Performance Limits of Tunable Servers with Finite Buffer Capacity and a Packet-Drop Probability Constraint

Akshit Kumar[1] Parikshit Hegde[1] Rahul Vaze[2]

[1]Department of Electrical Engineering
Indian Institute of Technology, Madras

[2]School of Technology and Computer Science
Tata Institute of Fundamental Research,Mumbai

VSRP Talk, 2017

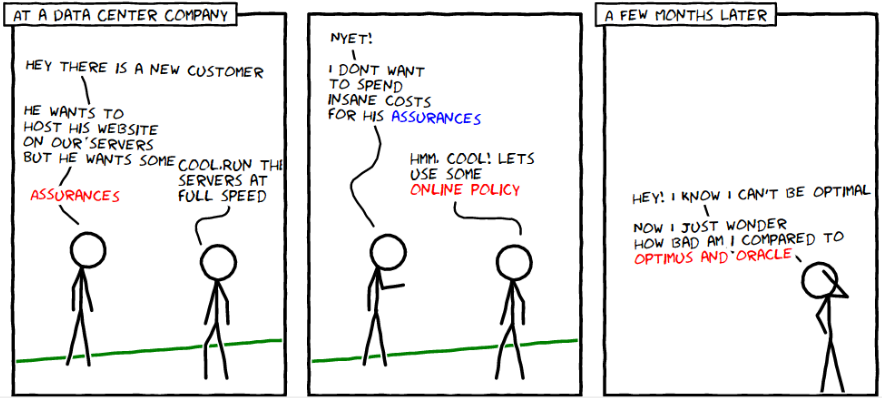
Motivation: Immigration Queue

Motivation: Immigration Queue

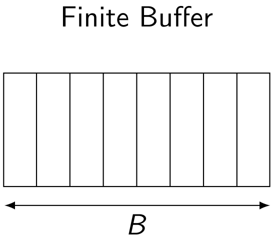
Motivation: Immigration Queue

Motivation:Web Hosting

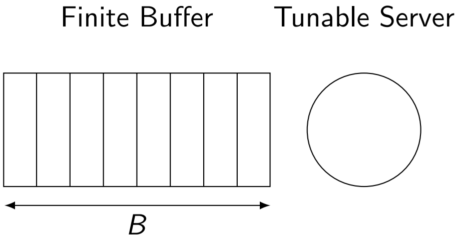
Motivation:Web Hosting



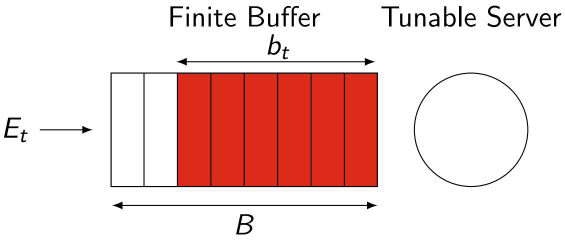
System Model



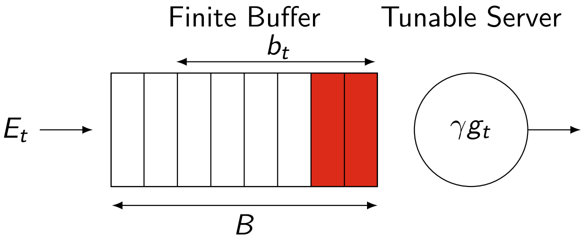
System Model



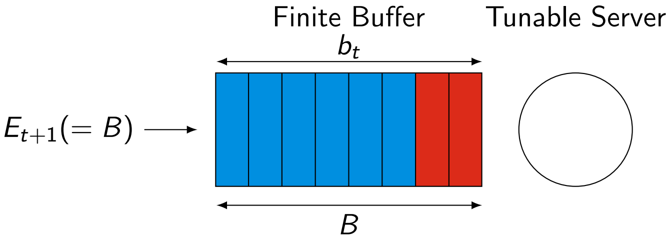
System Model



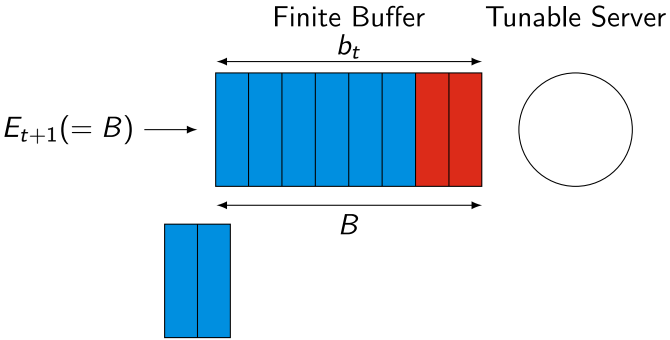
System Model



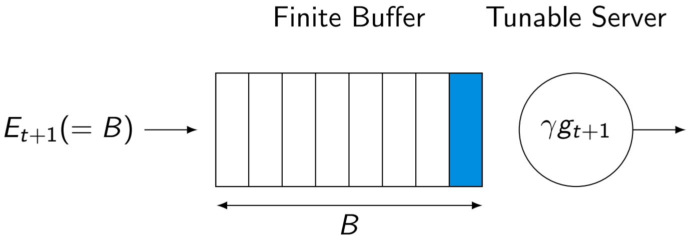
System Model



System Model



System Model



Problem Formulation

- $c_t \propto f(\gamma g_t)$, f is a convex function

Problem Formulation

- $c_t \propto f(\gamma g_t)$, f is a convex function
- $\mu = \mathbb{E}\left[E_t\right]$

Problem Formulation

- $c_t \propto f(\gamma g_t)$, f is a convex function
- $\mu = \mathbb{E}[E_t]$
- Good choice for servers is $c_t = g_t^2$ *[Wierman et. al.]*

Problem Formulation

- $c_t \propto f(\gamma g_t)$, f is a convex function
- $\mu = \mathbb{E}[E_t]$
- Good choice for servers is $c_t = g_t^2$ [*Wierman et. al.*]

So our objective is to

