Single-Server Queue

Hui Chen, Ph.D.

Dept. of Engineering & Computer Science

Virginia State University

Petersburg, VA 23806
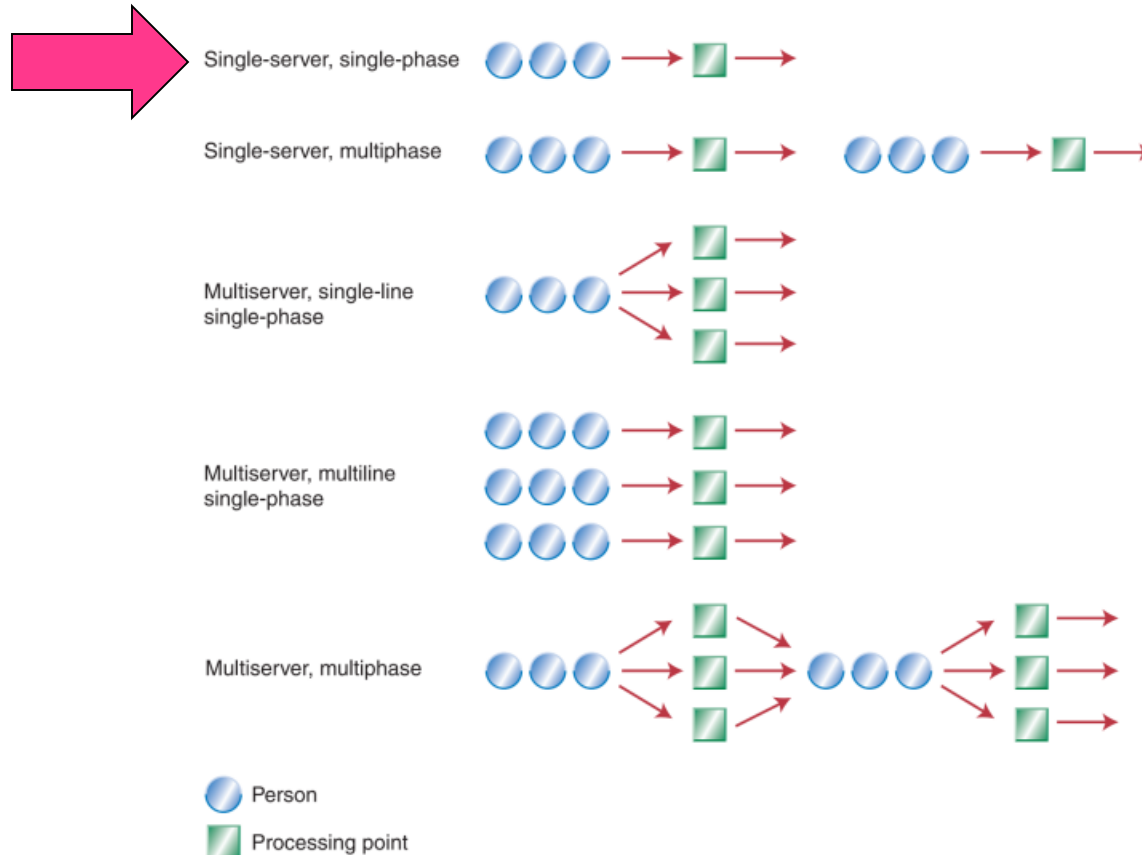
Outline

☐ Discussion on project and paper proposal

☐ Single-server queue

  ■ Concept model

  ■ Specification model

  ■ Simulation model and program

  ■ Numerical examples (Test cases for simulation program)

  ■ Job-averaged statistics

 ■ Time-averaged statistics

  ■ Applications

Single-Server Queue

◻ A single-server service node consists of a server plus its queue

◻ Example Applications
  ▪ Switches & routers
    ◻ Telephony switching
    ◻ Frame/packet forwarding (switching & routing)
  ▪ Blanket paging in PCS
  ▪ Single-CPU server
  ▪ Single elevator building
  ▪ Drive-by restaurant with a single waiter
  ▪ ……

Single-Server Queue

Single-server, single-phase

Single-server, multiphase

Multiserver, single-line single-phase

Multiserver, multiline single-phase

Multiserver, multiphase

Person

Processing point

- From "Dear Mona, Which Is The Fastest Check-Out Lane At The Grocery Store? " by Mona Chalabi, originally appears in Operations Management, 5th Edition by "R. Dan Reid, Nada R. Sanders", 2010

