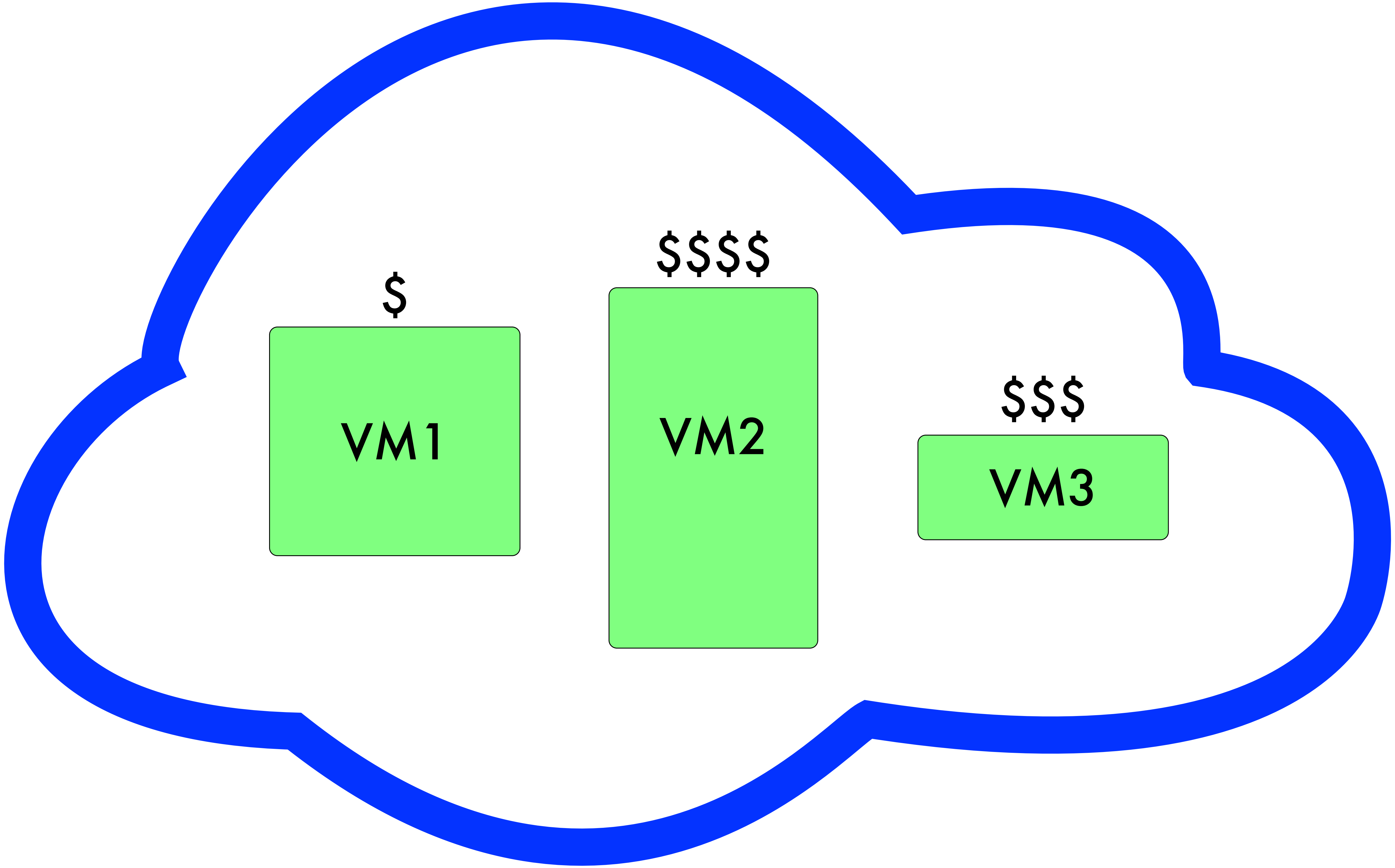
On Budgeting and Quality

With an Application to Safety-Critical Real-Time Systems

Bader Alahmad & Sathish Gopalakrishnan

The University of British Columbia


$
VM1
$$$$
VM2
$$$
VM3

Budgeting vs. Quality

Goal: Maintain a certain QoS level for the long-term operation of the system with high confidence

modulo perhaps initial transient behavior

Task Model

Task Model

- **Recurrent** tasks with given periods

Task Model

- **Recurrent** tasks with given periods
- Task has random execution-time **demand** X_e

