Fluid models for ad hoc networks

Michel Mandjes

CWI, Amsterdam, the Netherlands

&

Korteweg-de Vries Institute for Mathematics, University of Amsterdam, the Netherlands

Workshop on Extremes and Fluid Queues

EURANDOM, March 13–14, 2006

Thanks to: Frank Roijers

Ad hoc communication networks

Traditional networks: there is a notion of *sources* (or: *flows*), and a notion of *queue*.

For instance: on-off sources feeding into a queue, drained at constant rate, say C.

Ad hoc communication networks

Traditional networks: there is a notion of *sources* (or: *flows*), and a notion of *queue*.

For instance: on-off sources feeding into a queue, drained at constant rate, say C.

Ad hoc network: no clear division between flows and queue.

If n flows present, then

- each flow can use $C/(n+1)$ to send traffic into the queue;
- queue is drained at rate $C/(n+1)$.

Coupled input and output!

Some remarks on stylized model

Assume

- the number of flows simultaneously in the system is smaller than $N_{\rm m}$;
- flows arrive according to a Possion process with rate λ;
- flow size (in 'bytes') is exponential with mean μ;
- buffer has unlimited capacity.

Some remarks on stylized model

Assume

- the number of flows simultaneously in the system is smaller than $N_{\rm m}$;
- flows arrive according to a Possion process with rate λ;
- flow size (in 'bytes') is exponential with mean μ;
- buffer has unlimited capacity.

Notice that queue only drains when there are no flows present

(i.e., during an exp(λ) distributed time).

Goals

Distribution of:

- Workload distribution;
- Queueing delay distribution;
- Flow transfer delay distribution
  (i.e., the time it takes for the flow to transmit its traffic into the buffer);
- Sojourn time distribution
  (i.e., flow transfer delay + queueing delay last particle).

Stability condition

N_t: number of flows present at time t; Markov chain on $\{0, \ldots, N_{\mathrm{m}}\}$; generator matrix

$$Q := \begin{pmatrix} -\lambda & \lambda & & & & \\ \mu_1 C & -\mu_1 C - \lambda & \lambda & & & \\ & \mu_2 C & -\mu_2 C - \lambda & \lambda & & \\ & & \ddots & \ddots & \ddots & \\ & & & & \mu_{N_{\mathrm{m}}} C & -\mu_{N_{\mathrm{m}}} C \end{pmatrix},$$

$\mu_n := \mu n/(n+1)$.

When $N_t = n$, input rate is $r_{I,n} := Cn/(n+1)$, output rate is $r_{O,n} := C/(n+1)$.

Net rate of change of the queue is 0 when $Q_t = N_t = 0$, and otherwise, for $n \in \{0, \ldots, N_{\mathrm{m}}\}$,

$$r_{A,n} := r_{I,n} - r_{O,n} = C\,\frac{n-1}{n+1}.$$

Define $R_I := \mathrm{diag}\{r_I\}$, $R_O := \mathrm{diag}\{r_O\}$, and $R_A := R_I - R_O$.

Stability condition, ctd.

Balance equations

$$\pi_n \mu_n C = \pi_{n-1} \lambda, \quad n = 1, \ldots, N_{\mathrm{m}}.$$

Lead to ($\varrho := \lambda/(C\mu)$):

$$\pi_n = \frac{\varrho^n (n+1)}{\sum_{k=0}^{N_{\mathrm{m}}} \varrho^k (k+1)}.$$

The equilibrium condition of the fluid model is $\sum_{n=0}^{N_{\mathrm{m}}} \pi_n r_{A,n} < 0$, or

$$\frac{-1 + 2\varrho - \varrho^{N_{\mathrm{m}}+1} N_{\mathrm{m}} + \varrho^{N_{\mathrm{m}}+2}(N_{\mathrm{m}} - 1)}{1 - \varrho^{N_{\mathrm{m}}+1}(N_{\mathrm{m}} + 2) + \varrho^{N_{\mathrm{m}}+2}(N_{\mathrm{m}} + 1)} \cdot C < 0.$$

Stability condition, ctd.