System Diagram

arrivals → queue | server → departures

service node
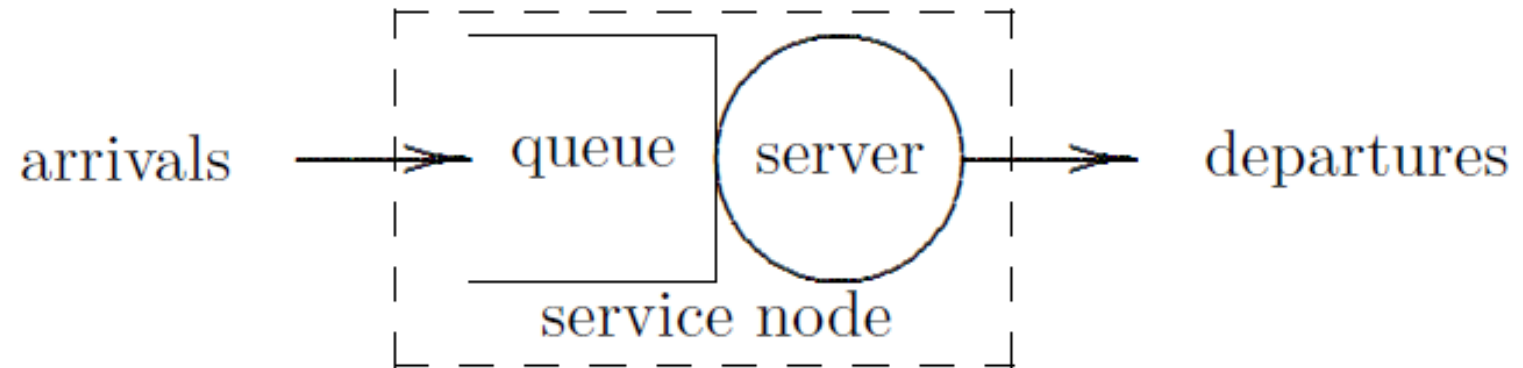
Queue and Service Model

- ◻ Queue
  - ◼ Queuing discipline: how to select a job from the queue
    - ◻ FIFO/FCFS: first in, first out/first come, first serve
    - ◻ LIFO: last in, first out
    - ◻ SIRO: serve in random order
    - ◻ Priority: e.g., shortest job first (SJF)
  - ◼ Capacity
  - ◼ Unless otherwise noted, assume FIFO with *infinite* queue capacity
- ◻ Service model
  - ◼ Non-preemptive
    - ◻ Once initiated, service of job will continue until completed
  - ◼ Conservative
    - ◻ Server will never remain idle if there is any job in the service node

Specification

- *Arrival* time: a_i

- *Delay* in queue (queuing delay): d_i

- Time that service begins: $b_i = a_i + d_i$

- *Service* time: s_i

- *Wait* in the node (total delay): $w_i = d_i + s_i$

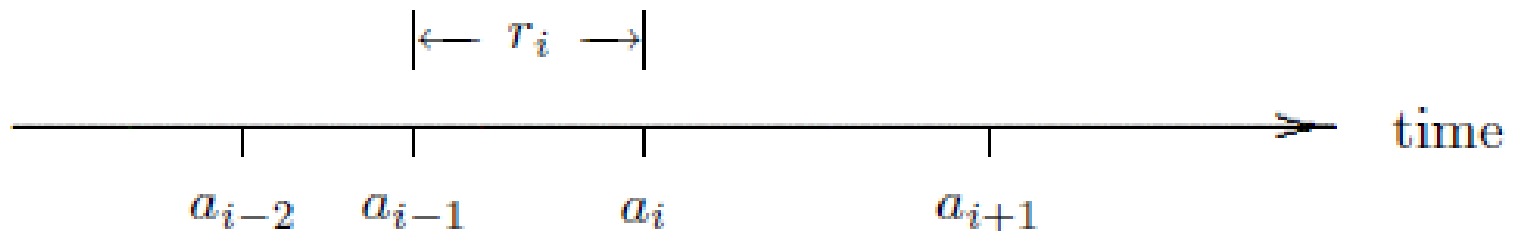- *Departure* time: $c_i = a_i + w_i$
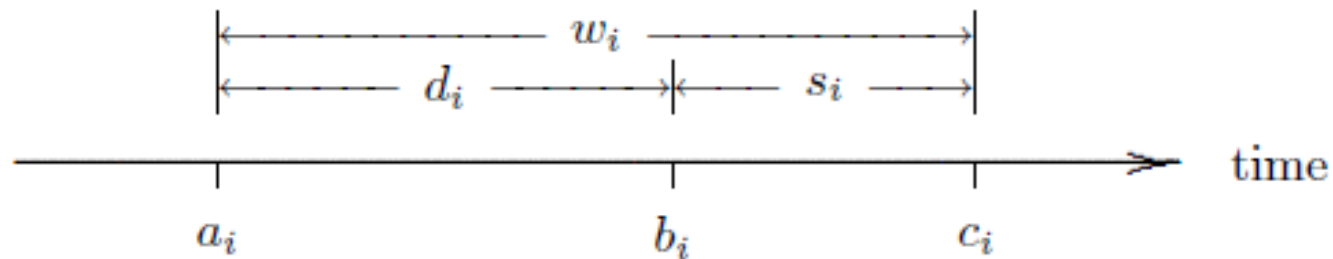
Arrivals

