F_i(x) = \pi_i$
  - And $\lim_{x\to 0} F_i(x) = 0$ if $d_i > 0$, gives the remaining constants
- Finite buffer case:
  - $\lim_{x\to 0} F_i(x) = 0$, if $d_i > 0$
  - $\lim_{x\to K} F_i(x) = \pi_i$, if $d_i < 0$

4 Applications to Communication Systems

In this section the following applications will be studied:

I Traffic Differentiation

II Traffic Shaping

III TCP Source

4.1 Traffic Differentiation (space priority)

- Source producing two types of fluid (e.g. voice and data)
- Both types of fluid are multiplexed in a single finite buffer of size K
- Source is regulated by Markov Process $u(t)$ with state space $\{0, \ldots, N\}$, infinitesimal generator matrix $\mathbf{Q}$ and limiting distribution π
- If $u(t) = i$ the source produces type j fluid with rate $r_{i,j}$
- Buffer sharing policy: accept type 2 fluid only if $Z(t) < K^*$, i.e. type 1 fluid is more important ($K^* < K$)
- Constant output rate c
- Net input:

$$\begin{aligned} d_i^{(1)} &= r_{i1} + r_{i2} - c, \quad && \text{when } Z(t) < K^* \\ d_i^{(2)} &= r_{i1} - c, && \text{when } Z(t) > K^* \end{aligned}$$

- Diagonal matrices:

$$\mathbf{D}^{(1)} = \begin{pmatrix} d_0^{(1)} & & & 0 \\ & d_1^{(1)} & & \\ & & \ddots & \\ 0 & & & d_n^{(1)} \end{pmatrix} \quad \text{and} \quad \mathbf{D}^{(2)} = \begin{pmatrix} d_0^{(2)} & & & 0 \\ & d_1^{(2)} & & \\ & & \ddots & \\ 0 & & & d_n^{(2)} \end{pmatrix}.$$

Foe,

$$F_i(x) = \lim_{t\to\infty} \mathrm{P}\{Z(t) \le x,\ u(t) = i\},$$

one obtains the following system of differential equations for $F_i(x)$:

$$\mathbf{D}^{(1)}\mathbf{F}'(x) = \mathbf{Q}^{\mathrm{T}}\mathbf{F}(x), \qquad 0 < x < K^* \tag{13}$$

$$\mathbf{D}^{(2)}\mathbf{F}'(x) = \mathbf{Q}^{\mathrm{T}}\mathbf{F}(x), \qquad K^* < x < K \tag{14}$$

From (13) alone: $\mathbf{F}^{(1)}(x) = \sum_{j=0}^{N} c_j^{(1)} \cdot \mathbf{v}_j^{(1)} e^{\lambda_j^{(1)} x}$, and from (14) alone: $\mathbf{F}^{(2)}(x) = \sum_{j=0}^{N} c_j^{(2)} \cdot \mathbf{v}_j^{(2)} e^{\lambda_j^{(2)} x}$.

Proposition 1 *The general solution for the above differential equation system is*

$$\mathbf{F}(x) = \begin{cases} \mathbf{F}^{(1)}(x) & \text{when } 0 < x < K^*, \\ \mathbf{F}^{(2)}(x) & \text{when } K^* < x < K. \end{cases}$$

How to determine $2(N+1)$ constants? $(c_j^{(1)}, c_j^{(2)},\ j = 0,\dots,N)$

Split the state space $S = \{0,\dots,N\}$ to three subsets $S^-, S^\pm, S^+$:

$$S^- = \{i \in S\ :\ r_{i1} + r_{i2} - c < 0\}$$
$$S^\pm = \{i \in S\ :\ r_{i1} + r_{i2} - c > 0,\ r_{i1} - c < 0\}$$
$$S^+ = \{i \in S\ :\ r_{i1} - c > 0\}$$

Boundary Conditions:

i) $F_i(0) = 0$ if $d_i^{(1)} > 0$, i.e. if $i \in S^\pm \cup S^+$

ii) $F_i(K^{*-}) = F_i(K^{*+})$ if $i \in S^- \cup S^+$

iii) $F_i(K) = \pi_i$ if $d_i^{(2)} < 0$, i.e. if $i \in S^- \cup S^\pm$

In total $2(N+1)$ boundary conditions and unknown constants $c_i^{(j)}$ can be determined.

