Optimal Congestion Control of TCP Flows for Internet Routers

K. Avrachenkov, U. Ayesta, **J. Doncel**, P. Jacko

INRIA Sophia-Antipolis, France
LAAS-CNRS, Toulouse, France
BCAM-Basque Centre for Applied Mathematics

MAMA 2012, London, June 15

Goals and Contributions

Fast and fair transmission of TCP data $\Rightarrow$ avoidance of network congestion

Goals and Contributions

Fast and fair transmission of TCP data $\Rightarrow$ avoidance of network congestion
The main contributions of this work are:

- Modeling the interaction between a TCP source and a bottleneck queue $\Rightarrow$ design optimal packet admission controls in the bottleneck queue

Goals and Contributions

Fast and fair transmission of TCP data ⇒ avoidance of network congestion
The main contributions of this work are:

- Modeling the interaction between a TCP source and a bottleneck queue ⇒ design optimal packet admission controls in the bottleneck queue
- Formulate AIMD TCP protocol as a Markov Decision Process (MDP)

Goals and Contributions

Fast and fair transmission of TCP data $\Rightarrow$ avoidance of network congestion
The main contributions of this work are:

- Modeling the interaction between a TCP source and a bottleneck queue $\Rightarrow$ design optimal packet admission controls in the bottleneck queue
- Formulate AIMD TCP protocol as a Markov Decision Process (MDP)
- Obtain approximate solution to MDP using index-policies

Goals and Contributions

Fast and fair transmission of TCP data $\Rightarrow$ avoidance of network congestion
The main contributions of this work are:

- Modeling the interaction between a TCP source and a bottleneck queue $\Rightarrow$ design optimal packet admission controls in the bottleneck queue
- Formulate AIMD TCP protocol as a Markov Decision Process (MDP)
- Obtain approximate solution to MDP using index-policies
- Validate the model with simulations in Network Simulator 3 (ns-3)

Index

1. Problem Description
   - Formulation of Markov Decision Process
2. Solution
   - Analitical Results
   - Numerical Results
3. Simulations in ns-3
4. Conclusions and Future Work

Problem Description: Notation

MDP elements:

- Action space: $\mathcal{A} = \{0, 1\}$, not transmitting or transmitting

Problem Description: Notation

MDP elements:

- Action space: $\mathcal{A} = \{0, 1\}$, not transmitting or transmitting
- State space: $\mathcal{N} = \{1, 2, ...N\}$, where N is the maximum congestion window

Problem Description: Notation

MDP elements:

- Action space: $\mathcal{A} = \{0, 1\}$, not transmitting or transmitting
- State space: $\mathcal{N} = \{1, 2, ...N\}$, where N is the maximum congestion window
- The transition matrix is $\mathbf{P}^a$, if the router takes decision a.

Problem Description: Notation

MDP elements:

- Action space: $\mathcal{A} = \{0, 1\}$, not transmitting or transmitting
- State space: $\mathcal{N} = \{1, 2, ...N\}$, where N is the maximum congestion window
- The transition matrix is $\mathbf{P}^a$, if the router takes decision a.
- The number of packets that the router transmits in state n if takes action a is W_n^a.

Problem Description: Notation

MDP elements:

- Action space: $\mathcal{A} = \{0, 1\}$, not transmitting or transmitting
- State space: $\mathcal{N} = \{1, 2, ...N\}$, where N is the maximum congestion window
- The transition matrix is $\mathbf{P}^a$, if the router takes decision a.
- The number of packets that the router transmits in state n if takes action a is W_n^a.
- The expected reward of W_n^a is R_n^a.

MDP elements:

- Action space: $\mathcal{A} = \{0, 1\}$, not transmitting or transmitting
- State space: $\mathcal{N} = \{1, 2, ...N\}$, where N is the maximum congestion window
- The transition matrix is $\mathbf{P}^a$, if the router takes decision a.
- The number of packets that the router transmits in state n if takes action a is W_n^a.
- The expected reward of W_n^a is R_n^a.

