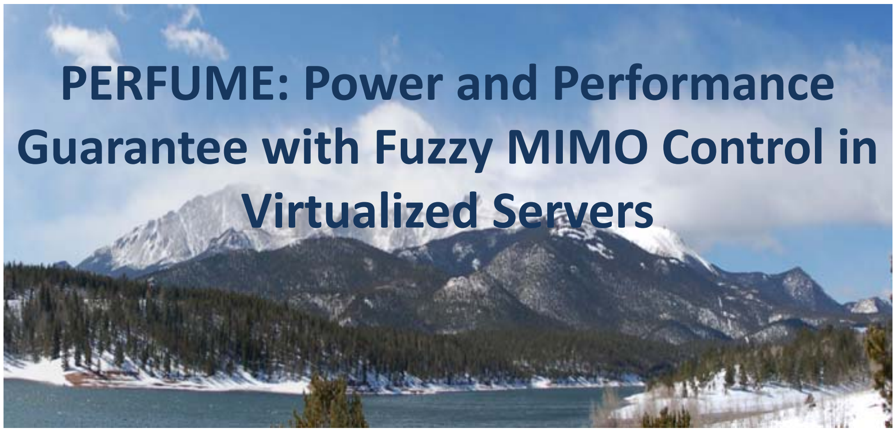
PERFUME: Power and Performance Guarantee with Fuzzy MIMO Control in Virtualized Servers



Palden Lama and Xiaobo Zhou
University of Colorado at Colorado Springs

Outline

- Background and Motivation
- Challenges
- Related Work
- Proposed Approach
- Performance Evaluation
- Conclusion
- Q & A

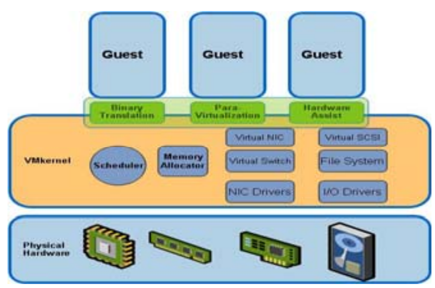
Data Center : Key Issues

- Multi-facet Challenges
  - Performance Assurance, Server Utilization, Power Consumption.
- Server utilization
  - built on the over-provisioning model.
  - dedicated servers for different applications.
  - most servers in a typical data center run at only 5-10 percent utilization
- Power consumption and Carbon footprint
  - According to the U.S. Department of Energy datacenters are the fastest-growing energy consumers in the United States today.
  - IEA (International Energy Agency) updated a warning in 5/2009 that Information and communication technology energy use could double by 2022, and triple by 2030
  - Data centers are responsible for the tens of millions of metric tons of carbon dioxide emissions annually more than 5% of the total global emissions.

3

Virtualized Data Centers

- Virtualization
  - abstracts physical resources into virtual machines (VMs).
  - diverse OS and applications share underlying server resources.
- Consolidation
  - improves server utilization
  - reduces power consumption
- Platform for Cloud Computing
  - Flexible and Fine-grained Resource Allocation
  - On-demand, pay-per-use service



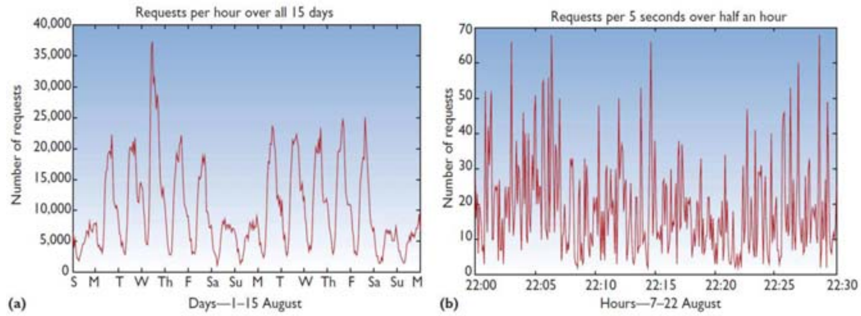
Power Management in Data Centers

- Power over-subscription
  - the sum of the possible peak power consumptions of all the servers combined is greater than the provisioned capacity
  - Power budgeting mechanism (DVS) on each server, to ensure total power stays below capacity.
- Hardware power budgeting
  - does not respect the isolation among virtual machines with different performance requirements.
- Need for a holistic view of power and performance management in data centers.

