A formulation and theory for delay guarantees in wireless

P. R. Kumar

With I-Hong Hou and V. Borkar

Dept. of Electrical and Computer Engineering, and Coordinated Science Lab
University of Illinois, Urbana-Champaign

Email: prkumar@illinois.edu
Web: http://decision.csl.illinois.edu/~prkumar/
Consider an open queueing network with I single-server stations and K customer classes. Each customer class requires service at a specified station, and customers change class after service in a Markovian fashion. (With K allowed to be arbitrary, this routing structure is almost perfectly general.) There is a renewal input process and general service time distribution for each class. The correspondence between customer classes and service stations is in general many to one, and the service discipline (or scheduling protocol) at each station is left as a matter for dynamic decision making.

Assuming that the total load imposed on each station is approximately equal to its capacity, we show how to approximate the queueing system by a Brownian network, a type of crude but relatively tractable stochastic system model. The data of the approximating Brownian model are calculated in terms of the queueing system’s parameters, including both the first and second moments of the interarrival and service time distributions. (The Brownian approximation is insensitive to specific distributional forms.) We do not attempt a rigorous convergence proof to justify the proposed approximation, but the argument given in support of the approximation amounts to a broad outline for such a proof.

The Brownian approximation is initially developed for a network of reliable single-server stations, then generalized to allow server breakdown, and finally extended to closed networks and networks with controllable inputs. In all cases examined thus far, the approximating Brownian network is more tractable than the conventional model it replaces, whether the objective is performance evaluation or system optimization.

1. Introduction

In a future paper [8] a class of crude but relatively tractable stochastic system models, called Brownian networks, will be systematically discussed. The three notions that are taken as primitive in such models are resources (indexed by \( k = 1, \ldots, I \)), activities (indexed by \( j = 1, \ldots, J \)) and stocks (indexed by \( b = 1, \ldots, B \)). The system dynamics and policy constraints of a Brownian network are compactly expressed by the following relations:

1. \( Z(t) = X(t) + BY(t) + S \) for all \( t \geq 0 \), and

2. \( U(t) = AY(t) \) is a non-decreasing process with \( U(0) = 0 \).

*Research supported by NSF grant ECS 8503445.
Challenge of providing delay guarantees for wireless

- Increasing use of wireless networks for serving traffic with delay constraints:
  - VoIP
  - Interactive Video
  - Networked Control
- Yet delay guarantees are not supported

- How to formulate a mathematical framework for delay-based QoS?
- Relevant: Jointly deal with several QoS issues
  - Deadlines
  - Delivery ratios
  - Channel unreliabilities
- Tractable: Provide solutions for QoS support
  - Admission control policies for flows
  - Packet Scheduling policies
A wireless system with an Access Point serving $N$ clients

- Time is slotted
- One slot = One packet

AP indicates which client should transmit in each time slot
Model of unreliable channels

- **Unreliable channels**

- **Packet transmission in each slot**
  - Successful with probability $p_n$
  - Fails with probability $1 - p_n$
  - So packet delivery time is a geometrically distributed random variable $\gamma_n$ with mean $1/p_n$

- **Non-homogeneous link qualities**
  - $p_1, p_2, \ldots, p_N$ can be different
QoS model

- Clients generate packets with fixed period $\tau$
- Packets expire and are dropped if not delivered in the period
- Delay of successfully delivered packet is therefore at most $\tau$
- Delivery ratio of Client $n$ should be at least $q_n$

$$\liminf_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} 1(\text{Packet delivered to Client } n \text{ in } t\text{-th period}) \geq q_n \text{ a.s.}$$
Multiple-time scale QoS requirements

- **Unreliable channels**
  - Short time scale: Slots

- **Arrivals and Deadlines**
  - Medium time scale:
    - Period $\tau$ arrivals
    - Relative Deadline $\tau$

- **Delivery ratio requirements**
  - Long time scale:

\[
\lim \inf_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} 1(\text{Packet of client } n \text{ delivered in } t\text{-th period}) \geq q_n \ a.s.
\]
Protocol for operation

