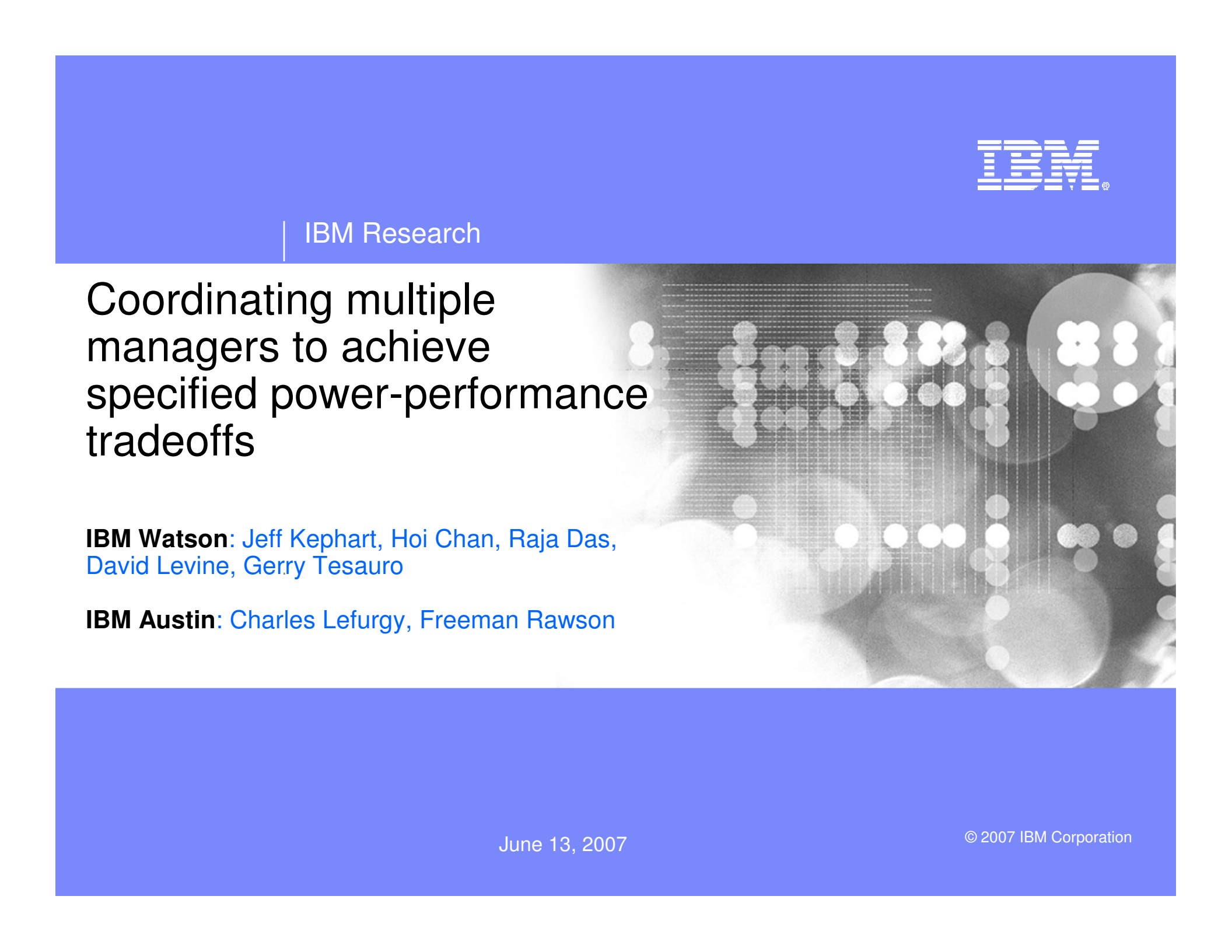


IBM Research

Coordinating multiple managers to achieve specified power-performance tradeoffs

IBM Watson: Jeff Kephart, Hoi Chan, Raja Das, David Levine, Gerry Tesauro

IBM Austin: Charles Lefurgy, Freeman Rawson

June 13, 2007

© 2007 IBM Corporation

Team Members

David Levine

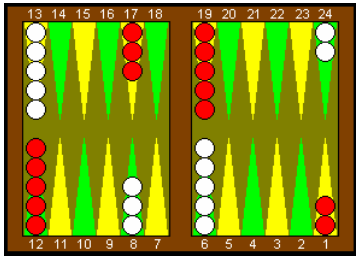
Hoi Chan

Jeff Kephart

Charles Lefurgy

Freeman Rawson

Raja Das



Gerry Tesauro

Austin Lab

Watson Lab

© 2007 IBM Corporation

Background

- The costs and constraints associated with electrical power are becoming an increasing concern for data center operation
  - 50% of data centers will have insufficient power and cooling capabilities by 2008 (Gartner)
  - Power will be the second-highest operating cost (after labor) in 70% of data centers
  - According to Berkeley RAD lab, 30% reduction in power would save $15B and 100M metric tons of CO_2 emissions per year in US (1.7% of total emissions)
- Power efficiency is becoming the subject of energy and environmental regulations by governments around the world
  - EPA report to Congress due in mid-2007; Energy Star standards for systems, data centers
  - Other agencies such as DoE and other governments involved as well
- Non-governmental organizations & industry groups actively studying the problem
  - Green Grid: consortium of IT companies developing best practices for reducing power consumption in data centers
    - AMD, Dell, HP, IBM, Sun Microsystems, Microsoft, VMWare, …
  - SPEC working on a power/performance benchmark
- Industry is aggressively developing and marketing power-conserving hardware and software solutions

© 2007 IBM Corporation

Agenda

- **Background**
- **Power-Performance Research**
  - Algorithms
  - Results
- **Commercialization**

© 2007 IBM Corporation

Universal algorithms for managing power and performance

- **Universally Optimal Power Management Algorithm**

```
Algorithm MaxPowerSavings()
For each object O in DataCenter
    TurnOff(O)
```

- **Universal Performance Management Algorithm**

```
Algorithm MaxPerformance()
For each object O in DataCenter
    TurnOnCompletely(O)
    PerformanceManage(O)
```