Task Model

- **Recurrent** tasks with given periods
- Task has random execution-time **demand** X_e
- Available **budget** is random variable X_b

Task Model

- **Recurrent** tasks with given periods
- Task has random execution-time **demand** X_e
- Available **budget** is random variable X_b
- X_e and X_b **not** necessarily independent

Task Model

- **Recurrent** tasks with given periods
- Task has random execution-time **demand** X_e
- Available **budget** is random variable X_b
- X_e and X_b **not** necessarily independent
- **Bad** situation: $X_e > X_b$, central quantity: $\mathbb{P}(X_e > X_b)$

Problem Description

$F_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_e^i > X_b^i\}$: Fraction of first n jobs where budget is insufficient

Problem Description

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_e^i > X_b^i\}$$: Fraction of first n jobs where budget is insufficient

- The process $\{(X_e^t, X_b^t)\}_{t\in\mathbb{R}_+}$ is iid, sampled at **job release times** $iP, i \in \{0, 1, \ldots\}$

Problem Description

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_e^i > X_b^i\}$$: Fraction of first n jobs where budget is insufficient

- The process $\{(X_e^t, X_b^t)\}_{t\in\mathbb{R}_+}$ is iid, sampled at **job release times** $iP, i \in \{0, 1, \ldots\}$
- **Given:**

Problem Description

$F_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_e^i > X_b^i\}$: Fraction of first n jobs where budget is insufficient

- The process $\{(X_e^t, X_b^t)\}_{t\in\mathbb{R}_+}$ is iid, sampled at **job release times** $iP, i \in \{0, 1, \ldots\}$
- **Given:**
  - **QoS level** $\alpha \in (0, 1]$: Tolerable fraction of jobs that might demand more than available budget

Problem Description

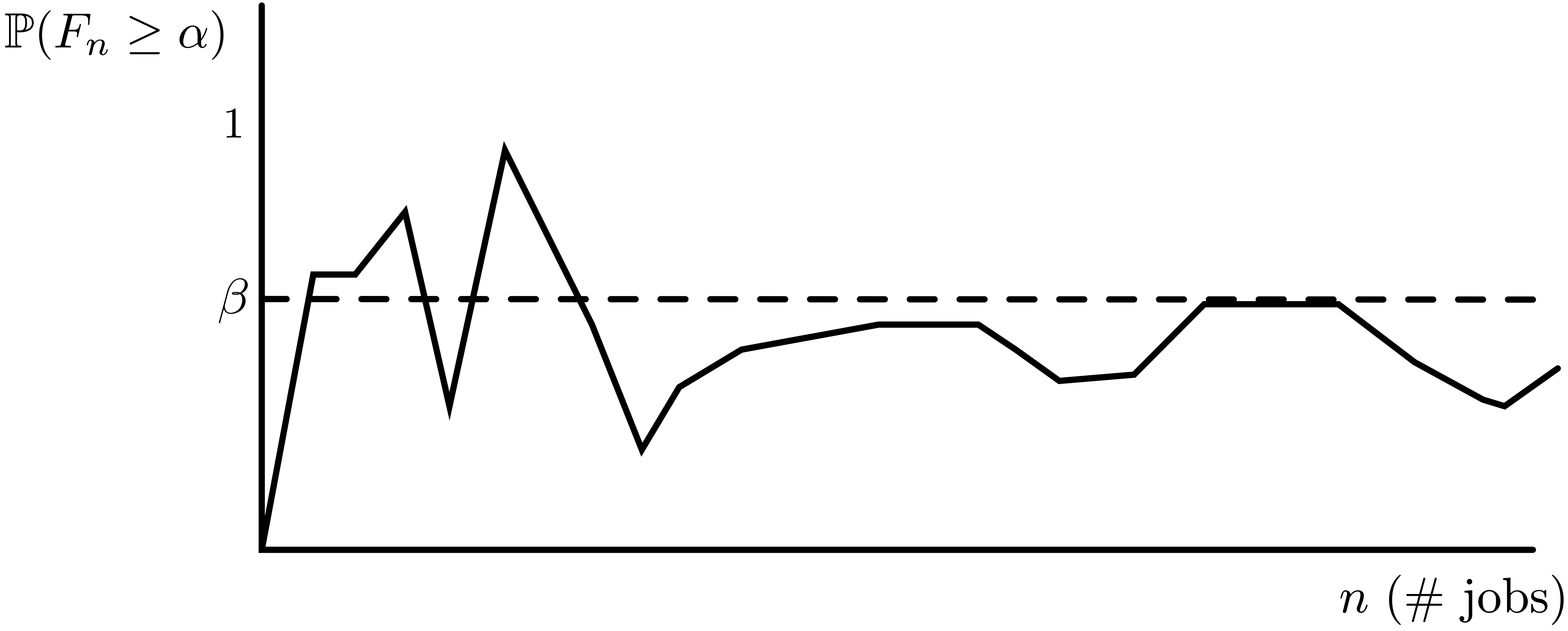
$$F_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_e^i > X_b^i\}$$: Fraction of first n jobs where budget is insufficient