Joint Power and Performance Control

- Power oriented vs. performance oriented
  - Controlling either power or performance while achieving the other objective in best-effort manner.
  - No explicit co-ordination between power and performance.
- Effect of workload dynamics (highly dyanamic and bursty)
  - Control accuracy
  - System stability
- Percentile based performance metric
  - Most previous works focus on average performance guarantee. (not suitable for interactive applications)
  - A percentile response time introduces much stronger nonlinearity to the system, making it difficult to derive an accurate performance model.

Challenges : Workload Dynamics



• Workload Variation at multiple time scales demands self-adaptive and robust techniques for power and performance management.

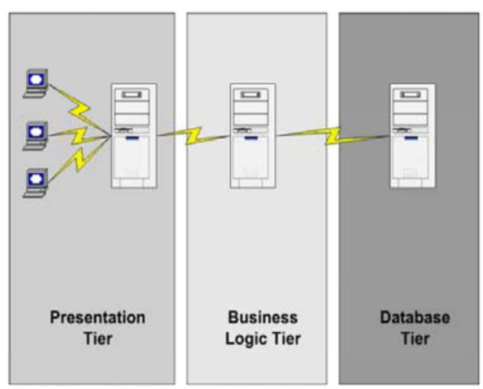
• System stability should be guaranteed to avoid oscillatory behavior in system states that result in poor power and performance assurance.

Challenges: Multi-tier architecture

- **Cross-tier dependencies**
- **Bottleneck switching**

- Performance is the result of a complex interaction of workloads in a very complex underlying computer system.

- Power usage of different tiers of one application may vary with workload.



RELATED WORK

- X. Wang, M. Chen, and X. Fu. **MIMO power control for high-density servers in an enclosure**. *IEEE Trans. On Parallel and Distributed Systems, 21(10):1412–1426, 2010.*
  - *MIMO control for cluster-level power control using DVFS.*
  - *Not applicable to virtualized servers*
  - *Power-oriented: no performance guarantee*
- R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu. **No "power" struggles: coordinated multi-level power management for the data center**. In ASPLOS'08. ACM, 2008
  - *Coordination of power controllers at various levels (Enclosure, Server & VM)*
  - *Power-oriented : no performance guarantee*

RELATED WORK

- R. Nathuji, C. Isci, and E. Gorbatov. **Exploiting platform heterogeneity for power efficient data centers**. In *Proc. IEEE Int'l Conf. on Autonomic Computing (ICAC), 2007.*
  - *Maps workloads to best suited platforms for power efficiency*
  - *Primary objective: meeting service level agreement of applications.*
  - *Lacks explicit control on power consumption.*
- Y. Wang, X. Wang, M. Chen, , and X. Zhu. **Partic: Power-aware response time control for virtualized web servers**. IEEE Trans. on Parallel and Distributed Systems, 21(4), 2010.
  - *Two-layer control architecture*
    - *primary control: VM resource allocation for balancing their relative perf. level.*
    - *secondary control: reducing power consumption by manipulating CPU frequency.*
  - *Power consumption is reduced in best-effort manner.*

RELATED WORK

- X. Wang and Y. Wang. **Co-con: Coordinated control of power and application performance for virtualized server clusters**. In *Proc. IEEE Int'l Workshop on Quality of Service (IWQoS), 2009.*
  - *Co-ordinated two-level controller for power and performance control*
  - *May not adapt to workload changes.*

- J. Gong and C.-Z. Xu. **vPnP: Automated coordination of power and performance in virtualized datacenters**. In *Proc. IEEE Int'l Workshop on Quality of Service (IWQoS), 2010.*
  - *Allows flexible tradeoff between power and performance objectives*
  - *Reduces performance relative deviation by 17% compared with two layer feedback controller (as in Co-Con).*
  - *lacks the guarantee on stability and performance of the server system especially in the face of highly dynamic and bursty workloads.*
  - *performance relative deviation may degrade in case of percentile-based performance metric.*

PERFUME System

- Flexible tradeoffs
  - It guarantees both power and performance targets with user specified tradeoffs.

- Well-suited to virtualized environments
  - It enforces power budgeting by controlling CPU usage limits of VMs instead of throttling CPU frequency of physical server.

