Dynamic scheduling and maintenance for a two-class queue with a deteriorating server

Jefferson Huang

School of Operations Research & Information Engineering
Cornell University

June 28, 2018

American Control Conference

Milwaukee, WI

Joint work with Douglas Down (McMaster), Mark Lewis (Cornell), Cheng-Hung Wu (National Taiwan University)

Talk Outline

1. Research Motivation
2. A Joint Scheduling and Maintenance Model
3. Scheduling Under Deterioration
4. Structure of Optimal Scheduling and Maintenance Policies

Motivation: Semiconductor Manufacturing

- Typical process flow has $\sim$1000 steps.
- Each step is handled by one of $\sim$100 machines.
- Each machine is
  1. flexible (i.e., can handle more than one job type), and
  2. subject to deterioration.

Allocate the machine to a waiting job?

or

Perform (preventive) maintenance?

Research Questions

1. Given a choice between prioritizing scheduling or maintenance, where should a decision-maker focus his/her efforts?

2. Under what conditions can classic scheduling/maintenance results be used to create useful heuristics?

Joint Scheduling & Maintenance Model

Controlled **2-Class G/M/1 Queue** with:

- server state $s \in \mathcal{S} := \{0, 1, \ldots, B\}$ evolving according to a continuous-time Markov chain
  - current state is $s \implies$ next state is $(s-1) \pmod{B+1}$
- class-k service rate μ_k^s when the server state is $s \in \mathcal{S}$
  - $s = 0 \implies$ server is down for maintenance ($\mu_k^0 = 0$)
  - $s \geqslant 1 \implies$ server is operational ($\mu_k^s > 0$)
- class-k holding cost rate c_k
- fixed maintenance costs K_C and K_M for (resp.) corrective and preventive maintenance
  - corrective = forced maintenance when the server fails
  - preventive = elective maintenance initiated when the server hasn't failed yet

Objective Function: long-run expected average cost

Special Case: Scheduling Under Deterioration

Assume there is no preventive maintenance.

$c\mu$-Rule: If the server state is $s \geqslant 1$, prioritize class 1 (resp. class 2) jobs if

$$c_1\mu_1^s \geqslant (\text{resp. } <) c_2\mu_2^s.$$

- If there is no deterioration, this rule is optimal (Nain, 1989).

Theorem (H. et al., 2018)

When there is deterioration, the $c\mu$-rule may not be optimal.

- The average cost under the $c\mu$-rule may be infinite, while a policy with finite average cost exists.

Special Case: Scheduling Under Deterioration

Under what conditions does the $c\mu$-rule work well?

Assumption (Constant-Ratio)

$$\frac{\mu_1^{s-1}}{\mu_1^s} = \frac{\mu_2^{s-1}}{\mu_2^s} \qquad \forall s \geqslant 1.$$

- The relative service rates remain constant.

Theorem (H. et al., 2018)

If the Constant-Ratio assumption holds, then the $c\mu$-rule is optimal.

- Proved with an interchange argument based on (Nain, 1989).

Sufficiency of the $c\mu$-Rule

Theorem (H. et al., 2018)

Suppose

1. *the Constant-Ratio assumption holds, and*
2. *in making maintenance decisions, queue-length information cannot be used.*

Then there exists an optimal policy that uses the $c\mu$-rule for scheduling jobs.

- Allowable maintenance policies include e.g., age-based, job-based, server state threshold policies.
- Can focus on finding good maintenance decisions.

Can the restriction on maintenance policies be removed?

- Yes, if "anticipative" joint scheduling & maintenance policies are allowed.

Can the suboptimality of using the $c\mu$-rule be bounded?

- Found to be within 1.8% of optimality on average, for test problems where the Constant-Ratio assumption is violated (H. et al., 2018).

Structure of Optimal Maintenance Policies

Can something analogous to the preceding results on scheduling be said about optimal maintenance decisions?

Theorem (H. et al., 2018)

Suppose

1. *the arrival processes are independent homogeneous Poisson processes,*
2. *the maintenance times are independent and identically distributed with positive first moment, and*
3. *the "average-cost optimality inequality" holds (Sennott, 1989).*

Then there is an optimal policy that is "monotone".

- "monotone" = for every fixed number of class 1 & class 2 jobs, preventive maintenance is initiated iff. the server state is below a threshold.
- Simple maintenance policies (e.g., based on 1 or 2 thresholds) may be suboptimal (Kaufman & Lewis, 2007).

Talk Summary

1. Provided conditions under which the decision-maker can focus on making good maintenance decisions.

2. The (heuristic) policy suggested by our results seems to still work well when the conditions do not hold.

Open Questions:

- Bounds on the suboptimality of the heuristic?
- Maintenance heuristics that are also near-optimal?