GENERALIZED MULTI-SERVER QUEUEING SYSTEMS

by

Nicholas Bambos and Jean Walrand

Memorandum No. UCB/ERL M89/61

26 May 1989

COVER PAGE

GENERALIZED MULTI-SERVER QUEUEING SYSTEMS

by

Nicholas Bambos and Jean Walrand

Memorandum No. UCB/ERL M89/61

26 May 1989

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

TITLE PAGE

GENERALIZED MULTI-SERVER QUEUEING SYSTEMS

by

Nicholas Bambos and Jean Walrand

Memorandum No. UCB/ERL M89/61

26 May 1989

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

GENERALIZED MULTI-SERVER QUEUEING SYSTEMS

Nicholas Bambos and Jean Walrand

Dept. of EE. and CS. and
Electronics Research Lab,
University of California, Berkeley.

* Note:

This is an outline of the work, including all the basic ideas and results.

* Research supported by: MICRO (state) 442427-55221 EB37

GENERALIZED MULTI-SERVER QUEUEING SYSTEMS

Multi-channel Communication Link.

- m identical processors (channels). Service rate 1.

- INPUT : $\mathbf{N} = \{ (t_j , [K_j , \sigma_j] , j \in Z \}$

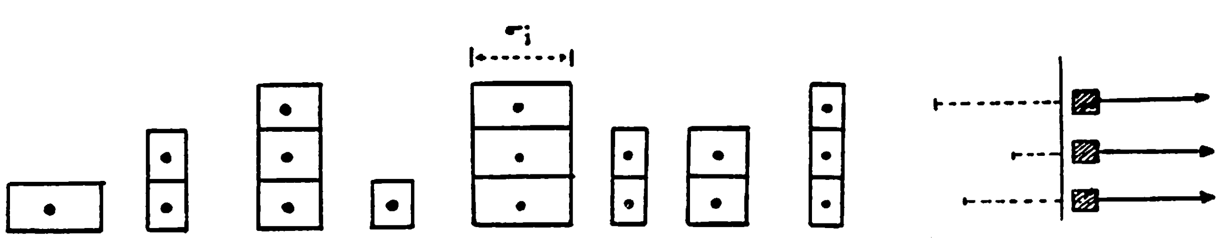
  - t_j = arrival time of j-th job (message).

  - $K_j \in \{1, 2, \ldots m\}$ = # of processors (channels) needed concurrently, for processing (transmission) of j-th job (message).

  - σ_j = processing time of j-th job (message).

- $\mathbf{N}$: STATIONARY ⊕ ERGODIC.

- Ex. $m = 3$, $K_j \in \{1, 2, 3\}$



- Processing Scheme (Allocation Policy): depends on processing times and is not fixed.

- PROBLEMS :

1. Does the system exhibit "Threshold Behavior" (Stable ↔ Unstable) w.r.t. arrival rate λ ?

2. What is the "Maximal Throughput" $\bar{\lambda}$?

3. What is the Processing Scheme that achieves the maximal throughput ("Optimal Operation") ?

THE OPTIMAL SCHEDULING PROBLEM.

- $\mathbf{N} = \{ (t_j, [K_j, \sigma_j]), \ j \in Z \}$

- <u>Define</u> process $\{ \xi_{st}(\mathbf{N}), \ s<t, \ t,s \in R \}$ by:

1. Gather jobs with arrival times in $(s,t]$ $(t_j \in (s,t])$ in some buffer, without operating on them.

2. Schedule jobs so as to minimize total execution time.
   * "Optimal Schedule"

3. ξ_{st} = Time to process this group under "Optimal Schedule".

- $\xi_{st} \leq \xi_{sx} + \xi_{xt}$, $s<x<t$ $\Rightarrow$ ξ: SUBADDITIVE

- Subad. Erg. Th. $\Rightarrow$ $\lim_{t\to\infty} \left[\frac{\xi_{ot}}{t} \right] = \gamma$

- Optimal Scheduling (*) is NP-complete in the # of jobs.

