IEOR 4615: Service Engineering

Ward Whitt

Lecture 2, January 22

Stochastic Queueing Models for Service Systems viewed as Queueing Systems

Steps in **Narrowing Focus**
Steps in Abstraction

1. Service Systems

2. Queueing Systems

3. Stochastic Queueing Models

- Analyze Performance
  - Mathematical solution
  - Simulate

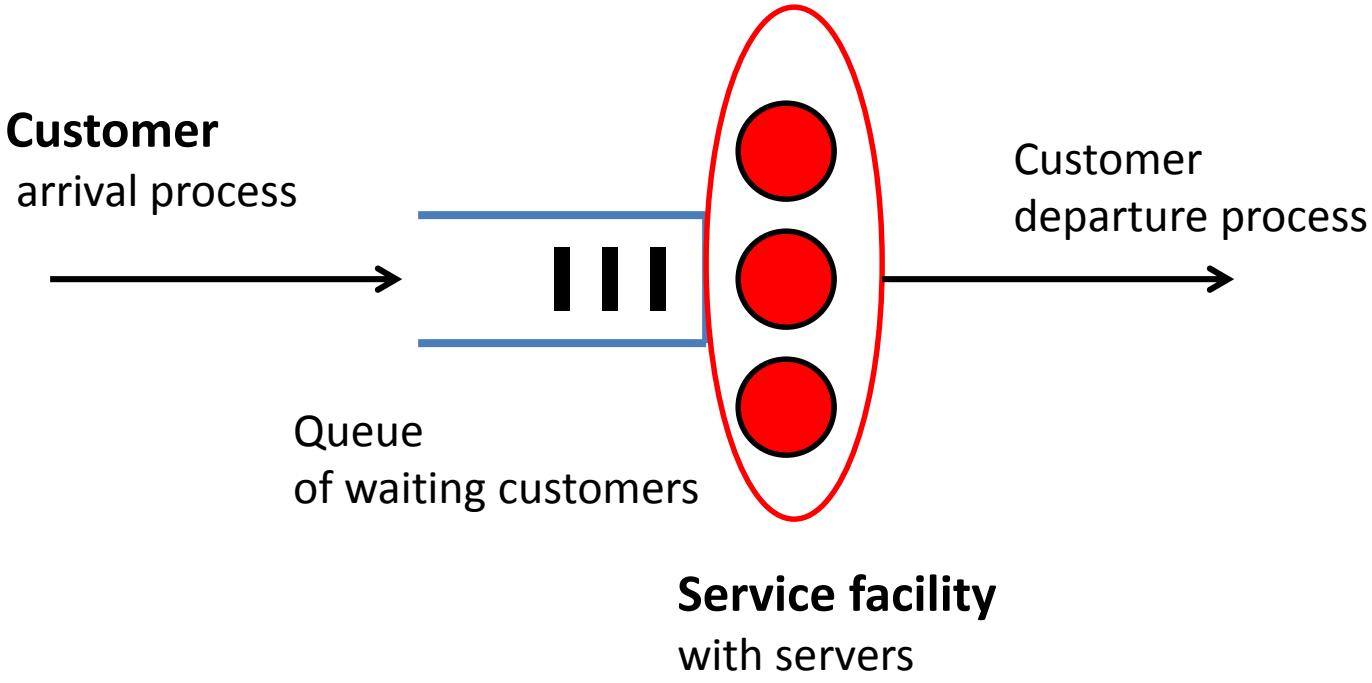
Reading (posted):
Chapter 1: Introduction

Queueing Methods for Services and Manufacturing

Randolph W. Hall, 1991

Recommended, but not required

Elements of Queueing Systems



Elements of Queueing Systems

- **Customer Characteristics**
  - Customer Classes (1 or more?)
    - Routing (if network of queues)
  - Arrival process
  - Abandonment Behavior
    - Balking
    - Reneging
    - Jockeying
- **The Service Mechanism and Queue**
  - Number and configuration of service facilities
  - Number of servers in each
  - Service-time distributions
  - Service Discipline

