LECTURE #14
M/G/1

204528

Queueing Theory and Applications in Networks

Assoc. Prof. Anan Phonphoem, Ph.D. (รศ.ดร. อนันต์ ผลเพิ่ม)

Computer Engineering Department, Kasetsart University

Outline

- More on M/G/1
- Busy period and its duration
- M/G/1 with Vacations

M/G/1

$$W_q = \frac{\lambda \overline{X^2}}{2(1-\rho)}$$

$$N_q = \frac{\lambda^2 \overline{x^2}}{2(1-\rho)}$$

$$W_T = \overline{X} + \frac{\lambda \overline{X^2}}{2(1-\rho)}$$

$$N_T = \rho + \frac{\lambda^2 \overline{x^2}}{2(1-\rho)}$$

The coefficient of variation (C_V)

- A normalized measure of dispersion of a probability distribution
- Defined as the ratio of the standard deviation to the mean

$$C_v = \frac{\sigma}{\mu}$$

- Only defined for *non-zero* mean
- C_V should only be computed for data measured on a **ratio scale**

dispersion = การกระจาย

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

The coefficient of variation (C_V)

- Example (from wikipedia)
  - For a group of temperatures
  - An object changes its temperature by 1 K also changes its temperature by 1 C
  - The standard deviation does not depend on whether the Kelvin or Celsius scale
  - However, the mean temp would differ in each scale by 273
  - So, the coefficient of variation would differ
- Investment Dictionary (http://www.answers.com/topic/coefficient-of-variation)
  - The lower the ratio, the better your risk-return tradeoff
  - The higher the ratio, the higher the risk

The coefficient of variation (C_V)

http://en.wikipedia.org/wiki/Coefficient_of_variation

- In queueing theory
  - **Exponential Dist.** is often more important than the **Normal Dist.**
- For $C_V = 1$
  - E.g. Exponential distribution ➔ ($\sigma = \mu$)
- For $C_V < 1$ ➔ low-variance
  - E.g. Erlang distribution [r-stage Erlangian server (E_r)]
- For $C_V > 1$ ➔ high-variance
  - E.g. Hyper-Exponential distribution [r-Stage Parallel Servers (H_r)]
- Some formulas are expressed using **squared coefficient of variation** (SCV)

More on P-K Formula

From: J. Virtamo, 38.3143 Queueing Theory / The *M/G/1* queue

Squared coefficient of variation C_v^2

$$C_v^2 = \frac{\mathrm{Var}[X]}{\left(\overline{X}\right)^2}$$

$$\overline{X^2} = \mathrm{Var}[X] + \left(\overline{X}\right)^2$$

$$= (1 + C_v^2) \bullet \left(\overline{X}\right)^2$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

More on P-K Formula

$$W_q = \frac{\lambda \overline{X^2}}{2(1-\rho)} = \frac{1 + C_v^2}{2} \cdot \frac{\rho}{1-\rho} \cdot \overline{X}$$

$$W_T = \overline{X} + \frac{\lambda \overline{X^2}}{2(1-\rho)} = \left(1 + \frac{1 + C_v^2}{2} \cdot \frac{\rho}{1-\rho}\right) \cdot \overline{X}$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

More on P-K Formula

$$N_q = \frac{\lambda^2 \overline{x^2}}{2(1-\rho)} = \frac{1 + C_v^2}{2} \bullet \frac{\rho^2}{1-\rho}$$

$$N_T = \rho + \frac{\lambda^2 \overline{x^2}}{2(1-\rho)} = \rho + \frac{1 + C_v^2}{2} \bullet \frac{\rho^2}{1-\rho}$$

More on P-K Formula

- Mean values depend only on the expectation E[X] and variance Var[X] of the service time distribution but not on higher moments.
- Mean values increase linearly with the variance.
- Randomness, ‘disarray’, leads to an increased waiting time and queue length.
- The formula are similar to those of the *M/M/1* queue;
  - the only difference is the extra factor $\dfrac{1 + C_v^2}{2}$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

M/G/1 Steady-State

From www.cse.msu.edu/~cse807/notes/slides/queueing.ppt

$$W_q = \frac{\lambda(\overline{X^2} + \sigma^2)}{2(1-\rho)}$$

$$N_q = \frac{\lambda^2(\overline{X^2} + \sigma^2)}{2(1-\rho)}$$

$$W_T = \overline{X} + \frac{\lambda(\overline{X^2} + \sigma^2)}{2(1-\rho)}$$

$$N_T = \rho + \frac{\lambda^2(\overline{X^2} + \sigma^2)}{2(1-\rho)}$$

$$p_0 = 1 - \rho$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

M/G/1 Example

- There are two workers competing for a job.
- Pual claims that an average service time is **faster** than Jar's
- But Jar claims to be **more consistent**, if not as fast.
- The arrivals is a Poisson process at a rate of $\lambda = 2$ per hour. (1/30 per minute).
- Paul's service statistics are an average service time of 24 minutes with a standard deviation of 20 minutes.
- Jar's service statistics are an average service time of 25 minutes, but a standard deviation of only 2 minutes.

M/G/1 Example

- If the <u>average length of the queue</u> is the criterion for hiring, which worker should be hired (Paul or Jar)?

M/G/1 Example

- For Paul,

$\lambda = 1/30$ (per min)

$1/\mu = 24$ min.

$\rho = \lambda/\mu = 24/30 = 4/5$

$\sigma^2 = 20^2 = 400$ min^2

$$N_q = \frac{\lambda^2(\overline{X}^2 + \sigma^2)}{2(1-\rho)} = \frac{(1/30)^2\,(24^2 + 400)}{2(1 - 4/5)}$$

$= 2.711$ customers

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

M/G/1 Example

- For Jar,

$\lambda = 1/30$ (per min)

$1/\mu = 25$ min.

