A Method for Cross-layer Analysis of Transmit Buffer Delays in Message Index Domain

WoNeCa

Sami Akın and Markus Fidler

Institute of Communications Technology
Leibniz Universität Hannover

February 28, 2018

Outline

Motivation

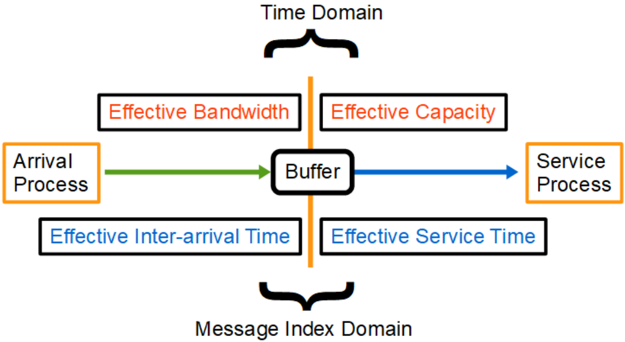
- Need for faster and more reliable data transmission techniques
- Effective bandwidth and capacity concepts
- An approximation for the steady-state buffer overflow probability

- Delay violation probability of first-come first-served systems when either the data arrival or service rate is constant
- Difficult to obtain effective bandwidth and capacity in certain scenarios without resorting to numerical techniques or particular assumptions

Contributions

- Analysis in the message index domain
- Waiting time (buffering delay) characterization

- Waiting time violation probability



State-of-the-art

Assuming that the buffer is unbounded (very big), the backlog (queue length) at time instant t is

$$q(t) = [q(t-1) + a_q(t) - s_q(t)]^+$$

$a_q(t)$: arrival rate, $s_q(t)$: service rate
$[\cdot]^+ = \max\{0, \cdot\}$

$q(\infty)$ will converge in distribution to a finite random variable when both $a_q(t)$ and $s_q(t)$ are stationary and ergodic, and when $\mathbb{E}\{a_q(t)\} < \mathbb{E}\{s_q(t)\}$

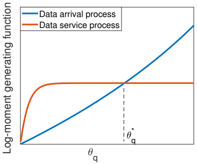
State-of-the-art

When $a_q(t)$ and $s_q(t)$ are independent of each other, there exists a unique $\theta_q^\star$ such that

$$\lim_{t\to\infty} \frac{\log \mathbb{E}\left\{e^{\theta_q^\star \sum_{\tau=1}^{t} a_q(\tau)}\right\}}{t} = -\lim_{t\to\infty} \frac{\log \mathbb{E}\left\{e^{-\theta_q^\star \sum_{\tau=1}^{t} s_q(\tau)}\right\}}{t}$$

and

$$\theta_q^\star = -\lim_{q_{\text{th}}\to\infty} \frac{\log \Pr\{q(\infty) \geq q_{\text{th}}\}}{q_{\text{th}}}$$



For large q_{th}, $\Pr\{q(\infty) \geq q_{\text{th}}\} \approx e^{-\theta_q^\star q_{\text{th}}}$

For $\theta_q > 0$

Effective bandwidth: $$\bar{a}_q(\theta_q) = \lim_{t\to\infty} \frac{\log \mathbb{E}\left\{e^{\theta_q \sum_{\tau=1}^{t} a_q(\tau)}\right\}}{\theta_q t}$$

Effective capacity: $$\bar{s}_q(-\theta_q) = \lim_{t\to\infty} \frac{\log \mathbb{E}\left\{e^{-\theta_q \sum_{\tau=1}^{t} s_q(\tau)}\right\}}{-\theta_q t}$$

Given a constant service or arrival rate;

$$\Pr\{d(\infty) \geq d_{\text{th}}\} \approx \varepsilon e^{-\theta_q \bar{a}_q(\theta_q) d_{\text{th}}} \quad \text{or} \quad \varepsilon e^{-\theta_q \bar{s}_q(-\theta_q) d_{\text{th}}}$$

where

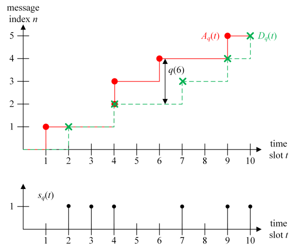
$$d_{\text{th}} = \frac{q_{\text{th}}}{\bar{a}_q(\theta_q)} \text{ or } \frac{q_{\text{th}}}{\bar{s}_q(-\theta_q)}, \text{ and } d(\infty) = \frac{q(\infty)}{\bar{a}_q(\theta_q)} \text{ or } \frac{q(\infty)}{\bar{s}_q(-\theta_q)}$$

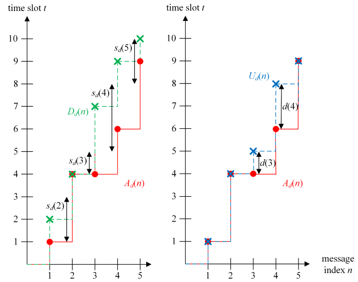
Objectives and approach

- $\Pr\{d(\infty) \geq d_{th}\}$ when both arrival and service processes are stochastic
- Communication scenarios where effective bandwidth and effective capacity are not straightforward

The pseudo-inverse functions of cumulative arrival and service processes: the number of messages and the number of time instants are positioned on the x-axis and y-axis, respectively