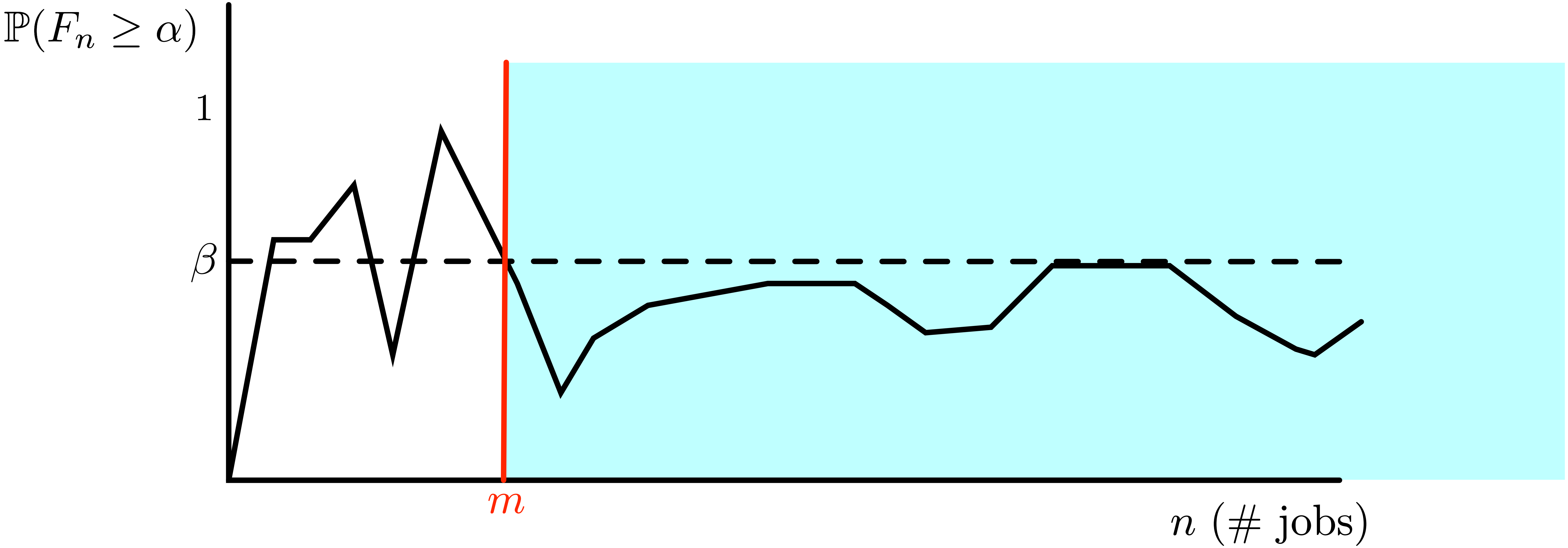
- The process $\{(X_e^t, X_b^t)\}_{t\in\mathbb{R}_+}$ is iid, sampled at **job release times** $iP, i \in \{0, 1, \ldots\}$
- **Given:**
  - **QoS level** $\alpha \in (0, 1]$: Tolerable fraction of jobs that might demand more than available budget
  - **Confidence** parameter $\beta \in (0, 1)$

Problem Description

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_e^i > X_b^i\}$$: Fraction of first n jobs where budget is insufficient

- The process $\{(X_e^t, X_b^t)\}_{t\in\mathbb{R}_+}$ is iid, sampled at **job release times** $iP, i \in \{0, 1, \ldots\}$
- **Given:**
  - **QoS level** $\alpha \in (0, 1]$: Tolerable fraction of jobs that might demand more than available budget
  - **Confidence** parameter $\beta \in (0, 1)$
- **Question:** Is there integer $m \geq 0$ s.t. $\mathbb{P}(F_n \geq \alpha) \leq \beta \;\; \forall n \geq m$ jobs of the task?