$\rho = \lambda/\mu = 25/30 = 5/6$

$\sigma^2 = 2^2 = 4 \text{ min}^2$

$$N_q = \frac{\lambda^2(\overline{X^2} + \sigma^2)}{2(1-\rho)} = \frac{(1/30)^2\,(25^2 + 4)}{2(1 - 5/6)}$$

= 2.097 customers

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

M/G/1 Example

- Although working faster on the average,
  - Paul's greater service variability results in an average queue length about 30% greater than Jar's.
- On the other hand, the proportion of arrivals who would find Paul idle and thus experience no delay is $p_0 = 1 - \rho = 1/5 = 20\%$
- While the proportion who would find Jar idle and thus experience no delay is $p_0 = 1 - \rho = 1/6 = 16.7\%$.
- On the basis of average queue length, Nq,
  - Jar wins.

Busy and Idle period

- To derive the distribution for the M/G/1 queue
  - the length of the Idle period
  - the length of the busy period

Busy and Idle period

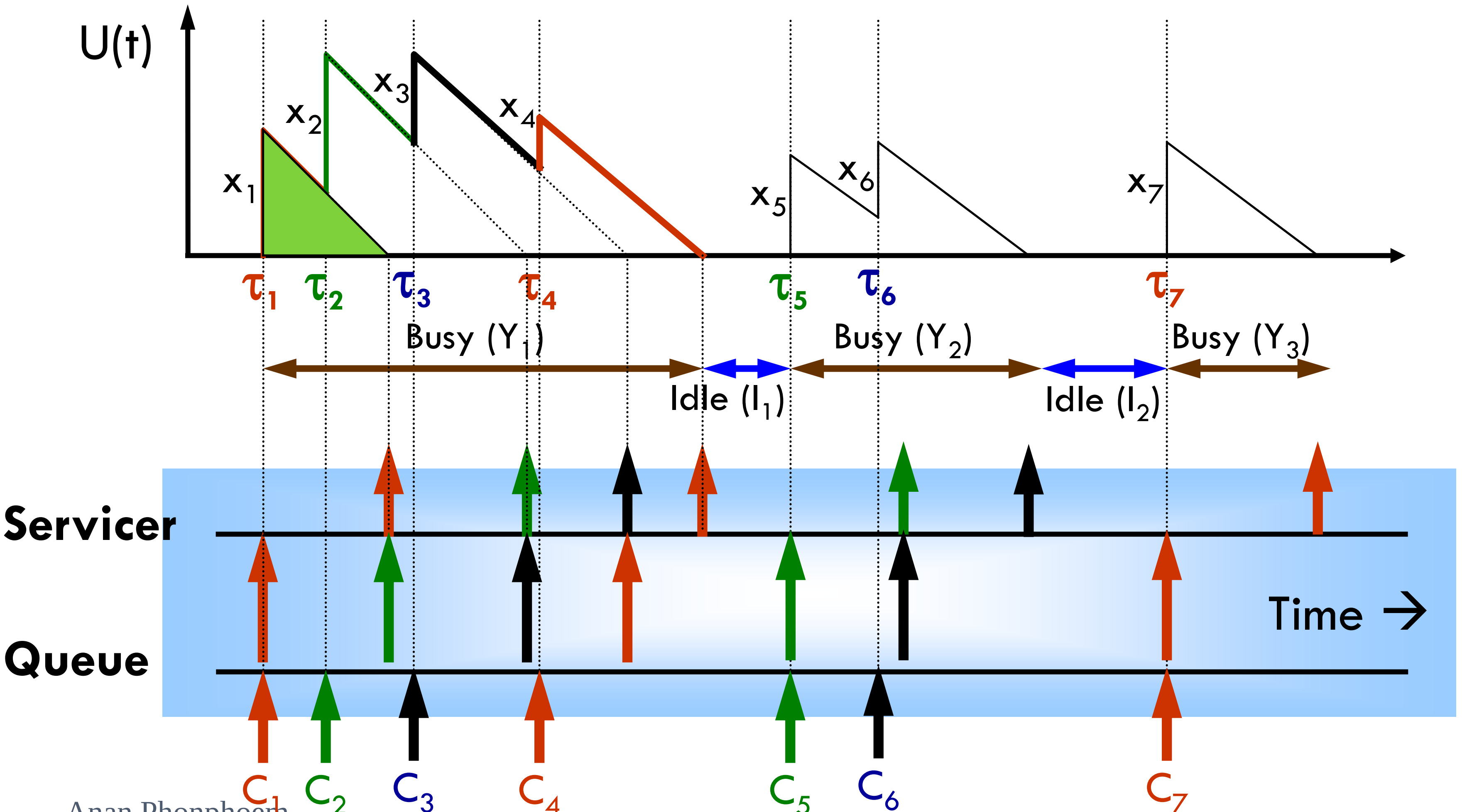
- C_n = the n^{th} customer to enter the system
- τ_n = arrival time of C_n
- $t_n = \tau_n - \tau_{n-1}$ = interarrival time between C_{n-1} and C_n
- x_n = Service time of C_n

Busy and Idle period

- U(t) = Unfinished work in the system
  = Virtual waiting time at time t
- Y_n = Busy period
- I_n = Idle period

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

The Unfinished Work (FCFS)



Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

M/G/1 (FCFS)