$$\underset{s}{\text{minimize}} \qquad \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} c_t\right]$$

Problem Formulation

- $c_t \propto f(\gamma g_t)$, f is a convex function
- $\mu = \mathbb{E}[E_t]$
- Good choice for servers is $c_t = g_t^2$ [*Wierman et. al.]*

So our objective is to

$$\begin{aligned} \underset{s}{\text{minimize}} \quad & \lim_{n \to \infty} \mathbb{E}\left[\frac{1}{n} \sum_{t=1}^{n} c_t\right] \\ \text{subject to} \quad & P_{dropping} \leq \alpha \end{aligned} \tag{1}$$

What's The Best I can do ?

What's The Best I can do ?

The probability constraint can be equivalently written as:

What's The Best I can do ?

The probability constraint can be equivalently written as:

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{t=1}^{n} g_t + B\right] \geq (1-\alpha)\,\mathbb{E}\left[\sum_{t=1}^{n} E_t\right]$$

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{\sum_{t=1}^{n} g_t}{n} + \cancelto{0}{\frac{B}{n}}\right] \geq (1-\alpha)\,\mathbb{E}\left[\frac{\sum_{t=1}^{n} E_t}{n}\right]$$

$$= (1-\alpha)\mu$$

What's The Best I can do ?

The probability constraint can be equivalently written as:

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{t=1}^{n} g_t + B\right] \geq (1-\alpha)\,\mathbb{E}\left[\sum_{t=1}^{n} E_t\right]$$

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{\sum_{t=1}^{n} g_t}{n} + \cancelto{0}{\frac{B}{n}}\right] \geq (1-\alpha)\,\mathbb{E}\left[\frac{\sum_{t=1}^{n} E_t}{n}\right]$$

$$= (1-\alpha)\mu$$

Now, the performance of an arbitrary policy is given by:

What's The Best I can do ?