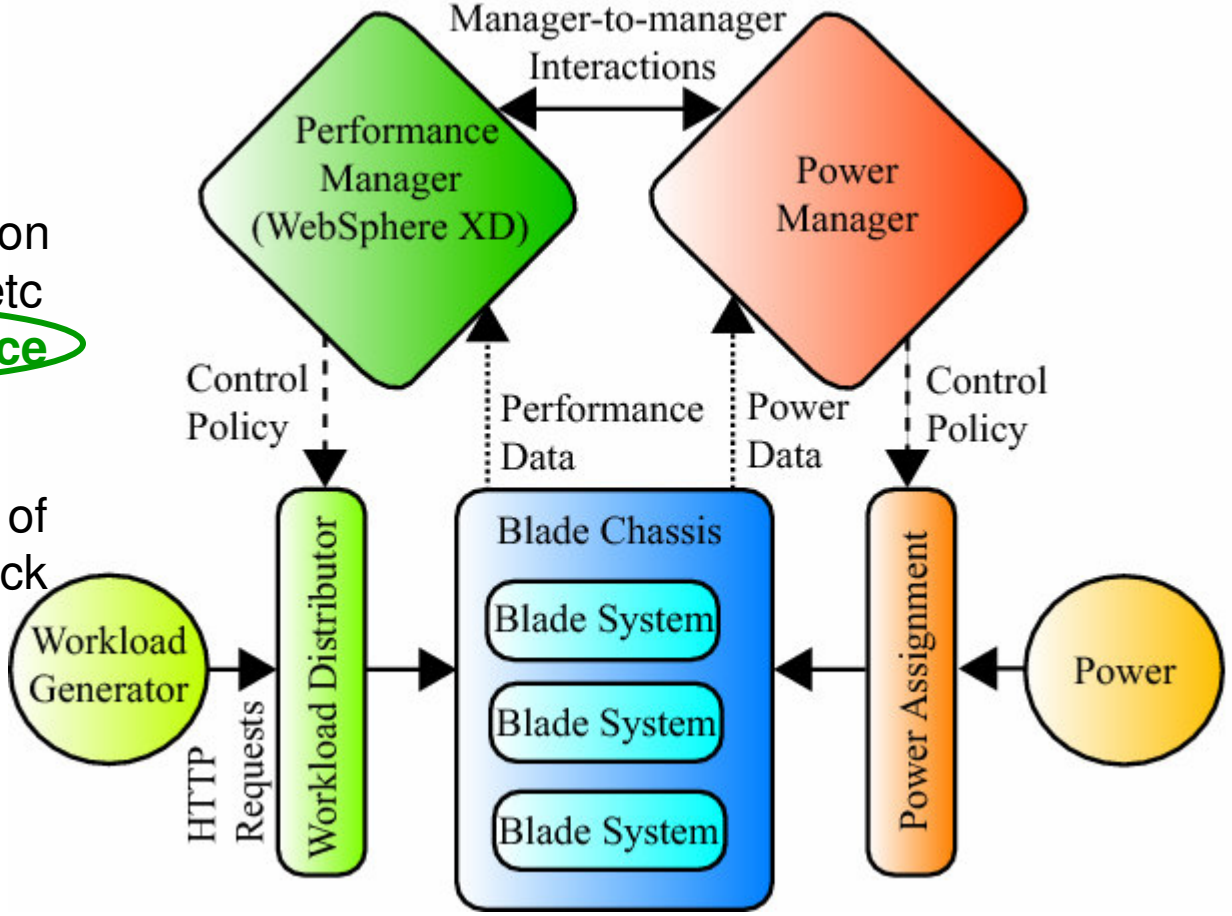
Now all we need to do is combine these algorithms somehow…?!?

© 2007 IBM Corporation

Power and Performance Management

- **Performance agent**
  - IBM WebSphere middleware adjusts load balancing, CPU application placement parameters ,etc to **maximize performance**
- **Power agent**
  - Autonomic Management of Energy throttles CPU clock to slow down processor and **save power**

An inherent conflict!

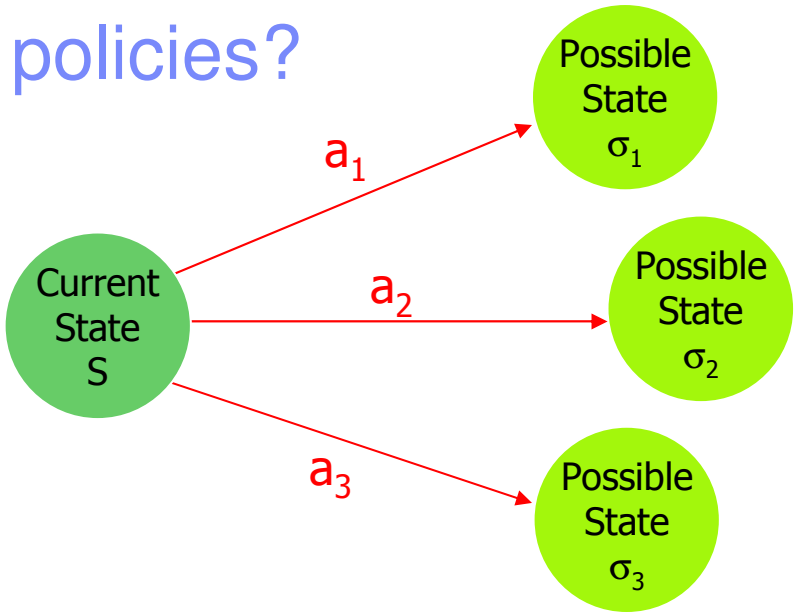


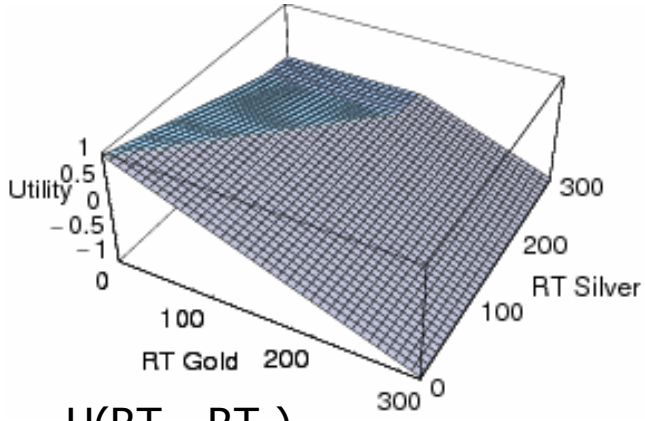
© 2007 IBM Corporation

Kephart and Walsh, Policy04

How to represent high-level policies?

- ***Utility functions*** map any possible state of a system to a scalar value
- They can be derived from
  - a service level agreement
  - preference elicitation techniques
  - simple templates, e.g. specify response time thresholds and "importance" levels
- They are a generally useful representation for high-level objectives, e.g.
  - Minimize power while meeting SLA
  - Maximize performance while meeting power constraint
  - Range in between