Balance equations

$$\pi_n \mu_n C = \pi_{n-1}\lambda, \quad n = 1, \dots, N_{\mathrm{m}}.$$

Lead to ($\varrho := \lambda/(C\mu)$):

$$\pi_n = \frac{\varrho^n (n+1)}{\sum_{k=0}^{N_{\mathrm{m}}} \varrho^k (k+1)}.$$

The equilibrium condition of the fluid model is $\sum_{n=0}^{N_{\mathrm{m}}} \pi_n r_{A,n} < 0$, or

$$\frac{-1 + 2\varrho - \varrho^{N_{\mathrm{m}}+1} N_{\mathrm{m}} + \varrho^{N_{\mathrm{m}}+2}(N_{\mathrm{m}} - 1)}{1 - \varrho^{N_{\mathrm{m}}+1}(N_{\mathrm{m}} + 2) + \varrho^{N_{\mathrm{m}}+2}(N_{\mathrm{m}} + 1)} \cdot C < 0.$$

Special case of $N_{\mathrm{m}} \to \infty$*:* $2\varrho < 1$.

Every flow has to be processed twice!

Workload distribution

Introduce: $B \stackrel{\mathrm{d}}{=} B_1$, with

$$B_n := \inf\{t \geq 0 : N_t = n - 1 \mid N_0 = n\}.$$

Also, with $A(s,t) := \int_s^t r_{A,N_u} \mathrm{d}u$, we have $T \stackrel{\mathrm{d}}{=} T_1$, with

$$T_n \stackrel{\mathrm{d}}{=} A(0, B_n);$$

Workload distribution

Introduce: $B \stackrel{\mathrm{d}}{=} B_1$, with

$$B_n := \inf\{t \geq 0 : N_t = n-1 \mid N_0 = n\}.$$

Also, with $A(s,t) := \int_s^t r_{A,N_u} \mathrm{d}u$, we have $T \stackrel{\mathrm{d}}{=} T_1$, with

$$T_n \stackrel{\mathrm{d}}{=} A(0, B_n);$$

Recursion! For $n = 1, \ldots, N_{\mathrm{m}} - 1$,

$$\mathbb{E}e^{-sT_n} = \frac{\lambda}{\lambda + \mu_n C + r_{A,n} s} \mathbb{E}e^{-sT_{n+1}} \mathbb{E}e^{-sT_n} + \frac{\mu_n C}{\lambda + \mu_n C + r_{A,n} s},$$

while for $n = N_{\mathrm{m}}$:

$$\mathbb{E}e^{-sT_{N_{\mathrm{m}}}} = \frac{\mu_{N_{\mathrm{m}}} C}{\mu_{N_{\mathrm{m}}} C + r_{A,N_{\mathrm{m}}} s}.$$

Can be solved iteratively!

Workload distribution, ctd.

Reich: $W^\star$ is steady-state workload,

$$W^\star \stackrel{\mathrm{d}}{=} M := \sup_{t\geq 0} A(-t,0) \stackrel{\mathrm{d}}{=} \sup_{t\geq 0} A(0,t),$$

(reversibility of $(N_t)_{t\in\mathbb{R}}$!)

Define

$$M_i := \sup_{t\geq 0}\{A(0,t) \mid N_0 = i\};$$

clearly $\mathbb{E}e^{-sM} = \sum_{n=0}^{N_{\rm m}} \pi_n \mathbb{E}e^{-sM_n}$;

hence we have to find $\mathbb{E}e^{-sM_n}$, for $n = 0,\ldots,N_{\rm m}$.

Workload distribution, ctd.

Well,

$$M_n \stackrel{\mathrm{d}}{=} T_n + T_{n-1} + \cdots T_1 + M_0,$$

with $B_n, B_{n-1}, \ldots, B_1, M_0$ independent.

Hence

$$\mathbb{E}e^{-sM_n} = \mathbb{E}e^{-sM_0} \cdot \prod_{i=1}^{n} \mathbb{E}e^{-sT_i};$$

Known: $\mathbb{E}e^{-sT_i}$ (previous slide).

Left: $\mathbb{E}e^{-sM_0}$.

Workload distribution, ctd.

Starting in 0, maximum $(t \geq 0)$ of $A(0,t)$ equals maximum $(i = 0, 1, \ldots)$ of $\sum_{j=0}^{i}(X_j - Y_j)$, with the $X_j =_{\rm d} T$, and the $Y_j =_{\rm d} \exp(C/\lambda)$.