4.2 Traffic Shaping

Bursty traffic is bad for the network performance, hence traffic shapers can be used.

Basic types:

i) Spacer

ii) Leaky Bucket

iii) 2-level Shaper

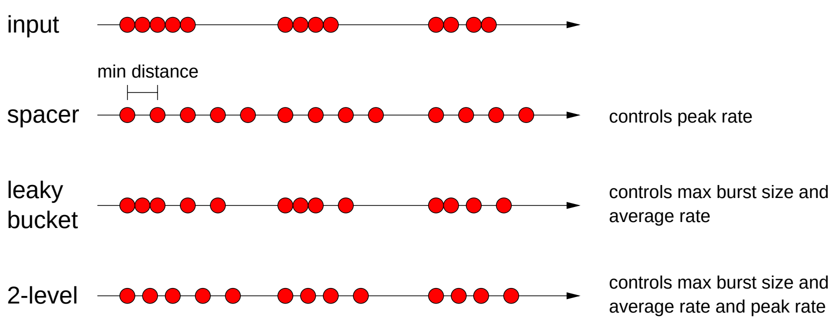


Figure 4: Behaviour of the different traffic shapers.

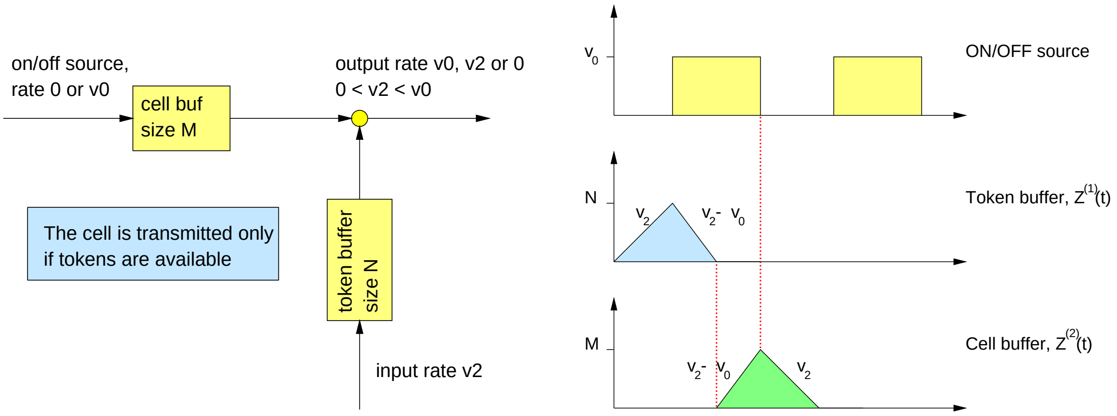
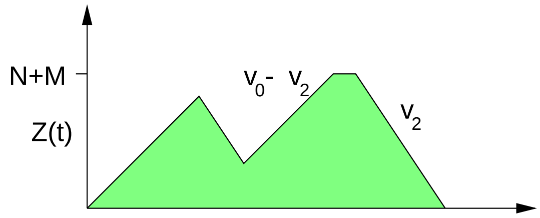


Figure 5: Fluid model for leaky bucket traffic shaper.

4.2.1 Fluid Model for Leaky Bucket

As either buffer is always empty in leaky-bucket shaper, the system can be reduced to one dimension: $Z(t) = Z^{(2)}(t) - Z^{(1)}(t) + N$ and we get a basic fluid model with input rate 0 or v_0 and output rate v_2.

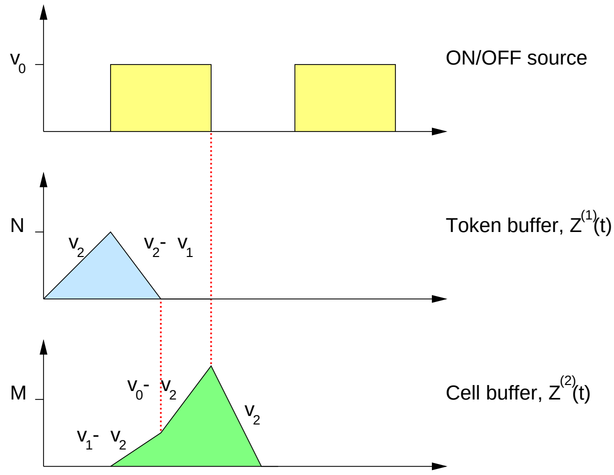


4.2.2 2-level Shaper

For a 2-level shaper the output rate equals to v_1, v_2 or 0, where

$$0 < \overbrace{v_2}^{\text{average rate}} < \overbrace{v_1}^{\text{peak rate}} < v_0.$$