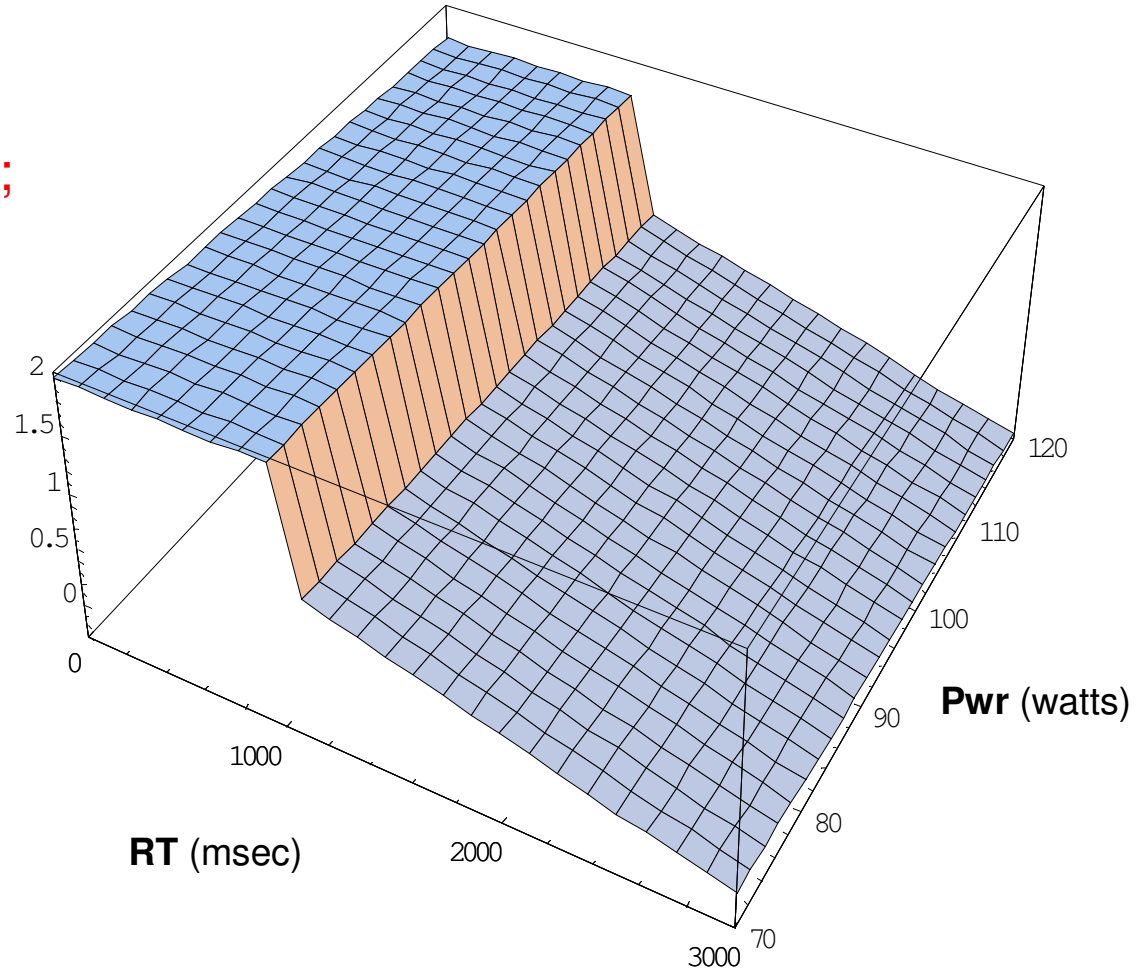
$U(RT_G, RT_S)$

 © 2007 IBM Corporation

Power-performance utility functions

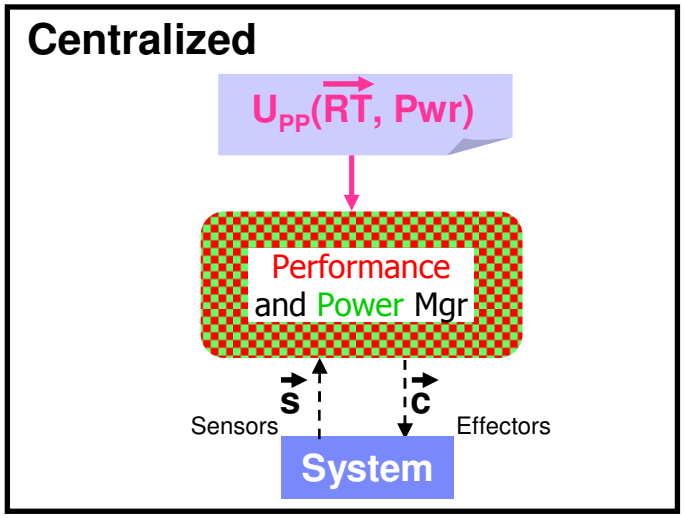
- U(perf, pwr) = U(perf) – ε Pwr;
  Pwr < Pwr_{Max}

- U(perf, pwr) = U(perf)/Pwr;
  Pwr < Pwr_{Max}



 © 2007 IBM Corporation

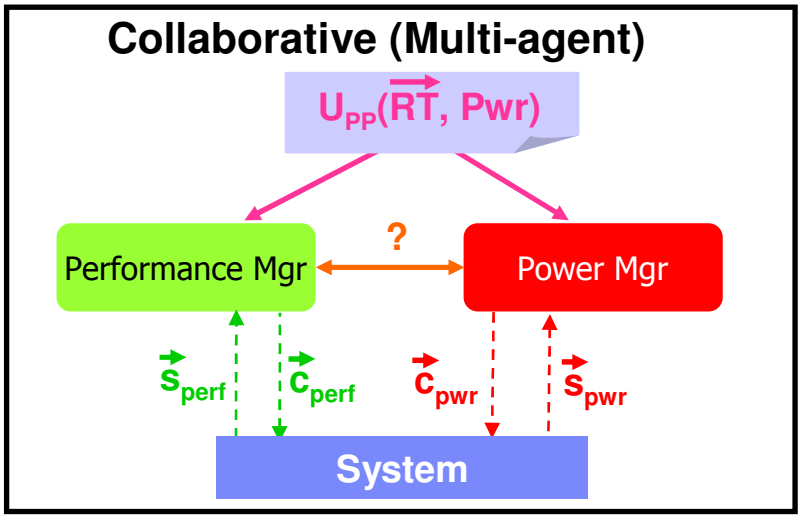
Multiagent approach to Power-Performance optimization



Set effectors c to maximize U_{PP}

Conceptually easiest

Not very practical!



What info should be exchanged, and how?

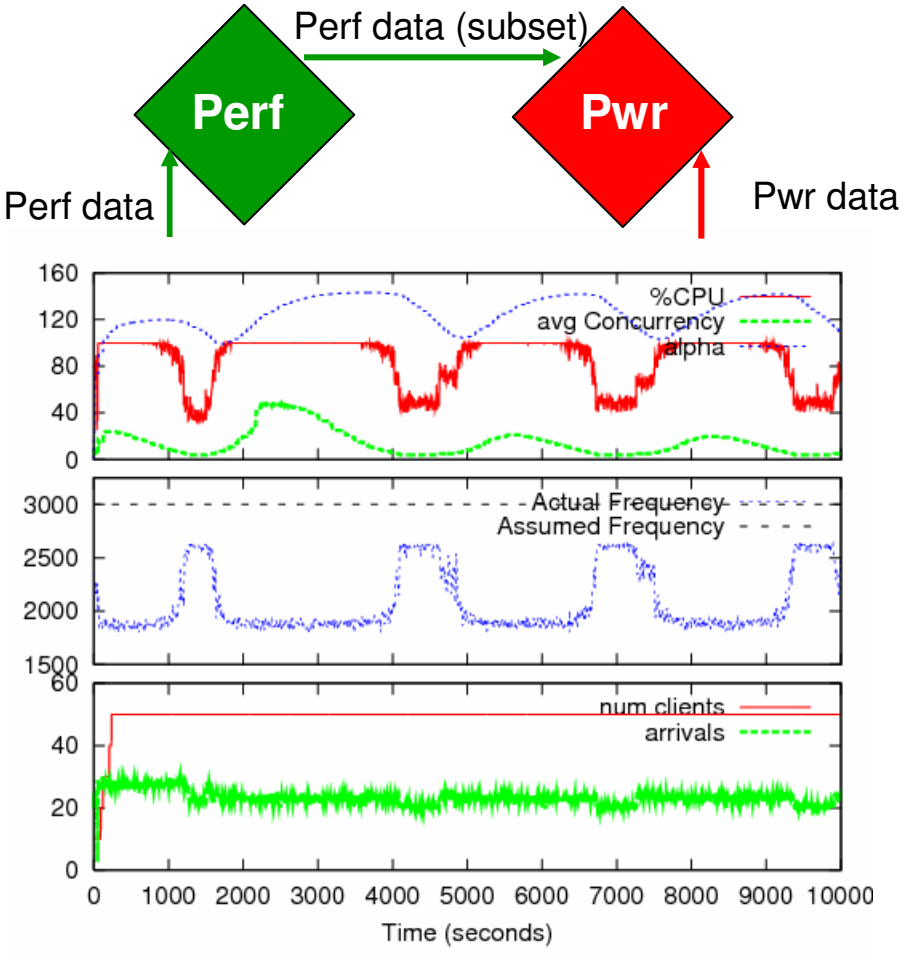
- Do we need negotiation?
- Do we need mediation?
- What are the right power control knobs?