$\mathbb{P}(F_n \geq \alpha)$
1
β
n (# jobs)


$\mathbb{P}(F_n \geq \alpha)$
1
β
m
n (# jobs)

What's in it for the system designer?

Resource Dimensioning

If we know the execution-time requirement X_e, how should the budget X_b look like so that for some m, $\mathbb{P}(F_n \geq \alpha) \leq \beta$ for all $n \geq m$?

What's in it for the system designer?

What's in it for the system designer?

» **Recast.** If we let $\boldsymbol{p} \equiv \mathbb{P}(\boldsymbol{X}_e > \boldsymbol{X}_b)$: Derive a bound $\boldsymbol{p}^+$ on $\boldsymbol{p}$ so that if $\boldsymbol{p} \leq \boldsymbol{p}^+$, then QoS requirement satisfied.

What's in it for the system designer?

- **Recast.** If we let $p \equiv \mathbb{P}(X_e > X_b)$: Derive a bound p^+ on p so that if $p \leq p^+$, then QoS requirement satisfied.
- Then any budget X_b satisfying $p \leq p^+$ is okay

What's in it for the system designer?

- **Recast.** If we let $p \equiv \mathbb{P}(X_e > X_b)$: Derive a bound p^+ on p so that if $p \leq p^+$, then QoS requirement satisfied.
- Then any budget X_b satisfying $p \leq p^+$ is okay
  - We have a sufficient range $(0, p^+]$

What's in it for the system designer?

- **Recast.** If we let $p \equiv \mathbb{P}(X_e > X_b)$: Derive a bound p^+ on p so that if $p \leq p^+$, then QoS requirement satisfied.
- Then any budget X_b satisfying $p \leq p^+$ is okay
  - We have a sufficient range $(0, p^+]$
  - The larger the p^+, the more the flexibility the system designer has in allocating resources (budgets)

Large Deviation bounds

Large Deviation bounds

» **Requirement:** Strongest possible bound

Large Deviation bounds

- **Requirement:** Strongest possible bound
  - $\mathbb{P}(F_n \geq \alpha)$ decays exponentially quickly in n

Large Deviation bounds

- **Requirement:** Strongest possible bound
  - $\mathbb{P}(F_n \geq \alpha)$ decays exponentially quickly in n
  - Gives the largest possible sufficient bound p^+ on $p \equiv \mathbb{P}(X_e > X_b)$

Large Deviation bounds

- **Requirement:** Strongest possible bound
  - $\mathbb{P}(F_n \geq \alpha)$ decays exponentially quickly in n
  - Gives the largest possible sufficient bound p^+ on $p \equiv \mathbb{P}(X_e > X_b)$
- **Requirement made precise:** Is there **strictly positive** function I (that might depend on α) such that

$$\mathbb{P}(F_n \geq \alpha) \approx e^{-nI} \,?$$

Chernoff Theorem

Let $Y_1, \ldots, Y_n$ be independent random variables such that Y_i always lies in the interval $[0, 1]$. Define $S_n = \sum_{i=1}^n Y_i$, and let $\mu = \mathbb{E}(S_n)$. Then for any $\delta > 0$,

$$\mathbb{P}\big(S_n \geq (1+\delta)\mu\big) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right)$$

Chernoff bound: $\mathbb{P}\big(S_n \geq (1+\delta)\mu\big) \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right) \quad (\delta > 0)$

$$\textbf{Chernoff bound: } \mathbb{P}\big(S_n \geq (1+\delta)\mu\big) \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right) \quad (\delta > 0)$$

» $Y_i = \mathbf{1}\{X_e^i > X_b^i\}$

$$\textbf{Chernoff bound: } \mathbb{P}\Big(S_n \geq (1+\delta)\mu\Big) \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right) \quad (\delta > 0)$$

» $Y_i = \mathbf{1}\{X_e^i > X_b^i\}$

» $S_n = \sum_{i=1}^{n} Y_i$

Chernoff bound: $\mathbb{P}\big(S_n \geq (1+\delta)\mu\big) \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right) \quad (\delta > 0)$

