IEEE GlobeCom 2005

Autonomic Power Management Schemes for Internet Servers and Data Centers

L. Mastroleon, N. Bambos, C. Kozyrakis, D. Economou

December 2005
St Louis, MO

Outline

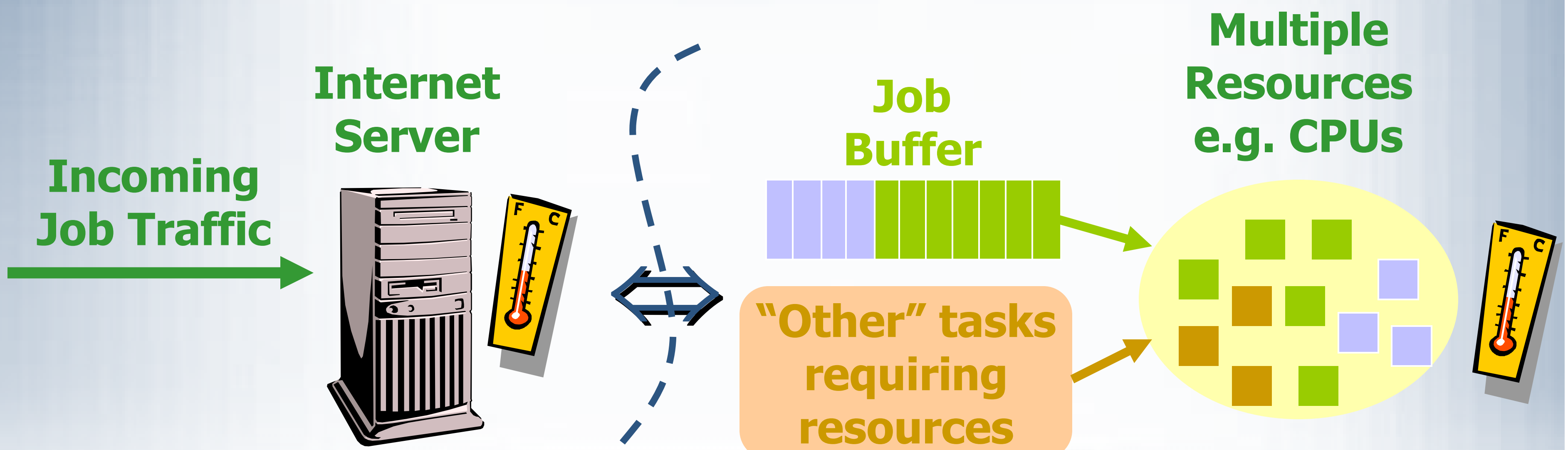
- Motivation
- The System Model and the DP Formulation
- A Justified Heuristic
- Simulation and Evaluation
- Conclusions & Future Work

Outline

- Motivation
  - ✓ Typical Internet Server
  - ✓ Typical Data Center
- The System Model and the DP Formulation
- A Justified Heuristic
- Simulation and Evaluation
- Conclusions & Future Work