Strategy: Try the simplest method first

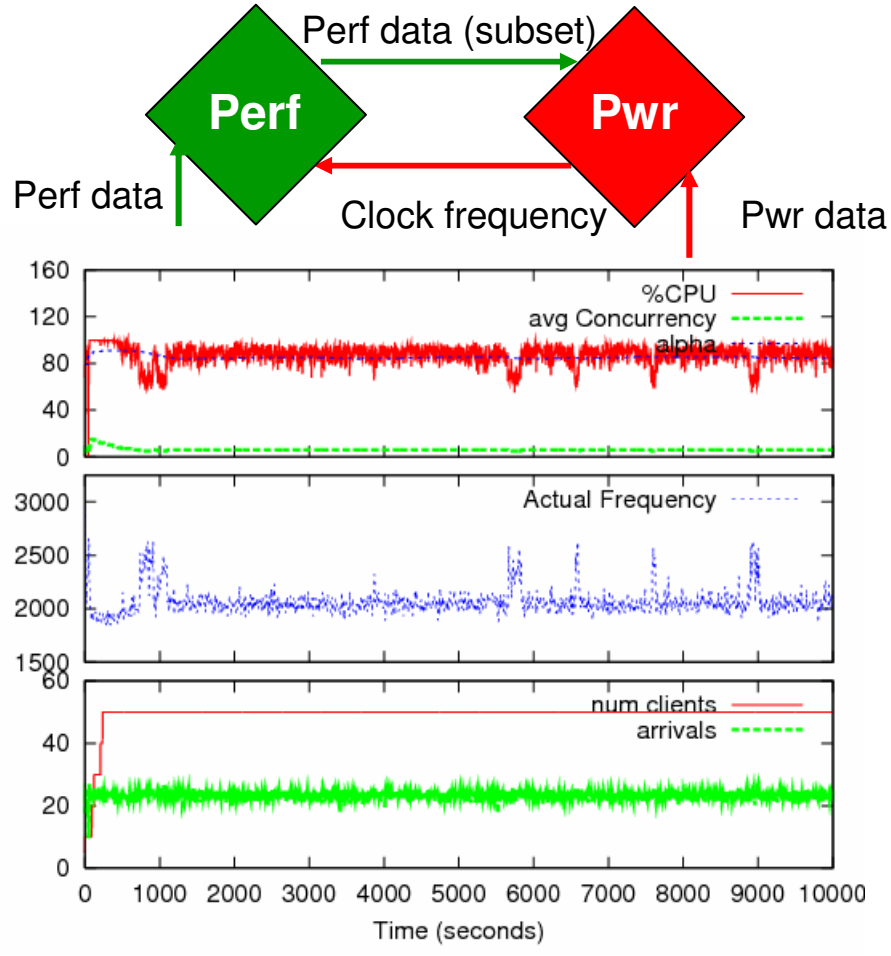
- No negotiation or mediation
- Power control knobs = power cap settings
- Minimize changes to WXD
- Add complexity only if/when really needed

© 2007 IBM Corporation

What needs to be communicated?



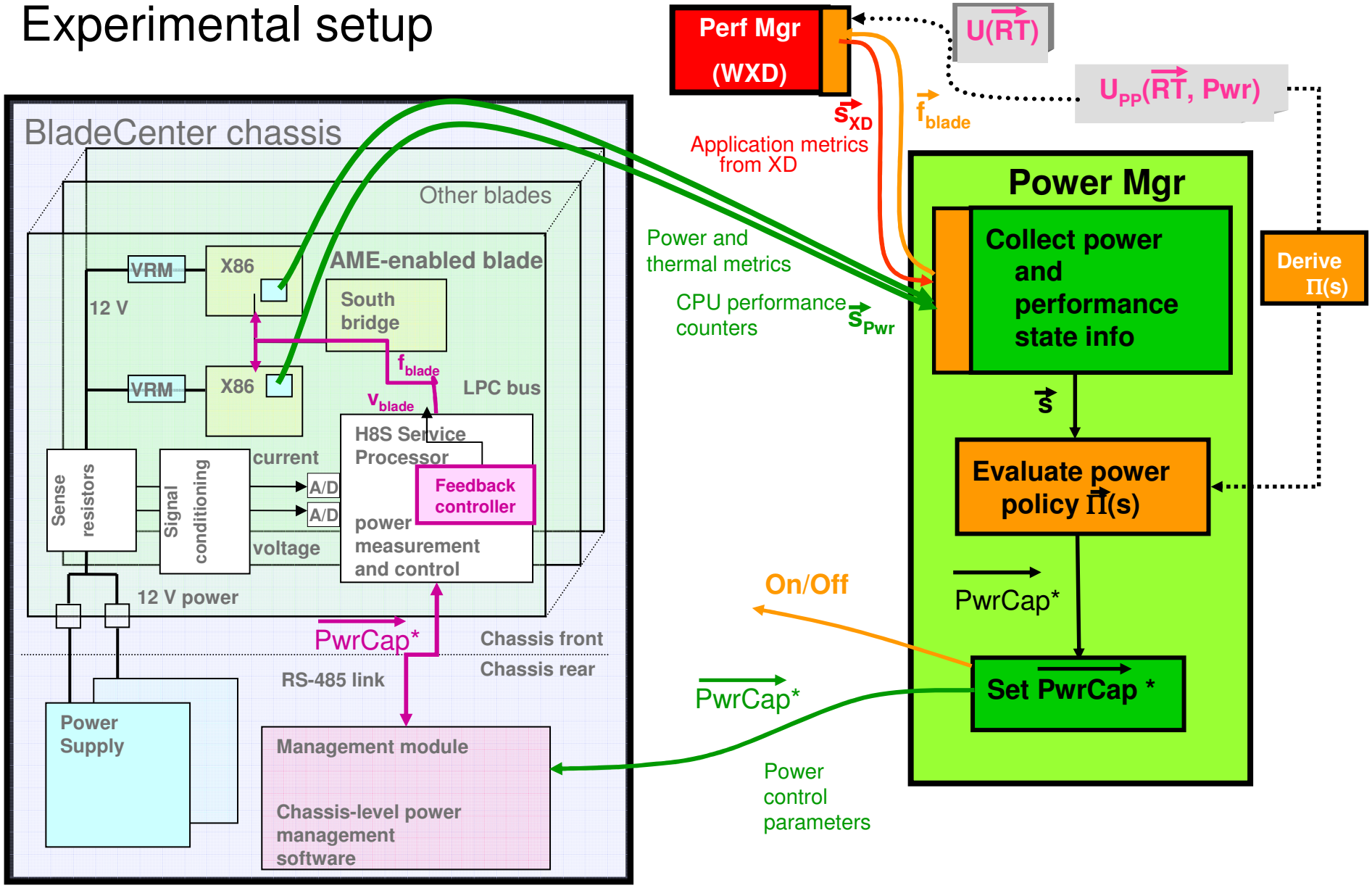
No frequency feedback



With frequency feedback

© 2007 IBM Corporation

Experimental setup



BladeCenter product with added software agents

© 2007 IBM Corporation

Deriving a Powercap Policy

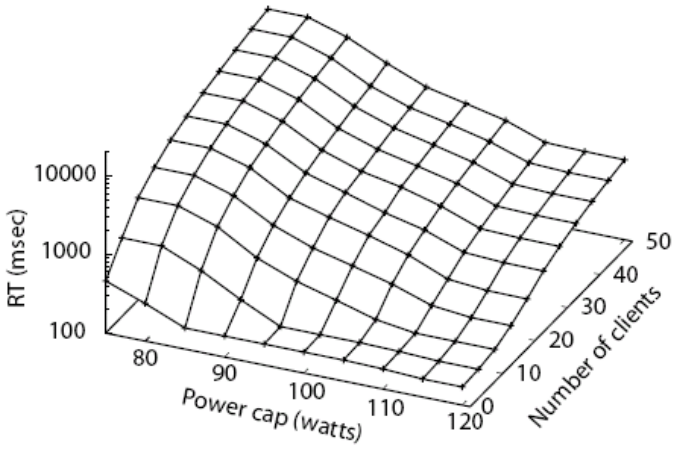
- Measure offline how performance and power use depend upon system state s and power cap Pcap for each blade
  - *~1 week of data collection*
- Use interpolation or regression to generate models
  - RT(s, Pcap)
  - Pwr(s, Pcap)

Models

- Substitute models into utility function
  - U_{pp}(RT(s,Pcap), Pwr(s,Pcap) = U'(s, Pcap)
- For given state s, determine Pcap that maximizes U'
  - Pcap*(s) = $argmax_{Pcap}$ U'(s, Pcap)

Optimization

RT(n_c, Pcap)



Pwr(n_c, Pcap)