Objectives and approach





- $A_q(t)$: Cumulative arrival process
- $D_q(t)$: Cumulative departure process

- $A_d(n)$: Message n arrives
- $D_d(n)$: Message n departs
- $U_d(n)$: Message n starts service

Waiting time characterization

The waiting time:

$$d(n) = [d(n-1) + s_d(n) - a_d(n)]^+ \text{ for } n \geq 1$$

$d(n)$: waiting time of message n
$a_d(n)$: Inter-arrival time between messages $n-1$ and n
$s_d(n)$: Service time of message $n-1$

Time message n starts service:

$$U_d(n) = \max\{A_d(n), U_d(n-1) + s_d(n)\}$$

$A_d(n) = \sum_{i=1}^{n} a_d(i)$ is the time message n arrives at the buffer

Steady-State Waiting Time Analysis

$d(\infty)$ will converge in distribution to a finite random variable when both $s_d(n)$ and $a_d(n)$ are stationary and ergodic, and when $\mathbb{E}\{s_d(n)\} < \mathbb{E}\{a_d(n)\}$

The service process is work-conserving, and the arrival and service processes are independent of each other;

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left\{e^{\theta_d^\star \sum_{i=1}^{n} s_d(i)}\right\} = -\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left\{e^{-\theta_d^\star \sum_{i=1}^{n} a_d(i)}\right\}$$

where

$$\theta_d^\star = -\lim_{d_{\text{th}}\to\infty} \frac{\log \Pr\{d(\infty) \geq d_{\text{th}}\}}{d_{\text{th}}}$$

and large d_{th}

$$\Pr\{d(\infty) \geq d_{\text{th}}\} \approx e^{-\theta_d^\star d_{\text{th}}}$$

Effective service time

For $\theta_d > 0$

$$\bar{s}_d(\theta_d) = \lim_{n\to\infty} \frac{1}{n\theta_d} \log \mathbb{E}\left\{ e^{\theta_d \sum_{i=1}^{n} s_d(i)} \right\}$$

When the service time samples are independent and identically distributed;

$$\bar{s}_d(\theta_d) = \frac{1}{\theta_d} \log \mathbb{E}\left\{ e^{\theta_d s_d(n)} \right\}$$

$$\lim_{\theta_d\to 0} \bar{s}_d(\theta_d) = \mathbb{E}\{s_d(n)\} \quad \text{and} \quad \lim_{\theta_d\to\infty} \bar{s}_d(\theta_d) = \max\{s_d(n)\}$$

Effective inter-arrival time

For $\theta_d > 0$

$$\bar{a}_d(-\theta_d) = \lim_{n\to\infty} \frac{-1}{n\theta_d} \log \mathbb{E}\left\{ e^{-\theta_d \sum_{i=1}^{n} a_d(i)} \right\}$$

When the inter-arrival time samples are independent and identically distributed;

$$\bar{a}_d(-\theta_d) = -\frac{1}{\theta_d} \log \mathbb{E}\left\{ e^{-\theta_d a_d(n)} \right\}$$

$$\lim_{\theta_d\to 0} \bar{a}_d(-\theta_d) = \mathbb{E}\{a_d(n)\} \text{ and } \lim_{\theta_d\to\infty} \bar{a}_d(-\theta_d) = \min\{a_d(n)\}$$

Message index domain $\Longleftrightarrow$ Time domain

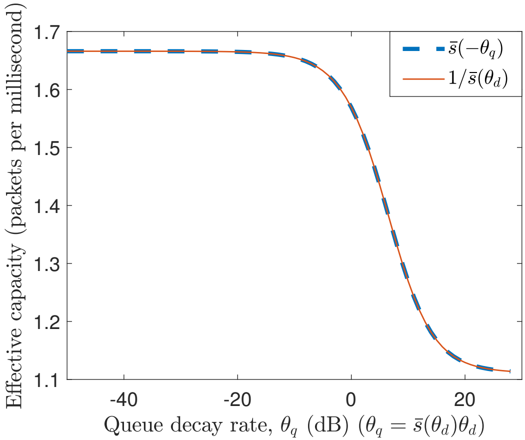
Given that the effective service time of a data service process is $\bar{s}_d(\theta_d)$ for $\theta_d > 0$ and the effective capacity of the same service process is $\bar{s}_q(-\theta_q)$ for $\theta_q > 0$, there exists a reciprocal relation between $\bar{s}_d(\theta_d)$ and $\bar{s}_q(-\theta_q)$ such that

$$\bar{s}_d(\theta_d)\bar{s}_q(-\theta_q) = 1 \quad \text{and} \quad \theta_q = \bar{s}_d(\theta_d)\theta_d$$

Given that the effective inter-arrival time of a data arrival process is $\bar{a}_d(-\theta_d)$ for $\theta_d > 0$ and the effective bandwidth of the same arrival process is $\bar{a}_q(\theta_q)$, there exists a reciprocal relation between $\bar{a}_d(-\theta_d)$ and $\bar{a}_q(\theta_q)$ such that

$$\bar{a}_d(-\theta_d)\bar{a}_q(\theta_q) = 1 \quad \text{and} \quad \theta_q = \bar{a}_d(-\theta_d)\theta_d$$

Message index domain $\Longleftrightarrow$ Time domain



Conclusion

Thank You

&

Questions