Interleaving Workloads with Performance Guarantees in Storage Clusters

Alma Riska
Department of Computer Science
College of William and Mary

In collaboration with
Evgenia Smirni
Department of Computer Science
College of William and Mary

Storage Clusters

- Increasingly more complex
  - Larger in size
  - Heterogeneous devices
  - Support a variety of applications simultaneously
- Fulfill conflicting goals
  - Performance
  - Reliability/availability
  - Power consumption
- Key to scaling effectively such systems
  - Adaptive and robust management of resources

Utilization in Storage Systems

2 hour long idle periods	READ Idle	WRITE Idle	Disk Idle
Portion of drives	**79%**	**90%**	**79%**

[Sigmetrics-09]

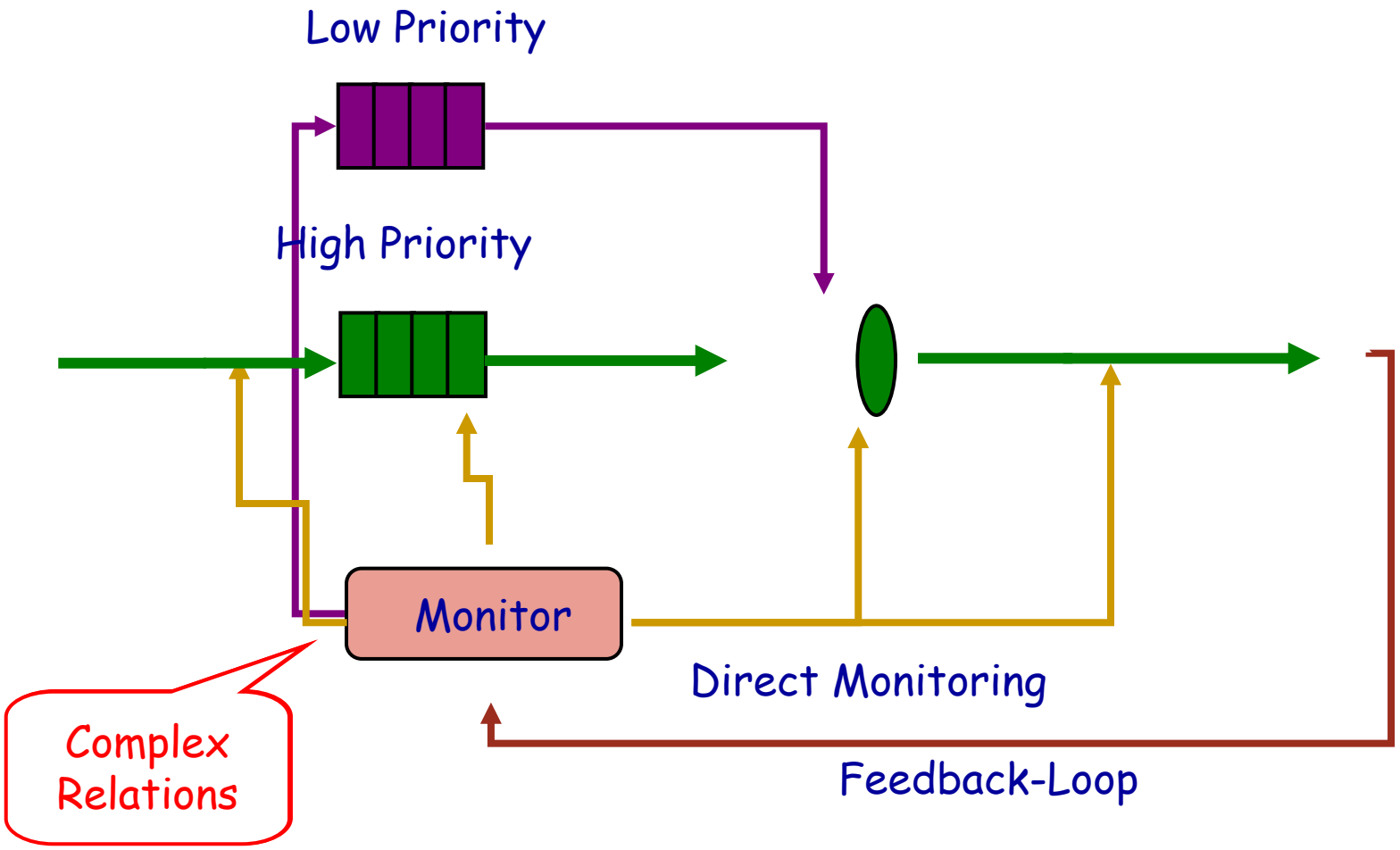
	Idleness	Avg. Idle Period Length	Avg. Service Time
Web	96 %	274 ms	7 ms
E-mail	92 %	625 ms	5 ms
Application Development	94 %	119 ms	5 ms
User Accounts Server	98 %	183 ms	5 ms