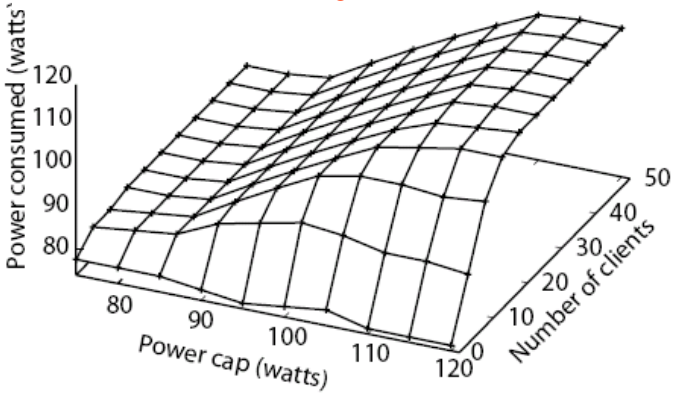
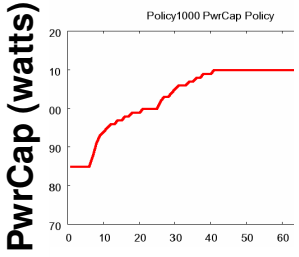


Figure 4: a) Experimentally-determined models of a) RT(p_κ, n_c) (note log scale) and b) Pwr(p_κ, n_c) for GB.

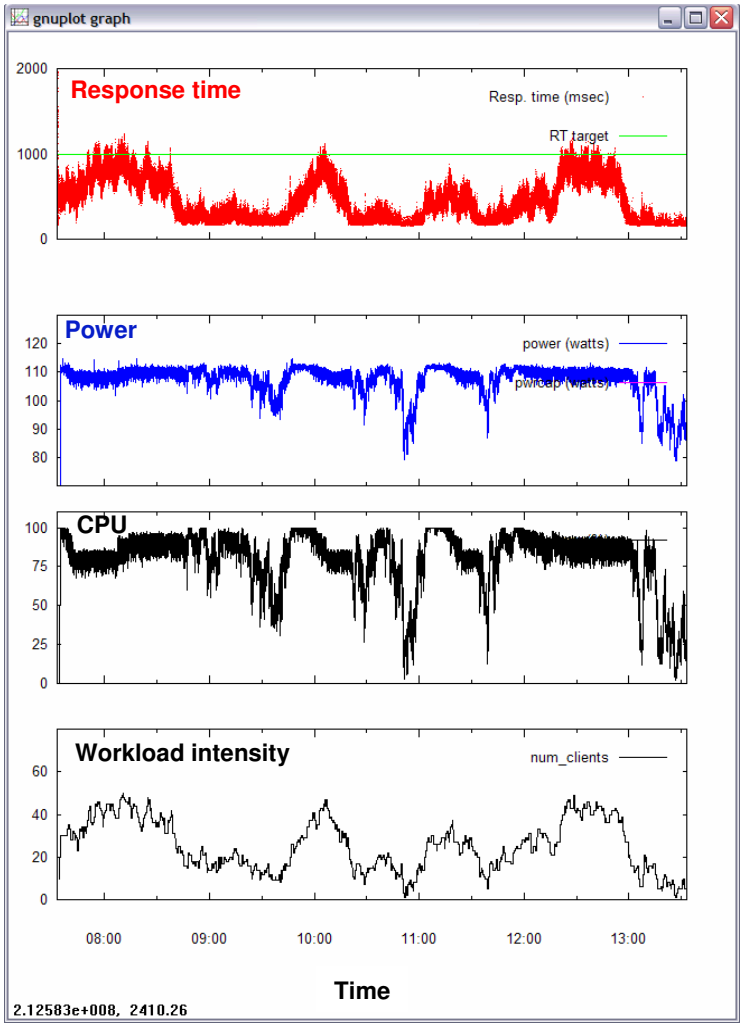
PwrCap Poli



© 2007 IBM Corporation

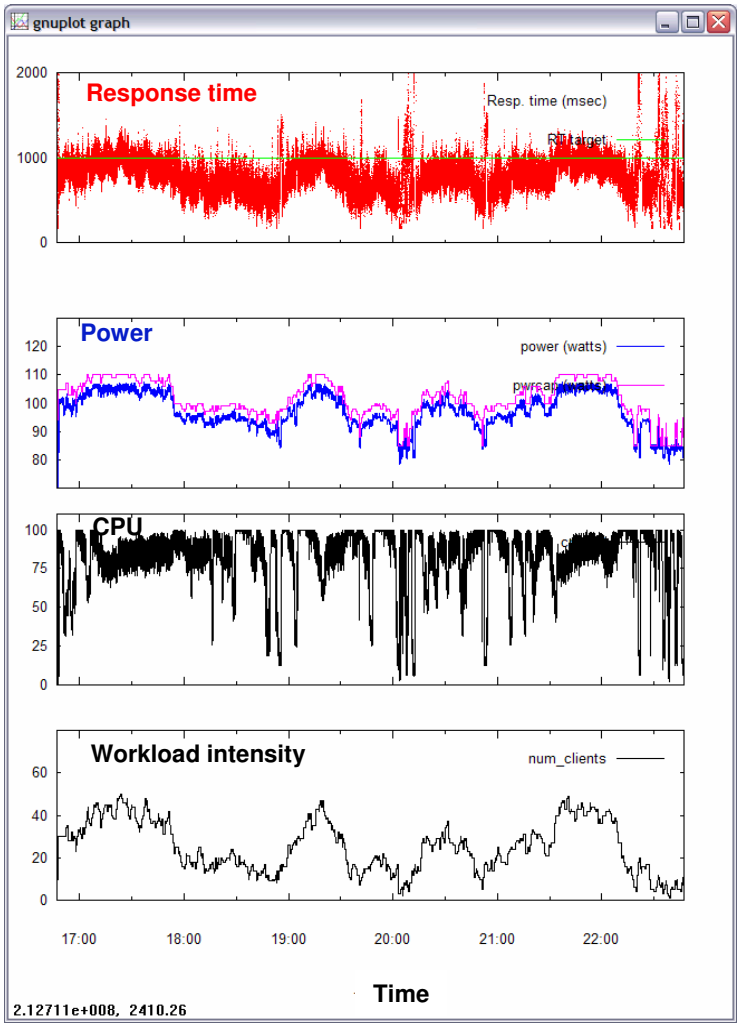
Experiment: Hand-crafted Power Policy

No power management



Avg power = 107.9 watts

Power management, using Policy HC01



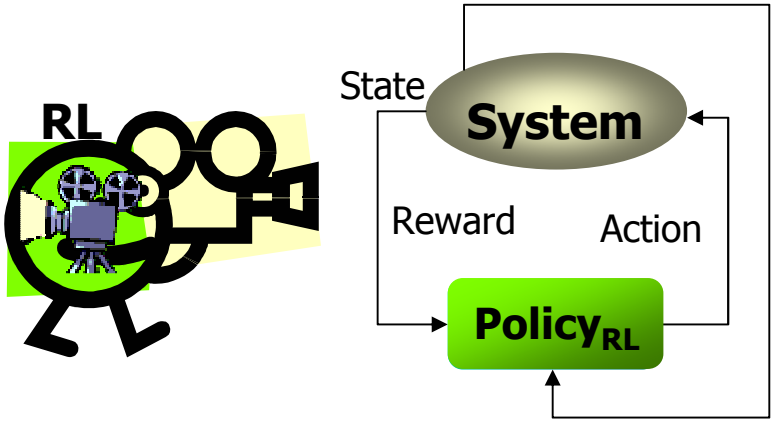
Avg power = 96.6 watts (savings: 11.3 watts = 10.5%)