- $Y_i = \mathbf{1}\{X_e^i > X_b^i\}$
- $S_n = \sum_{i=1}^{n} Y_i$
- $p = \mathbb{E}(Y_i) = \mathbb{P}(X_e^i > X_b^i)$

Chernoff bound: $\mathbb{P}\big(S_n \geq (1+\delta)\mu\big) \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right) \quad (\delta > 0)$

- $Y_i = \mathbf{1}\{X_e^i > X_b^i\}$
- $S_n = \sum_{i=1}^n Y_i$
- $p = \mathbb{E}(Y_i) = \mathbb{P}(X_e^i > X_b^i)$
- $\mu = \mathbb{E}(S_n) = np$

Chernoff bound: $\mathbb{P}\big(S_n \geq (1+\delta)\mu\big) \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right) \quad (\delta > 0)$

- » $Y_i = \mathbf{1}\{X_e^i > X_b^i\}$
- » $S_n = \sum_{i=1}^{n} Y_i$
- » $p = \mathbb{E}(Y_i) = \mathbb{P}(X_e^i > X_b^i)$
- » $\mu = \mathbb{E}(S_n) = np$
- » $F_n = S_n / n$

Chernoff bound: $\mathbb{P}\big(S_n \geq (1+\delta)\mu\big) \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right) \quad (\delta > 0)$

- $Y_i = \mathbf{1}\{X_e^i > X_b^i\}$
- $S_n = \sum_{i=1}^n Y_i$
- $p = \mathbb{E}(Y_i) = \mathbb{P}(X_e^i > X_b^i)$
- $\mu = \mathbb{E}(S_n) = np$
- $F_n = S_n/n$
- If $0 < p < \alpha$, then setting $\delta = -1 + \alpha/p$ gives $\delta > 0$ 😬

Chernoff bound: $\mathbb{P}\big(S_n \geq (1+\delta)\mu\big) \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right) \quad (\delta > 0)$

- $Y_i = \mathbf{1}\{X_e^i > X_b^i\}$
- $S_n = \sum_{i=1}^n Y_i$
- $p = \mathbb{E}(Y_i) = \mathbb{P}(X_e^i > X_b^i)$
- $\mu = \mathbb{E}(S_n) = np$
- $F_n = S_n / n$
- If $0 < p < \alpha$, then setting $\delta = -1 + \alpha/p$ gives $\delta > 0$ 😬
  - $\mathbb{P}\big(S_n \geq (1 - 1 + \alpha/p)np\big) = \mathbb{P}\big(F_n \geq \alpha\big) \leq \exp\left(-\frac{(\alpha-p)^2}{\alpha+p}n\right)$

Our Large Deviation Bound

$$\text{If } 0 < p < \alpha, \text{ then } \mathbb{P}\big(F_n \geq \alpha\big) \leq e^{-nI}, \quad I = \frac{(\alpha - p)^2}{\alpha + p}$$

Upper bound on p-value $\mathbb{P}(X_e > X_b)$

Let $\gamma(\beta, m) = \ln(1/\beta)/m$. If $\alpha > \gamma(\beta, m)$, then

$$p^+ = p^+(m) = \frac{\gamma(\beta, m) + 2\alpha - \sqrt{\gamma(\beta, m)^2 + 8\alpha\gamma(\beta, m)}}{2} > 0.$$

Observe: $p^+(m) < \alpha$ for every m, and $p^+(m) \uparrow \alpha$ as $m \uparrow \infty$

System-wide QoS

Dependent Tasks

Dependent tasks & System-wide QoS

Dependent tasks & System-wide QoS

- **Given:**

Dependent tasks & System-wide QoS

- **Given:**
  - System-wide confidence $Q > 0$;

Dependent tasks & System-wide QoS

- **Given:**
  - System-wide confidence $Q > 0$;
  - Task QoS levels $\alpha_1, \ldots, \alpha_N$;

Dependent tasks & System-wide QoS

- **Given:**
  - System-wide confidence $Q > \mathbf{0}$;
  - Task QoS levels $\alpha_1, \ldots, \alpha_N$;
  - "limiting" number of job releases m_j of task j, $j \in [N]$;

Dependent tasks & System-wide QoS

- **Given:**
  - System-wide confidence $Q > \mathbf{0}$;
  - Task QoS levels $\alpha_1, \ldots, \alpha_N$;
  - "limiting" number of job releases m_j of task j, $j \in [N]$;
  - **That at most d tasks are dependent**