[USENIX-06]

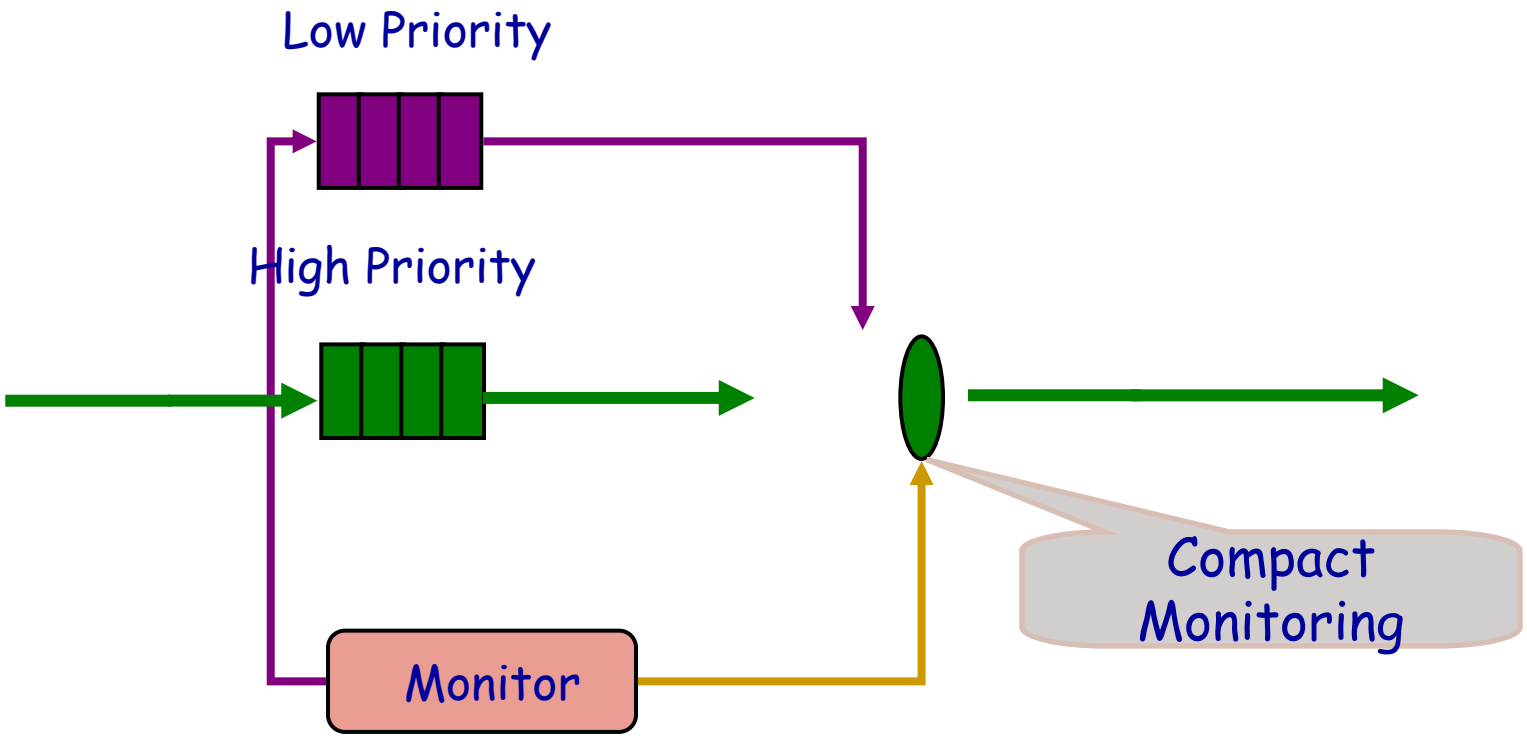
Storage Clusters

- Better utilization by interleaving workloads
  - Different priorities
    - Lower/higher/dynamic
  - Different requirements
    - Performance/reliability/power consumption
- Challenge: Seamless interleaving
  - Meet workload's quality-of-service constraints

Monitoring for Self Management



Monitoring for Self Management



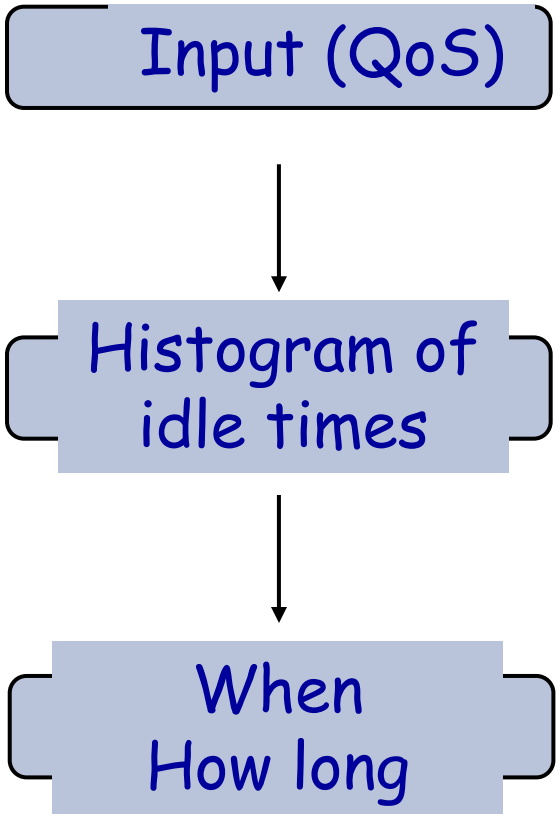
Compact Monitoring

- Monitor outcome processes rather than input processes
  - E.g. resource availability rather than
    - Arrival process
    - Service process
    - Utilization
    - Scheduling policy
    - Workload characteristics (R/W, access pattern)
    - ….

Informative Monitoring

- Commonly monitored metrics
  - Mean, standard deviation
  - Higher moments of a process
- More informative metrics
  - Histograms
    - Compact lists
    - High accuracy
- Incorporate such metrics into decision making