The probability constraint can be equivalently written as:

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{t=1}^{n} g_t + B\right] \geq (1-\alpha)\,\mathbb{E}\left[\sum_{t=1}^{n} E_t\right]$$

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{\sum_{t=1}^{n} g_t}{n} + \cancelto{0}{\frac{B}{n}}\right] \geq (1-\alpha)\,\mathbb{E}\left[\frac{\sum_{t=1}^{n} E_t}{n}\right]$$

$$= (1-\alpha)\mu$$

Now, the performance of an arbitrary policy is given by:

$$\begin{aligned}
\mathcal{J} &= \mathbb{E}\left[\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} f(g_t)\right] \\
&\geq f\left(\mathbb{E}\left[\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} g_t\right]\right) && \text{Using Jensen's Inequality} \\
&= f((1-\alpha)\mu) = \mathcal{J}*
\end{aligned}$$

Near Optimal Online Policy?

- *Perkins & Srikant,1999* : Finding the optimal policy is hard

Near Optimal Online Policy?

- *Perkins & Srikant,1999* : Finding the optimal policy is hard
- Hence we look for a "near"- optimal policy

Near Optimal Online Policy?

- *Perkins & Srikant,1999* : Finding the optimal policy is hard
- Hence we look for a "near"- optimal policy
- Fix packet arrival distribution to **Bernoulli Distribution**

Near Optimal Online Policy?

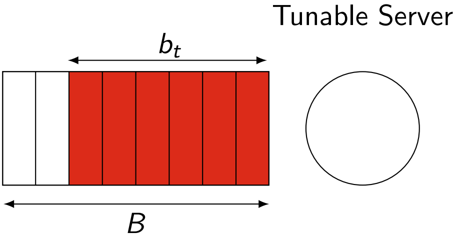
- *Perkins & Srikant,1999* : Finding the optimal policy is hard
- Hence we look for a "near"- optimal policy
- Fix packet arrival distribution to **Bernoulli Distribution**

$$E_t = \begin{cases} B, & \text{w.p } p \implies \text{Queue gets full} \\ 0, & \text{w.p } 1-p \end{cases} \tag{2}$$

- Bernoulli distribution gives the worst performance (*Shaviv et al.,2016*)

λ-Fraction Policy

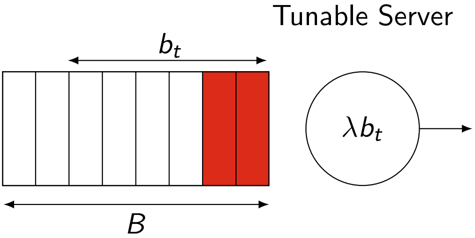
- If the number of packets in the buffer are b_t at time t, then



service: $g_t = \lambda b_t$

λ-Fraction Policy

- If the number of packets in the buffer are b_t at time t, then



service: $g_t = \lambda b_t$

λ-Fraction Policy

- If the number of packets in the buffer are b_t at time t, then service: $g_t = \lambda b_t$

λ-Fraction Policy

- If the number of packets in the buffer are b_t at time t, then service: $g_t = \lambda b_t$
- The Probability of Dropping can be calculated as:

$$P = \frac{\mathbb{E}_N \left[B(1-\lambda)^N \right]}{B}$$

λ-Fraction Policy

- If the number of packets in the buffer are b_t at time t, then service: $g_t = \lambda b_t$
- The Probability of Dropping can be calculated as:

$$P = \frac{\mathbb{E}_N \left[B(1-\lambda)^N \right]}{B}$$

- Setting $P = \alpha$ in the above equation we get:

$$\lambda = \frac{p(1-\alpha)}{p(1-\alpha)+\alpha}$$

Performance of the λ-Fraction Policy

Performance of the λ-Fraction Policy

- The performance J_λ is given by:

$$J_\lambda = \frac{\mathbb{E}\left[\sum_{t=1}^{N} g_t^2\right]}{\mathbb{E}\, N} = \frac{B^2 p^2 (1-\alpha)^2}{1-(1-p)(1-\alpha)^2}$$