- Inter-arrival time between jobs *i-1* and *i*

  $r_i = a_i - a_{i-1}$

  where $a_i = 0$

- Note

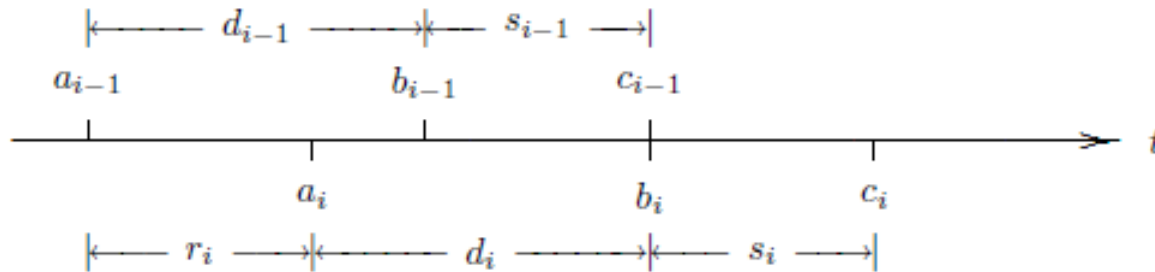 $a_i = a_{i-1} + r_i = r_1 + r_2 + \ldots + r_i$
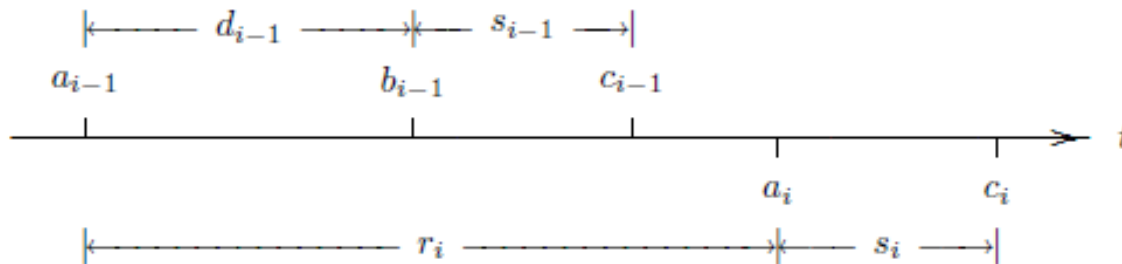
Algorithmic Question

□ Given the arrival times and service times, how may the delay times be computed?

How do jobs experience delay?