© 2007 IBM Corporation

Hybrid Reinforcement Learning

Tesauro et al., ICAC 2006

- Reinforcement learning methods learn to make good decisions by
  - observing <state, action, reward> tuples
  - learning long-range value functions V(state, action)
  - Abiding by optimal policy: when in state s, take action a that maximizes V(s, a)
- Typically, they learn by updating V(state, action) starting from random assumptions
- This can take a long time, and performance can be very poor during the learning phase
- We invented a new RL technique, Hybrid RL, that starts from an existing policy, and improves upon it
- Very general method that automatically improves ***any*** existing systems management policy
  - No knowledge engineering needed



State = {power, performance metrics}

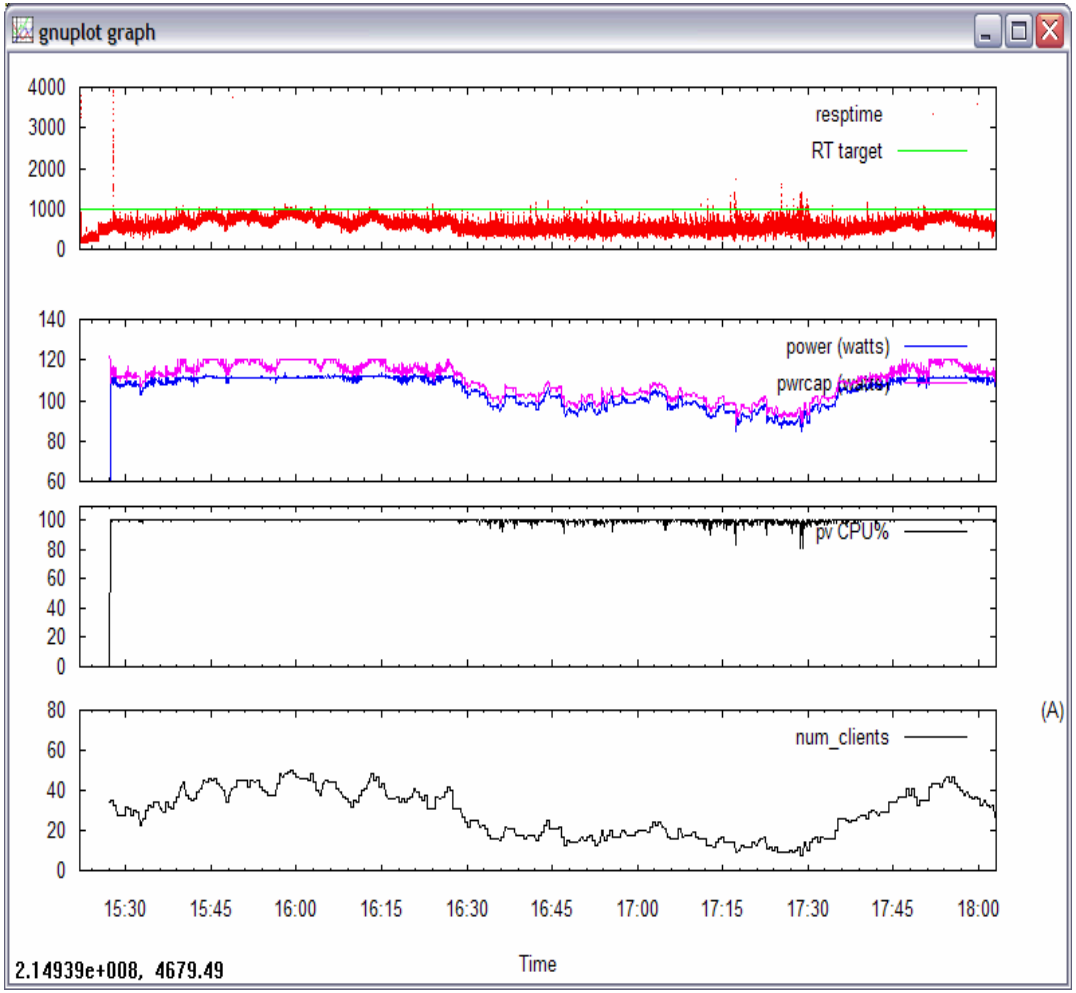
Action = {powercaps}

Reward = U(perf, pwr)

 © 2007 IBM Corporation

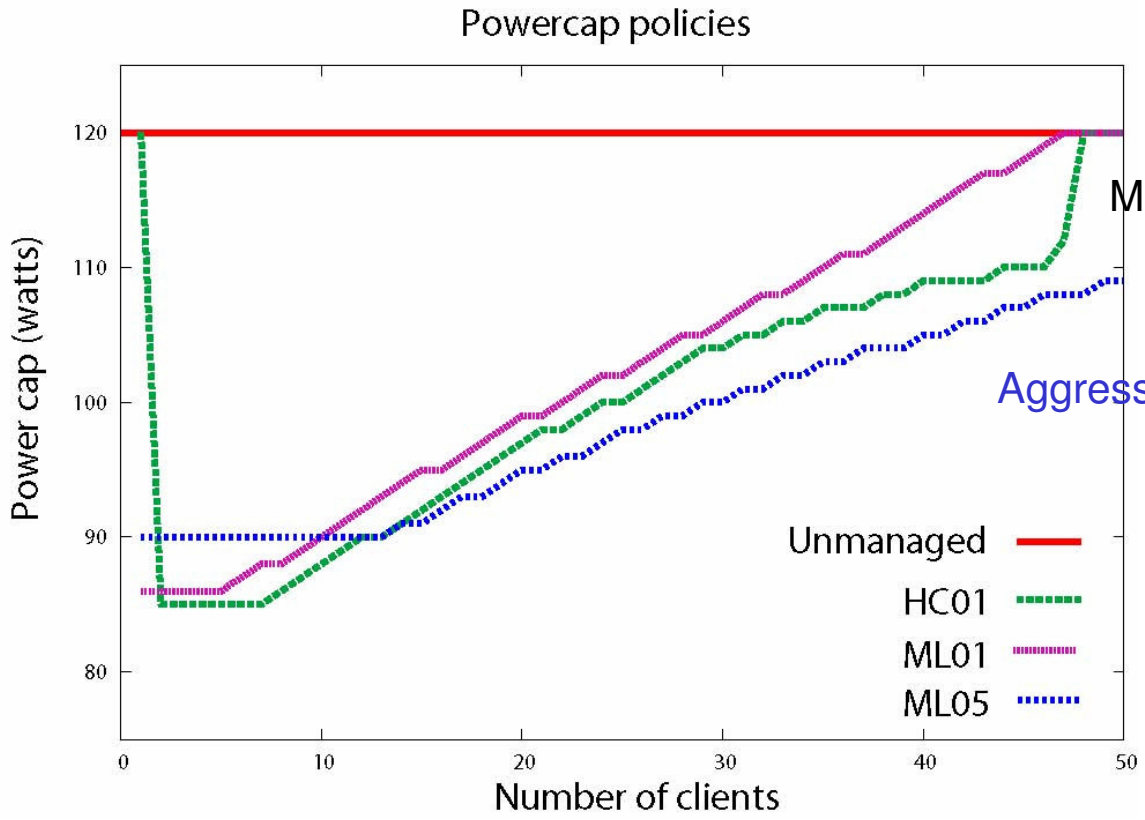
Hybrid RL improves our initial power policy!