Performance of the λ-Fraction Policy

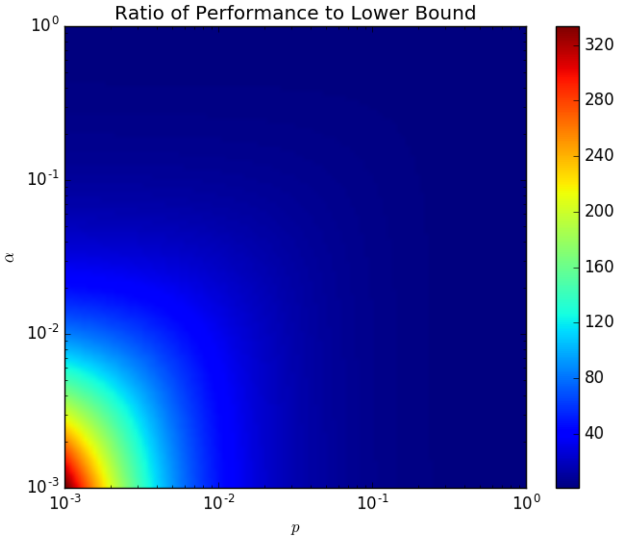
- The performance J_λ is given by:

$$J_\lambda = \frac{\mathbb{E}\left[\sum_{t=1}^{N} g_t^2\right]}{\mathbb{E}\, N} = \frac{B^2 p^2 (1-\alpha)^2}{1-(1-p)(1-\alpha)^2}$$

- Recollect that the lower bound $\mathcal{J}^*$ is given by:

$$\mathcal{J}^* = B^2 p^2 (1-\alpha)^2 \tag{3}$$

Performance of the λ-Fraction Policy



Small α,Small p Regime

Proposition

For $\alpha << 1, p << 1$, there exists no online policy that has a cost which is lower than $\Theta\left(\frac{B^2 p^2}{p+\alpha}\right)$

Small α,Small p Regime

Proposition

For $\alpha << 1, p << 1$, there exists no online policy that has a cost which is lower than $\Theta\left(\frac{B^2 p^2}{p+\alpha}\right)$

$$
\begin{aligned}
\mathcal{J}_\lambda &= \frac{B^2 p^2 (1-\alpha)^2}{1-(1-p)(1-\alpha)^2} \\
&\approx \frac{B^2 p^2}{p+2\alpha} \qquad [\alpha << 1, p << 1]
\end{aligned}
$$

Small α,Small p Regime

Simple Case: $\alpha << p$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$P(N = 1)$$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$P(N = 1)[B - g_1]$$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$P(N = 1)[B - g_1] \leq B\alpha$$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$P(N = 1)[B - g_1] \leq B\alpha$$
$$g_1 \geq B\left(1 - \frac{\alpha}{p}\right)$$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$\begin{aligned} P(N=1)[B-g_1] &\leq B\alpha \\ g_1 &\geq B\left(1-\frac{\alpha}{p}\right) \\ &\approx B \end{aligned}$$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$g_1 \approx B$$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$g_1 \approx B$$

$$\mathcal{J} = \frac{\mathbb{E}\left[\sum_{t=1}^{N} g_t^2\right]}{\mathbb{E}\, N}$$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$g_1 \approx B$$

$$\begin{aligned} \mathcal{J} &= \frac{\mathbb{E}\left[\sum_{t=1}^{N} g_t^2\right]}{\mathbb{E}\, N} \\ &\geq \frac{g_1^2}{\mathbb{E}\, N} \end{aligned}$$

Small α, Small p Regime

Simple Case: $\alpha << p$

$$g_1 \approx B$$

$$\begin{aligned}
\mathcal{J} &= \frac{\mathbb{E}\left[\sum_{t=1}^{N} g_t^2\right]}{\mathbb{E}\,N} \\
&\geq \frac{g_1^2}{\mathbb{E}\,N} \\
&\approx B^2 p
\end{aligned}$$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$g_1 \approx B$$