Single Node Estimation



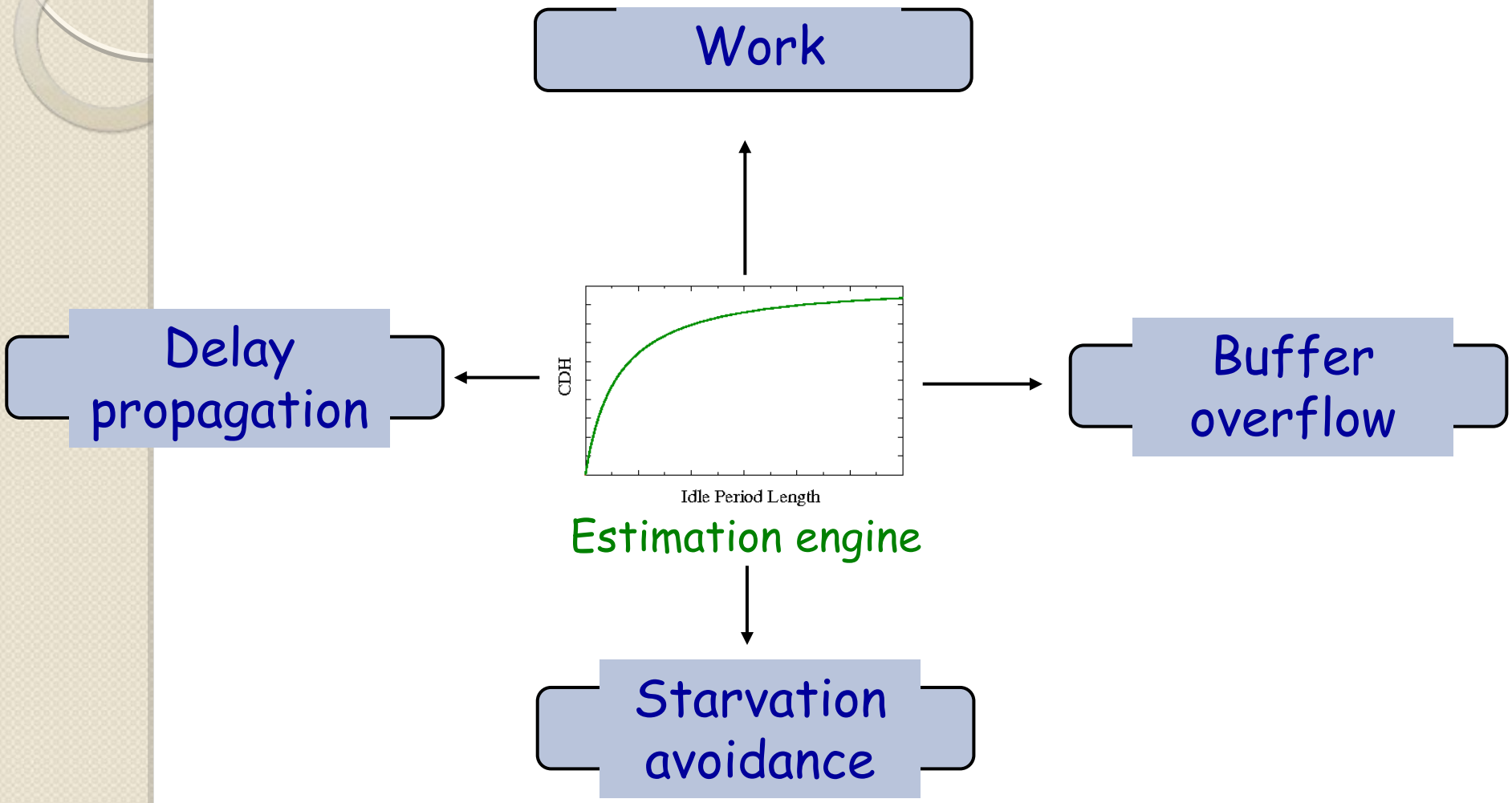
Scheduling parameters that meet QoS

Standard queuing theory concepts

Select the pair that satisfies all constraints

Estimation engine enabled by the histogram

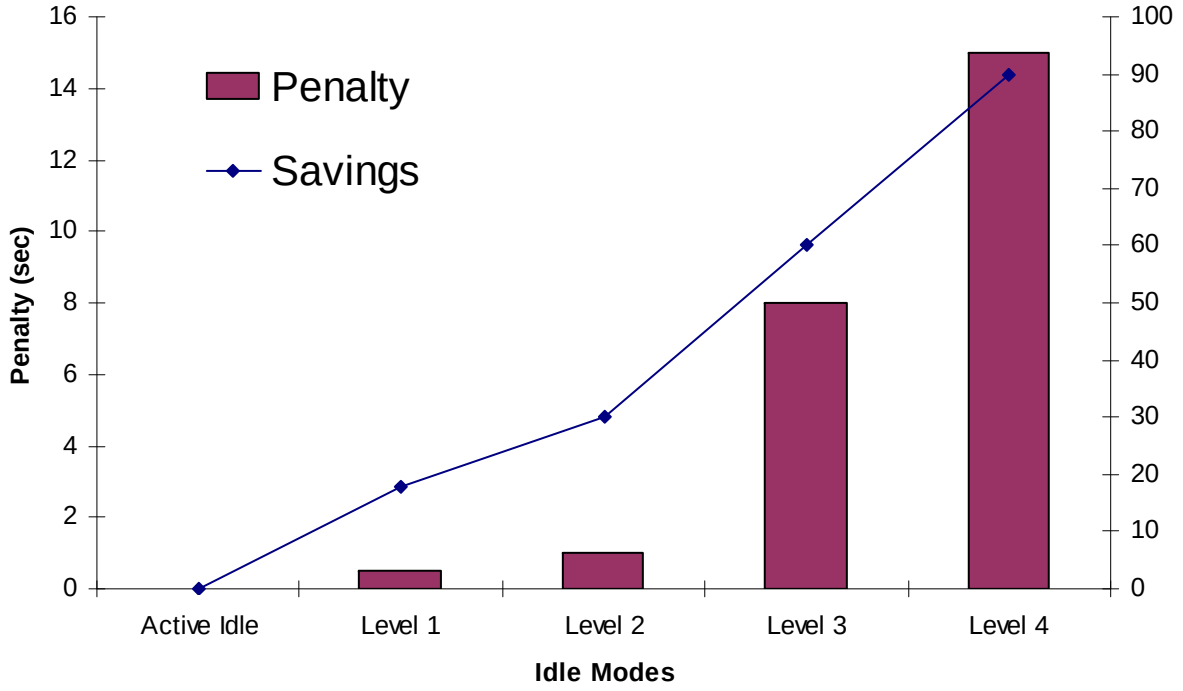
Estimation Engine



Example – Power Consumption

- In storage clusters, idle times explored to spin down disks to conserve power
  - Modeled as a low priority workload
- Interleaving this workload with user workload in a storage device requires to determine
  - When to start a power saving mode
  - How long a device should stay in a power saving mode
  - Without violating user workload performance