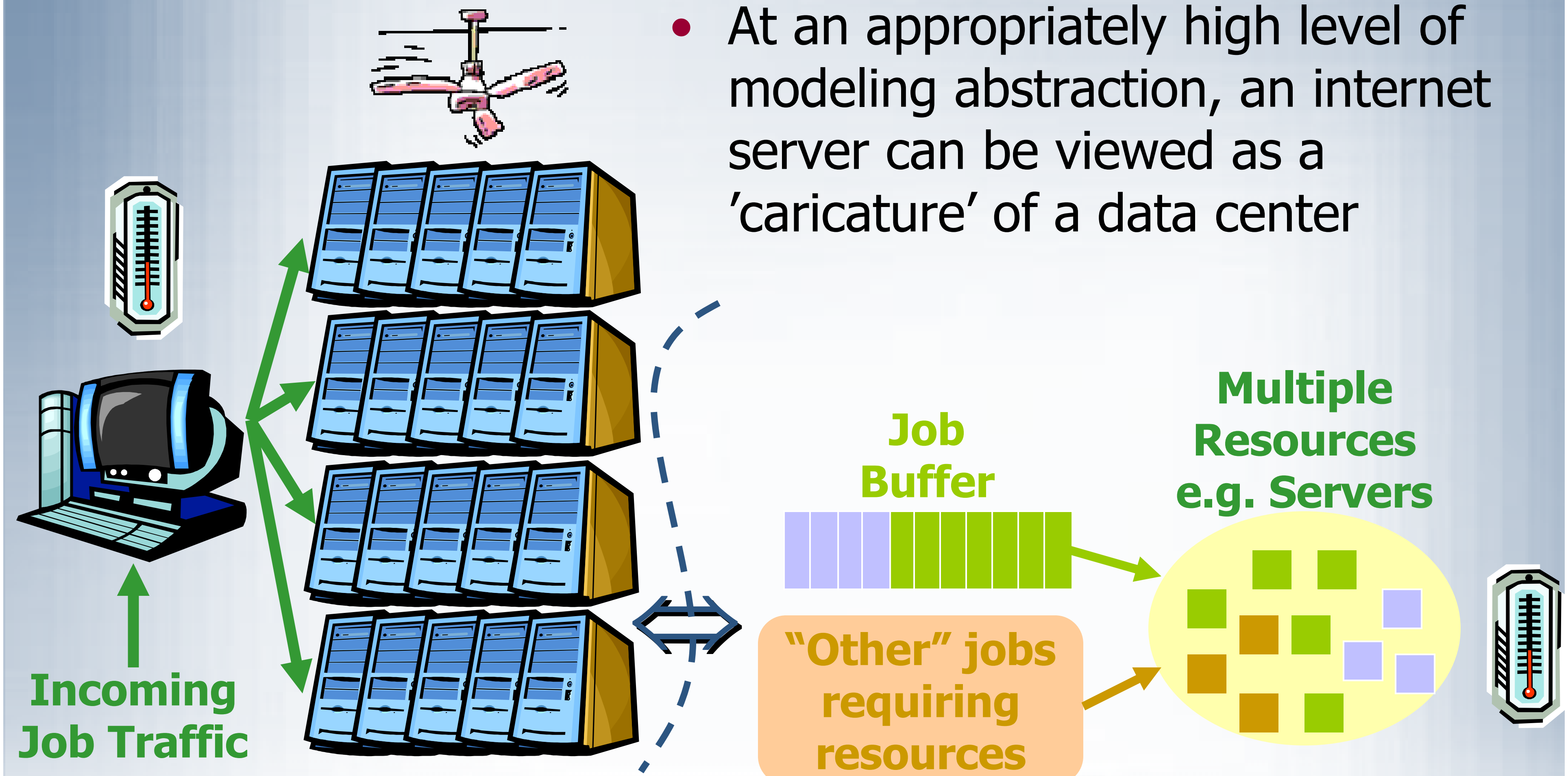
Typical Internet Server

- Internet Server has multiple resources (e.g. CPUs)
- Incoming jobs are placed in a buffer
- "Other" tasks need also to be completed
- Server has to appropriately allocate its resources
- Temperature of operation is important



Typical Data Center

- At an appropriately high level of modeling abstraction, an internet server can be viewed as a 'caricature' of a data center