- $A(t) = P[t_n \leq t] = 1 - e^{-\lambda t}$ $\quad t \geq 0$
- $B(x) = P[x_n \leq x]$
- A(t) and B(x) are independent on n
- F(y) = Idle period distribution

  $= P[I_n \leq y]$
- G(y) = Busy period distribution

  $= P[Y_n \leq y]$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

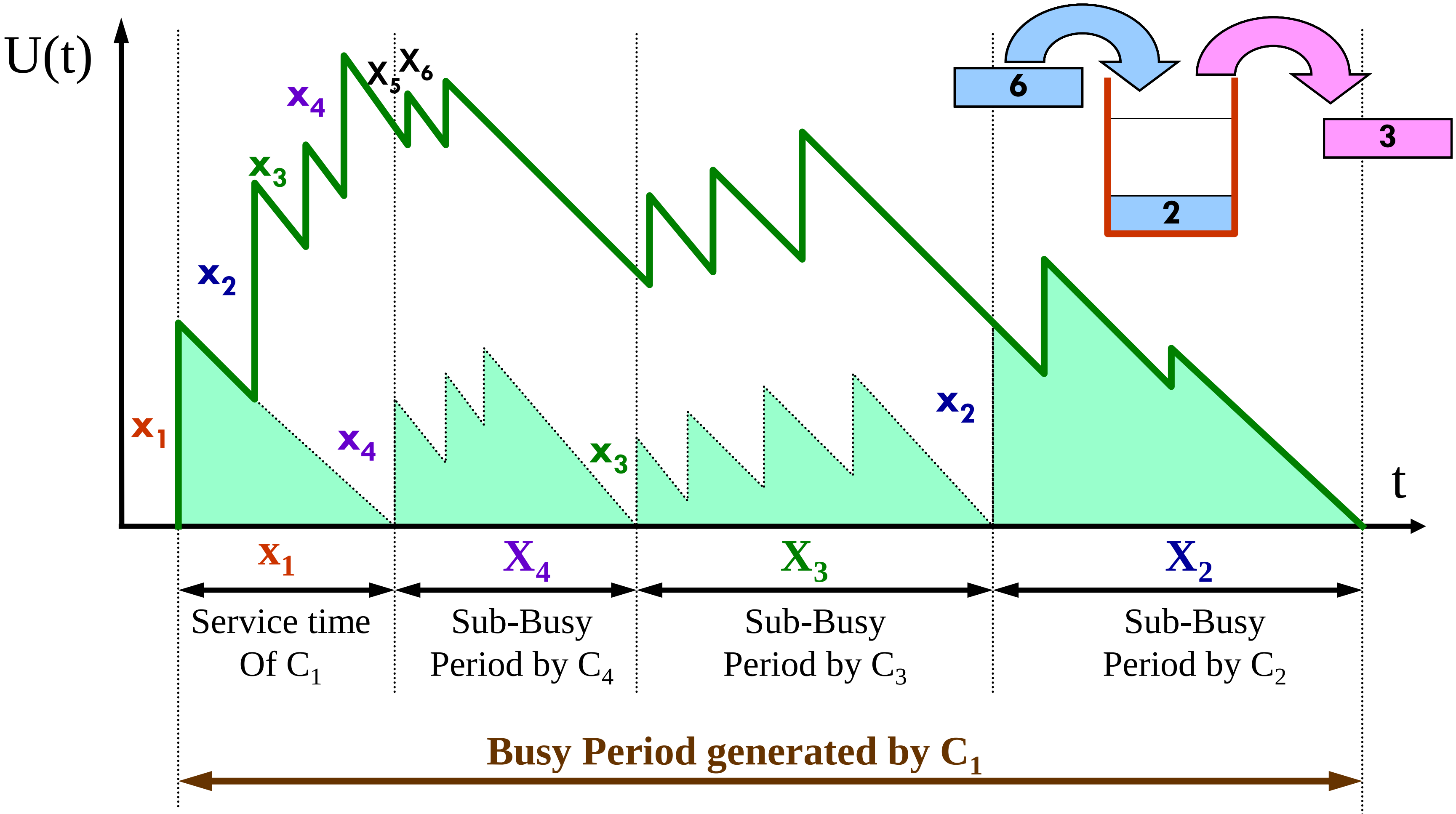
M/G/1 (FCFS)

- For Idle period F(y)
  - After busy period ➔ start the idle period
  - A new idle period will stop immediately when the new customer arrives
  - Therefore, from the memoryless distribution, Idle period distribution F(y)