Workload distribution, ctd.

Starting in 0, maximum $(t \geq 0)$ of $A(0,t)$ equals maximum $(i = 0, 1, \ldots)$ of $\sum_{j=0}^{i}(X_j - Y_j)$,

with the $X_j =_{\mathrm{d}} T$, and the $Y_j =_{\mathrm{d}} \exp(C/\lambda)$.

Pollaczek-Khinchine!

$$\mathbb{E}e^{-sM_0} = \left(1 - \frac{\lambda \mathbb{E}T}{C}\right) \frac{s}{s - (\lambda/C)(1 - \mathbb{E}e^{-sT})}.$$

Hence,

$$\mathbb{E}e^{-sW^\star} = \mathbb{E}e^{-sM} = \sum_{n=0}^{N_{\mathrm{m}}} \pi_n \left(1 - \frac{\lambda \mathbb{E}T}{C}\right) \frac{s}{s - (\lambda/C)(1 - \mathbb{E}e^{-sT})} \left(\prod_{i=1}^{n} \mathbb{E}e^{-sT_i}\right).$$

Workload distribution, ctd.

Some ramifications:

- *Joint* distribution workload $W^\star$ and number of flows $N^\star$:

$$\mathbb{E}(e^{-sW^\star}1\{N^\star = n\}) = \pi_n \left(1 - \frac{\lambda \mathbb{E}T}{C}\right) \frac{s}{s - (\lambda/C)(1 - \mathbb{E}e^{-sT})} \left(\prod_{i=1}^{n} \mathbb{E}e^{-sT_i}\right).$$

Workload distribution, ctd.

Some ramifications:

- *Joint* distribution workload $W^\star$ and number of flows $N^\star$:

$$\mathbb{E}(e^{-sW^\star}1\{N^\star = n\}) = \pi_n \left(1 - \frac{\lambda \mathbb{E}T}{C}\right) \frac{s}{s - (\lambda/C)(1 - \mathbb{E}e^{-sT})} \left(\prod_{i=1}^{n} \mathbb{E}e^{-sT_i}\right).$$

- *Mean workload*:

$$\mathbb{E}W^\star = \left(\frac{1}{2}\frac{\lambda \mathbb{E}T^2}{C - \lambda \mathbb{E}T}\right) + \left(\sum_{n=0}^{N_\mathrm{m}} \left(\pi_n \sum_{i=1}^{n} \mathbb{E}T_i\right)\right).$$

Queueing delay distribution

$O(0,t)$: *output* capacity available in the interval $[0,t)$. Then,

$$\begin{aligned}\mathbb{E}e^{-sD^\star} &= \int_0^\infty e^{-st}\mathbb{P}(D^\star = t)\mathrm{d}t = \int_0^\infty e^{-st}\mathbb{P}(W^\star = O(0,t))\mathrm{d}t \\ &= \sum_{n=0}^{N_\mathrm{m}} \int_0^\infty e^{-st}\mathbb{P}(W^\star = O(0,t), N^\star = n)\mathrm{d}t.\end{aligned}$$

Queueing delay distribution

$O(0,t)$: *output* capacity available in the interval $[0,t)$. Then,

$$\begin{aligned}\mathbb{E}e^{-sD^\star} &= \int_0^\infty e^{-st}\mathbb{P}(D^\star = t)\mathrm{d}t = \int_0^\infty e^{-st}\mathbb{P}(W^\star = O(0,t))\mathrm{d}t \\ &= \sum_{n=0}^{N_\mathrm{m}} \int_0^\infty e^{-st}\mathbb{P}(W^\star = O(0,t), N^\star = n)\mathrm{d}t.\end{aligned}$$

Now define, for $z \geq 0$,

$$\tau_z := \inf\{t \geq 0 : O(0,t) = z\} = \inf\left\{t \geq 0 : \int_0^t r_{O,N_s}\mathrm{d}s = z\right\};$$

notice that $O(0,t)$ is increasing in t.

Queueing delay distribution, ctd.