- AP indicates which client should transmit in each time slot

- Downlink
  - DATA
  - ACK
  - \( p_n = \text{Prob( Both DATA and ACK are delivered) } \)
Protocol for operation

- AP indicates which client should transmit in each time slot

- Downlink
  - DATA
  - ACK
  - \( p_n = \text{Prob( Both DATA and ACK are delivered) } \)

- Uplink
  - POLL (e.g., CF-POLL in 802.11 PCF)
  - DATA
  - \( p_n = \text{Prob( Both POLL and DATA are delivered ) } \)
  - No need for ACK
Feasibility of a set of clients
Implied workload

◆ Workload due to Client $n$

$$\omega_n = \frac{E(\# \text{ deliveries per period}) \cdot E(\# \text{ slots per delivery})}{\text{# of slots of per period}}$$

$$q_n \cdot \frac{1}{p_n} = \frac{q_n}{p_n \tau}$$

◆ The proportion of time slots needed by Client $n$ is

$$\omega_n = \frac{q_n}{p_n \tau}$$
Necessary condition for feasibility of QoS requirements

- Necessary condition from classical queueing theory
  \[ \sum_{n=1}^{N} w_n \leq 1 \]

- But not sufficient

- Reason: Unavoidable idle time
  - No queueing: At most one packet

![Diagram showing necessary condition for QoS feasibility]
Let \( I(1, 2, \ldots, N) := \text{Unavoidable idle time after serving } \{1, 2, \ldots, N\} \)

\[
I(1, 2, \ldots, N) = \frac{1}{\tau} E \left[ \left( \tau - \sum_{n=1}^{N} \gamma_n \right)^+ \right] \quad \text{where } \gamma_n \sim \text{Geom}(p_n)
\]

Stronger necessary condition

\[
\sum_{n=1}^{N} w_n + I(1, 2, \ldots, N) \leq 1
\]

Sufficient?

Still not sufficient!
Counterexample

- Two clients: Period $\tau = 3$

- Client 1
  - $p_1 = 0.5$
  - $q_1 = 0.876$
  - $w_1+I_1 = 3.002/3 > 1$ $\times$

- Client 2
  - $p_2 = 0.5$
  - $q_2 = 0.45$
  - $w_2+I_2 = 2.15/3 < 1$ $\checkmark$

- Clients $\{1,2\}$
  - $w_1+w_2+I_{\{1,2\}} = 2.902/3 < 1$ $\checkmark$

\[
\begin{align*}
  w_1 &= \frac{q_1}{p_1\tau} = \frac{1.752}{3} = \frac{1.25}{3} \\
  I_1 &= \frac{1}{3} \left( 2p_1 + (1 - p_1)p_1 \right) = \frac{1.25}{3} \\
  w_2 &= \frac{q_2}{p_2\tau} = \frac{0.9}{3} = \frac{1.25}{3} \\
  I_2 &= \frac{1}{3} \\
  w_{\{1,2\}} &= w_1 + w_2 = \frac{2.652}{3} \\
  I_{\{1,2\}} &= \frac{p_1p_2}{3} = \frac{0.25}{3}
\end{align*}
\]
Even stronger necessary condition

- Every subset of clients $S \subseteq \{1, 2, \ldots, N\}$ should also be feasible.

- Let $I(S) := \frac{1}{\tau} E \left[ \left( \tau - \sum_{n \in S} \gamma_n \right)^+ \right] = \text{Idle time if only serving } S$

- Stronger necessary condition: $\sum_{n \in S} w_n + I(S) \leq 1, \quad \forall S \subseteq \{1, 2, \ldots, N\}$

- Not enough to just evaluate for the whole set $\{1, 2, \ldots, N\}$

- Theorem (Hou, Borkar & K ’09)
  Condition is necessary and sufficient for a set of clients to be feasible.
Scheduling policy
Debt-based scheduling policies

- Compute “debt” owed to each client at beginning of period
- A client with higher debt gets a higher priority on that period
Two definitions of debt