Dependent tasks & System-wide QoS

- **Given:**
  - System-wide confidence $Q > 0$;
  - Task QoS levels $\alpha_1, \ldots, \alpha_N$;
  - "limiting" number of job releases m_j of task j, $j \in [N]$;
  - **That at most d tasks are dependent**
- **Bad** event for task j: $B_j \equiv \left\{F_n^j \geq \alpha_j\right\}$, where $F_n^j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_{e,j}^i > X_{b,j}^i\}$

Dependent tasks & System-wide QoS

- **Given:**
  - System-wide confidence $Q > 0$;
  - Task QoS levels $\alpha_1, \ldots, \alpha_N$;
  - "limiting" number of job releases m_j of task j, $j \in [N]$;
  - **That at most d tasks are dependent**
- **Bad** event for task j: $B_j \equiv \{F_n^j \geq \alpha_j\}$, where $F_n^j = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_{e,j}^i > X_{b,j}^i\}$
- **Question:** Under what condition(s) is $\mathbb{P}(B_1^c \cap \cdots \cap B_N^c) \geq Q$ for all $n_j \geq m_j$ releases of every task task T_j, $j \in [N]$?

Lovász Local Lemma (LLL)

Let $B_1, B_2, \ldots, B_N$ be a sequence of events such that each event occurs with probability **at most** f and such that each event is independent of all the other events except for at most d of them.

If $efd \leq 1$, then $\mathbb{P}(B_1^c \cap \cdots \cap B_N^c) \geq (1-f)^N > 0$

System-wide QoS bounds

System-wide QoS bounds

- **Bad** event for task j: $B_j \equiv \left\{F_n^j \geq \alpha_j\right\}$, where $F_n^j = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\left\{X_{e,j}^i > X_{b,j}^i\right\}$

System-wide QoS bounds

- **Bad** event for task j: $B_j \equiv \{F_n^j \geq \alpha_j\}$, where $F_n^j = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_{e,j}^i > X_{b,j}^i\}$
- **Question:** Under what condition(s) is $\mathbb{P}(B_1^c \cap \cdots \cap B_N^c) \geq Q$ for all $n_j \geq m_j \; \forall j \in [N]$?

System-wide QoS bounds

» **Bad** event for task j: $B_j \equiv \left\{F_n^j \geq \alpha_j\right\}$, where $F_n^j = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\left\{X_{e,j}^i > X_{b,j}^i\right\}$

» **Question:** Under what condition(s) is $\mathbb{P}(B_1^c \cap \cdots \cap B_N^c) \geq Q$ for all $n_j \geq m_j \ \forall j \in [N]$?

» By **LLL** and LD bound, when

$$e^{-n_j I_j} \leq \min\left\{1/(de), 1 - \sqrt[N]{Q}\right\} \quad \forall j \in [N]$$

System-wide QoS bounds

- **Bad** event for task j: $B_j \equiv \{F_n^j \geq \alpha_j\}$, where $F_n^j = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_{e,j}^i > X_{b,j}^i\}$
- **Question:** Under what condition(s) is $\mathbb{P}(B_1^c \cap \cdots \cap B_N^c) \geq Q$ for all $n_j \geq m_j \; \forall j \in [N]$?
- By **LLL** and LD bound, when

$$e^{-n_j I_j} \leq \min\{1/(de), 1 - \sqrt[N]{Q}\} \quad \forall j \in [N]$$

- When $1 - \sqrt[N]{Q} < 1/(de)$, if $\alpha_j > -\ln\big(1 - \sqrt[N]{Q}\big)/m_j \equiv d(m_j)$, then it's sufficient that

$$p_j \leq p_j^+(m_j) \equiv \frac{d(m_j) + 2\alpha_j - \sqrt{d(m_j)^2 + 8\alpha_j d(m_j)}}{2}$$