□ If $a_i < c_{i-1}$, job i arrives before job $i\text{-}1$ completes

□ If $a_i \geq c_{i-1}$, job i arrives after job $i\text{-}1$ completes

Algorithm 1.2.1 Delay of Each Job (Single-Server FIFO Service Node with Infinite Capacity)

```
c_0 = 0.0;                          /* assumes that a_0 = 0.0 */
i = 0;
while ( more jobs to process ) {
    i++;
    a_i = GetArrival();
    if (a_i < c_{i-1})
        d_i = c_{i-1} - a_i;
    else
        d_i = 0.0;
    s_i = GetService();
    c_i = a_i + d_i + s_i;
}
n = i;
return d_1, d_2, ..., d_n;
```
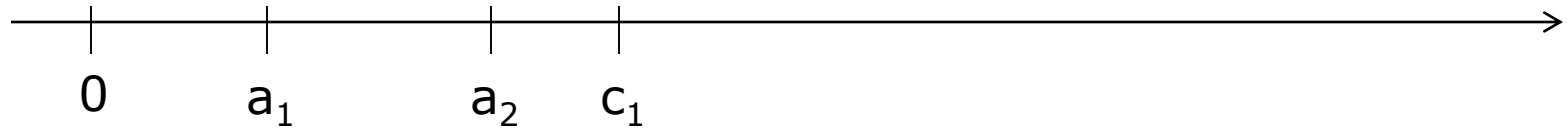
Trace-driven Simulation

- Simulation driven by external data (i.e., a trace)
- Trace can be a running record of a real system

Algorithm 1.2.1 Processing 10 Jobs

i	1	2	3	4	5	6	7	8	9	10
read from file a_i	15	47	71	111	123	152	166	226	310	320
from algorithm d_i	0	11	23	17	35	44	70	41	0	26
read from file s_i	43	36	34	30	38	40	31	29	36	30

☐ Running algorithm manually

■ $a_1 = 15, s_1 = 43, d_1 = ?$

0   a_1 a_2 c_1

■ $a_2 = 47, d_2 = ?$

Output Statistics

- Gain insight from various statistics!
- Examples
  - Job/Customer perspective: waiting time
  - Managing perspective: utilization
- Job-averaged statistics
- Time-average statistics

Job-Averaged Statistics (1)

☐ Average inter-arrival time

$$\bar{r} = \frac{1}{n} \sum_{i=1}^{n} r_i = \frac{a_n}{n}$$

- Arrival rate: inverse of average inter-arrival time

☐ Average service time

$$\bar{s} = \frac{1}{n} \sum_{i=1}^{n} s_i$$

- Service rate: inverse of average service time

Algorithm 1.2.1 Processing 10 Jobs: Exercise L1-1

i	1	2	3	4	5	6	7	8	9	10
read from file a_i	15	47	71	111	123	152	166	226	310	320
from algorithm d_i	0	11	23	17	35	44	70	41	0	26
read from file s_i	43	36	34	30	38	40	31	29	36	30

- Average inter-arrival time?
- Average service time?
- Arrival rate?
- Service rate?
- *What conclusion can you draw from the above statistics?*
  - *Hint: compare arrival rate and service rate*

Job-Averaged Statistics (2)

□ Average delay

$$\overline{d} = \frac{1}{n} \sum_{i=1}^{n} d_i$$

□ Average wait

$$\overline{w} = \frac{1}{n} \sum_{i=1}^{n} w_i$$

□ Since $w_i = d_i + s_i$

$$\overline{w} = \frac{1}{n} \sum_{i=1}^{n} w_i = \frac{1}{n} \sum_{i=1}^{n} (d_i + s_i) = \frac{1}{n} \sum_{i=1}^{n} d_i + \frac{1}{n} \sum_{i=1}^{n} s_i = \overline{d} + \overline{s}$$
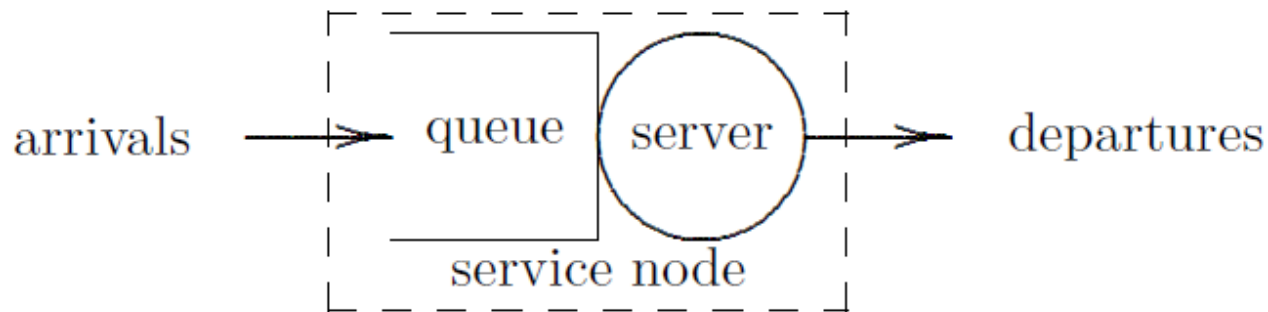
Algorithm 1.2.1 Processing 10 Jobs: Exercise L1-2

i	1	2	3	4	5	6	7	8	9	10
read from file a_i	15	47	71	111	123	152	166	226	310	320
from algorithm d_i	0	11	23	17	35	44	70	41	0	26
read from file s_i	43	36	34	30	38	40	31	29	36	30