Well,

$$
\begin{aligned}
\mathbb{E}e^{-sD^\star} &= \sum_{n=0}^{N_\mathrm{m}} \int_0^\infty e^{-st}\mathbb{P}(\tau_{W^\star} = t, N^\star = n)\mathrm{d}t \\
&\overset{\mathrm{A}}{=} \sum_{n=0}^{N_\mathrm{m}} \int_0^\infty \int_0^\infty e^{-st}\mathbb{P}(W^\star = z, N^\star = n)\mathbb{P}(\tau_z = t \mid N^\star = n)\mathrm{d}z\mathrm{d}t \\
&\overset{\mathrm{B}}{=} \sum_{n=0}^{N_\mathrm{m}} \int_0^\infty \mathbb{E}(e^{-s\tau_z} \mid N^\star = n)\mathbb{P}(W^\star = z, N^\star = n)\mathrm{d}z;
\end{aligned}
$$

Queueing delay distribution, ctd.

Well,

$$
\begin{aligned}
\mathbb{E}e^{-sD^\star} &= \sum_{n=0}^{N_\mathrm{m}} \int_0^\infty e^{-st}\mathbb{P}(\tau_{W^\star} = t, N^\star = n)\mathrm{d}t \\
&\stackrel{\mathrm{A}}{=} \sum_{n=0}^{N_\mathrm{m}} \int_0^\infty \int_0^\infty e^{-st}\mathbb{P}(W^\star = z, N^\star = n)\mathbb{P}(\tau_z = t \mid N^\star = n)\mathrm{d}z\mathrm{d}t \\
&\stackrel{\mathrm{B}}{=} \sum_{n=0}^{N_\mathrm{m}} \int_0^\infty \mathbb{E}(e^{-s\tau_z} \mid N^\star = n)\mathbb{P}(W^\star = z, N^\star = n)\mathrm{d}z;
\end{aligned}
$$

A: remark that $O(0,t)$ depends on $(W^\star, N^\star)$ just through $N^\star$;

B: interchange order of integrations.

Expression for $\mathbb{E}(e^{-s\tau_z} \mid N^\star = n)$?

Queueing delay distribution, ctd.

Auxiliary result:

$$\xi_n(s,z) := \mathbb{E}(e^{-s\tau_z} \mid N^\star = n), \quad \text{and} \quad \xi(s,z) = (\xi_1(s,z),\ldots,\xi_{N_\mathrm{m}}(s,z))^\mathrm{T},$$

and $\mathbf{1}$ an $(N_\mathrm{m}+1)$-dimensional vector with 1's,

$$\xi(s,z) = \exp((R_O^{-1}Q - sR_O^{-1})z)\mathbf{1}.$$

In addition, the eigenvalues $\delta_0(s),\ldots,\delta_{N_\mathrm{m}}(s)$ of $R_O^{-1}Q - sR_O^{-1}$ are real, negative, and unique $(s > 0)$.

Queueing delay distribution, ctd.

Auxiliary result:

$$\xi_n(s,z) := \mathbb{E}(e^{-s\tau_z} \mid N^\star = n), \quad \text{and} \quad \xi(s,z) = (\xi_1(s,z),\ldots,\xi_{N_\mathrm{m}}(s,z))^\mathrm{T},$$

and $\mathbf{1}$ an $(N_\mathrm{m}+1)$-dimensional vector with 1's,

$$\xi(s,z) = \exp((R_O^{-1}Q - sR_O^{-1})z)\mathbf{1}.$$

In addition, the eigenvalues $\delta_0(s),\ldots,\delta_{N_\mathrm{m}}(s)$ of $R_O^{-1}Q - sR_O^{-1}$ are real, negative, and unique $(s>0)$.

Proof: set up system of de's + 'Geršgorin'.

Queueing delay distribution, ctd.

Hence, for constants γ_{mn} with $m, n = 0, \ldots, N_{\mathrm{m}}$,

$$\mathbb{E}(e^{-s\tau_z} \mid N^\star = n) = \sum_{m=0}^{N_{\mathrm{m}}} \gamma_{mn} e^{\delta_m(s)\, z}.$$

Queueing delay distribution, ctd.