$$F(y) = 1 - e^{-\lambda y} \qquad y \geq 0$$

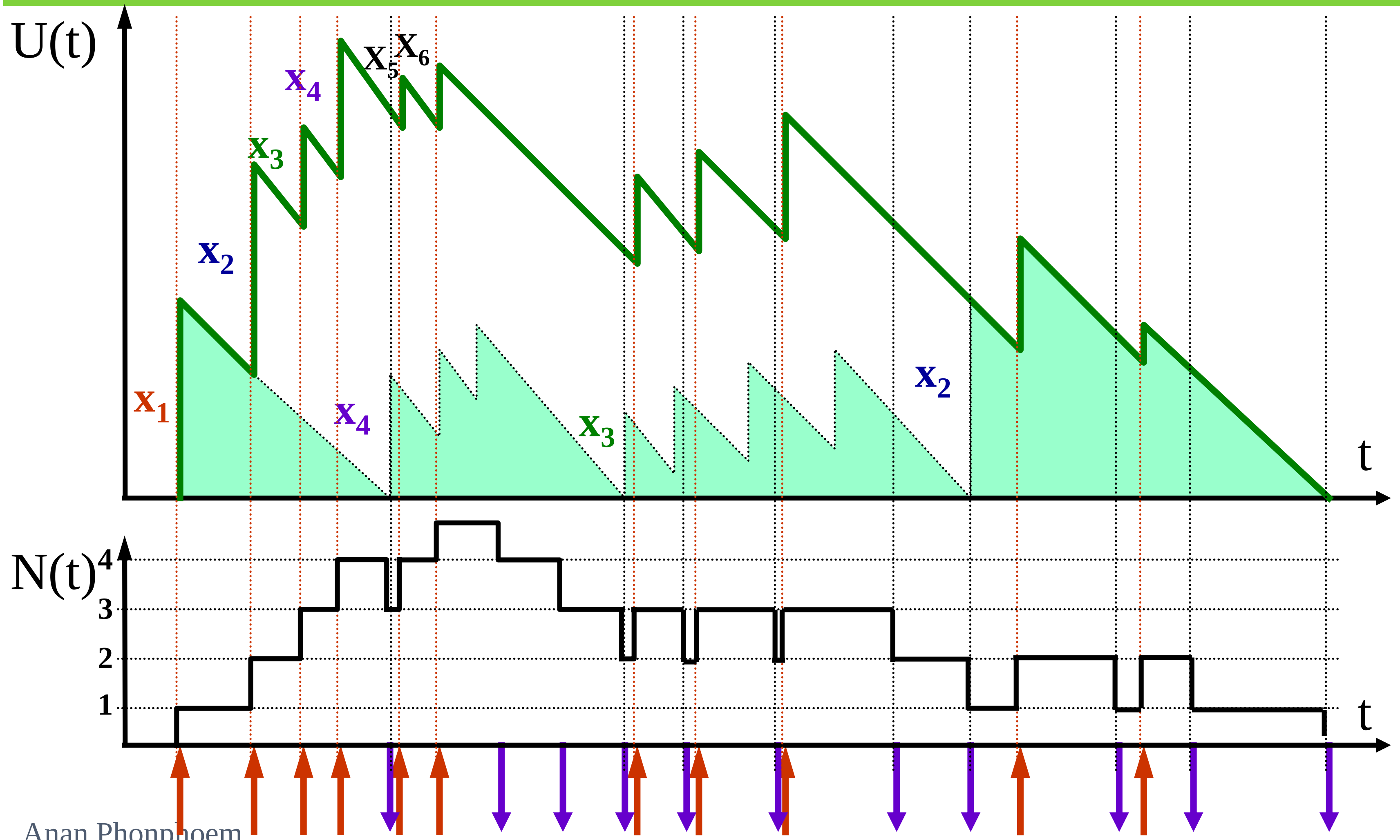
Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

The Unfinished Work (LCFS)



Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Number of Users (LCFS)



Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

M/G/1 (LCFS)

- Each sub-busy period behaves statistically the same as the major busy period
- The duration of busy period Y

$$Y = X_1 + X_{v+1} + X_{v+2} + \ldots + X_3 + X_2$$

- X_v = sub-busy period
- v = an RV = # of customer arrives during C_1 service interval

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

M/G/1 (LCFS)

- For Busy period G(y)

  $G(y) = P[Y_n \leq y] \qquad\qquad y \geq 0$

- Transform of M/G/1 busy-period distribution

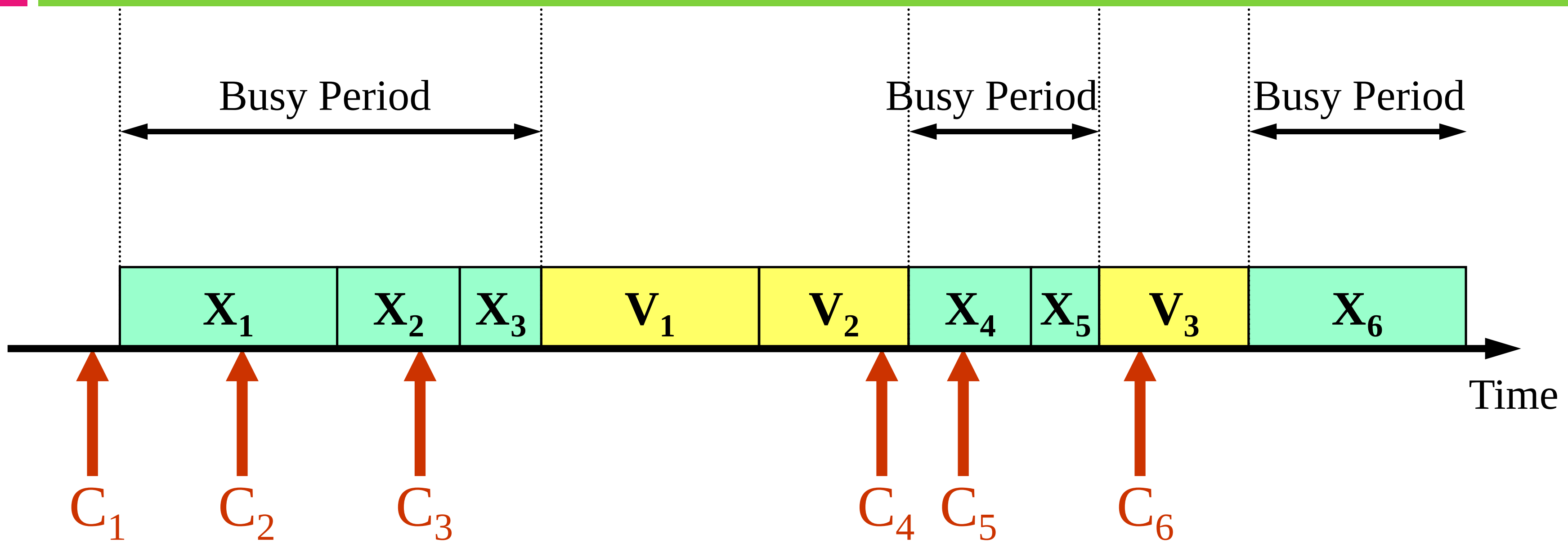
  $G^*(s) = B^*[s + \lambda - \lambda G^*(s)]$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