As can be seen from the Figure, both buffers can be non-empty at the same time and the process cannot be reduced to one dimension.
Adan and Resing discretize one of the fluid buffers using the stochastic discretization technique, and the system of pde's becomes a system of ode's which can be solved.



4.3 TCP Source

$u(t)$ = state of a TCP source, if $u(t) = i$ then the output rate is $r \cdot i$.

Buffer sends positive/negative feedback signals depending on the buffer content $Z(t)$.

$Z(t) < K$	Positive feedback signals, $u(t)$ increases by one Positive signals occur at rate λ
$Z(t) = K$	Negative feedback signals, $u(t)$ increases by factor 2: $u(t) \leftarrow \lfloor u(t)/2 \rfloor$ Negative signals occur at rate μ.

"Feedback fluid system", $u(t)$ regulates $Z(t)$, but also $Z(t)$ regulates $u(t)$!

Example: TCP source with $N = 5$ and $2r < c < 3r$

Notation: $d_i = r \cdot i - c$, and

$\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_5)$.

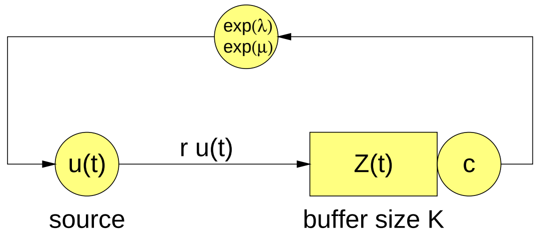


Figure 6: Diagram of the TCP source model.

States: $S^- = \{i \,:\, d_i < 0\}$ and $S^+ = \{i \,:\, d_i > 0\}$. (again, $d_i \neq 0$ for all i).

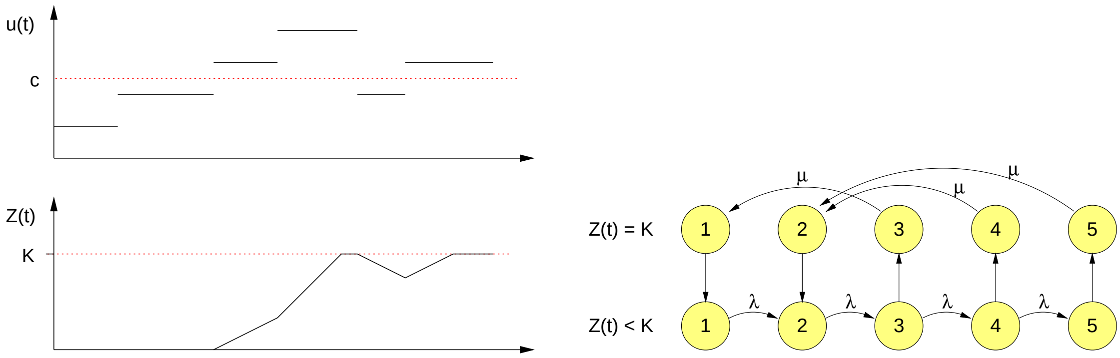


Figure 7: Example realization of the TCP source (left) and the state space of the system (right).

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & -\lambda & \lambda & \\ & & & -\lambda & \lambda \\ & & & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \tilde{\mathbf{Q}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \mu & 0 & -\mu & 0 & 0 \\ 0 & \mu & 0 & -\mu & 0 \\ 0 & \mu & 0 & 0 & -\mu \end{pmatrix}.$$

System of differential equations:

$$\mathbf{D} \cdot \mathbf{F}'(x) = \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{F}(x) \qquad \text{(like ordinary fluid queue)}$$

as long as $Z(t) < K$. The general solution,

$$\mathbf{F}(x) = \sum_{i=1}^{N} c_i \cdot \mathbf{v}_i \cdot e^{\lambda_i x}.$$

Normally the boundary conditions are,

$$F_i(0) = 0, \text{ if } i \in S^+ \text{ and } F_i(K^-) = F_i(K), \text{ if } i \in S^-.$$

But what are boundary conditions in this case? (ordinary $F_i(K) = \pi_i$)

Define $G_i = F_i(K) - F_i(K^-)$, i.e. $G_i = \lim_{t\to\infty} \mathrm{P}\{Z(t) = K, u(t) = i\}$.