Hence, for constants γ_{mn} with $m, n = 0, \ldots, N_{\rm m}$,

$$\mathbb{E}(e^{-s\tau_z} \mid N^\star = n) = \sum_{m=0}^{N_{\rm m}} \gamma_{mn} e^{\delta_m(s)\, z}.$$

Consequently, for $s > 0$,

$$\mathbb{E}e^{-sD^\star} = \sum_{n=0}^{N_{\rm m}} \sum_{m=0}^{N_{\rm m}} \gamma_{mn} \mathbb{E}(e^{\delta_m(s) W^\star} 1\{N^\star = n\}),$$

where the expression for $\mathbb{E}(e^{-sW^\star} 1\{N^\star = n\})$ was already available.

Flow transfer delay distribution

$(Z_i)_{i\in\mathbb{N}}$: number of flows present at (i.e., *just after*) arrival epochs.

$\textsc{Pasta}$-property:

$$\pi_n^Z := \frac{\pi_{n-1}}{\sum_{m=0}^{N_\mathrm{m}-1}\pi_m}.$$

Flow transfer delay distribution

$(Z_i)_{i\in\mathbb{N}}$: number of flows present at (i.e., *just after*) arrival epochs.

$\textsc{Pasta}$-property:

$$\pi_n^Z := \frac{\pi_{n-1}}{\sum_{m=0}^{N_\mathrm{m}-1} \pi_m}.$$

F: the transfer delay of a tagged flow that arrives at, say, time 0.

$$\phi_{nm}(s) := \mathbb{E}(e^{-sF} 1\{N_{F+} = m\} \mid N_0 = n).$$

(N_0 includes tagged flow).

Goal: compute distribution of F.

Flow transfer delay distribution, ctd.

As in [Borst, Boxma, Hegde]:

$$\phi_{nm}(s) = \frac{1}{\lambda + \mu_n C + s}\left(\lambda\phi_{n+1,m}(s) + \frac{n-1}{n}\mu_n C\,\phi_{n-1,m}(s) + \frac{1}{n}\mu_n C\,1\{n-1=m\}\right).$$

Also

$$\phi_{N_\mathrm{m}m}(s) = \frac{1}{\mu_{N_\mathrm{m}}C + s}\left(\frac{N_\mathrm{m}-1}{N_\mathrm{m}}\,\mu_{N_\mathrm{m}}C\,\phi_{N_\mathrm{m}-1,m}(s) + \frac{1}{N_\mathrm{m}}\,\mu_{N_\mathrm{m}}C\,1\{N_\mathrm{m}-1=m\}\right).$$

Flow transfer delay distribution, ctd.

As in [Borst, Boxma, Hegde]:

$$\phi_{nm}(s) = \frac{1}{\lambda + \mu_n C + s}\left(\lambda\phi_{n+1,m}(s) + \frac{n-1}{n}\mu_n C\,\phi_{n-1,m}(s) + \frac{1}{n}\mu_n C\,1\{n-1=m\}\right).$$

Also

$$\phi_{N_\mathrm{m}m}(s) = \frac{1}{\mu_{N_\mathrm{m}}C + s}\left(\frac{N_\mathrm{m}-1}{N_\mathrm{m}}\,\mu_{N_\mathrm{m}}C\,\phi_{N_\mathrm{m}-1,m}(s) + \frac{1}{N_\mathrm{m}}\,\mu_{N_\mathrm{m}}C\,1\{N_\mathrm{m}-1=m\}\right).$$

Linear system; for any $s > 0$ diagonally dominant and thus non-singular, and hence there is a unique solution:.

$$\mathbb{E}e^{-sF} = \sum_{n=1}^{N_\mathrm{m}}\sum_{m=0}^{N_\mathrm{m}-1}\pi_n^Z\,\phi_{nm}(s).$$

Sojourn time distribution

Recall: sojourn time $=$ flow transfer time $+$ queueing delay last particle.

Sojourn time distribution

Recall: sojourn time $=$ flow transfer time $+$ queueing delay last particle.

Long sojourn time can be due to a combination of

- System being full at flow arrival epoch;
- Long (tagged) flow;
- Large amount of traffic entering during flow transfer time;
- Large amount of fluid entering after flow transfer time.

Complex!

Sojourn time distribution, ctd.

First consider the epoch of arrival of flow: PASTA.