Measures of Performance

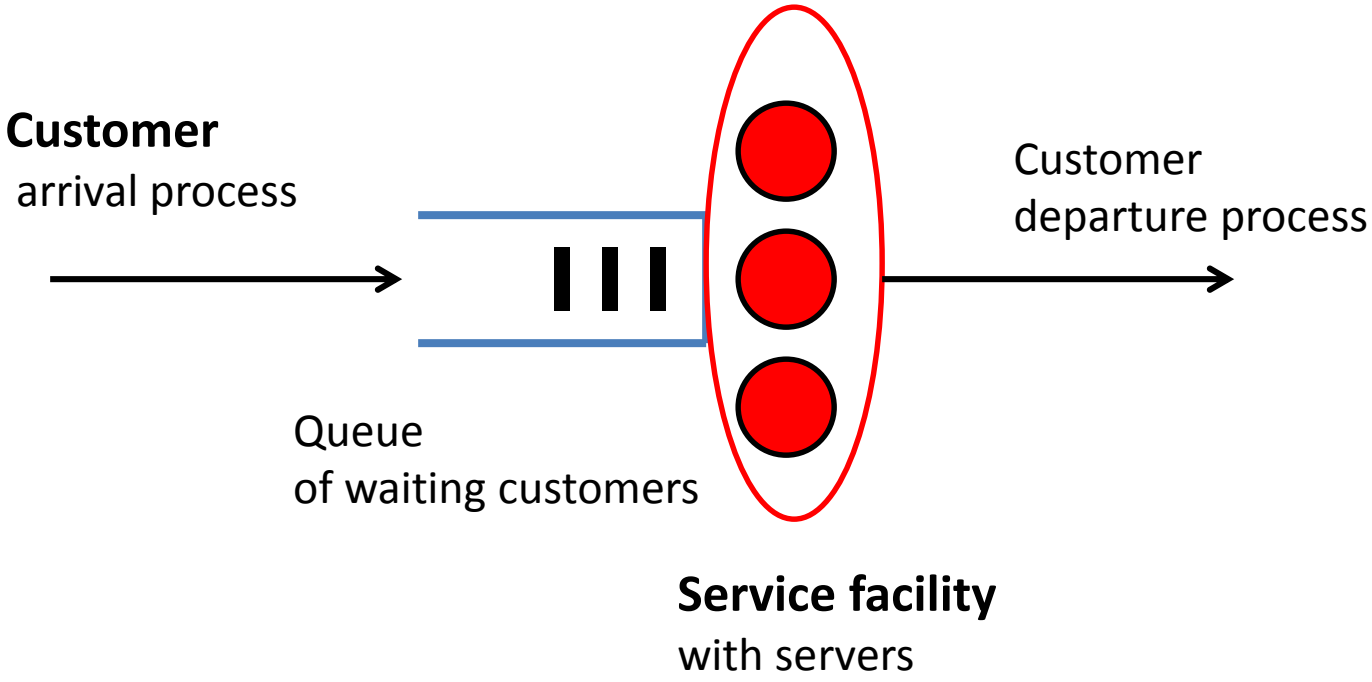
- **Randomness** (need stochastic models and statistics)
  - Arrivals, service times and abandonment
- **Different Points of View**
  - Customer View (Benefit, needs met, revenue)
    - Quality of service received ("first-call resolution")
    - Waiting time (until start service), service time, response time (from arrival to departure) - distributions
  - Service Agent View
    - Job satisfaction, not overworked
  - Management View (Efficiency: Cost)
    - Throughput (rate of completed services)
    - Utilization of servers (service capacity not wasted)

Formalized Measures of Performance

- Service Level Agreements: **Call centers**
  - P(Wait < 20 seconds) > 0.80 (80/20 rule)
  - P(Abandonment) < 0.05
- Regulations: **Hospitals** (UK, Singapore), Call Centers (Israel)
  - P(Wait to be assigned to bed > 6 hours) < 0.05

Measure are **indicators** of the operational status of the system, and **correlated** to higher level goals