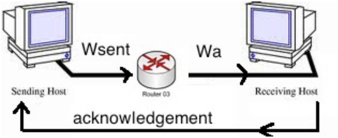


Figure: Example of one user sending TCP data

Problem Description: Formulation of MDP

R_n is function of W_n:

$$R_n^a := \begin{cases} \dfrac{(1+W_n^a)^{1-\alpha}-1}{1-\alpha}, & \text{if } \alpha \neq 1, \\ \log(1+W_n^a), & \text{if } \alpha = 1; \end{cases}$$

Problem Description: Formulation of MDP

R_n is function of W_n:

$$R_n^a := \begin{cases} \frac{(1+W_n^a)^{1-\alpha}-1}{1-\alpha}, & \text{if } \alpha \neq 1, \\ \log(1+W_n^a), & \text{if } \alpha = 1; \end{cases}$$

AIMD TCP as a Markov Chain definition:

- We consider *additive increasing* always

$$\text{cwnd}(t+1) = \text{cwnd}(t) + 1$$

- Multiplicative decrease factor ($\gamma \in [0,1)$)

$$\text{cwnd}(t+1) = max\{floor(\gamma * \text{cwnd}(t)), 1\}$$

Problem Description: Formulation of MDP

R_n is function of W_n:

$$R_n^a := \begin{cases} \dfrac{(1+W_n^a)^{1-\alpha}-1}{1-\alpha}, & \text{if } \alpha \neq 1, \\ \log(1+W_n^a), & \text{if } \alpha = 1; \end{cases}$$

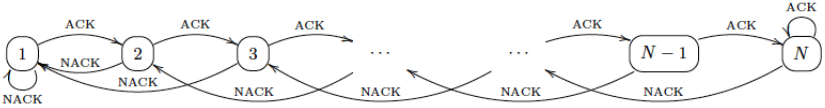
AIMD TCP as a Markov Chain definition:

- We consider *additive increasing* always

$$\text{cwnd}(t+1) = \text{cwnd}(t) + 1$$

- Multiplicative decrease factor $(\gamma \in [0,1))$

$$\text{cwnd}(t+1) = max\{floor(\gamma * \text{cwnd}(t)), 1\}$$



Maximization Problem

- Maximizing the multiflow problem

$$\max_{\pi \in \Pi} \lim_{T \to \infty} \frac{1}{T} \mathbb{E}^{\pi}_{\mathbf{n},B_0} \left[\sum_{t=0}^{T-1} \sum_{m \in \mathcal{M}(t)} R^{a_m(t)}_{m,X_m(t)} \right]$$

- Subject to limited bandwidth and buffer space

$$\lim_{T \to \infty} \frac{1}{T} \mathbb{E}^{\pi}_{\mathbf{n},B_0} \left[\sum_{t=0}^{T-1} \sum_{m \in \mathcal{M}(t)} W^{a_m(t)}_{m,X_m(t)} \right] \leq \overline{W}$$

$$B(t) + \sum_{m \in \mathcal{M}(t)} W^{a_m(t)}_{m,X_m(t)} \leq B, \text{ for all } t$$

Maximization Problem

- Maximizing the multiflow problem

$$\max_{\pi \in \Pi} \lim_{T \to \infty} \frac{1}{T} \mathbb{E}^{\pi}_{\mathbf{n}, B_0} \left[\sum_{t=0}^{T-1} \sum_{m \in \mathcal{M}(t)} R^{a_m(t)}_{m, X_m(t)} \right]$$

- Subject to limited bandwidth and buffer space

$$\lim_{T \to \infty} \frac{1}{T} \mathbb{E}^{\pi}_{\mathbf{n}, B_0} \left[\sum_{t=0}^{T-1} \sum_{m \in \mathcal{M}(t)} W^{a_m(t)}_{m, X_m(t)} \right] \leq \overline{W}$$

$$B(t) + \sum_{m \in \mathcal{M}(t)} W^{a_m(t)}_{m, X_m(t)} \leq B, \text{ for all } t$$

- Hard to solve

Relaxation of the problem

1. Relax (omit) the buffer constraint

Relaxation of the problem

1. Relax (omit) the buffer constraint
2. The standard solution is by solving for each ν,

$$\max_{\pi \in \Pi} \mathbb{E}_{\mathbf{n}}^{\pi} \left[\sum_{t=0}^{\infty} \sum_{k \in \mathcal{K}} \beta^t \left(R_{k,X_k(t)}^{a_k(t)} - \nu W_{k,X_k(t)}^{a_k(t)} \right) \right] + \nu \frac{\overline{W}}{1-\beta} \quad (1)$$

where ν is the Lagrangian parameter (per-packet *transmission cost*).