- PROBLEM: Is there a "SIMPLE" Scheduling that is ASYMPTOTICALLY $(t\to\infty)$ <u>as good as</u> the Optimal one?

o ASYMPTOTICALLY OPTIMAL SCHEDULE (for $m = 3$)

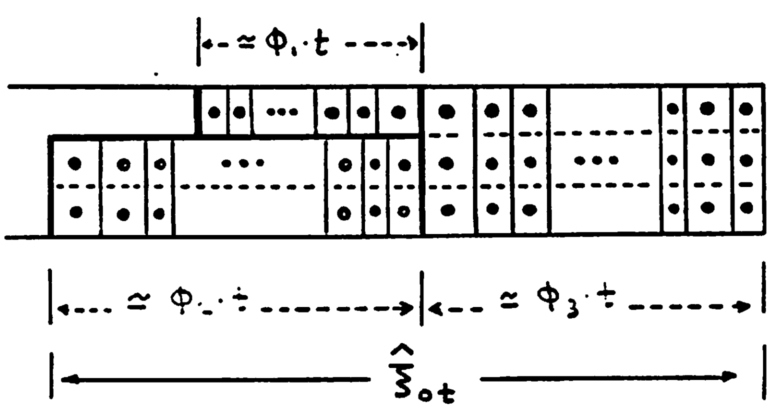
- Define:

$$\phi_k \overset{a.s.}{=} \lim_{t\to\infty} \left[\frac{\sum_j \sigma_j \, \mathbb{1}\{k_j = k,\ t_j \in (0,t]\}}{t} \right] \qquad k = 1, 2, 3$$

$$= E\left[\sum_j \sigma_j \, \mathbb{1}\{k_j = k,\ t_j \in (0,1]\} \right]$$

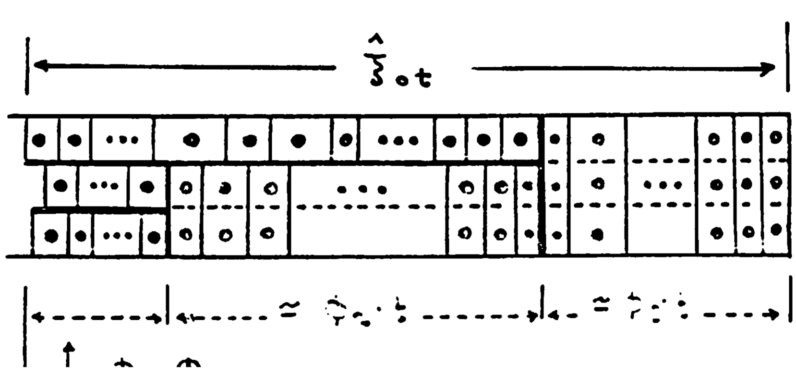
- ASYMPT. OPT. SCHED. / "SIMPLE". ($m = 3$)

1. $\phi_1 < \phi_2$



$$\lim_{t\to\infty} \left[\frac{\hat{\xi}_{ot}}{t} \right] = \phi_3 + \phi_2$$

2. $\phi_1 > \phi_2$



$$\lim_{t\to\infty} \left[\frac{\hat{\xi}_{ot}}{t} \right] = \phi_3 + \phi_2 + \frac{\phi_1 - \phi_2}{3}$$

- LEMMA :

$$\lim_{t\to\infty} \left[\frac{\bar{\zeta}_{0t}}{t} \right] = \lim_{t\to\infty} \left[\frac{\hat{\bar{\zeta}}_{0t}}{t} \right] = \gamma \underset{\substack{\uparrow \\ m=3}}{=} \Phi_3 + \Phi_2 + \frac{[\Phi_1 - \Phi_2]^+}{3}$$

Pf. : Tech. in gen. case.