For example for $i = 4$ one then gets,

$$\begin{aligned} \mathrm{P}\{Z(t+\Delta t) = K, u(t+\Delta t) = 4\} =& \mathrm{P}\{Z(t) = K, u(t) = 4\} \cdot (1 - \mu\Delta t) \\ &+ \mathrm{P}\{Z(t) \in (K - d_4 \cdot \Delta t, K), u(t) = 4\} \cdot (1 - \lambda\Delta t) \end{aligned}$$

and as Δt goes to zero,

$$0 = -\mu G_4 + d_4 F_4'(K^-).$$

Repeating the same steps for each i, it turns out that the boundary conditions are of form:

$$\tilde{\mathbf{Q}}^{\mathrm{T}} \cdot \mathbf{G} + \mathbf{D} \cdot \mathbf{F}'(K^-) = 0.$$

Alternatively this can be written as,

$$\tilde{\mathbf{Q}}^{\mathrm{T}} \cdot \mathbf{G} + \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{F}(K^-) = 0.$$

In total we have $2N$ unknowns: $\{c_i\}, \{G_i\}$.

Boundary conditions:

i)	$F_i(0) = 0, \quad \text{when } i \in S^+$ $G_i = 0, \quad \text{when } i \in S^-$	gives N conditions in total
ii)	$\tilde{\mathbf{Q}}^{\mathrm{T}} \cdot \mathbf{G} + \mathbf{D} \cdot \mathbf{F}'(K^-) = 0,$	gives another N conditions but one depends on the other $\Rightarrow$ $N-1$ conditions
iii)	$\sum_{i=1}^{N} \underbrace{F_i(K^-)}_{\text{not full}} + \underbrace{G_i}_{\text{full}} = 1,$	i.e. normalization

Hence, we have in total $2N$ equations and the unknown constants can be solved.

5 Fluid Models and Heavy Tails

Motivation: file sizes in the Internet have heavy tails, i.e.

$$\mathrm{P}\{X > t\} \approx C \cdot t^{-\nu}, \qquad \text{where } 1 < \nu < 2 \text{ typically.}$$

Question: what is the effect of heavy tailed file sizes to the buffer content or waiting times?

5.1 Model to be studied

We study the basic fluid model with the following exceptions:

- ON-periods $X_1, X_2, \ldots$ are heavy tailed
- OFF-periods $Y_1, Y_2, \ldots$ are still exponentially distributed

The basic fluid model was solved with 3 different approach:

I used the fact that everything was exponential,

III also used the fact that everything was exponential.

Hence we are left with the approach II, i.e. the discrete time analysis using the results from the queueing theory.

5.2 Heavy-tailed Random Variable

There are several formulations for heavy tailedness. Here we use so called *regularly varying* random variables.

Def 2 (regularly varying) *A function $f:\mathbb{R}\to\mathbb{R}$ is said to be regularly varying with index α, if,*

$$\lim_{x\to\infty}\frac{f(xt)}{f(x)}=t^{\alpha}.$$

Def 3 *A random variable X is said to be regularly varying with index $-\nu$ if $G(x)=\mathrm{P}\{X>x\}$ is a regularly varying with index $-\nu$, and is denoted with $RV(-\nu)$.*

Regularly varying random variables have the following properties:

[RV1] If X is RV$(-\nu)$, then X_{Res}, X_{Elaps} and X_{Curr} are RV$(1-\nu)$, i.e. have even heavier tail.

[RV2] If X_1 and X_2 are both RV$(-\nu)$ and independent, then X_1+X_2 is also RV$(-\nu)$.