- Average delay?
- Average wait?
- Consistency check (part of verification)

$$\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i = \frac{1}{n}\sum_{i=1}^{n}(d_i + s_i) = \frac{1}{n}\sum_{i=1}^{n} d_i + \frac{1}{n}\sum_{i=1}^{n} s_i = \overline{d} + \overline{s}$$
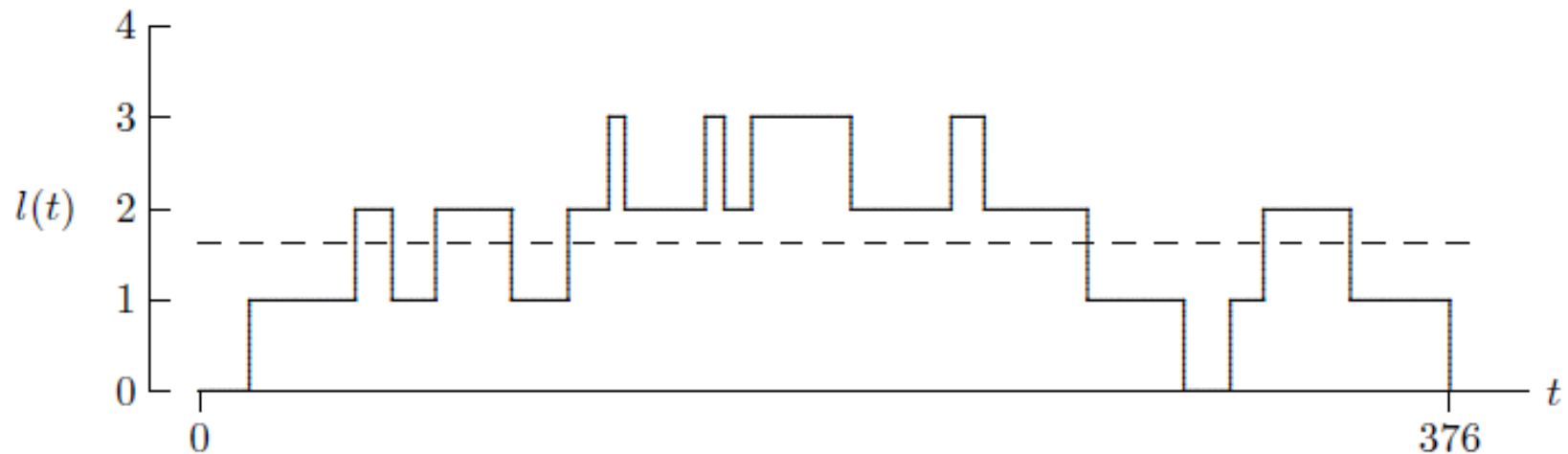
Time-Averaged Statistics (1)

❑ Defined by the area under a curve (integral)

❑ Single-Server Queue: Start with *statistics at time t*

■ *l(t)*: number of jobs in the service node at time *t*

■ *q(t)*: number of jobs in the queue at time *t*

■ *x(t)*: number of jobs in service at time *t*

❑ By definition: *l(t) = q(t) + x(t)*

Time-Averaged Statistics: Example of *l(t)*

i	1	2	3	4	5	6	7	8	9	10
read from file a_i	15	47	71	111	123	152	166	226	310	320
from algorithm d_i	0	11	23	17	35	44	70	41	0	26
read from file s_i	43	36	34	30	38	40	31	29	36	30

Time-Averaged Statistics (2)