- **The time debt** of Client $n$

  $$w_n - \text{Actual proportion of transmission slots given to Client } n$$

- **The weighted delivery debt** of Client $n$

  $$q_n - \text{Actual delivery ratio of client } n$$

  $$p_n$$

- **Theorem (Hou, Borkar & K ’09)**

  Both largest debt first policies fulfill every set of clients that can be fulfilled
Proof
Blackwell’s theory of approachability

- Period $t$
- Action $u(t)$
- Reward vector $r(t) \in R^N$
  - Distribution of $r(t)$ depends only on $u(t)$
  - Mean reward $= E(r \mid u)$
- Let $\rho(T) := \frac{1}{T} \sum_{t=1}^{T} r(t)$ = Time average of rewards up to stage $T$
- Consider a set $A \subseteq R^N$

- Definition of an approachable set $A$
  For some policy, for every $\varepsilon > 0$ and $\delta > 0$, there is a $T_0$ such that
  
  $$P(\text{Dist}(\rho(T), A) < \varepsilon \text{ for all } T \geq T_0) > 1 - \delta$$
Blackwell’s sufficient condition for approachability

- **Theorem (Blackwell ‘56)**

- Suppose $A$ is closed

- Suppose for every $x \notin A$, there is an action $u$ such that the mean reward $E(r | u)$ lies on the other side of the hyperplane $H$ passing through $y$, the point in $A$ closest to $x$, and perpendicular to the line $xy$

- Then $A$ is approachable under this policy, where any action can be taken when $x \in A$. 
Proof that time-debt policy is feasibility optimal

- Period $t$
- Action $u =$ Priority determined by time-debt policy
- $r_n =$ $n$-th component of resulting reward
  
  
  $:= \tau w_n$ - Time spent on Client $n$ in period

- Time average of rewards up to stage $T =$ Time-debt
  - Want all debts non-positive

- $A =$ Non-positive orthant of $R^N$
Sufficient condition for approachability of \( A \)

- **Order** \( x = (x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_N) \) with \( x_n \downarrow \) in \( n \) 
  \[ >0 \leq 0 \]

- **Then** \( y = (0, 0, \ldots, 0, x_{m+1}, \ldots, x_N) \)

- **Hyperplane is** \( H := \{ z : (z - y)^T (x - y) = 0 \} \)
  \[ = \{ z : \sum_{n=1}^{m} z_n x_n = 0 \} \]

- **Now** \( x \) is in the positive half-space of \( H \)

- **So we only need to show** that \( \sum_{n=1}^{m} E(r_n | u) x_n \leq 0 \)

- **I.e., we only need to show** \( \sum_{n=1}^{m} E(\tau w_n - \text{Time spent on Client } n \text{ in period}) \cdot x_n \leq 0 \)
  \[ := B_n \]
Approachability proof

- Let $B_n = \text{Busy time spent on Client } n$

- Priority order for service is $1, 2, 3, \ldots, k, \ldots, m, m+1, \ldots, N$

- So for $1 \leq k \leq m$, $\tau - \sum_{n=1}^{k} E(B_n) = \tau I(1,2,\ldots,k) \leq \tau - \tau \sum_{n=1}^{k} w_n$

- So $\alpha^k := \sum_{n=1}^{k} (\tau w_n - E(B_n)) \leq 0$ for all $1 \leq k \leq m$

- Then $\sum_{n=1}^{m} E(\tau w_n - B_n) \cdot x_n = \sum_{n=1}^{m} (\alpha^n - \alpha^{n-1})x_n$

  $\leq \sum_{n=1}^{m} (\alpha^n x_n - \alpha^{n-1} x_{n-1})$ (Since $x_n \searrow$ in $n$).

  $= \alpha^m x_m$

  $\leq 0$.

  (Set $\alpha^0 := 0$ and $x_0 := x_1$).