[RV3] If X_1 is RV$(-\nu_1)$ and X_2 is RV$(-\nu_2)$ and independent, then X_1+X_2 is RV$(\max\{-\nu_1,\,-\nu_2\})$.

From earlier we remember that the analogy was::

- X_i's were interarrival times, and
- Y_i's were service times

First we need the following important result from the queueing theory.

Theorem 2 *In M/G/1 queue, where service times are RV$(-\nu)$, the waiting time W is RV$(1-\nu)$.*

This means that the tails of the waiting times are even heavier than the tails of the service times.

Idea of the proof:

$$\begin{aligned}\text{P-K formula: } \tilde{W}(s) &= \tfrac{(1-\rho)s}{\lambda\tilde{S}(s)+s-\lambda} = (1-\rho)\frac{1}{1-\underbrace{\lambda\mathrm{E}[S]}_{=\rho}\underbrace{\dfrac{1-\tilde{S}(s)}{s\mathrm{E}[S]}}_{=\tilde{S}_{\text{Res}}(s)}}\\ &= (1-\rho)\tfrac{1}{1-\rho\tilde{S}_{\text{Res}}(s)} = (1-\rho)\textstyle\sum_{n=0}^{\infty}\rho^n\left[\tilde{S}_{\text{Res}}(s)\right]^n,\end{aligned}$$

and thus,

$$\mathrm{P}\{W>t\}=\sum_{n=0}^{\infty}(1-\rho)\rho^n\mathrm{P}\{S^{(1)}_{\text{Res}}+S^{(2)}_{\text{Res}}+\ldots+S^{(n)}_{\text{Res}}>t\}$$

and using [RV1] and [RV2] we get that W is RV$(1-\nu)$.

5.3 From M/G/1 to Fluid Model

For buffer content Z we have,

$$Z=\begin{cases} W+X_{\text{Elaps}}, & \text{w.p. } \frac{\mathrm{E}[X]}{\mathrm{E}[X]+\mathrm{E}[Y]},\\ \max\{T-Y_{\text{Elaps}},0\}=W, & \text{w.p. } \frac{\mathrm{E}[Y]}{\mathrm{E}[X]+\mathrm{E}[Y]}.\end{cases}$$

Both waiting time W and X_{Elaps} are RV$(1-\nu)$, and thus the buffer content Z must be RV$(1-\nu)$ as well.

References

[AMS82] D. Anick, D. Mitra, and M.M. Sondhi. Stochastic theory of a data handling systems with multiple sources. *The Bell System Technical Journal*, 61:1871–1894, 1982.

[AR96] I. Adan and J. Resing. A simple analysis of a fluid queue driven by an M/M/1 queue. *Queueing Systems 22*, pages 171–174, 1996.

[AR00] I. Adan and J. Resing. A two-level traffic shaper for an on-off source. *Performance Evaluation 42*, pages 279–298, 2000.

[BD98] O.J. Boxma and V. Dumas. Fluid queues with heavy-tailed activity period distributions. *Computer Communications 21*, pages 1509–1529, 1998.

[BGT87] N.H. Bingham, C. Goldie, and J. Teugels. *Regular Variation*. Cambridge University Press, Cambridge, UK, 1987.

[Box96] O.J. Boxma. Fluid queues and regular variation. *Performance Evaluation 27 & 28*, pages 699–712, 1996.

[Kle75] L. Kleinrock. *Queueing Systems, Vol. I: Theory*. Wiley, New York, 1975.

[Kul] V.G. Kulkarni. Lecture notes of a tutorial on fluid queues.

[Kul97] V.G. Kulkarni. Fluid models for single buffer systems. *In: Frontiers of Queueing Systems*, pages 321–339, 1997.

[KW92] O. Kella and W. Whitt. A storage model with a two-state random environment. *Operations Research 40*, pages S257–S262, 1992.

[Sch98] W. Scheinhardt. *Markov-modulated and feedback fluid queues*. Ph.d. thesis, University of Twente, 1998.

[vFMS01] N. van Foreest, M. Mandjes, and W. Scheinhardt. Analysis of a feedback fluid model for TCP with heterogeneous sources, 2001. University of Twente.

[Zwa01] A.P. Zwart. *Queueing Systems with Heavy Tails*. Ph.d. thesis, Eindhoven University of Technology, 2001.