Relaxation of the problem

1. Relax (omit) the buffer constraint
2. The standard solution is by solving for each ν,

$$\max_{\pi \in \Pi} \mathbb{E}_{\mathbf{n}}^{\pi} \left[\sum_{t=0}^{\infty} \sum_{k \in \mathcal{K}} \beta^t \left(R_{k,X_k(t)}^{a_k(t)} - \nu W_{k,X_k(t)}^{a_k(t)} \right) \right] + \nu \frac{\overline{W}}{1-\beta} \quad (1)$$

   where ν is the Lagrangian parameter (per-packet *transmission cost*).
3. Lagrangian theory: there exists ν^*, for which the Lagrangian relaxation (1) achieves optimum of the above problem.

Relaxation of the problem

1. Relax (omit) the buffer constraint
2. The standard solution is by solving for each ν,

$$\max_{\pi \in \Pi} \mathbb{E}_{\mathbf{n}}^{\pi} \left[\sum_{t=0}^{\infty} \sum_{k \in \mathcal{K}} \beta^t \left(R_{k,X_k(t)}^{a_k(t)} - \nu W_{k,X_k(t)}^{a_k(t)} \right) \right] + \nu \frac{\overline{W}}{1-\beta} \tag{1}$$

where ν is the Lagrangian parameter (per-packet *transmission cost*).

3. Lagrangian theory: there exists ν^*, for which the Lagrangian relaxation (1) achieves optimum of the above problem.
4. We can decompose (1) into K individual-flow

$$\max_{\pi_k \in \Pi_k} \mathbb{E}_{n_k}^{\pi_k} \left[\sum_{t=0}^{\infty} \beta^t \left(R_{k,X_k(t)}^{a_k(t)} - \nu W_{k,X_k(t)}^{a_k(t)} \right) \right] \tag{2}$$

Relaxation of the problem

1. Relax (omit) the buffer constraint
2. The standard solution is by solving for each ν,

$$\max_{\pi\in\Pi} \mathbb{E}^{\pi}_{\mathbf{n}} \left[\sum_{t=0}^{\infty} \sum_{k\in\mathcal{K}} \beta^t \left(R^{a_k(t)}_{k,X_k(t)} - \nu W^{a_k(t)}_{k,X_k(t)} \right) \right] + \nu \frac{\overline{W}}{1-\beta} \quad (1)$$

where ν is the Lagrangian parameter (per-packet *transmission cost*).

3. Lagrangian theory: there exists ν^*, for which the Lagrangian relaxation (1) achieves optimum of the above problem.
4. We can decompose (1) into K individual-flow

$$\max_{\pi_k\in\Pi_k} \mathbb{E}^{\pi_k}_{n_k} \left[\sum_{t=0}^{\infty} \beta^t \left(R^{a_k(t)}_{k,X_k(t)} - \nu W^{a_k(t)}_{k,X_k(t)} \right) \right] \quad (2)$$

If for a given parameter ν, each policy π^*_k for $k \in \mathcal{K}$ optimizes the individual-flow problem then π^* optimizes the multi-flow problem (1).

Relaxation of the problem

1. For finite-state finite-action MDPs there exists an optimal policy that is deterministic, stationary, and independent of the initial state
    - we narrow our focus to those policies
    - policy $\mathcal{S}$ prescribes to *transmit* in states in $\mathcal{S}$ and *warn* in states in $\mathcal{S}^{\complement} := \mathcal{N} \setminus \mathcal{S}$

Relaxation of the problem

1. For finite-state finite-action MDPs there exists an optimal policy that is deterministic, stationary, and independent of the initial state
   - we narrow our focus to those policies
   - policy $\mathcal{S}$ prescribes to *transmit* in states in $\mathcal{S}$ and *warn* in states in $\mathcal{S}^{\complement} := \mathcal{N} \setminus \mathcal{S}$
2. Combinatorial problem $\max_{\mathcal{S} \subseteq \mathcal{N}} \mathbb{R}_n^{\mathcal{S}} - \nu \mathbb{W}_n^{\mathcal{S}}$, where

$$\mathbb{R}_n^{\mathcal{S}} := \mathbb{E}_n^{\mathcal{S}} \left[\sum_{t=0}^{\infty} \beta^t R_{X(t)}^{a(t)} \right], \qquad \mathbb{W}_n^{\mathcal{S}} := \mathbb{E}_n^{\mathcal{S}} \left[\sum_{t=0}^{\infty} \beta^t W_{X(t)}^{a(t)} \right]$$

Solution: Indexability

Definition

We say that the above problem is indexable, if it exists real numbers ν_n, $n \in \mathcal{N}$ such that for all states the following holds:

1. if $\nu_n \geq \nu$, is optimal transmitting in state n
2. if $\nu_n \leq \nu$ is not optimal transmitting in state n

The function $n \rightarrow \nu_n$ is called *index* and ν_n is the *index value of n*.