Application to safety-critical systems

Monitors + Isochronous execution

Application to safety-critical systems

Monitors + Isochronous execution

- **Main task** is high quality but unpredictable with execution-time demand X_e

Application to safety-critical systems

Monitors + Isochronous execution

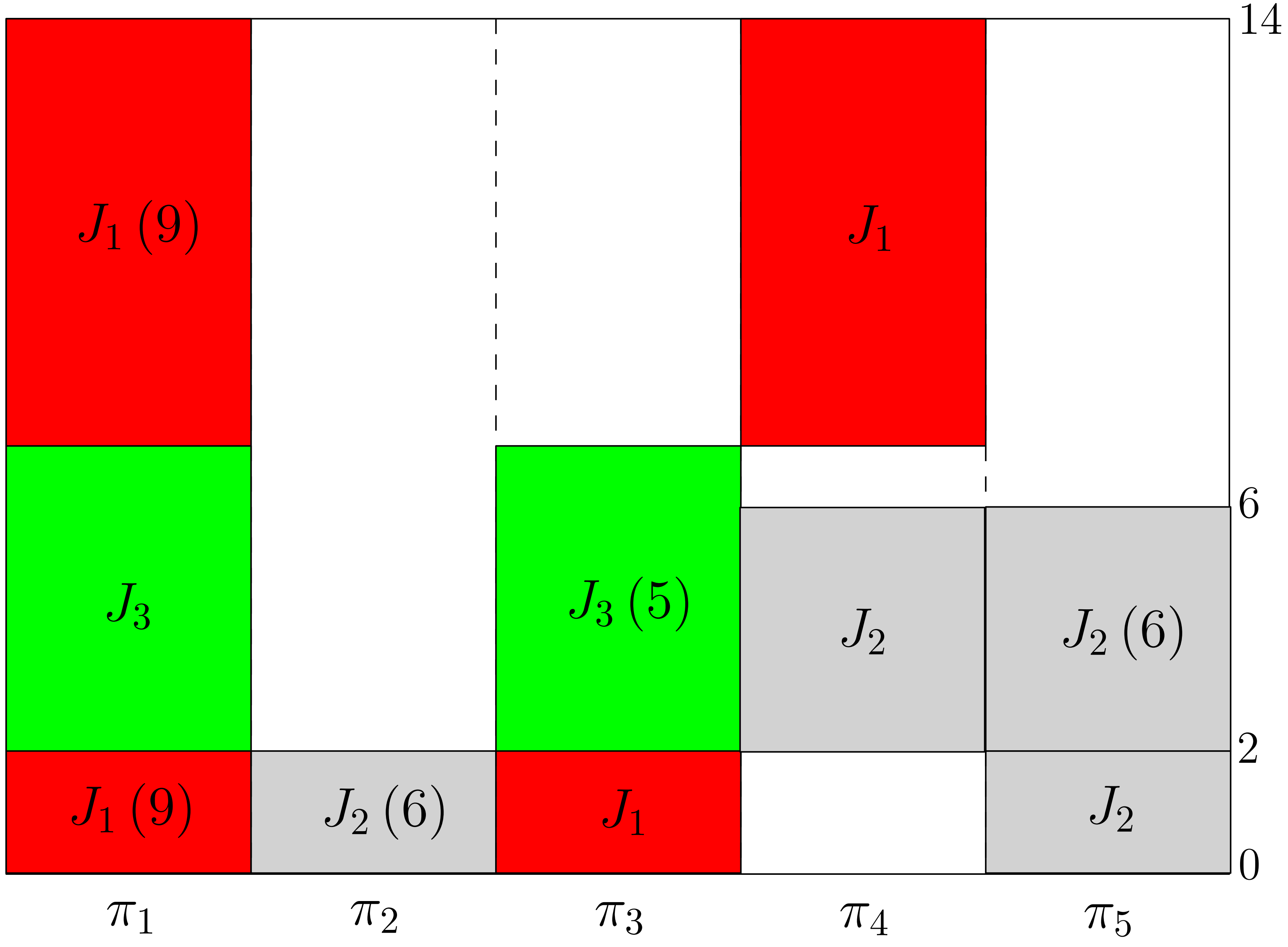
- **Main task** is high quality but unpredictable with execution-time demand X_e
- **Monitor** is lower-quality and with deterministic WCET $c > 0$ (to be determined)

Application to safety-critical systems

Monitors + Isochronous execution

- **Main task** is high quality but unpredictable with execution-time demand X_e
- **Monitor** is lower-quality and with deterministic WCET $c > 0$ (to be determined)
- Here $p = \mathbb{P}(X_e > c)$

Isochronous Execution



Two problems to be solved

Two problems to be solved

- **Problem1:** Derive **lower** bounds on monitor WCETs $c_1, \ldots, c_N$ so that each task achieves the desired QoS

Two problems to be solved

- **Problem1:** Derive **lower** bounds on monitor WCETs $c_1, \ldots, c_N$ so that each task achieves the desired QoS
- **Problem 2:** Derive **upper** bounds on monitor WCETs $c_1, \ldots, c_N$ so that all N tasks meet their hard deadlines under isochronous execution while achieving QoS requirements (from **Problem 1**).