- **Good power savings**
  - 8.9% less power than for unmanaged case
  - Was 10.5% for hand-tuned policy
- **Reduced SLA violations**
  - 1.5% of response times exceed threshold
  - Was 21% for hand-tuned policy



© 2007 IBM Corporation

Powercap policies



Moderate power saving

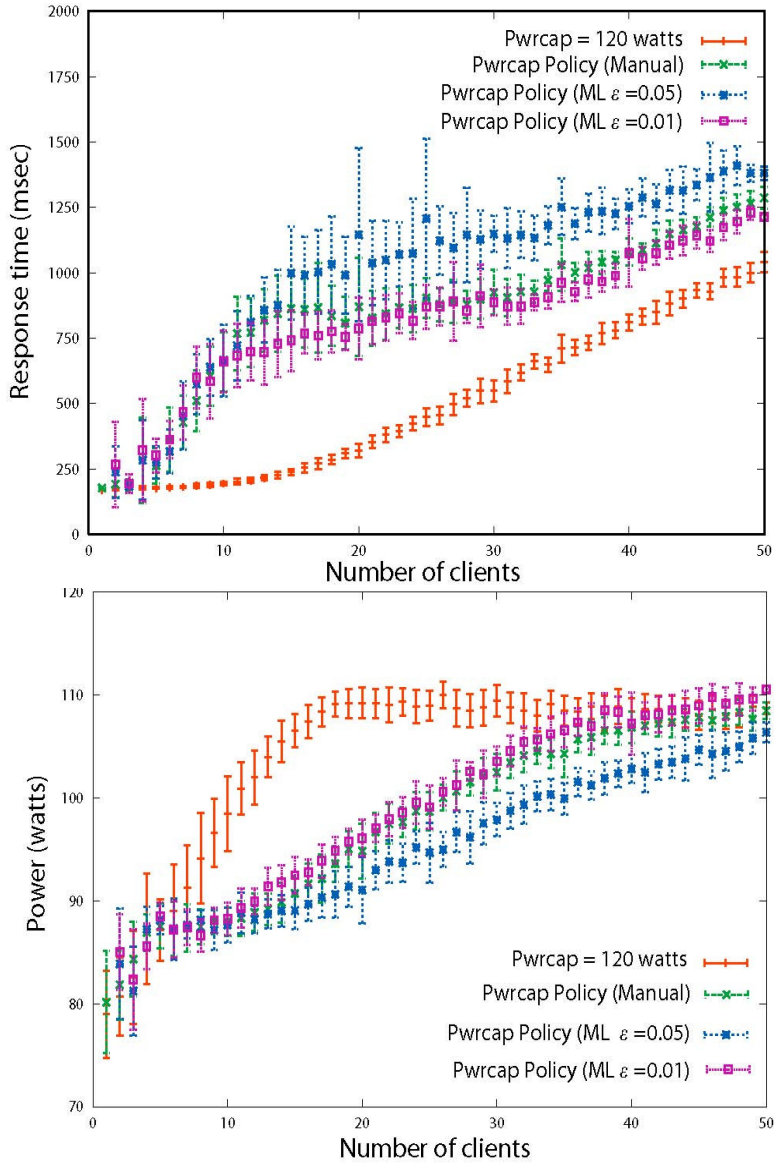
Aggressive power saving

© 2007 IBM Corporation

Where is the power being saved?

- **Power is saved primarily when the number of clients is moderate**
  - For low workloads, power consumption is not constrained by powercaps
  - At high workloads, utility is maximized by setting high powercaps

Policy	Avg watts	Savings
Unmanaged	104.9	0%
HC01	95.3	9.2%
ML01	96.1	8.4%
ML05	92.7	11.6%



© 2007 IBM Corporation

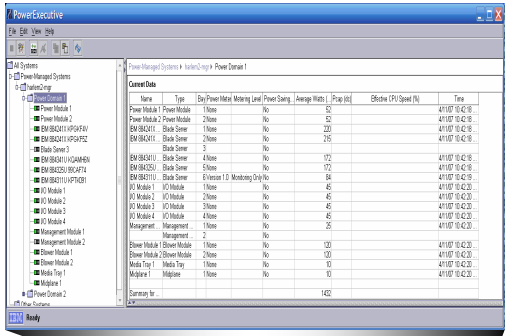
Agenda

- **Introduction**
- **Power-Performance Research**
  - Algorithms
  - Results
- **Commercialization**

© 2007 IBM Corporation

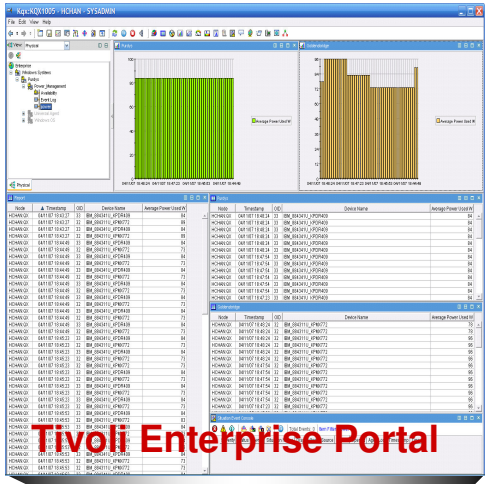
ITM Power Agent

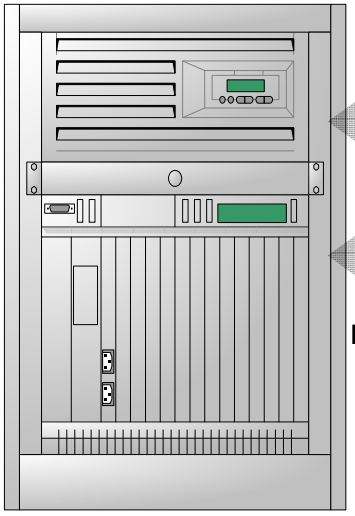
PowerExecutive is an IBM Director plug-in that interacts with the hardware management module to monitor power usage and temperature, and control power caps.

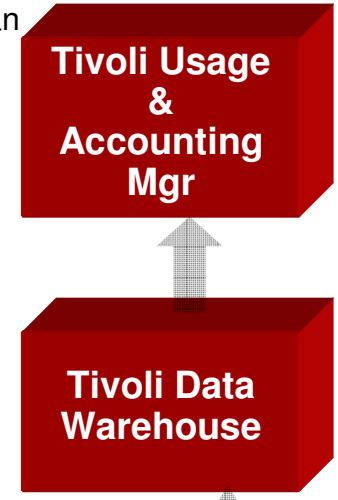


TUAM and other Tivoli managers can consume the historical aggregated performance and power data stored in **TDW**.

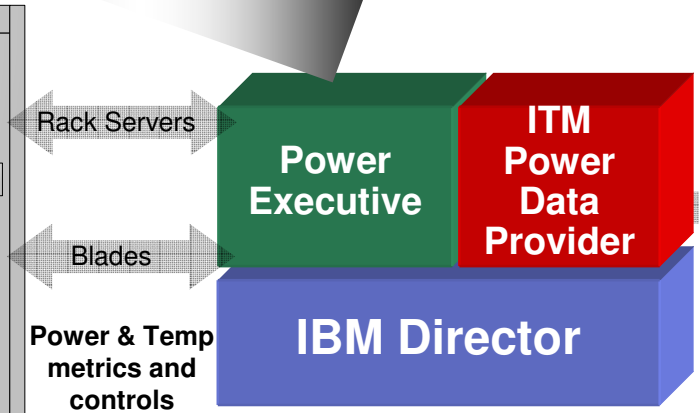
Admins can view and act upon live aggregated data and alerts on **TEP**.