- DECOMPOSITION OF γ

$$\gamma = \lim_{t\to\infty} \frac{\bar{\zeta}_{0t}}{\sum_j \mathbb{1}_{\{t_j \in (0,t]\}}} \cdot \lim_{t\to\infty} \frac{\sum_j \mathbb{1}_{\{t_j \in (0,t]\}}}{t} = \bar{\delta}^{\circ} \cdot \lambda = \frac{\lambda}{\bar{\lambda}}$$

$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$\bar{\delta}^{\circ} \qquad\qquad\qquad\qquad \lambda$$

Depends only on statistics of σ_j's and K_j's. — Arrival Rate

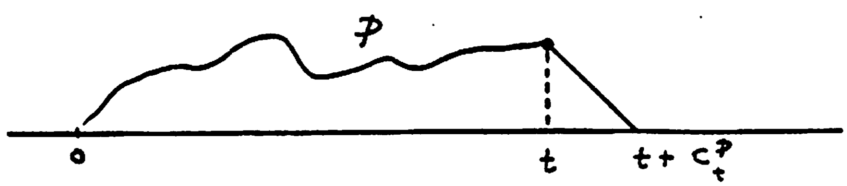
- $\frac{1}{\bar{\delta}^{\circ}} = \bar{\lambda}$

UPPER BOUND ON THROUGHPUT.

- Start processing **N** at time $t=0$, according to some Processing Scheme $\mathcal{P}$

- System is initially ($t=0$) empty.

- <u>Define</u> : $c_t^{\mathcal{P}}$ = Time to empty the system, if no Job is accepted after time t.



- THEOREM : For ANY Processing Scheme $\mathcal{P}$

$$\bar{\lambda} < \lambda \quad (\gamma > 1) \quad \Rightarrow \quad \lim_{t\to\infty} c_t^{\mathcal{P}} = +\infty \qquad \text{(System : UNSTABLE)}$$

Pf.:

$$c_t^{\mathcal{P}} \geq \zeta_{0t} - t \Rightarrow \liminf_{t\to\infty} \left[\frac{c_t^{\mathcal{P}}}{t} \right] \geq \lim \frac{\zeta_{0t}}{t} - 1 \geq \frac{\lambda}{\bar{\lambda}} - 1 > 0$$

- QUESTION:

For any $\lambda < \bar{\lambda}$, is there a Processing Scheme P_λ°, for which the system is STABLE ?

(i.e. is the bound on Throughput <u>Sharp</u> ?)

Ans. Yes!

- <u>Remark</u>: $\bar{\lambda}$ can be exactly computed by the Asymptotically Optimal Schedule.

STABILIZING PROCESSING SCHEME $(\lambda < \bar{\lambda})$

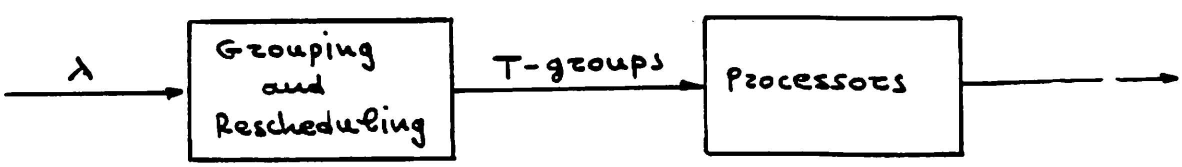
- Pick $T > 0$:

$$E[\hat{\xi}_{0T}] < T \qquad \left(\lim_{t \to \infty} \frac{E[\hat{\xi}_{0t}]}{t} = \gamma = \frac{\lambda}{\bar{\lambda}} < 1 \right)$$

- T - Processing Scheme.

1. Jobs arriving in $((k-1)T, kT]$ (k-th T-group) are scheduled (rearranged) according to the Asymptotically Optimal Schedule (Simple!)

2. Then, the k-th T-group is processed, only after all the previous T-groups have been processed and have left.

- 

- STABILITY :

The system is a G/G/1 queue with :

Arrival Times : $kT \quad k \in \{0, 1, 2 \dots\}$

Service Times : $E[\hat{\xi}_{(k-1)T, kT}] = E[\hat{\xi}_{0T}] < T$

$\Rightarrow$ STABLE

(Loynes 1961)