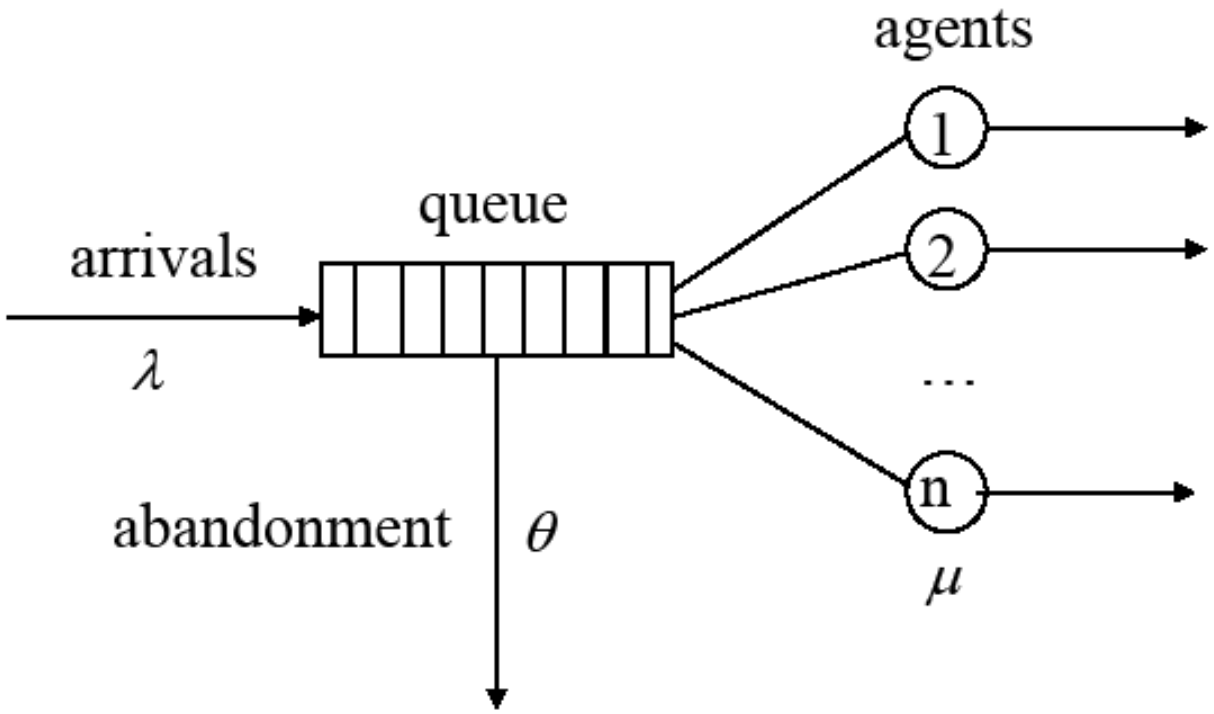
Basic Queueing Model



One queue with multiple homogeneous servers working in parallel

A Large Call Center

The Erlang A Model (*M*/*M*/*n*+*M*)
M for "Markov" – special assumptions

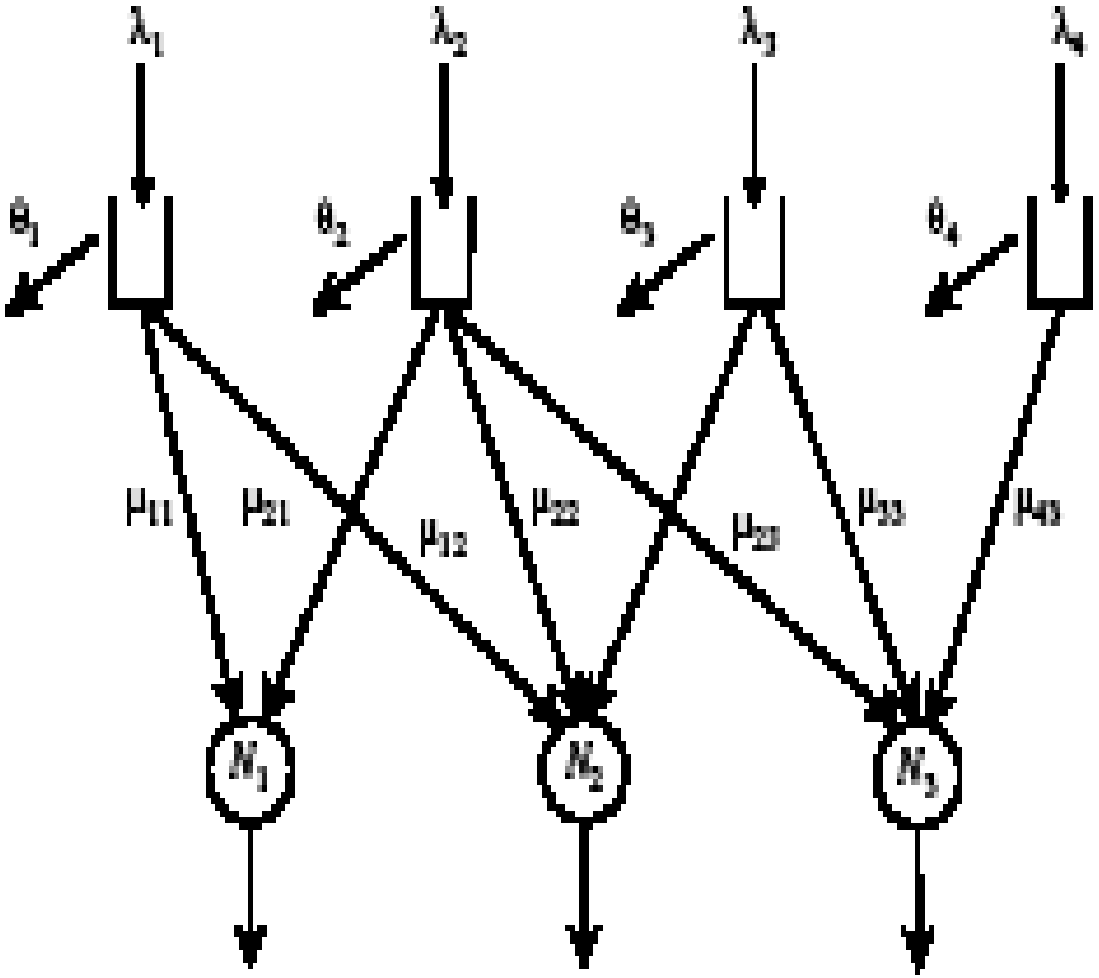


Stochastic Model Heierarchy

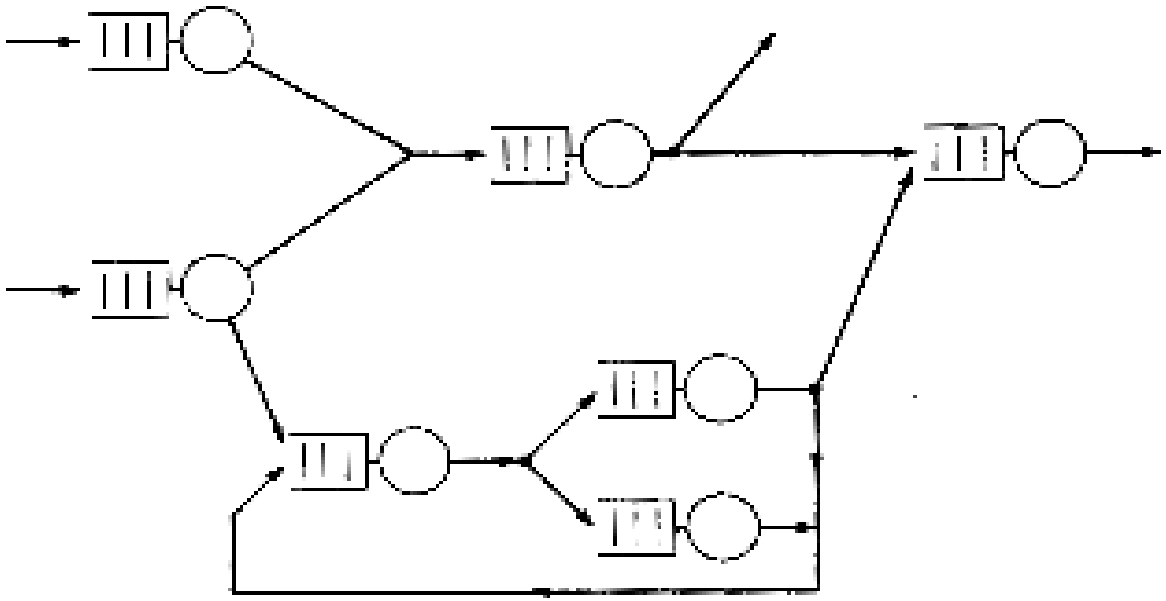
- Continuous-Time Markov Chains
  - See Ross (2014) and Whitt (2014)
- Birth-and-Death Stochastic Processes
  - Sec 6.3 of Ross (2014) and Sec 4 of Whitt (2014)
- Erlang Models
  - Erlang B (loss) M/M/n/0
  - Erlang C (delay) M/M/n/∞
  - Erlang A (abandonment) M/M/n+M
    - Poisson arrival process
    - IID exponential service times
    - IID exponential patience times
      - (IID = independent and identially distributed)