Associating time 0 with the accepted flow arrival,

$$
\begin{aligned}
\chi_n(s) &:= \mathbb{E}(e^{-sW_0} 1\{N_0 = n\}) \\
&= \frac{\pi_{n-1}}{\sum_{m=0}^{N_{\mathrm{m}}-1} \pi_m} \left(1 - \frac{\lambda \mathbb{E} T}{C}\right) \frac{s}{s - (\lambda/C)(1 - \mathbb{E} e^{sT})} \left(\prod_{i=1}^{n} \mathbb{E} e^{sT_i}\right).
\end{aligned}
$$

Sojourn time distribution, ctd.

ΔW: increment of the workload during flow transfer delay.

Note that $\Delta W \geq 0$ *a.s.*

Define joint transfrom of F and ΔW:

$$\psi_{nm}(\vec{s}) := \mathbb{E}(e^{-s_1 F - s_2 \Delta W} 1\{N_{F+} = m\} \mid N_0 = n),$$

with $\vec{s} \equiv (s_1, s_2)$.

The distribution of ΔW depends on the past only through N_0

(importantly, the value of W_0 does not play a role).

Sojourn time distribution, ctd.

The $\psi_{nm}(\vec{s})$ satisfy, for $n = 1, \ldots, N_{\mathrm{m}} - 1$, the following system of equations:

$$\psi_{nm}(\vec{s}) = \frac{1}{\lambda + \mu_n C + s_1 + r_{A,n} s_2} \left(\lambda \psi_{n+1,m}(s) + \frac{n-1}{n} \mu_n C \, \psi_{n-1,m}(s) + \frac{1}{n} \mu_n C \, 1\{n-1 = m\} \right).$$

Sojourn time distribution, ctd.

The $\psi_{nm}(\vec{s})$ satisfy, for $n = 1, \ldots, N_{\mathrm{m}} - 1$, the following system of equations:

$$\psi_{nm}(\vec{s}) = \frac{1}{\lambda + \mu_n C + s_1 + r_{A,n} s_2} \left(\lambda \psi_{n+1,m}(s) + \frac{n-1}{n} \mu_n C \, \psi_{n-1,m}(s) + \frac{1}{n} \mu_n C \, 1\{n - 1 = m\} \right).$$

We also have

$$\psi_{N_{\mathrm{m}}m}(\vec{s}) = \frac{1}{\mu_{N_{\mathrm{m}}} C + s_1 + r_{A,N_{\mathrm{m}}} s_2} \left(\frac{N_{\mathrm{m}} - 1}{N_{\mathrm{m}}} \, \mu_{N_{\mathrm{m}}} C \, \psi_{N_{\mathrm{m}}-1,m}(s) + \frac{1}{N_{\mathrm{m}}} \, \mu_{N_{\mathrm{m}}} C \, 1\{N_{\mathrm{m}} - 1 = m\} \right).$$

For fixed m and $\vec{s}$, again non-singular.

Sojourn time distribution, ctd.

So what do we know so far:

- How we find the system;
- What happens till the flow is completely put into the buffer.

What is left is the queueing delay of the last particle.

At the moment the flow is completely put into the buffer, the buffer content is $W_0 + \Delta W$.

Sojourn time distribution, ctd.

$$
\begin{aligned}
\mathbb{E}e^{-sS} &= \mathbb{E}\exp(-sF - s\tau_{W_0+\Delta W}) \\
&= \int_0^\infty \int_0^\infty \sum_{n=1}^{N_\mathrm{m}} \sum_{m=0}^{N_\mathrm{m}-1} \mathbb{P}(W_0 = x, N_0 = n) \\
&\qquad \mathbb{E}(e^{-sF} 1\{\Delta W = y, N_{F+} = m\} \mid N_0 = n) \mathbb{E}(e^{-s\tau_{x+y}} \mid N_0 = m) \mathrm{d}x\mathrm{d}y \\
&= \int_0^\infty \int_0^\infty \sum_{n=1}^{N_\mathrm{m}} \sum_{m=0}^{N_\mathrm{m}-1} \mathbb{P}(W_0 = x, N_0 = n) \\
&\qquad \mathbb{E}(e^{-sF} 1\{\Delta W = y, N_{F+} = m\} \mid N_0 = n) \sum_{k=0}^{N_\mathrm{m}} \gamma_{km} e^{\delta_k(s)\,(x+y)} \mathrm{d}x\mathrm{d}y,
\end{aligned}
$$