M/G/1 with Vacations

- At the end of busy period
  - The server goes on “vacation”
  - The vacation period = random interval of time
  - A new arrive during vacation has to wait until the end of vacation period
  - If the system is idle after vacation, a new vacation starts right away

M/G/1 with Vacations



- V_n = Vacation period with $\overline{V}$ and $\overline{V^2}$
  = IID random variable and independent of customer interarrival and service time
- X_n = Service period

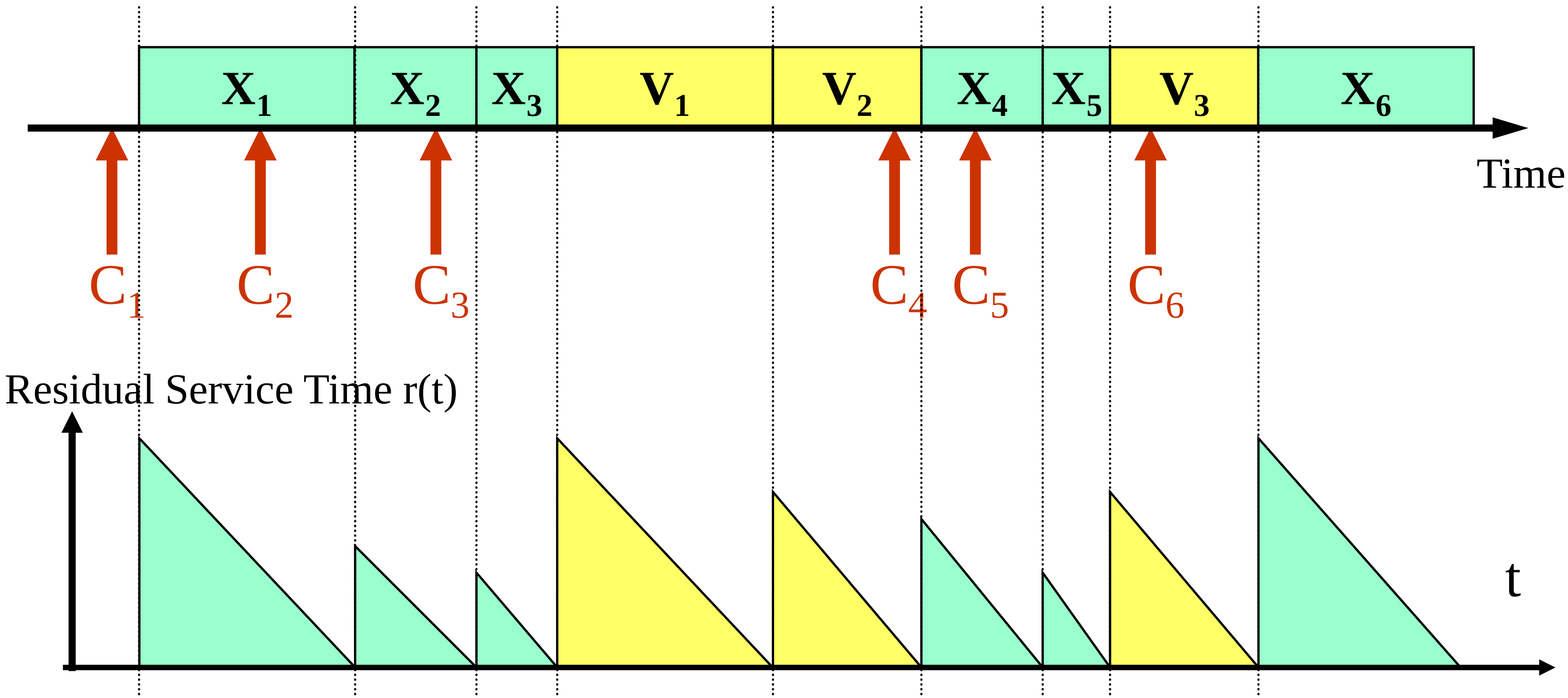
M/G/1 with Vacations

- A new customer is Poisson arrival and service time is general distribution
- The waiting time for customer is W

$$W = \frac{R}{1 - \rho}$$

- R = Residual Time

M/G/1 with Vacations



Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

M/G/1 with Vacations

$$R = \frac{1}{t}\int_0^t r(\tau)\,d\tau = \frac{1}{t}\sum_{i=1}^{M(t)} \frac{1}{2}X_i^2 + \frac{1}{t}\sum_{i=1}^{L(t)} \frac{1}{2}V_i^2$$

$$= \frac{M(t)}{t}\,\frac{\sum_{i=1}^{M(t)} \frac{1}{2}X_i^2}{M(t)} + \frac{L(t)}{t}\,\frac{\sum_{i=1}^{L(t)} \frac{1}{2}V_i^2}{L(t)}$$