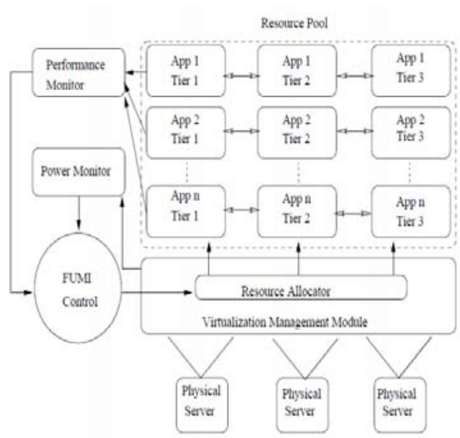
- Stability and control accuracy (Fuzzy MIMO Control)
  - FUMI applies Model Predictive control (MPC) technique to control CPU usage limits of various multi-tier applications hosted in virtualized servers.
  - To apply MPC technique, it generates fuzzy models that capture power and performance behavior of multi-tier applications hosted in virtualized servers.
  - It adapts the fuzzy models at run-time in response to changes in workload.

- It is able to control both average and percentile-based performance metric due to its Fuzzy modeling

PERFUME System Architecture

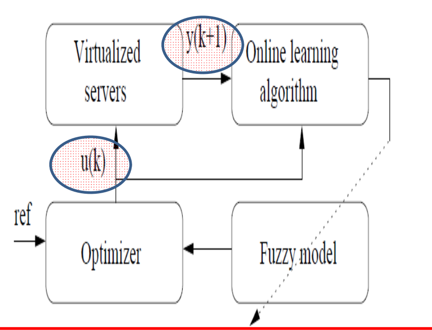
Testbed :

- HP ProLiant BL460C G6 blade server modules with a HP EVA storage area network.
- 10 Gbps Ethernet and 8 Gbps Fibre/iSCSI dual channels.
- Virtualized with Vmware ESX 4.1
- Hosting multi-tier application benchmark, RUBiS



FUMI Control Interface

- Fuzzy model:
  - represents an arbitrarily complex system by a combination of inter-linked subsystems with simple functional dependencies.
  - Accurately capture the non-linearity of computer systems: response time vs. resource allocation
- Optimizer:
  - Formulates MIMO control problem as a constrained optimization.



Minimize:

$$V(k) = \sum_{i=1}^{H_p} ||r1 - y_1(k+i)||_P^2 + \sum_{i=1}^{H_p} ||r2 - y_2(k+i)||_Q^2 + \sum_{j=0}^{H_c-1} ||\Delta u(k+j)||_R^2.$$

Constraint: $\sum_{j=1}^{M} (\Delta u_j(k) + u_j(k)) \leq U_{max}$

Fuzzy Modeling

- Each controlled variable (power, performance) is represented by a fuzzy model

$$y(k+1) = R(\xi(k), u(k)).$$

Regression vector (current & previous power/perf measurement).

Current resource allocations

- Fuzzy model (R) is composed of a set of fuzzy rules.

$$R_i\text{: If } \xi_1(k) \text{ is } \Omega_{i,1} \text{ and .. } \xi_\varrho(k) \text{ is } \Omega_{i,\varrho} \text{ and } u_1(k) \text{ is } \Omega_{i,\varrho+1} \text{ and .. } u_m(k) \text{ is } \Omega_{i,\varrho+m} \text{ then}$$

$$y_i(k+1) = \zeta_i \xi_i(k) + \eta_i u(k) + \phi_i.$$

- Model's final output is sum of output given by each rule, weighted by its activation strength.
- Initial fuzzy model obtained by subtractive clustering and ANFIS (Artificial Neural Network Fuzzy Inference System) technique.
- At run time, wRLS method updates the fuzzy model parameters.