  ($x_k > 0$ for $1 \leq k \leq m$).
Computationally tractable admission control
Computationally tractable policy for admission control

- Admission control consists of determining feasibility

- We need to check: \[ \sum_{n \in S} w_n + I_S \leq 1, \quad \forall S \subseteq \{1, 2, ..., N\} \]

- Apparently \(2^N\) tests, so computationally complex, but

- **Theorem (Hou, Borkar & K ’09)**
  - Order the clients according to \(q_n\) in decreasing order
  
  - Then \(\{1, 2, ..., N\}\) infeasible \(\iff\) \[ \sum_{n=1}^{k} w_n + I(1, 2, ..., k) > 1 \] for some \(k\)
  
  - So we need only \(N\) tests to check \(\{1, 2, ..., k\}\) for \(1 \leq k \leq N\)
  
  - Polynomial time \(O\left(N \tau \log \tau\right)\) algorithm for admission control
Proof of polynomial time test

- Say a set $S$ is *bad* if $\sum_{n \in S} w_n + I(S) > 1$

- Suppose $S = \{1 \ 2 \ 3 \ 4 \ 5 \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \}\}$ is bad

- But $S-m = \{1 \ 2 \ 3 \ 4 \ 5 \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ N\}$ not bad

- Will show $S' := S+j = \{1 \ 2 \ 3 \ 4 \ 5 \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ \circ \ N\}$ is bad
  - $S'$ has one less hole than $S$

- Now prune $m, m-1, \ldots, m-n$ till we get a set that is not bad

- Repeat with $S'-(m-1)-(m-2)-(m-n)$
Proof of polynomial time test

- Give highest priority to $S-m$. Next priority to $m$. Last priority to $j$.

- Then $\sum_{n \in S+j} w_n + I(S+j) = \left( \sum_{n \in S} w_n + I(S) \right) + w_j - \frac{E(B_j)}{\tau}$

- So to show $S+j$ is bad, it is sufficient to show $w_j - \frac{E(B_j)}{\tau} \geq 0$.

- Now $w_j - \frac{E(B_j)}{\tau} = w_j - \frac{1}{\tau} \sum_{\sigma=1}^{\tau} P(m \text{ completed with } \sigma \text{ slots left}) \cdot E(B_j | \sigma \text{ slots left})$

  $\geq w_j - \frac{1}{\tau} \sum_{\sigma=1}^{\tau} P(m \text{ completed with } \sigma \text{ slots left}) \cdot E(B_j | \infty \text{ slots left})$

  $\geq w_j - \frac{1}{\tau p_j} \sum_{\sigma=1}^{\tau} P(m \text{ completed with } \sigma \text{ slots left})$

  $S$ is bad $\geq w_j - \frac{1}{\tau p_j} q_m = \frac{1}{\tau p_j} q_j - \frac{1}{\tau p_j} q_m \geq 0$.  

1 2 3 4 5 ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ ⊗ Concatenation of n subgraphs

\{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 \cdot 11 \cdot 12 \cdot 13 \cdot 14 \cdot 15 \cdot 16 \cdot 17 \cdot 18 \cdot 19 \cdot 20 \cdot 21 \cdot 22 \cdot 23 \cdot 24 \cdot 25 \cdot 26 \cdot 27 \cdot 28 \cdot 29 \cdot 30 \cdot 31 \cdot 32 \cdot 33 \cdot 34 \cdot 35 \cdot 36 \cdot 37 \cdot 38 \cdot 39 \cdot 40 \cdot 41 \cdot 42 \cdot 43 \cdot 44 \cdot 45 \cdot N \}
Complexity of admission control algorithm

- Order \( q_1, q_2, \ldots, q_N \) in decreasing order

- Evaluate \( w_1, w_2, \ldots, w_N \), where \( w_n = \frac{q_n}{\tau p_n} \)

- Evaluate \( g_m(t) := \text{Prob(Packets of 1,2,...,m are delivered by } t) \) by FFT

\[
g_m(t) = \sum_{s=1}^{\tau} (1 - p_m)^{s-1} p_m g_{m-1}(t - s)
\]

- Evaluate \( I(1,2,\ldots,m) = \frac{1}{\tau} \sum_{s \geq 1} P(\text{Number of idle slots } \geq s) = \frac{1}{\tau} \sum_{s \geq 1} g_m(\tau - s) \)

- So \( O(N \tau \log \tau + N \log N) \)
Proof that weighted delivery debt policy is feasibility optimal