Problem 1: Monitor WCET *lower* bounds

Problem 1: Monitor WCET *lower* bounds

- **Problem1:** Derive **lower** bounds on monitor WCETs $c_1, \ldots, c_N$ so that each task achieves the desired QoS

Problem 1: Monitor WCET *lower* bounds

- **Problem1:** Derive **lower** bounds on monitor WCETs $c_1, \ldots, c_N$ so that each task achieves the desired QoS
- **Solution:** It is sufficient that $c_j \geq H_{e,j}^{-}(1 - p_j^{+})$ for every $j \in [N]$

$$H_{e,j}^{-} : \textbf{Quantile function of demand distribution } \mathbb{P}_{X_{e,j}}$$

Problem 1: Monitor WCET *lower* bounds

- **Problem1:** Derive **lower** bounds on monitor WCETs $c_1, \ldots, c_N$ so that each task achieves the desired QoS
- **Solution:** It is sufficient that $c_j \geq H_{e,j}^{-}(1 - p_j^{+})$ for every $j \in [N]$

$$H_{e,j}^{-} : \text{Quantile function of demand distribution } \mathbb{P}_{X_{e,j}}$$

- *Is there a feasible isochronous schedule of the tasks with these monitor WCETs?*

Problem 2: Feasible Monitor WCET *Upper* Bounds

Problem 2: Feasible Monitor WCET *Upper* Bounds

- Derive **upper** bounds on monitor WCETs $c_1, \ldots, c_N$ so that all N tasks meet their hard deadlines under isochronous execution while achieving QoS requirements (from **Problem 1**).

Problem 2: Feasible Monitor WCET *Upper* Bounds

- Derive **upper** bounds on monitor WCETs $c_1, \ldots, c_N$ so that all N tasks meet their hard deadlines under isochronous execution while achieving QoS requirements (from **Problem 1**).
- **Recall:** Given monitor WCETs $c_1, \ldots, c_N$, a feasible isochronous schedule exists if the optimal solution $(x_1^*, \ldots, x_M^*)$ to the LP below is such that $\sum_{i=1}^{M} x_i^* \leq 1$:

$$\begin{aligned} \min_{x \in \mathbb{R}^M} \quad & \sum_{i=1}^{M} x_i \\ \textbf{subject to:} \quad & \sum_{i \in F_j} x_i \geq \frac{c_j}{P_j}, \qquad && j \in [N] \\ & x_i \geq 0, && i \in [M]. \end{aligned}$$

Problem 2: Feasible Monitor WCET Upper Bounds

Problem 2: Feasible Monitor WCET Upper Bounds

- **One possible solution:**

$$
\begin{aligned}
\max_{c \in \mathbb{R}^N,\, x \in \mathbb{R}^M} \quad & \sum_{j=1}^{N} c_j \\
\text{subject to:} \quad & \sum_{i=1}^{M} x_i \leq 1 \\
& \sum_{i \in F_j} x_i \geq \frac{c_j}{P_j}, && j \in [N] \\
& c_j \geq H^{-}_{e,j}(1 - p^{+}_j), && j \in [N] \\
& x_i \geq 0, && i \in [M]
\end{aligned}
$$

Problem 2: Feasible Monitor WCET Upper Bounds

- **One possible solution:**

$$
\begin{aligned}
\max_{c \in \mathbb{R}^N,\, x \in \mathbb{R}^M} \quad & \sum_{j=1}^{N} c_j \\
\text{subject to:} \quad & \sum_{i=1}^{M} x_i \leq 1 \\
& \sum_{i \in F_j} x_i \geq \frac{c_j}{P_j}, && j \in [N] \\
& c_j \geq H_{e,j}^{-}(1 - p_j^{+}), && j \in [N] \\
& x_i \geq 0, && i \in [M]
\end{aligned}
$$

- Gives **feasible range** $[H_{e,j}^{-}(1 - p_j^{+}), c_j^{*}]$ for task j's monitor WCET if instance is feasible

Do we always need monitors?

- Task may not need a monitor if its demand is unbounded