M/G/1 with Vacations

$$t \to \infty, \qquad \frac{M(t)}{t} = \lambda \quad and \quad \frac{L(t)}{t} = \frac{(1-\rho)}{\overline{V}}$$

$$R = \frac{1}{2}\lambda\overline{X^2} + \frac{(1-\rho)\overline{V^2}}{2\overline{V}}$$

$$W = \frac{\lambda\overline{X^2}}{2(1-\rho)} + \frac{\overline{V^2}}{2\overline{V}}$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Examples

- Slotted M/D/1
- FDM
- Slotted FDM
- TDM

Slotted M/D/1

Modified from "Eytan Modiano", MIT

- Packet transmission starts only @beginning of the slot
- Each slot, only one packet will be transmitted with rate $1/\mu$
- If no packet is waiting in queue, the slot is idle (Server is on vacation)

Slotted M/D/1

$$W = \frac{\lambda \overline{X^2}}{2(1-\rho)} + \frac{\overline{V^2}}{2\overline{V}}$$

- $\overline{X} = \overline{V} = 1/\mu$
- $\overline{X^2} = \overline{V^2} = 1/\mu^2$

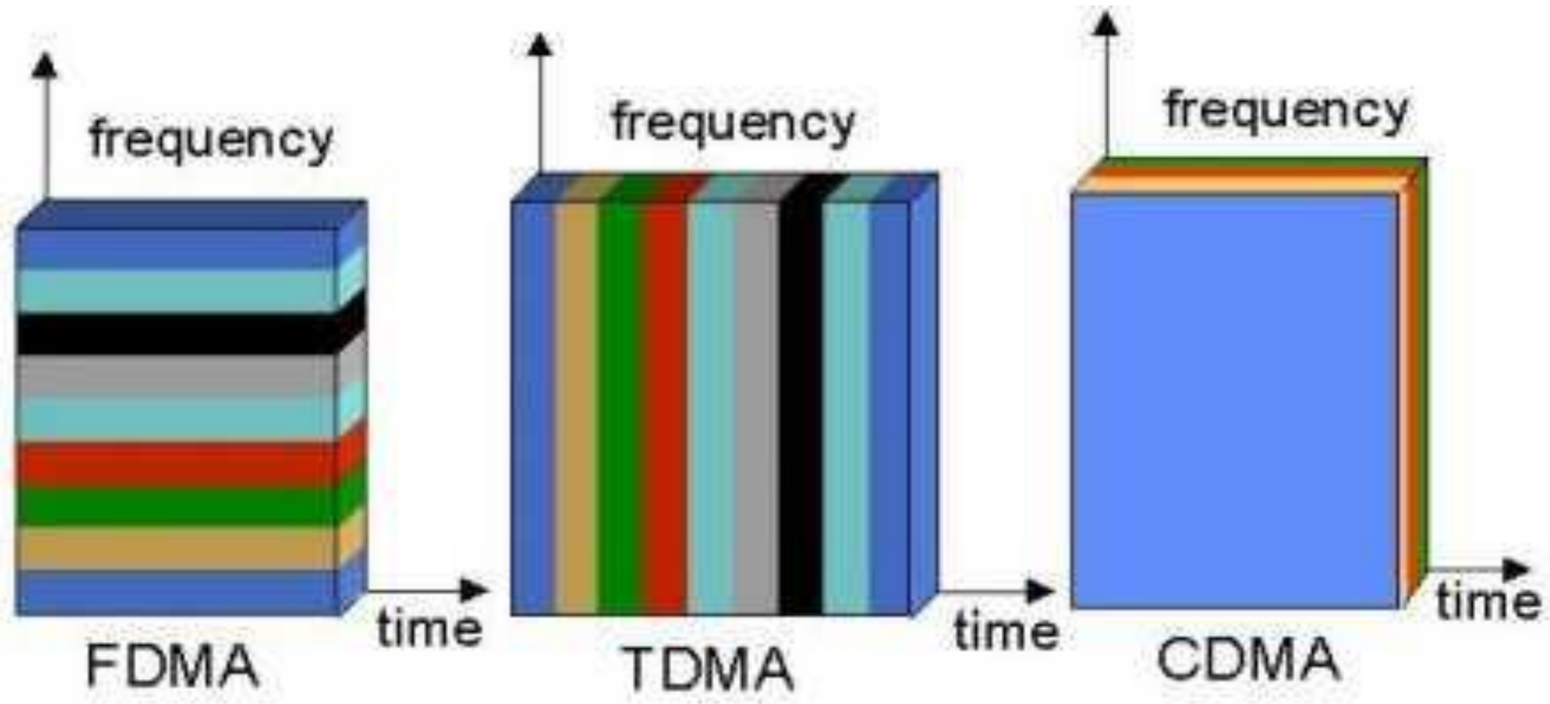
$$W = \frac{\lambda/\mu^2}{2(1-\lambda/\mu)} + \frac{1/\mu^2}{2/\mu}$$