- Period $t$

- Action $u = \text{Priority determined by weighted delivery debt policy}$

- $r_n = n$-th component of resulting reward

\[
:= \frac{q_n}{p_n} - \frac{1(\text{n is successfully delivered})}{p_n}
\]

- Time average of rewards up to stage $T = \text{Weighted delivery debt}$

- $A = \text{Non-positive orthant of } R^N$
Sufficient condition for approachability of $A$ for weighted delivery debt policy

- As before order $x = (x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_N)$ with $x_n \downarrow$ in $n$

- Now we need to show

$$\sum_{n=1}^{m} E\left( \frac{q_n - 1(n \text{ is successfully delivered})}{p_n} \right) \cdot x_n \leq 0$$

- If $\pi_n := P(n \text{ is successfully delivered})$

- We need to show

$$\sum_{n=1}^{m} \left( \frac{q_n - \pi_n}{p_n} \right) \cdot x_n \leq 0$$

- As before it suffices to show

$$\sum_{n=1}^{k} \frac{q_n - \pi_n}{p_n} \leq 0 \quad \text{for} \quad 1 \leq k \leq m$$
Approachability proof for weighted-delivery debt policy

- Note \( \frac{\pi_1}{p_1} = \frac{1 - (1 - p_1)^T}{p_1} = 1 + (1 - p_1) + \ldots + (1 - p_1)^{T-1} = P(B_1 \geq 1) + \ldots + P(B_1 \geq T) = E(B_1) \)

- Similarly, conditioned on \( \{1,2,\ldots,k-1\} \) completing \( \sigma \) slots before end of period

\[
P(k \text{ is successful} | \sigma \text{ slots left when } \{1,2,\ldots,k-1\} \text{ completes}) = \frac{1 - (1 - p_1)^{\sigma}}{p_1}
\]

\[
= 1 + (1 - p_1) + \ldots + (1 - p_1)^{\sigma-1} = E(B_k | \sigma \text{ slots left to serve } k)
\]

- So \( \frac{\pi_k}{p_k} = E(B_k) \)

- Hence \( \sum_{n=1}^{k} \frac{q_n - \pi_n}{p_n} = \sum_{n=1}^{k} \tau w_n - E(Busy \text{ time serving } \{1,2,\ldots,k\}) \)

\[
= \sum_{n=1}^{k} \tau w_n - (\tau - I(1,2,\ldots,k)) \leq 0
\]
Main results

- **Theorem (Hou, Borkar & K ’09)**
  - A set of clients \(\{1, 2, \ldots, N\}\) is feasible
  \[\begin{align*}
  \sum_{n=1}^{k} \frac{q_n}{p_n \tau} + \frac{1}{\tau} E \left[ \left( \tau - \sum_{n=1}^{k} \gamma_n \right)^+ \right] \leq 1 \quad \text{for all } k = 1, 2, \ldots, N
  \end{align*}\]
  - The weighted delivery debt policy satisfies all the clients
  \[\begin{align*}
  \sum_{n=1}^{k} \frac{q_n}{p_n \tau} + \frac{1}{\tau} E \left[ \left( \tau - \sum_{n=1}^{k} \gamma_n \right)^+ \right] \leq 1 \quad \text{for all } k = 1, 2, \ldots, N
  \end{align*}\]
  - An \(O(N \tau \log \tau + N \log N)\) Admission Control Policy
Simulation testing
Simulation testing on ns-2

- Implement on IEEE 802.11 Point Coordination Function (PCF)
  - Point Coordinator (PC) assigns transmission opportunities to clients
  - Packets should be sent by broadcasting to avoid ACKs
  - Compatible with Distributed Coordination Function (DCF)

- Application: VoIP standard

<table>
<thead>
<tr>
<th>64 kbp data rate</th>
<th>20 ms period</th>
</tr>
</thead>
<tbody>
<tr>
<td>160 Byte packet</td>
<td>11 Mb/s transmission rate</td>
</tr>
<tr>
<td>610 μs time slot</td>
<td>32 time slots in a period</td>
</tr>
</tbody>
</table>

