Traffic Engineering
EN3370

Chathuranga Weeraddana

March 2021

Lecture 11: Flow Control II

Flow Control Methods

- window strategies [see previous set of slides]
  - end-to-end
  - node-to-node
- rate control schemes
  - time window flow control
  - leaky-bucket scheme

Rate Control Schemes [1]

[1] Based on § 6.3 of [Bertsekas-Gallager-1992].

Why Rate Control Schemes?

- window flow control is not well suited for high-speed sessions
- large propagation delays necessitate large window sizes
- end-to-end packet delays are not regulated
- no guarantees on a minimum data rate

Why Rate Control Schemes?

- an alternative form of flow control $\rightarrow$ give each session a guaranteed rate as needed
- specified rate depend on the session type

How to Implement a Rate r?

- admit 1 packet each $1/r$ seconds
  - not good with bursty traffic (large delays possible)
- admit as many as W packets every W/r seconds
  - good with bursty traffic
  - there are several variations

Time Window Flow Control

Time Window Flow Control

- r is the rate assigned to the session
- an allocation of W packets to the system is permitted
- a count x of the unused portion of the allocation is known at the origin

Time Window Flow Control

- packets from the session are admitted into the network as long as $x > 0$
- each time a packet is admitted, the count x is decremented by $1/W$
- W/r seconds later the count x is incremented by 1

Leaky Bucket Scheme

Leaky Bucket Scheme

- the scheme can be viewed by associating each session with
  - a queue of packets without a permits
  - a bucket of permits which are are generated at the desired input rate r, 1 permit in $1/r$ sec

Leaky Bucket Scheme

- the packet at the head of the packet queue obtains a permit once one is available in the permit bucket
- simultaneously the packet joins set of packets waiting to be transmitted
- permits are generated as long as the number in the permit bucket does not exceed a certain threshold W

Leaky Bucket Scheme

- rate adjustment schemes, usually by means of leaky buckets
- the choice and adjustment of the leaky bucket parameters (e.g., W) should carefully be done

Rate Adjustment Algorithms [2]

[2] Based on § 6.5 of [Bertsekas-Gallager-1992].

Algorithmic Aspects

- some algorithmic aspects of rate control schemes are discussed

Problem Formulations

- proper balance between high throughput and average delay
- fairness with an acceptable level of average delay per packet

Optimal Routing Problem [3]

[3] Based on § 5.4 of [Bertsekas-Gallager-1992].

Network

- consider a network with nodes indexed by $1, \ldots, N$
- communication from node i to j is possible $\rightarrow$ directed link denoted (i, j) exists
- let f_{ij} denote the flow of link (i, j), a 'measure' of congestion
- let f denote a vector whose elements are f_{ij}s

Cost Function

- an objective function $g_0(f)$ of the form

$$g_0(f) = \sum_{(i,j)\in\mathcal{L}} g_{ij}(f_{ij}) \tag{1}$$

  may be used to define a network-wide cost

- $\mathcal{L}$ is the set of all directed links

- $g_{ij}(\cdot)$ is a cost function associated with link (i, j), 'shape' depends on the context

Cost Function

- a cost function to model average number of data units in the system would be g_0 [see (1)] with

$$g_{ij}(f) = \frac{f_{ij}}{c_{ij} - f_{ij}} + d_{ij} f_{ij}$$

- d_{ij} is the propagation delay
- c_{ij} is the link capacity

Cost Function

- a cost function to model maximum link utilization would be

$$g_0(f) = \max_{(i,j)\in\mathcal{L}} \frac{f_{ij}}{c_{ij}}$$

Origin-Destination Pairs

- the network can contain many origin-destination node (O-D) pairs denoted by (o, d)
- let $\mathcal{W}$ denote the set of O-D pairs
- associated with each $(o, d) \in \mathcal{W}$
  - there is a input traffic arrival process with rate r_{od}
  - there is a set $\mathcal{P}_{od}$, the set of all directed paths from node o to node d
- the flow rate of path $p \in \mathcal{P}_{od}$ is denoted by x_p^{od}