More Complex Queueing Models with Network Structure

Call Center with Multiple Customer Classes
Skill-Based Routing



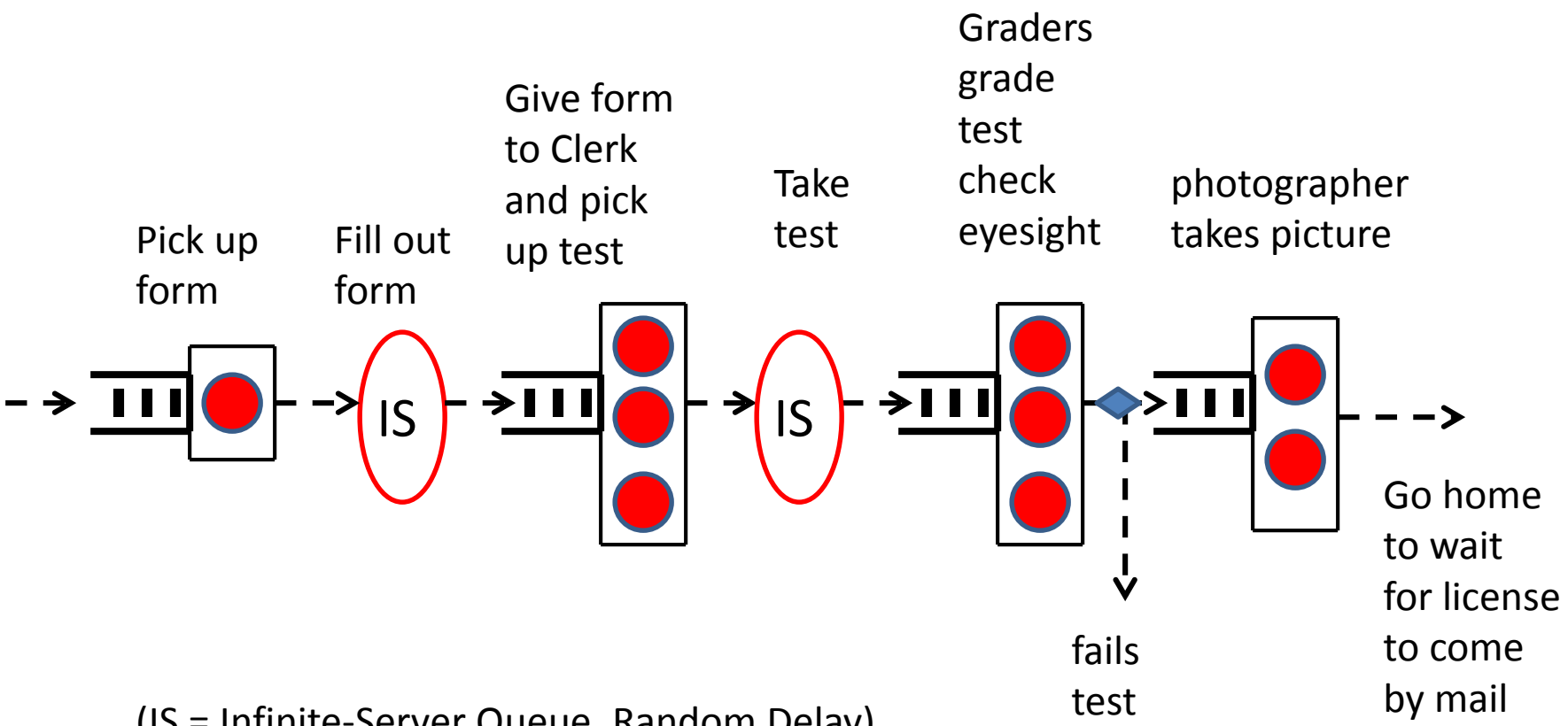
