State-dependent M/G/1 type queuing analysis for congestion control in data networks

Eitan Altman    Kostya Avrachenkov    Chadi Barakat    Rudesindo Nunez Queija

INRIA - Sophia Antipolis, France    CWI - The Netherlands

IEEE INFOCOM
Thursday, April 26, 2001
Anchorage, Alaska
Outline

- Congestion control modeling.
- Our model and its analysis.
- Numerical and experimental results.
- Concluding remarks.
Congestion control in data networks

☞ **Objective:** Efficient use and fair sharing of network resources.

☞ Widely used algorithm:

☞ Probe the network by linearly increasing the transmission rate.

☞ Divide the transmission rate by a constant factor (*typically* 2) when the network becomes congested.

☞ Congestion is detected either by an explicit signal sent by the network, or by inferring mechanisms at the sources.

☞ Henceforth, consider as a reference the TCP protocol, widely used for congestion control in the Internet …
TCP mechanisms

☞ A window-based flow control protocol.

☞ Use the loss of packets to detect network congestion.

![Diagram of TCP mechanisms]
- Slow start threshold
- Receiver window
- Congestion Avoidance
- Slow Start
- Updated slow start threshold
- Loss detection via duplicate ACKs (Fast Retransmit + Fast Recovery)
- Timeout (Stop of the ACK clock)
Congestion control modeling

☞ **Main objective:** Calculation of the average transmission rate (*the throughput*).

☞ Useful for understanding the behavior of the congestion control and for the design of new applications and protocols (e.g., mechanisms for routers).

☞ **Requirements:**

☞ Model for the variation of the transmission rate between congestion events and during congestion.

☞ Model for the appearance of congestion events.

☞ **Literature:**

☞ General models for congestion events, but

☞ No exact expressions of the throughput when the rate is limited ...
Limitation of the transmission rate

Caused by the receiver window in case of TCP ...

Michigan Univ.

Long-life TCP transfer
Receiver wnd = 32 Kbytes

INRIA - Sophia Antipolis
Our model for congestion control

☞ A fluid model for the variation of TCP window:

\[ W(t) \]

Linear increase at rate \( \alpha = \frac{1}{bRTT} \)

Congestion event

\[ W(t) \]

Use the techniques in [Altman, Avrachenkov, Barakat, SIGCOMM’00] to account for timeouts and the discrete nature of TCP.
Our model for congestion control (Ctd.)

☞ At the $n$-th congestion event, $N_n$ divisions of the window by a factor $\gamma > 1$

$$W(T_n+) = \frac{W(T_n-)}{\gamma^{N_n}} \quad \{N_n\} \text{ form an i.i.d. sequence with,}$$

$$Q(z) = \sum_{k=1}^{\infty} z^k P\{N_n = k\}$$

☞ Useful when the congestion lasts for multiple consecutive RTTs, or

☞ For future versions of TCP that reduce the window as a function of the congestion level (e.g., number of packets dropped during congestion).

☞ Congestion events (i.e., the instants $\{T_n\}$) occur according to a homogenous Poisson process of intensity $\lambda$. 
The analysis

☞ **Theorem:** The window of the TCP connection converges to the same stationary regime for any initial state.

☞ **Outputs of the analysis:**

☞ $F(x) = \text{PDF of } W(t) \text{ in the stationary regime.}$

☞ $E[W^k(t)], \ k \geq 1 = \text{Moments of } W(t) \text{ in the stationary regime.}$

☞ In particular:

TCP throughput: $\overline{X} = \lim_{t \to \infty} \frac{1}{t} \int_0^t \frac{W(\tau)}{RTT} d\tau = \frac{E[W(t)]}{RTT}$
Dual M/G/1 queuing model *

The problem of congestion control can be seen as a queuing model:

- The window corresponds to system workload ($U(t) = \frac{M - W(t)}{\alpha}$).
- Congestion events correspond to customer arrivals.
- Reduction of window corresponds to increase in workload.

Service time of a customer in the dual model:

$$x_n = \left(\frac{M}{\alpha} - U_n\right) \left(1 - \frac{1}{\gamma^{N_n}}\right)$$

Dependent of the workload.