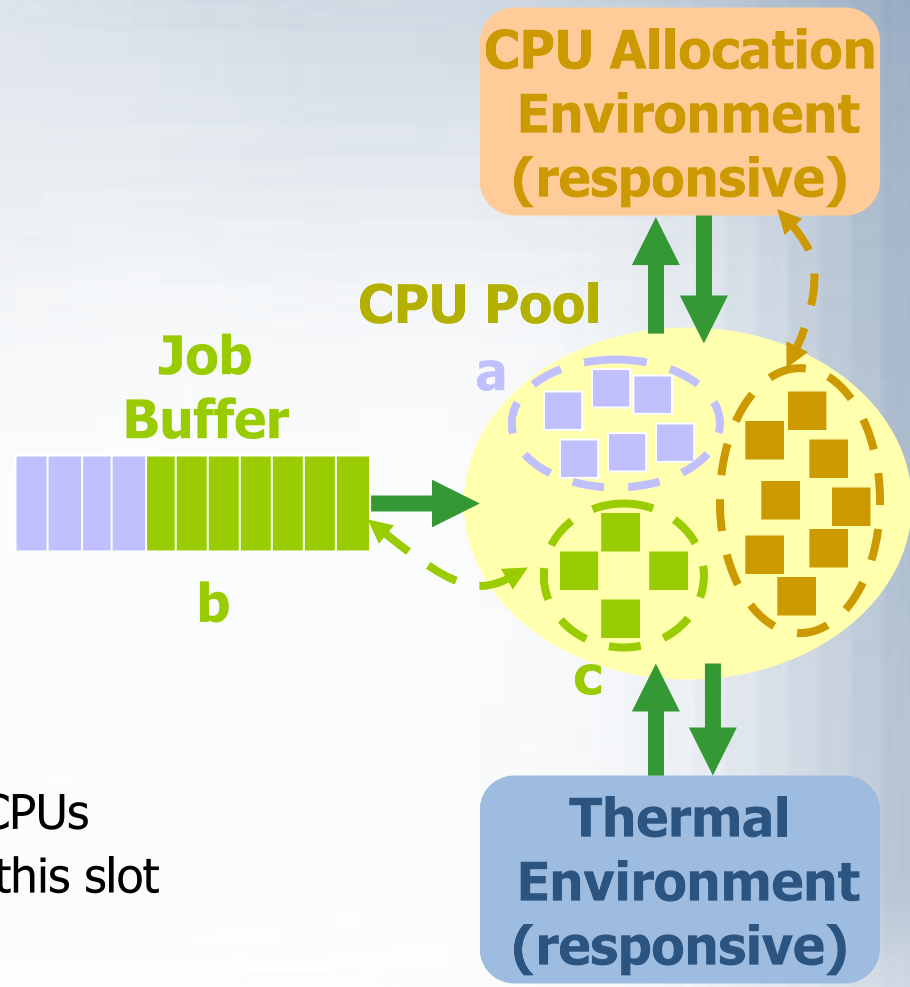


Outline

- Motivation
- The System Model and the DP Formulation
  - ✓ The Model
  - ✓ CPU Allocation Environment & Thermal Environment
  - ✓ Associated Costs
  - ✓ The Optimization Problem
  - ✓ Bellman's Equation
- A Justified Heuristic
- Simulation and Evaluation
- Conclusions & Future Work

The model

- Slotted time
- CPU Pool of Size ***M***
- Thermal State Set ***T***
- Job Buffer of Size ***B***
- At the beginning of a slot:
  - ✓ ***c*** is the# of CPUs used by buffer
  - ✓ ***a*** is the # of available CPUs
  - ✓ ***(M-a-c)*** is the # of unavailable CPUs
  - ✓ ***u*** is the # of CPUs to use during this slot



CPU Allocation Environment & Thermal Environment

- CPU Allocation Environment
  - ✓ At the beginning of a slot the state is $\boldsymbol{a}$
  - ✓ After decisions the state is $\boldsymbol{a^* = (a+c-u)}$
  - ✓ At the beginning of the next slot the state will be $\boldsymbol{a'}$
  - ✓ The transitions $\boldsymbol{a^* \rightarrow a'}$ will be Markovian $\boldsymbol{p_{a^*a'|u}}$

- Thermal Allocation Environment
  - ✓ State remains the same within a time slot
  - ✓ Current state is $\boldsymbol{t}$
  - ✓ Next state will be $\boldsymbol{t'}$
  - ✓ The transitions $\boldsymbol{t \rightarrow t'}$ will be Markovian $\boldsymbol{q_{tt'|(c,a,u)}}$

- Statistically Independent Environments

Associated Costs

- Costs incurred in each time slot:
  - ✓ Backlog Cost $C_b(b)$
    - ➢ Increasing in b
  - ✓ Power Cost (& Thermal Stress) $C_{ut}(u,t)$
    - ➢ Increasing in u & t (temperature)
  - ✓ Reconfiguration Cost $C_{uc}(u,c)$
    - ➢ Depends on the difference $(u-c)$

The Optimization Problem

- At time 0: System starts with a full Buffer
- Objective: Empty the buffer with minimum overall cost
- At any time slot:
  - ✓ The state is: ***(b, t, c, a)***
  - ✓ The management decision is: ***u***
  - ✓ At most one job will finish with probability ***s(u)***
- This is a transient problem
- The solution depends on the state but not the specific time slot

Bellman's Equation

- *J(b,t,c,a)* cost-to-go beginning from state *(b,t,c,a)*

$$J(b,t,c,a)=\min_u \{\xi C_b(b) + (1-\xi)(C_{ut}(u,t) + C_{uc}(u,c))$$
$$+s(u) \Sigma\, q_{tt'|(c,a,u)}\, p_{a*a'|u}\, J(b-1,t',u,a')$$
$$+(1-s(u)) \Sigma\, q_{tt'|(c,a,u)}\, p_{a*a'|u}\, J(b,t',u,a') \}$$

buffer state : $b \in \{0,..,B\}$
thermal state : $t \in T$
CPUs used : $c \in \{0,..,M\}$
CPUs available : $a \in \{0,..,M-c\}$

$\Rightarrow S=(b, t, c, a)$

$\xi \in [0,1]$

CPUs to use : $u \in \{0,..,c+a\}$ $\Rightarrow u$

Backlog cost : $C_b(b)$
Power cost : $C_{ut}(u,t)$
Reconfiguration cost : $C_{uc}(u,c)$