- ❑ Defined by the area under a curve (integral)
    - ■ Over the time interval $(0, \tau)$ the time-averaged number in the node

$$\bar{l} = \frac{1}{\tau} \int_0^\tau l(t)dt$$

    - ■ Over the time interval $(0, \tau)$ the time-averaged number in the queue

$$\bar{q} = \frac{1}{\tau} \int_0^\tau q(t)dt$$

    - ■ Over the time interval $(0, \tau)$ the time-averaged number in service

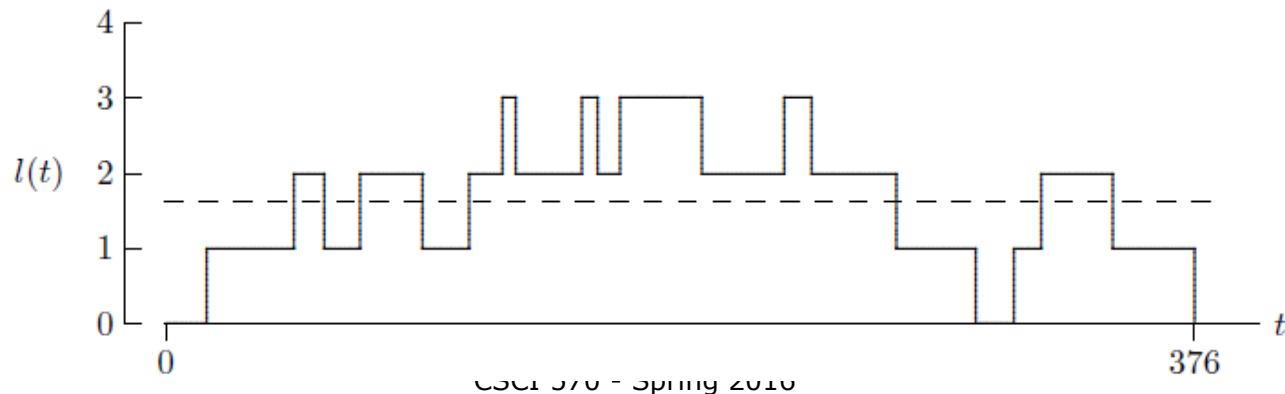$$\bar{x} = \frac{1}{\tau} \int_0^\tau x(t)dt$$

Time-Averaged Statistics (3)

- ❑ Defined by the area under a curve (integral)

  - ■ Over the time interval $(0, \tau)$

  $$\bar{l} = \frac{1}{\tau} \int_0^\tau l(t)dt \qquad \bar{q} = \frac{1}{\tau} \int_0^\tau q(t)dt \qquad \bar{x} = \frac{1}{\tau} \int_0^\tau x(t)dt$$

  - ■ Since $l(t) = q(t) + x(t)$ for all $t > 0$,

  $$\bar{l} = \bar{x} + \bar{q}$$

CSCI 570 - Spring 2016

Job-Averaged and Time-Averaged Statistics

- Little's Equations
- If
  - (a) queue discipline is FIFO
  - (b) service node capacity is infinite, and
  - (c) service is idle both at $t=0$ and $t=c_n$,
- Then

$$\int_0^{c_n} l(t)dt = \sum_{i=1}^{n} w_i$$

$$\int_0^{c_n} q(t)dt = \sum_{i=1}^{n} d_i$$

$$\int_0^{c_n} x(t)dt = \sum_{i=1}^{n} s_i$$

Exercise L1-3

i	1	2	3	4	5	6	7	8	9	10
read from file a_i	15	47	71	111	123	152	166	226	310	320
from algorithm d_i	0	11	23	17	35	44	70	41	0	26
read from file s_i	43	36	34	30	38	40	31	29	36	30

◻ Using Little's Equations to calculate

$$\bar{q}$$

$$\bar{l}$$

$$\bar{x}$$

Server Utilization

□ Sever utilization: time averaged number in service

■ Represents probability that the server is busy

$$\overline{x} = \frac{1}{\tau}\int_0^\tau x(t)dt$$

Traffic Intensity

□ Traffic intensity: ratio of arrival rate to service rate

$$\frac{1/\bar{r}}{1/\bar{s}} = \frac{\bar{s}}{\bar{r}} = \frac{\bar{s}}{a_n/n} = \left(\frac{c_n}{a_n}\right)\bar{X}$$

Large Trace?

◻ Write a program!

◻ Sample programs

 ◼ C/C++ version

 ◼ Java version

Case Study

- ❑ Sven and Larry's Ice Cream Shoppe
  - ■ Owners considering adding new flavors and cone options
  - ■ Concerned about resulting service times and queue length
- ❑ Can be modeled as a single-server queue
  - ■ ssq1.dat represents 1000 customer interactions
  - ■ Direct consequence of adding new flavors and cone options
    - ❑ Service time per customer increases
  - ■ What's the consequence?