Sojourn time distribution, ctd.

$$\begin{aligned}
\mathbb{E}e^{-sS} &= \mathbb{E}\exp(-sF - s\tau_{W_0+\Delta W}) \\
&= \int_0^\infty \int_0^\infty \sum_{n=1}^{N_{\mathrm{m}}} \sum_{m=0}^{N_{\mathrm{m}}-1} \mathbb{P}(W_0 = x, N_0 = n) \\
&\qquad \mathbb{E}(e^{-sF} 1\{\Delta W = y, N_{F+} = m\} \mid N_0 = n) \mathbb{E}(e^{-s\tau_{x+y}} \mid N_0 = m) \mathrm{d}x\mathrm{d}y \\
&= \int_0^\infty \int_0^\infty \sum_{n=1}^{N_{\mathrm{m}}} \sum_{m=0}^{N_{\mathrm{m}}-1} \mathbb{P}(W_0 = x, N_0 = n) \\
&\qquad \mathbb{E}(e^{-sF} 1\{\Delta W = y, N_{F+} = m\} \mid N_0 = n) \sum_{k=0}^{N_{\mathrm{m}}} \gamma_{km} e^{\delta_k(s)\,(x+y)} \mathrm{d}x\mathrm{d}y,
\end{aligned}$$

or, summarizing:

For $s > 0$,

$$\mathbb{E}e^{-sS} = \sum_{n=1}^{N_{\mathrm{m}}} \sum_{m=0}^{N_{\mathrm{m}}-1} \sum_{k=0}^{N_{\mathrm{m}}} \gamma_{km} \chi_n(-\delta_k(s)) \psi_{nm}(s, -\delta_k(s)).$$

Decay rates

Focus on tail probabilities of the four random variables.

$$\Lambda_A(\theta) := \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E} \exp(\theta A(0,t)),$$

and also $I_A(x) := \sup_\theta(\theta x - \Lambda_A(\theta))$; $I_A(\cdot)$ is convex, $I_A(m_A) = 0$ for mean rate m_A

Decay rates

Focus on tail probabilities of the four random variables.

$$\Lambda_A(\theta) := \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E} \exp(\theta A(0,t)),$$

and also $I_A(x) := \sup_\theta(\theta x - \Lambda_A(\theta))$; $I_A(\cdot)$ is convex, $I_A(m_A) = 0$ for mean rate m_A

Decay rate of $W^\star$:

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(W^\star > x) = -\inf_{m>0} \frac{I_A(m)}{m}.$$

'Cost per time unit' argument.

Decay rates, ctd.

Consider the event $\{D^\star > t\}$ that fluid particle arriving at time 0 has (approximately) virtual delay t.

- After time 0, the queue drains at rate m: *cost*: $I_O(m)$ per unit of time.
- To achieve delay t, the workload at time 0 should have been mt.

  Supposing that the queue built up at rate $m' > 0$ before time 0, with cost $I_A(m')$ per unit of time, this took $(m/m')t$ time.

Decay rates, ctd.

Consider the event $\{D^\star > t\}$ that a fluid particle arriving at time 0 has (approximately) virtual delay t.

- After time 0, the queue drains at rate m: *cost*: $I_O(m)$ per unit of time.
- To achieve delay t, the workload at time 0 should have been mt.

  Supposing that the queue built up at rate $m' > 0$ before time 0, with cost $I_A(m')$ per unit of time, this took $(m/m')t$ time.

This leads to

$$\inf_{m,m'>0}\left(I_A(m')\,\frac{mt}{m'} + I_O(m)\,t\right) = t\left(\inf_{m>0}\left(\theta^\star m + I_O(m)\right)\right),$$

where the equality is due to decay rate of $W^\star$.

Decay rates, ctd.

Similarly: decay rates of F and S.

Decay rates, ctd.

Similarly: decay rates of F and S.

(Latter is complex, due to four factors involved...)

Challenges

- Other allocation schemes;

Challenges

- Other allocation schemes;
- Other distributions;

Challenges

- Other allocation schemes;
- Other distributions;
- Networks;

Challenges

- Other allocation schemes;
- Other distributions;
- Networks;
- . . .