Control Solution

- Express MIMO Control Objective as Quadratic Programming Problem

$$\min_{\Delta \mathbf{u}} \left\{ \frac{1}{2} \Delta \mathbf{u}^T \cdot \mathbf{H} \cdot \Delta \mathbf{u} + \mathbf{f}^T \cdot \Delta \mathbf{u} \right\}$$

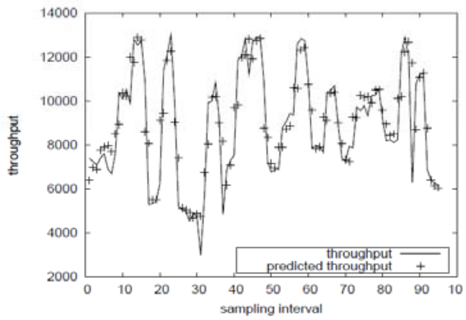
  - Linearize fuzzy model at each sampling interval, to extract state space model

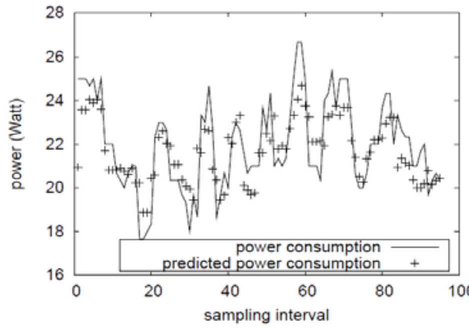
$$\bar{\mathbf{x}}_{\mathrm{lin}}(k+1) = \bar{\mathbf{A}}(k)\bar{\mathbf{x}}_{\mathrm{lin}}(k) + \bar{\mathbf{B}}(k)\Delta\mathbf{u}(k)$$
$$\mathbf{y}_{\mathrm{lin}}(k) = \bar{\mathbf{C}}(k)\bar{\mathbf{x}}_{\mathrm{lin}}(k).$$

  - The matrices *A(k),B(k) and C(k)* are constructed by freezing the parameters of the fuzzy model at a certain operating point y(k) and u(k).

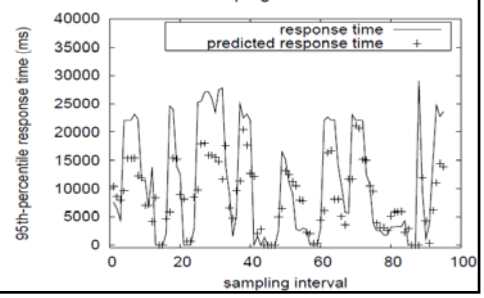
- Solve using any Quadratic solver software or MATLAB.

Prediction Accuracy



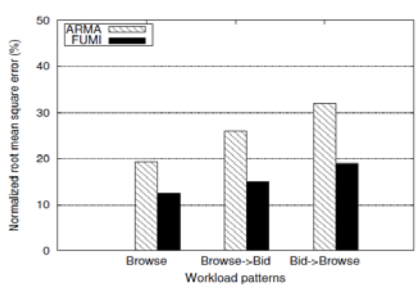


	through put	Power	95th percentile response time
NRMSE	12.5%	15.2%	17.6%

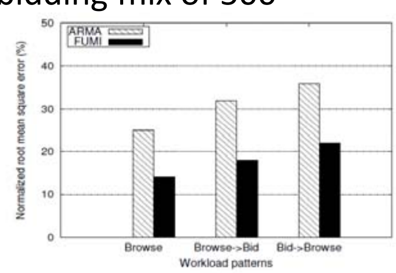


Self-Adaptiveness of Power/Perf Model

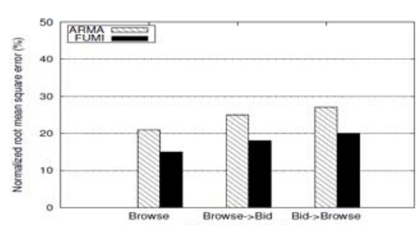
- Browsing mix of 1000 users to bidding mix of 500 concurrent users & vice versa
- Comparison with ARMA



(a) Throughput.



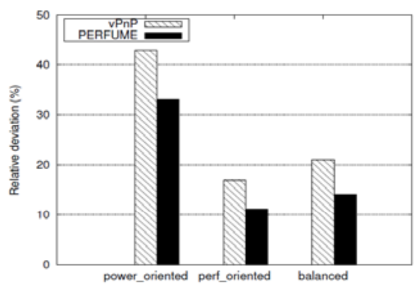
(b) The 95_{th}-percentile response time.



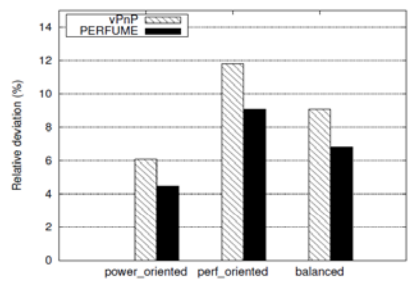
The power consumption.