Outline

- Motivation
- The System Model and the DP Formulation
- A Justified Heuristic
  - ✓ Key Insights
  - ✓ The Formula
- Simulation and Evaluation
- Conclusions & Future Work

Key Insights

- $u \leq (c+a)$
- $b=0 \Rightarrow u=0$
- $u \geq 1$ *if* $b>0$ *and* $(c+a) \geq 1$ (not necessarily true)
- $u(b+1,t,c,a) \geq u(b,t,c,a)$
- $u(b,t,c,a) \geq u(b,t,c,a')$ if $a' \geq a$ (not necessarily true)
- The heuristic should incorporate the parameter ξ

The formula

- The heuristic is described by the following formula:

$$u(b,t,c,a)=\min\{h(c+a,t)\,,\ 1_{\{b>0\}}(1+\mathrm{round}(f(\xi)h(c+a,t)\frac{b}{B}-\frac{a}{M-c+1}))\}$$

- $h(c+a,t)$ limits the CPUs based on t
- $f(\xi)=2^{(3\xi-1)}$
- Note $f([0.0\ 0.5\ 1.0])=[0.5\ 1.4\ 4.0]$

Outline

- Motivation
- The System Model and the DP Formulation
- A Justified Heuristic
- Simulation and Evaluation
  - ✓ Simulation Environment
  - ✓ Simulation Details
  - ✓ Simulation Results
- Conclusions & Future Work

Simulation Environment

- Evaluation with simulations using: OMNET++

- Two cases of arrivals:
  - ✓ *Constant Rate* – inter-arrival time ~ exp(2)
  - ✓ *Bursty* – sequential constant rate and idle cycles with length ~ exp(100)

- Executed our simulation for $\xi \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$

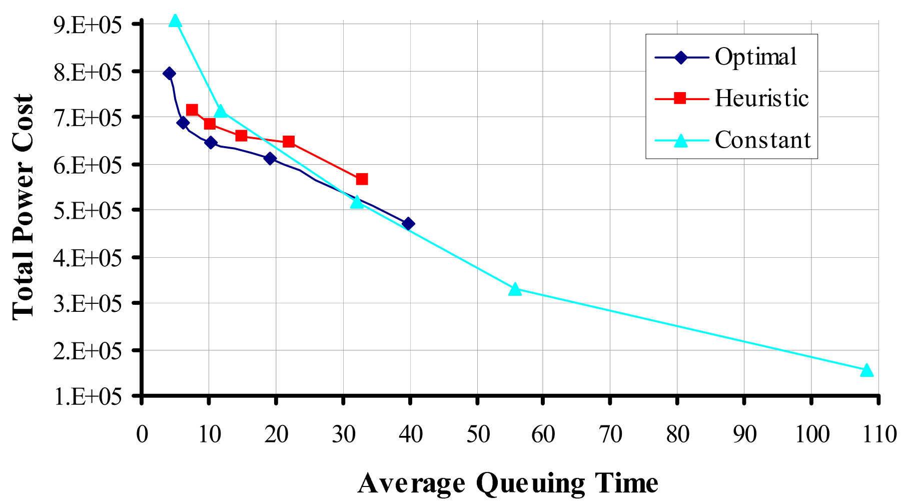
- Our benchmark was a constant CPU usage policy