A Network of Queues



Getting Driver's License

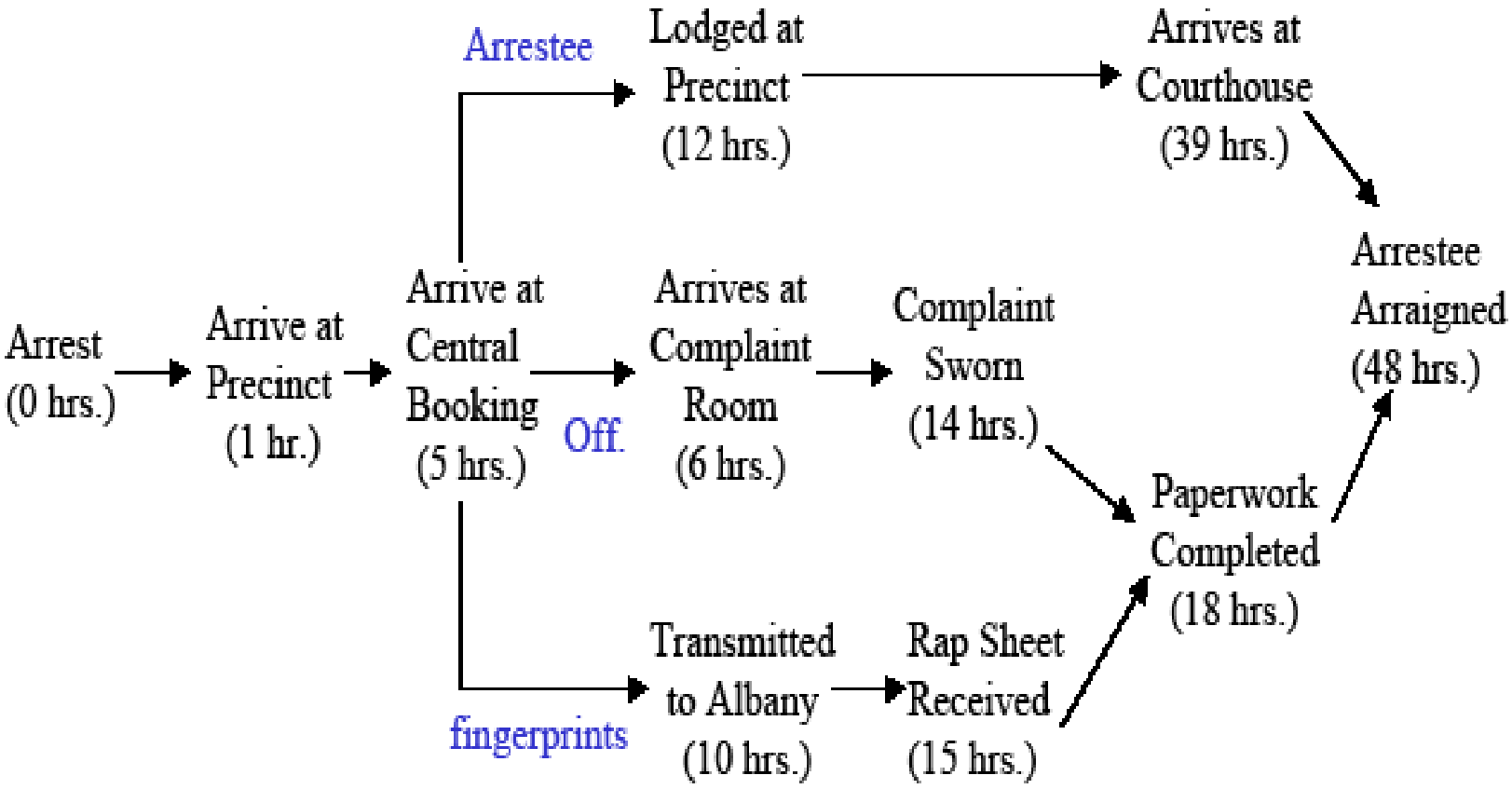
Department of Motor Vehicles (DMV)

A Network of Queues (Hall, Ch. 1)