Ice Cream Shoppe
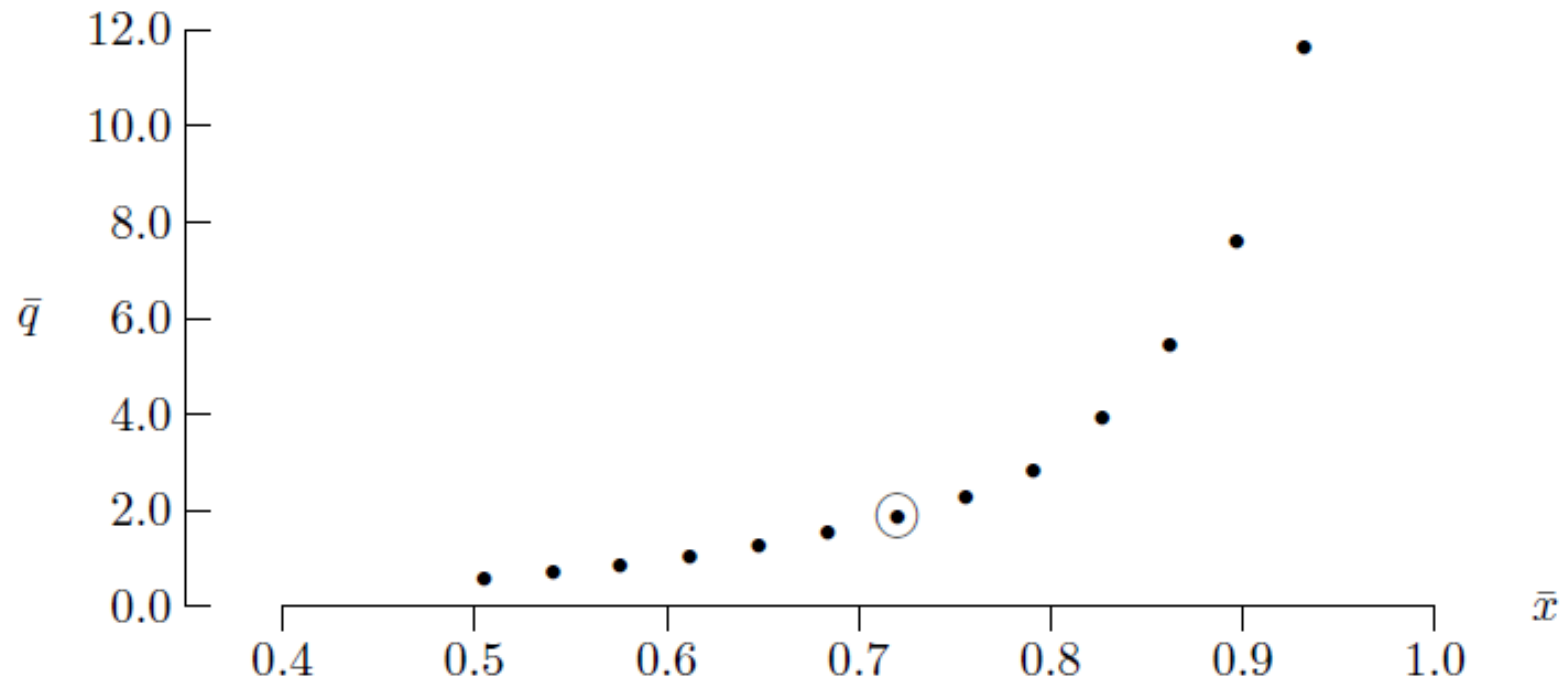
Exercise: L1-4

□ Run either C/C++ or Java program against the trace, submit the result.

Exercise: L1-5

- Modify program ssq1 to output the additional statistics

$$\bar{q} \qquad \bar{l} \qquad \bar{x}$$

- As in the case study (Sven and Larry's Ice Cream Shoppe), use this program to compute a table of the above three statistics for the traffic intensities that are 0.6, 0.7, 0.8, 0.9, 1.0, 1.1 and 1.2 times of original one in the input file

- Illustrate your result using either Matlab/Octave or Excel.

Summary

- □ Look for paper and form project topic
- □ Single-server queue
  - Concept model
  - Specification model
  - Simulation model and program
  - Numerical examples (Test cases for simulation program)
  - Job-averaged statistics
  - Time-averaged statistics
  - Applications