$W_{M/D/1}$

$$\frac{1/\mu}{2} = \frac{\overline{X}}{2}$$

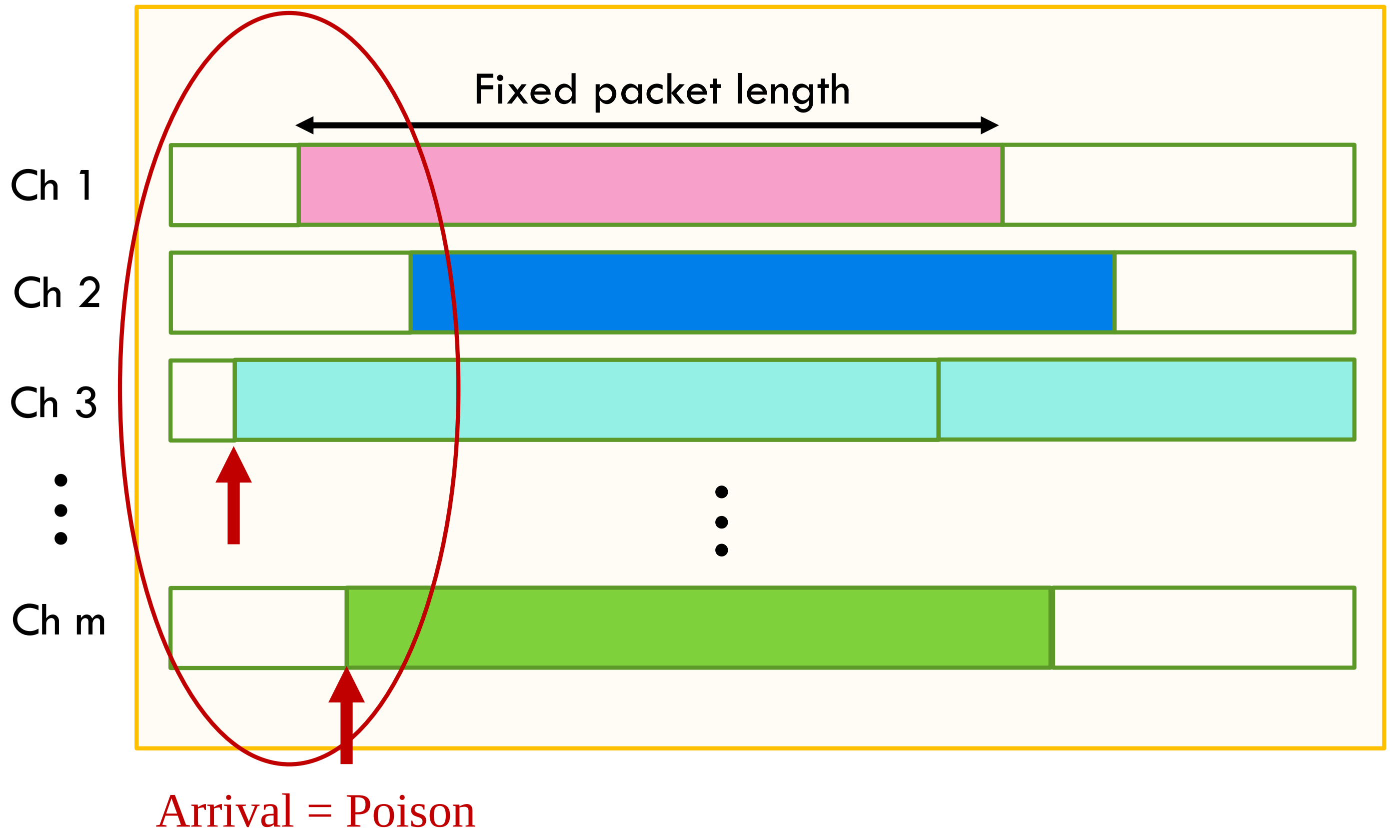
Wait for half slot

FDM, TDM, and CDMA



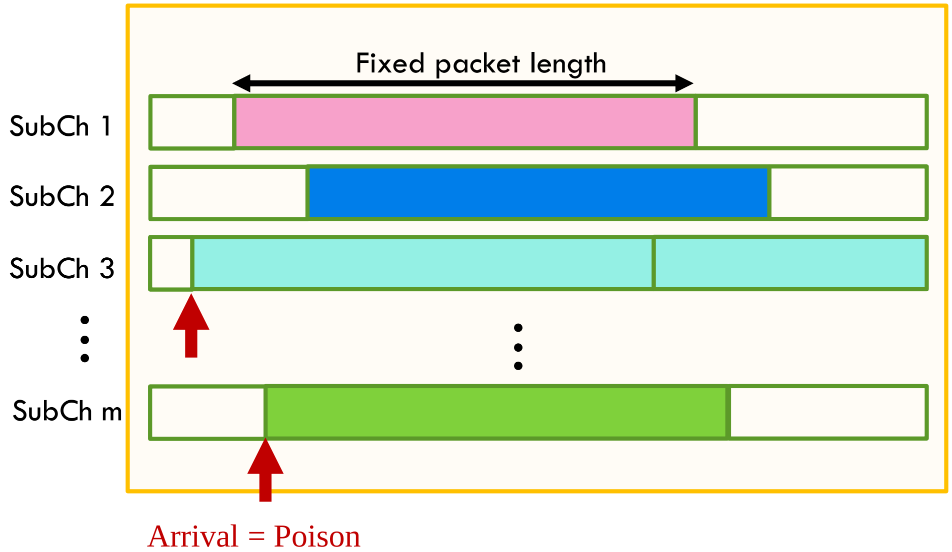
http://www.cs.ru.nl/~ths/a3/html/h2/h2.html

FDM



FDM

Modified from “Eytan Modiano”, MIT



- m Poisson streams of fixed length packets
- Arrival rate λ/m each multiplexed by FDM on m subchannels
- Total traffic = λ
- Each slot, only one packet will be transmitted with rate $\mu = 1/m$
- The total system load: $\rho = \lambda$