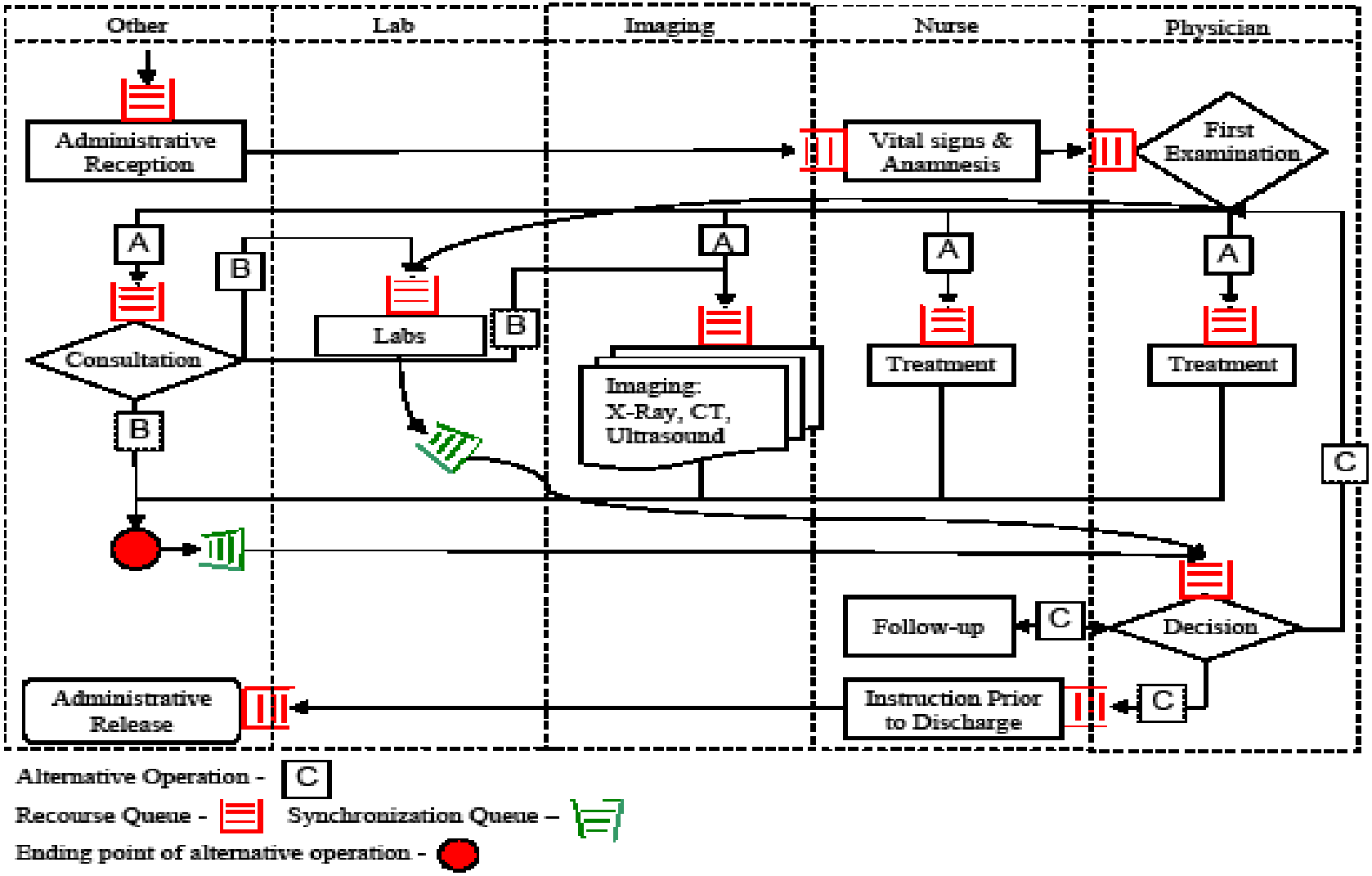


(IS = Infinite-Server Queue, Random Delay)

Arrest-to-Arraignment Process
Larson et al. (1993)



Patient Flow in an Emergency Department



References

- Hall, R. W. (1991) ***Queueing Models for Services and Manufacturing***, Prentice Hall, Ch. 1.
- Larson, R. C., M. F. Cahn, M. C. Shell (1993) Improving the New York City Arrest-to-Arraignment System. ***Interfaces***, vol. 23, 76-96.
- Ross, S. M. (2014) ***Introduction to Probability Models***, 11th ed., textbook for IEOR 3106/4106.
- Whitt, W. (2014) Lecture Notes on Continuous-Time Markov Chains, Dec. 6, 2014. (Used in IEOR 3106, posted in the readings on Courseworks)
- Whitt, W. (2014) Service Systems as Queueing Systems. Lecture notes for this class.

Another Reference

- Mor Armony, Shlomo Israelit, Avishai Mandelbaum, Yariv N. Marmor, Yulia Tseytlin, and Galit B. Yom-Tov, Patient Flow in Hospitals: A Data-Based Queueing-Science Perspective, Working paper, 2014. See page 17 of the longer 2011 draft for the network shown in slide 17. Both drafts are on the surveys link on the course web page: http://www.columbia.edu/~ww2040/4615S15/IEOR4615S15.html