Solution: Indexability

Definition

We say that the above problem is indexable, if it exists real numbers ν_n, $n \in \mathcal{N}$ such that for all states the following holds:

1. if $\nu_n \geq \nu$, is optimal transmitting in state n
2. if $\nu_n \leq \nu$ is not optimal transmitting in state n

The function $n \rightarrow \nu_n$ is called *index* and ν_n is the *index value of n*.

Definition

We say that the above problem can be solved under threshold policies if $\nu_1 \geq \nu_2 \geq ... \geq \nu_N$.

Main results: Analitical Results

From previous work, always indexable and solvable under threshold policies:

1. 1-state and 2-state TCP flows
2. 3-state TCP flow with decrease factor γ less than $\frac{2}{3}$

Main results: Analitical Results

From previous work, always indexable and solvable under threshold policies:

1. 1-state and 2-state TCP flows
2. 3-state TCP flow with decrease factor γ less than $\frac{2}{3}$

Proposition

Three state TCP flow with $\gamma > \frac{2}{3}$ is indexable and:

- *if $\alpha < 1$, the threshold policies are optimal and the values of the indices are* $\nu_{k,1} = \frac{R_{k,1}}{W_{k,1}}$, $\nu_{k,2} = \frac{R_{k,2} - \beta R_{k,1}}{W_{k,2} - \beta W_{k,1}}$, $\nu_{k,3} = \frac{R_{k,3} + \beta(R_{k,3} - R_{k,2})}{W_{k,3} + \beta(W_{k,3} - W_{k,2})}$.
- *if $\alpha \geq 1$, threshold policies are not optimal in general ($\nu_{k,1} > \nu_{k,3} > \nu_{k,2}$) and the values of the indices are* $\nu_{k,1} = \frac{R_{k,1}}{W_{k,1}}$, $\nu_{k,2} = \frac{R_{k,2} + \beta(R_{k,3} - R_{k,1}) + \beta^2(R_{k,3} - R_{k,2})}{W_{k,2} + \beta(W_{k,3} - W_{k,1}) + \beta^2(W_{k,3} - W_{k,2})}$, $\nu_{k,3} = \frac{R_{k,3} - \beta^2 R_{k,1}}{W_{k,3} - \beta^2 W_{k,1}}$.

Indexability of the problem tested over a large number of flows with different parameters ⇒ always indexable.
Conjecture: the scheme is always indexable.

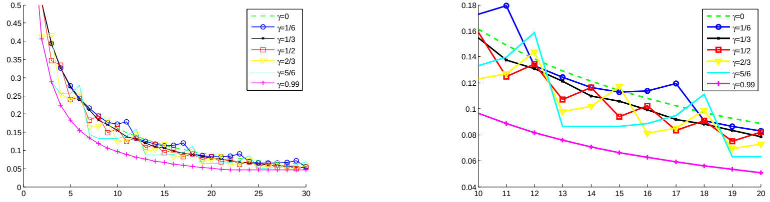


Figure: Seven Heterogeneous TCPs

Simulations Scenario Description

Network Simulator-3:

Simulations Scenario Description

Network Simulator-3:

Implementing the model:

- AIMD with no slow start

Simulations Scenario Description

Network Simulator-3:

Implementing the model:

- AIMD with no slow start

We compare the behaviour of this model with droptail and RED

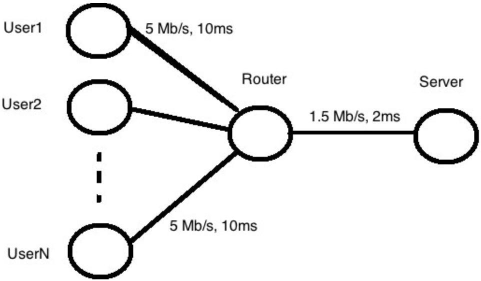
Simulations Scenario Description

Network Simulator-3:

Implementing the model:

- AIMD with no slow start

We compare the behaviour of this model with droptail and RED



Packet Size: 536 Bytes

Buffer size = Bandwidth-Delay Product = 14

Simulation scenario description

Packet-level heuristic index policy: Upon a packet arrival,

- if the buffer is not full, then accept the packet
- otherwise, drop the packet (either the new one or from the queue) with *smallest index* value
- in case of ties, drop the packet that has been the *longest* in the queue

Simulation scenario description

Packet-level heuristic index policy: Upon a packet arrival,

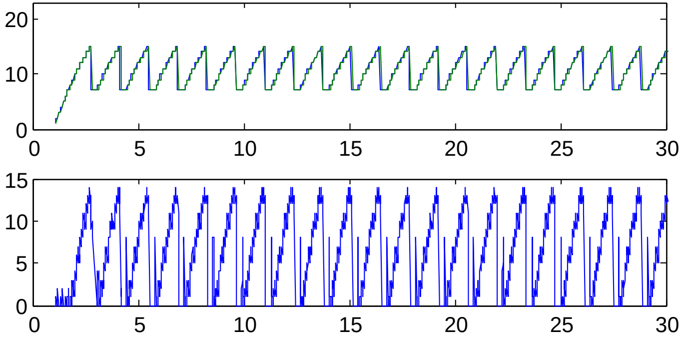
- if the buffer is not full, then accept the packet
- otherwise, drop the packet (either the new one or from the queue) with *smallest index* value
- in case of ties, drop the packet that has been the *longest* in the queue

Implementation in ns3:

1. We calculate the indices for each user when program starts.
2. We get the congestion window of the user that want to send a packet.
3. We send the packet with the corresponding index, according to the congestion window.
4. In the queue of the router the index is read and it is taken the decision of transmitting it or not.

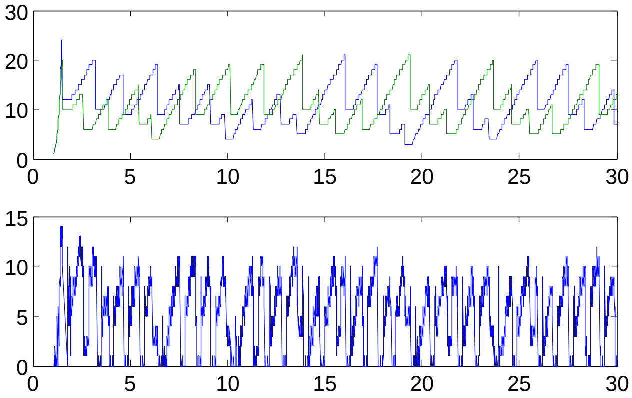
Simulation Results: 2 users and $\gamma = \frac{1}{2}$

Droptail policy



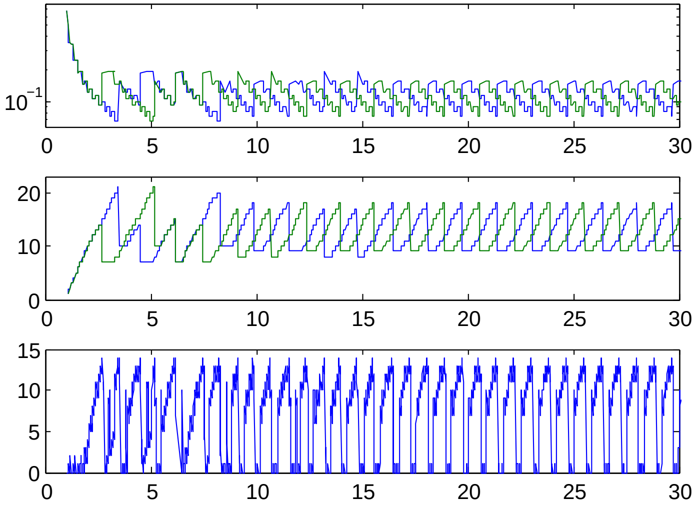
Simulation Results: 2 users and $\gamma = \frac{1}{2}$

RED



Simulation Results: 2 users and $\gamma = \frac{1}{2}$

Index policies model with $\alpha = 1$.



Conclusions

Main conclusions:

- Throughput increases
- More efficient buffer management
- Developed a packet implementation of index-policy

Future Work:

- Development new TCP models (Slow-start, users with different decrease factor...)
- Calculation of the index in the router $\Rightarrow$ not needed to assume compliant end-users (index estimating and learning techniques)
- Investigate more complicate topologies.

Thank you for your attention

Thank you!!!