FDM

M/G/1 $$W = \frac{\lambda \, \overline{X^2}}{2(1-\rho)}$$

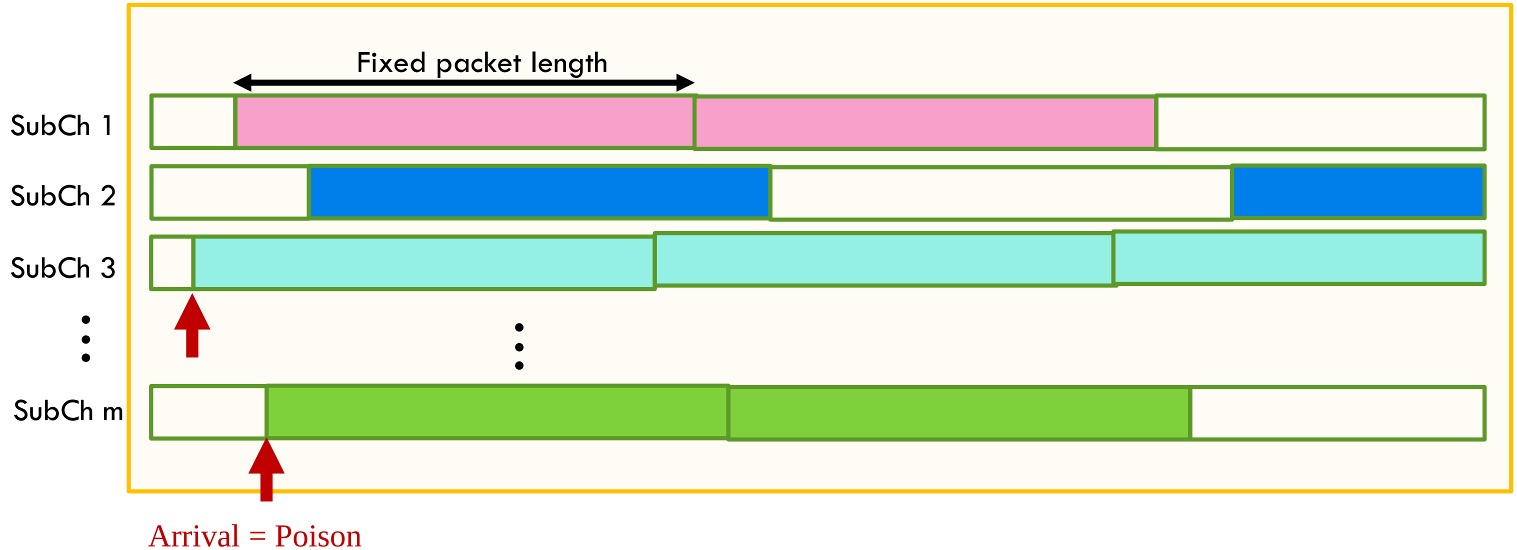
- $\overline{X} = 1/\mu = m$
- $\overline{X^2} = 1/\mu^2 = m^2$
- $\rho = \lambda$

$$W_{FDM} = \frac{(\lambda/m)\, m^2}{2\,(1-\rho)} = W_{M/D/1} = \frac{\rho\, m}{2\,(1-\rho)}$$

Anan Phonphoem
Dept. of Computer Enginerring, Kasetsart University, Thailand

Slotted FDM

Modified from "Eytan Modiano", MIT



- Packet transmission starts only @beginning of the slot
- Each slot, only one packet will be transmitted with rate $\mu=1/m$
- m Poisson streams of fixed length packets
- Arrival rate λ/m each multiplexed by FDM on m subchannels

Slotted FDM

$$W = \frac{\lambda \overline{X^2}}{2(1-\rho)} + \frac{\overline{V^2}}{2\overline{V}}$$

- $\overline{X} = \overline{V} = 1/\mu = m$
- $\overline{X^2} = \overline{V^2} = 1/\mu^2 = m^2$

$$W = \frac{\rho m}{2(1-\rho)} + \frac{m}{2}$$

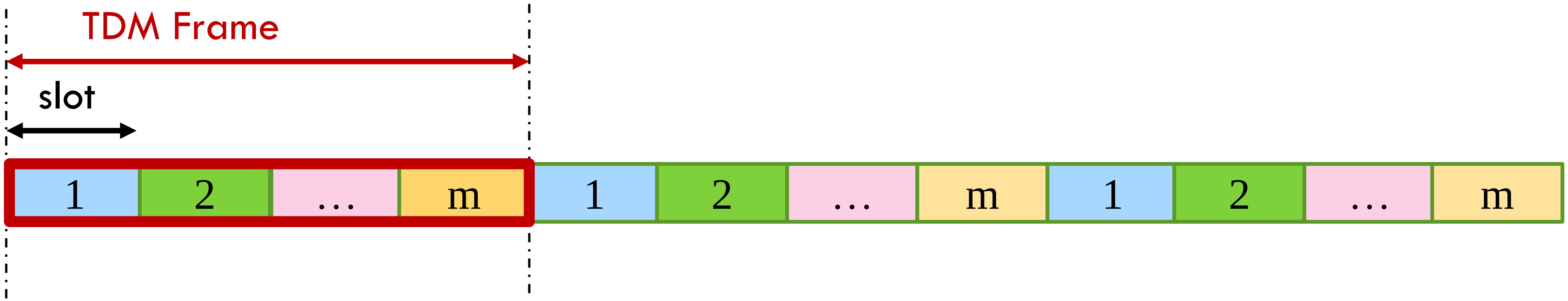
W_{FDM}

$$\frac{m^2}{2m} = \frac{\overline{X}}{2}$$

Wait for half slot

TDM

Modified from “Eytan Modiano”, MIT



- Packet transmission starts only @beginning of the slot
- Each slot, only one packet will be transmitted with rate $\mu=1/m$
- A stream(session) has to wait for its his own slot
- m Poisson streams of fixed length packets
- Arrival rate λ/m each multiplexed by TDM on m subchannels