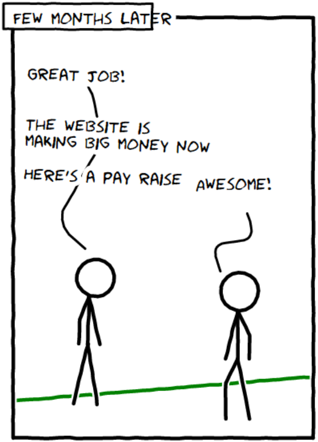
$$\begin{aligned}\mathcal{J} &= \frac{\mathbb{E}\left[\sum_{t=1}^{N} g_t^2\right]}{\mathbb{E}\, N} \\ &\geq \frac{g_1^2}{\mathbb{E}\, N} \\ &\approx B^2 p\end{aligned}$$

$$\boxed{\therefore \mathcal{J} \geq \Theta\left(\frac{B^2 p^2}{p+\alpha}\right)} \qquad [\text{For } \alpha << p]$$

Summary

- Tune server speed with convex cost function s.t., $P_{\text{drop}} \leq \alpha$
- Approximate online-policy for Bernoulli packet arrivals.
- Lower bounds for any online-policy;our policy is near optimal.
- Future Work:
  - Extend the policy to a general arrival distribution
  - Tighten the lower bound for small α for a general distribution

Thank You!


FEW MONTHS LATER
GREAT JOB!
THE WEBSITE IS MAKING BIG MONEY NOW
HERE'S A PAY RAISE
AWESOME!

Reference: I

James R. Perkins, R. Srikant
The Role of Queue Length Information In Congestion Control and Resource Pricing.
Proceedings of the 38th Conference on Decision and Control,1999.

Adam Wierman, Lachlan L.H. Andrew, Ao Tang
Power-aware speed scaling in processor sharing systems: Optimality and robustness
Performance Evaluation, 2012.

Small α,Small p Regime

Simple Case: $\alpha << p$

$$P(N = 1)$$

$$\begin{aligned} g_1 &\geq B\left(1 - \frac{\alpha}{p}\right) \\ &\geq B/2 \end{aligned}$$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$P(N = 1)[B - g_1]$$

$$\begin{aligned} g_1 &\geq B\left(1 - \frac{\alpha}{p}\right) \\ &\geq B/2 \end{aligned}$$

Small α,Small p Regime

Simple Case: $\alpha << p$

$$\begin{aligned} P(N=1)[B-g_1] &\leq B\alpha \\ g_1 &\geq B\left(1-\frac{\alpha}{p}\right) \\ &\geq B/2 \end{aligned}$$

Small α,Small p Regime

Case: $p \geq 2\alpha$

$$g_1 \geq B/2$$

$$\begin{aligned}
\mathcal{J} &= \frac{\mathbb{E}\left[\sum_{t=1}^{N} g_t^2\right]}{\mathbb{E}\, N} \\
&\geq \frac{g_1^2}{\mathbb{E}\, N} \\
&= \frac{B^2 p}{4} \qquad [\text{For } \alpha << 1, p \geq 2\alpha] \\
\therefore \mathcal{J} &\geq \Theta\left(\frac{B^2 p^2}{p+\alpha}\right)
\end{aligned}$$

- Recall: $\mathcal{J}_\lambda = \frac{B^2 p^2 (1-\alpha)^2}{1-(1-p)(1-\alpha)^2} \approx B^2 p \quad$ [For $\alpha << p$]

Small α,Small p Regime

$p \leq 2\alpha$

$$P[N \leq k]\left(B - \sum_{t=1}^{k} g_t\right) \leq B\alpha$$

$$\sum_{t=1}^{k} g_t \geq B\left(1 - \frac{\alpha}{pk}\right)$$

$$\sum_{t=1}^{k} g_t \geq B/2 \qquad \text{choosing } k = \left\lceil \frac{2\alpha}{p} \right\rceil$$

$$\begin{aligned}
\mathcal{J} &= \frac{\mathbb{E}\left[\sum_{t=1}^{N} g_t^2\right]}{\mathbb{E}\,N} \\
&\geq p \cdot P(N \geq k)[g_1^2 + g_2^2 + \cdots + g_k^2] \\
&\geq p(1 - pk) \cdot \frac{B^2}{4k} \\
&= \Theta\left(\frac{B^2 p^2}{\alpha}\right)
\end{aligned}$$

$$\therefore \mathcal{J} \geq \Theta\left(\frac{B^2 p^2}{\alpha + p}\right)$$