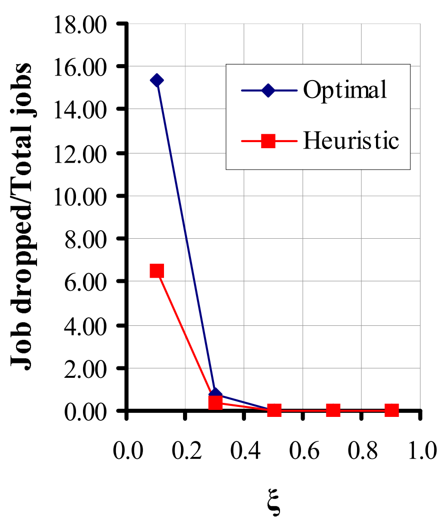
Simulation Details

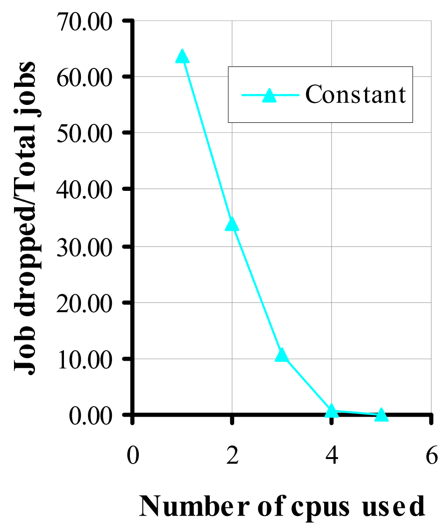
- In our simulation scenario we assumed the following:
  - ✓ Size of buffer : B=20
  - ✓ Set of thermal states : T={1,2,3,4,5}
  - ✓ # of CPUs : M=8
  - ✓ s(u) : s(u)=exp(-0.2u)
  - ✓ $C_b(b)$: Cb(b)=b
  - ✓ $C_{ut}(u,t)$: Cut(u,t)=u $\sqrt{t}$
  - ✓ $C_{uc}(u,c)$: Cuc(u,c)=$0.2[u-c]^{+}+0.1[u-c]^{-}$

Simulation Results

- Constant Rate arrivals

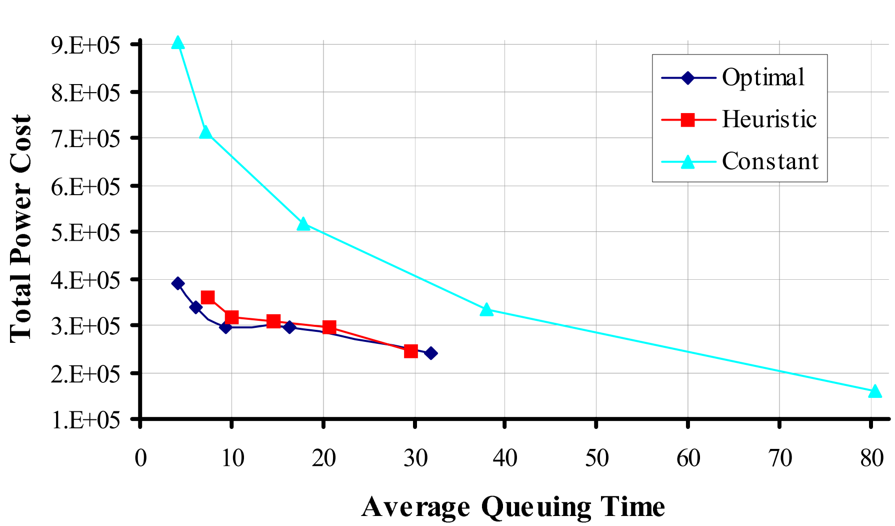


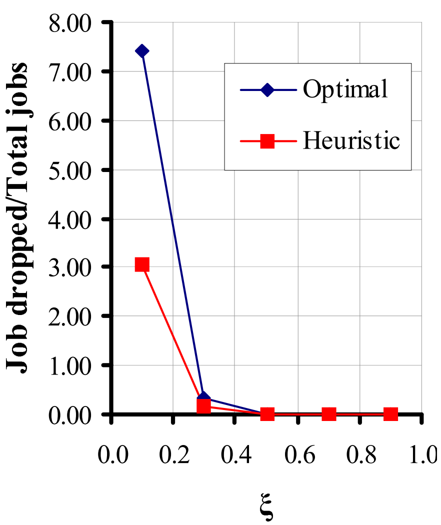


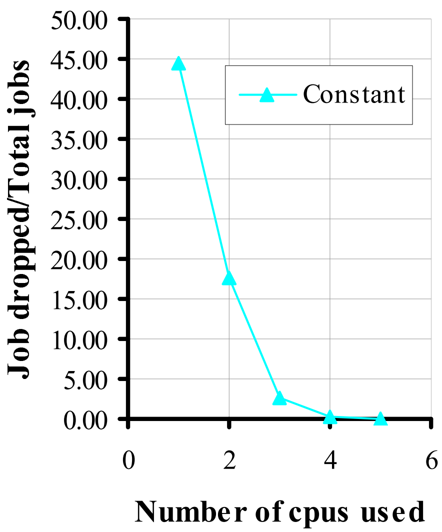


Simulation Results

- Bursty arrivals







Outline

- Motivation
- The System Model and the DP Formulation
- A Justified Heuristic
- Simulation and Evaluation
- Conclusions & Future Work
  - ✓ Conclusions
  - ✓ Future Work

Conclusions

- Presented a *novel power management model* for internet servers and data centers
- Formulated a *DP* that is agnostic to the arrival rate
- Created a *low complexity justified heuristic*
- *Simulated* for various arrival patterns
- *Substantial benefits* especially when the arrivals exhibited strong temporal variations

Future Work

- Future work will follow in two directions:
  - ✓Evaluation with actual system parameters
  - ✓Traffic estimation

Thank You!