* [Misra, Gong, Towsley, 1999]
Recursive equations for moments

Steady state Kolmogorov equation: Relation between \( F(x+dx) \) and \( F(x) \)

\[
\alpha \frac{dF(x)}{dx} = \lambda \sum_{k=1}^{\infty} P\{N = k\}F(\min(\gamma^k x, M)) - \lambda F(x)
\]

Applying Laplace Stieltjes Transform, then differentiating, we get

For \( k = 1, 2, \ldots \), \( E[W^k(t)] = \frac{k\alpha(E[W^{k-1}(t)] - P_M M^{k-1})}{\lambda(1 - Q(\gamma^{-k}))} \)

which gives all the moments of \( W(t) \), as a function of \( P_M = P\{W(t) = M\} \) ...
Window distribution

\( F(x) \) is continuous on the interval \((0, M)\) with a jump at \(M\) due to \(P_M\).

From the Kolmogorov equation we can write:

For \(x \in \left[ \frac{M}{\gamma^k}, \frac{M}{\gamma^{k-1}} \right], \ k \geq 1, \quad \overline{F}(x) = 1 - F(x) = P_M \sum_{i=1}^{k} c_i^{(k)} e^{-\frac{\lambda}{\alpha} \gamma^i x}

with \(c_i^{(k)}\) some constants that can be recursively determined.

**First method to calculate** \(P_M\): \( \lim_{x \to 0} F(x) = 0 \)

**But**, not efficient since the calculation of \(c_i^{(k)}\) for large \(k\) is not very accurate.
Efficient method to calculate $P_M$

\[
P_M = \frac{1/\lambda}{1/\lambda + S_M}
\]

$E[T(x)]$: Average time to return to $x$

$S_M = E[T(M)]$

$X(t)$

$M$

$1/\lambda$

$S_M$

Time

$X(t)$

$T(x)$

$T(y)$

$T(y')$

Time
Efficient method to calculate $P_M$ (Ctd.)

Derive an integral equation for $E[T(x)]$.

Apply Laplace Transforms, solve the equation and invert back:

$$E[T(x)] = \frac{1 - Q(\gamma^{-1})}{\lambda} \sum_{i=0}^{\infty} d_i \left( e^{-\frac{\gamma^{-i} \lambda}{\alpha} x} - 1 \right)$$

where $d_i$ are constants that can be recursively determined.

The algorithm is efficient since the infinite series converge very fast.
Extension to congestion limitation

☞ The model can be easily extended to the case when the window is reduced whenever it reaches $M$ (Congestion Limitation case).

☞ **Example:** $M$ corresponds to the available bandwidth.

![Workload Graph]

\[
F^{cl}(x) = \frac{F(x)}{1 - P_M}
\]

\[
E[(W^{cl})^k] = \frac{E[W^k] - P_M M^k}{1 - P_M}
\]
Model validation: Throughput

Long-life TCP transfer
Receiver wnd = 32 Kbytes
\( \gamma = 2 \)

INRIA - Sophia Antipolis

Univ. of South Australia

[Altman et al., SIGCOMM'00]
Model validation: Distribution

Michigan State Univ.

Long-life TCP transfer
Receiver wnd = $M$
$\gamma = 2$

INRIA - Sophia Antipolis

One-hour traces
Experimentation
Model

$M = 32$ Kbytes

Time between congestion events (s)

$M = 32$ Kbytes

Congestion window (bytes)
Model validation: Distribution (Ctd.)

- **Density function**
  - $M = 48$ Kbytes
  - $M = 64$ Kbytes

- **CDF**
  - Time between congestion events (s)
  - $M = 48$ Kbytes
  - $M = 64$ Kbytes

- **Congestion window (bytes)**
Conclusions and perspectives

☞ **Results:** Moments and distribution of transmission rate were obtained when there is a maximum limit and when the process of congestion events is close to Poisson.

Approximations of the throughput exist in the literature (e.g., the fixed-point approach by [Padhye, Firoiu, Towsley, and Kurose, SIGCOMM’98]).

☞ **Future work:**

☞ Consider more general processes for congestion events (e.g., MMPP).

☞ Validate the model on bursty paths where the congestion of the network lasts for multiple consecutive RTTs.

☞ Other rate increase policies (e.g., sub-linear increase).