- Four policies
  - DCF and PCF with randomly assigned priorities
  - Time-debt policy and Weighted-delivery debt policy
Traffic requirements: Test at edge of feasibility

- Two groups of clients
  - Group A requires 99% delivery ratio
  - Group B requires 80% delivery ratio
  - The $n^{th}$ client in each group has $(60+n)\%$ channel reliability

- Feasible set: **11** group A clients and **12** group B clients
- Infeasible set: **12** group A clients and **12** group B clients

- Evaluation Measure
  - $\text{DMR}(n) := (q_n \ - \ \text{percentage of actual delivered packets})^+$
  - DMR of system $= \sum_{n=1}^{N} \text{DMR}(n)$
Results

Feasible set

Infeasible set

Clients fulfilled

Clients not fulfilled
Extensions …
More general arrivals

- Theorem (Hou & K ’09)

- Suppose \( r(S) = \text{Probability that packets for set } S \text{ arrive in a period} \)
  - Packet need not arrive in every period. (Can extend to periodic arrivals)
  - Client arrivals can be correlated

- Then, a set of clients \( \{1,2,\ldots,N\} \) is feasible

\[
\sum_{n \in S} \frac{q_n}{p_n} + \sum_{G \subseteq \{1,2,\ldots,N\}} r(G)E \left[ \left( \tau - \sum_{n \in S \cap G} \gamma_n \right)^+ \right] \leq \tau \quad \text{for all } S \subseteq \{1,2,\ldots,N\}
\]

- The weighted delivery debt policy satisfies all the clients
Time varying channels, heterogeneous deadlines, rate adaptation, etc

- More general packet arrivals at beginning of periods
- Clients with different deadlines
- Either
  - No rate adaptation and unreliable channel, or
  - Rate adaptation with reliability
- Time varying channels

- Pseudo-debt $r_n(t)$: Clients fulfilled $\leftrightarrow \lim_{t \to \infty} \frac{r_n(t)}{t} \leq 0$

- **Theorem (Hou & K ’09)**
  - Let $\mu_n = $ Expected reduction in pseudo-debt for Client $n$
  - $\mu_n$ depends on the scheduling policy
  - Policy that maximizes $\sum_n \mu_n r_n^+(t)$ is feasibility optimal.
Utility maximization

System
\[
\max_{\{q_n\} \text{ feasible}} \sum_{n=1}^{N} U_n(q_n)
\]
\(U_n\) str. concave, str. Incr., \(U_n(0) = \text{right limit}\)

Client \(n\)
\[
\max_{0 \leq \rho_n \leq \psi_n} U_n\left(\frac{\rho_n}{\psi_n}\right) - \rho_n
\]

Price \(\psi_n\)

Payment \(\rho_n\)

Access Point
\[
\max_{\{q_n\} \text{ feasible}} \sum_{n=1}^{N} \rho_n \log q_n
\]

Achieved by Weighted Transmission Time Policy
Give priority to lowest \(u_n(t)/\rho_n\)
\(u_n(t) = \text{Number of slots in } [0, t] \text{ given to Client } n\)

Is weighted max-min fair
And weighted proportionally fair

(Hou & K ’09)
Conclusion

- A framework for delay-based QoS that deals with
  - deadlines
  - delivery ratios
  - channel unreliabilities
  - fading channels
  - general arrivals
  - rate adaptation
  - client utilities, etc

- Analytically tractable

- Implementable policies for admission control and scheduling
References


Thank you
Attribution-Noncommercial-No Derivative Works 3.0 Unported

You are free:
- to Share — to copy, distribute and transmit the work

Under the following conditions:
- Attribution — You must attribute this work to P. R. Kumar (with link).
  http://decision.csl.illinois.edu/~prkumar/talks.html
- Noncommercial — You may not use this work for commercial purposes.
- No Derivative Works — You may not alter, transform or build upon this work.

With the understanding that:
- Waiver — Any of the above conditions can be waived if you get permission from the copyright holder.
- Other Rights — In no way are any of the following rights affected by the license.
  - Your fair dealing or fair use rights;
  - The author's moral rights;
  - Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.
- Notice — For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.

This is a human-readable summary of the Legal Code (the full license).