Constraints

- path flow rates associated with each O-D pair $(o, d) \in \mathcal{W}$ should conform to

$$\sum_{p \in \mathcal{P}_{od}} x_p^{od} = r_{od}$$

- path flow rates are positive, i.e.,

$$x_p^{od} \geq 0, \quad (o, d) \in \mathcal{W}, \ p \in \mathcal{P}_{od}$$

Constraints

- flow rate of link (i, j) equals the sum of path flow rates through link (i, j) itself, i.e.,

$$\sum_{(p,(o,d))\in\mathcal{F}_{ij}} x_p^{od} = f_{ij}, \quad (i, j) \in \mathcal{L}$$

where

$$\mathcal{F}_{ij} = \{(p, (o, d)) \mid \exists\, (o, d) \in \mathcal{W},\ \text{path } p \in \mathcal{P}_{od} \text{ contains the link } (i, j)\}$$

Optimal Routing Problem

- the optimal routing problem is [4]

$$\begin{array}{ll} \text{minimize} & g_0(f) \\ \text{subject to} & \sum_{(p,(o,d))\in\mathcal{F}_{ij}} x_p^{od} = f_{ij}, \quad (i,j) \in \mathcal{L} \\ & \sum_{p\in\mathcal{P}_{od}} x_p^{od} = r_{od}, \ (o,d) \in \mathcal{W} \\ & x_p^{od} \geq 0, \ p \in \mathcal{P}_{od}, \ (o,d) \in \mathcal{W} \end{array}$$

where the variables are f_{ij} s and x_p^{od} s

[4] Unless $f_{ij} \leq c_{ij}$ is implicit from the domain of g_0, each constraint should explicitly be listed as a constraint in the problem.

Problem Formulations

- proper balance between high throughput and average delay
- fairness with an acceptable level of average delay per packet

High Throughput and Average Delay

A Related Problem

- a related problem is

$$\begin{aligned} \text{minimize} \quad & \textstyle\sum_{(i,j)\in\mathcal{L}} g_{ij}(f_{ij}) + \sum_{(o,d)\in\mathcal{W}} e_{od}(r_{od}) \\ \text{subject to} \quad & \textstyle\sum_{(p,(o,d))\in\mathcal{F}_{ij}} x_p^{od} = f_{ij}, \quad (i,j) \in \mathcal{L} \\ & \textstyle\sum_{p\in\mathcal{P}_{od}} x_p^{od} = r_{od}, \ (o,d) \in \mathcal{W} \\ & x_p^{od} \geq 0, \ p \in \mathcal{P}_{od}, \ (o,d) \in \mathcal{W} \end{aligned}$$

  where the variables are f_{ij} s and x_p^{od} s, and r_{od} s

- $e_{od}(\cdot)$ is a monotonicity decreasing function

- i.e., large input traffic arrival rates r_{od} are encouraged

Problem Formulations

- we can consider each $(o, d) \in \mathcal{W}$ as a session as well
- by changing e_{od}, different priorities can be imposed on different sessions

Fairness with Acceptable Levels of Average Delays

Problem Formulations

- keeping fairness among sessions are necessary in some applications
- fairness can be defined in a number of different ways
- one important way is the max-min fairness
- i.e., to maximize the 'allocation' for the most poorly treated session

A Related Problem

- the related problem to floe control in a fair manner is

$$\begin{aligned} \text{maximize} \quad & \min_{(o,d)\in\mathcal{W}} r_{od} \\ \text{subject to} \quad & \textstyle\sum_{(p,(o,d))\in\mathcal{F}_{ij}} x_p^{od} = f_{ij}, \quad (i,j) \in \mathcal{L} \\ & \textstyle\sum_{p\in\mathcal{P}_{od}} x_p^{od} = r_{od}, \ (o,d) \in \mathcal{W} \\ & x_p^{od} \geq 0, \ p \in \mathcal{P}_{od}, \ (o,d) \in \mathcal{W} \end{aligned}$$

  where the variables are f_{ij} s and x_p^{od} s, and r_{od} s

- problem finds $r = (r_{od})_{(o,d)\in\mathcal{W}}$ a vector of session rates, which has the max-min fair property

Max-Min Fair Vector?

- a vector $r \in \mathbb{R}^N$ is said to be max-min fair if it is feasible and for each $i \in \{1, \ldots, N\}$, r_i cannot be increased while maintaining feasibility without decreasing r_j for some session j for which $r_j \leq r_i$

Major Objectives of Flow Control

Major Objectives of Flow Control

- limiting average delay
- limiting buffer overflow
- treating sessions fairly