Performance, Power, Temp



Power & Temp metrics and controls

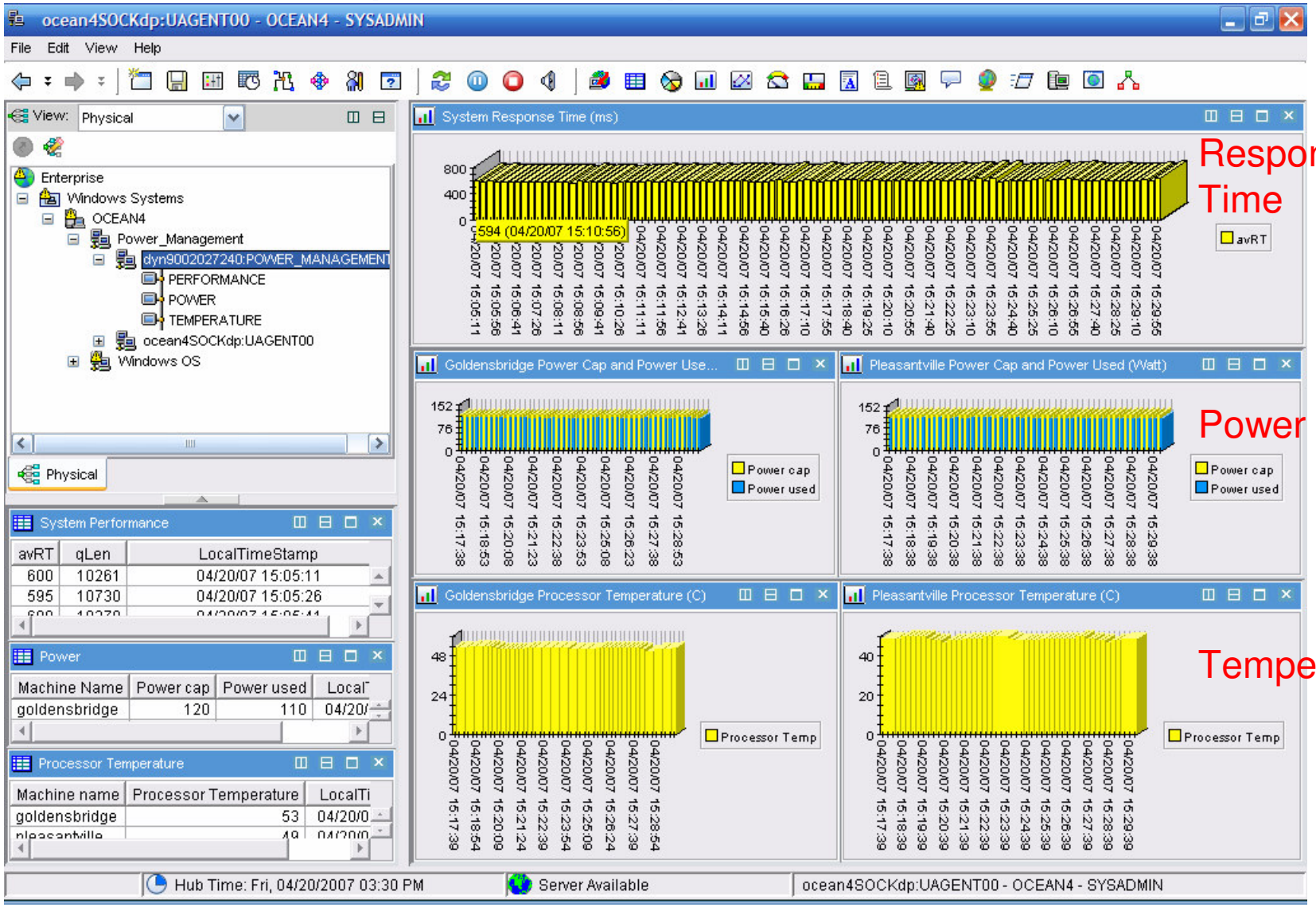
Power, Temp

ITM Power Data Provider is an IBM Director extension that feeds the **ITM Power Agent**.

ITM Power Agent augments performance data traditionally collected from performance managers and the OS with power and temperature data. All of these data are aggregated for consumption by **TEP** and **TDW**.

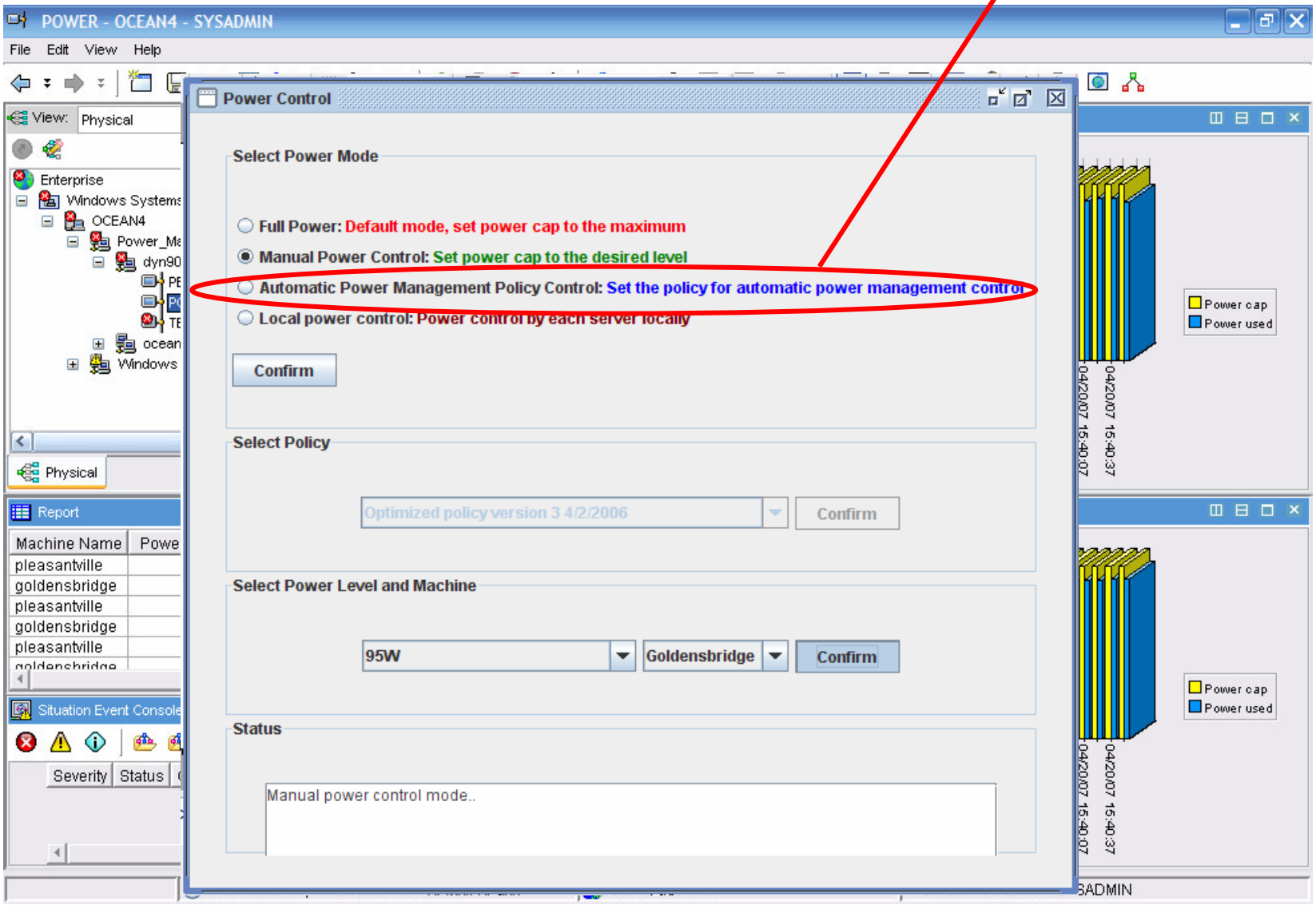
© 2007 IBM Corporation

Admin console (power and performance data)



© 2007 IBM Corporation

Power Management interface



© 2007 IBM Corporation

Summary and Future Directions

- **Thus far, we have achieved coordination across**
  - Multiple levels (from chip to application)
  - Two management disciplines (performance, power)
  - Approach
    - Express joint objectives in terms of utility functions
    - Combine modeling, optimization and state-of-art ML technologies
      - ML can save time and yield better policy
- **We are currently exploring**
  - Dynamic *voltage* and frequency scaling
  - Dynamic power-off of servers, exploiting virtualization and load balancing
    - **> 30% power savings in initial tests**
  - An additional management discipline: availability
- **Opportunities to extend to data center level**

© 2007 IBM Corporation