Power Savings Modes in Disks

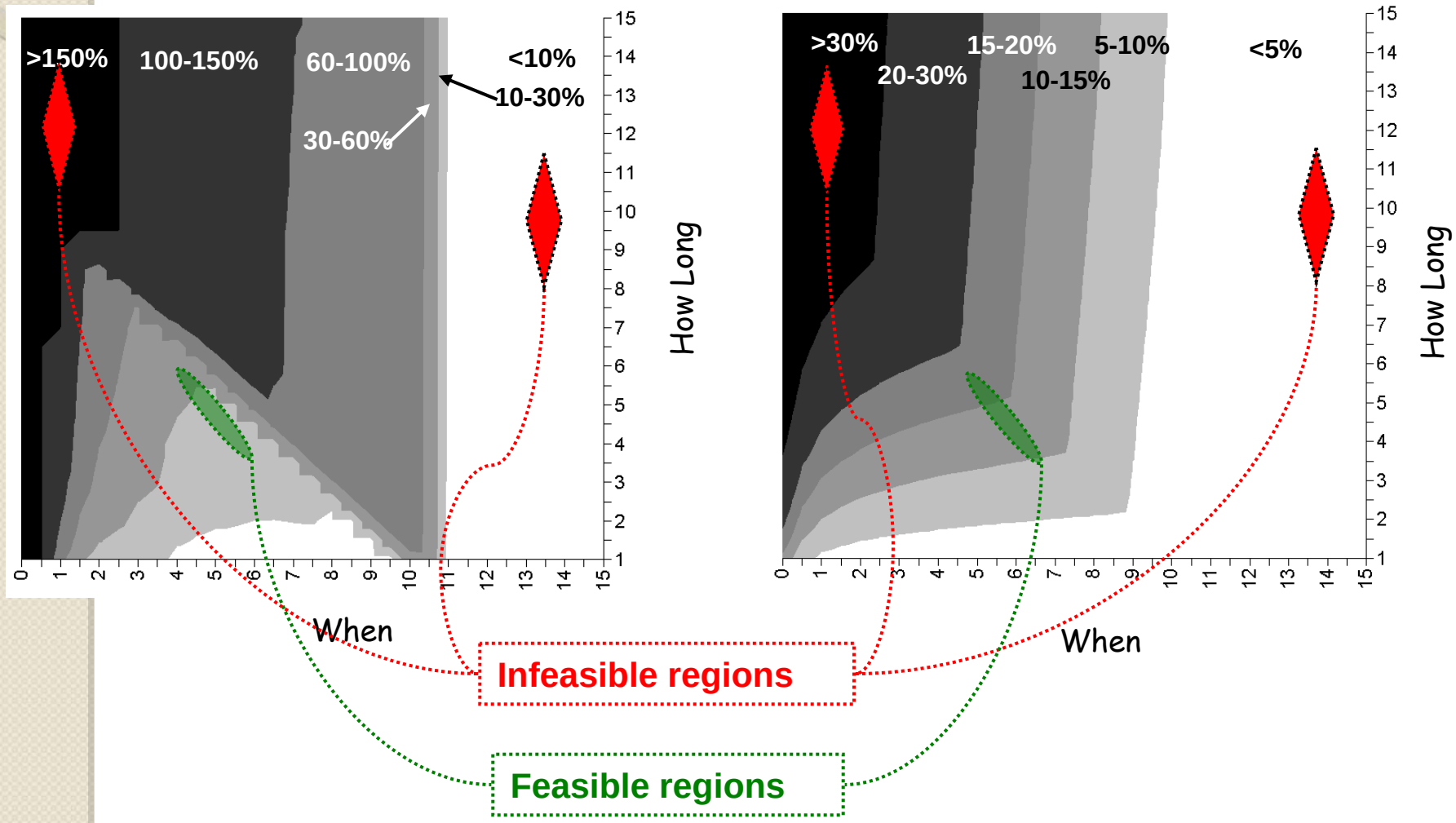


- Power savings modes in disk drives are associated with high penalties on user performance
  - The higher power savings the higher the performance penalty

Estimation of Power Saving Capabilities in Disk Drives

- The histogram-based estimation engine is used to identify
  - Power savings opportunities in a storage device given a QoS (slowdown in user workload)
  - Determine which power saving mode to use
  - Estimate if a workload shaping / reduction technique would be beneficial to use
  - Identify nodes in the cluster that would serve effectively to serve the workload shifted from the node to be placed in the power saving mode

Performance Slowdown vs. Power Savings



Interleaving Workloads in Storage Clusters

- Enhance estimation procedure at the single node level
- Scale single node estimations to a cluster of storage nodes

Single Node Workload Interleaving

- Improve accuracy
  - Identify additional histogram-like metrics
    - E.g., to capture burstiness
  - Critical for interleaving workloads with orders of magnitude difference in service demands
  - Seamlessly add these additional data structures in the current framework
- Extend the set of priorities of the interleaved workloads
  - Not just high/low
  - Dynamically changing priorities

Single Node Workload Interleaving

- Enhance the histogram-based estimation engine
  - Enable estimation of more complex metrics
    - Histograms of performance metrics
  - Add workload shaping into estimation
    - Remove portion of workload and estimate the effects
- Enhance the QoS guarantees
  - Meet percentiles rather than averages

Cluster-wide Workload Interleaving

- Incorporate single-node estimations into a middleware platform for efficient workload placement in a storage cluster
- Identify effective load balancing by
  - Estimating the benefits of placing a given workload in a node or set of nodes
  - Estimate the efficiency of various workload shaping techniques
    - Work-off loading
- Answer "What if" questions with regard to workload placements and interleaving