Flexible Tradeoffs

- Control accuracy for various tradeoffs.

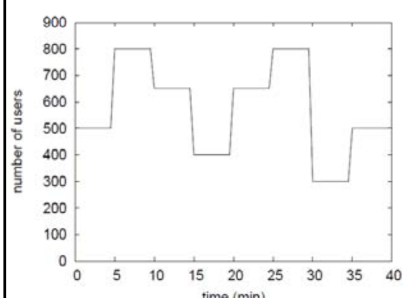


(b) Throughput.

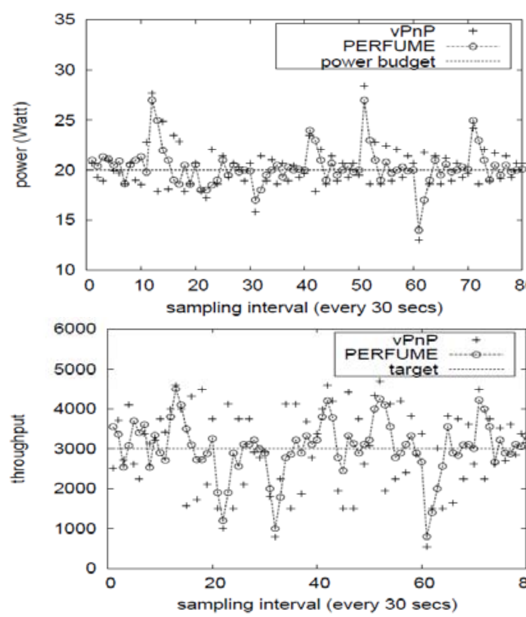


(c) Power consumption.

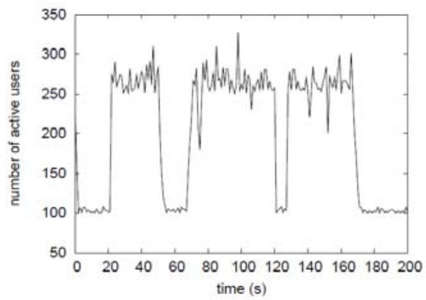
System Stability



(a) A highly dynamic workload.

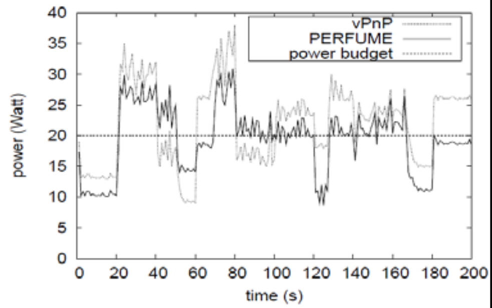


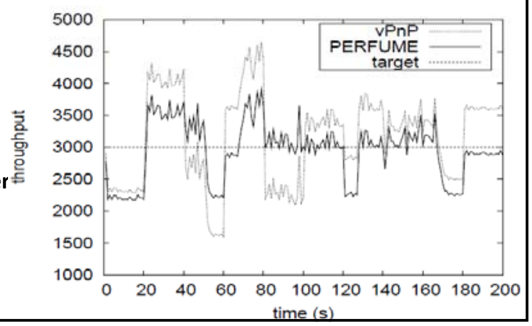
System Stability



(a) Bursty workload.

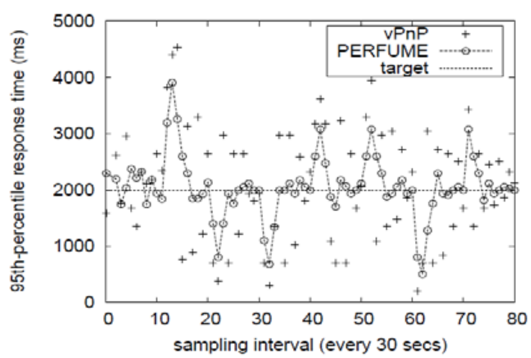
N. Mi, G. Casale, L. Cherkasova, and E. Smirni. **Injecting realistic burstiness to a traditional client-server benchmark**. In *Proc. IEEE Int' Conference on Autonomic Computing (ICAC), 2009.*





Percentile-based Response Time Guarantee

- Improvement of 40% in terms of relative deviation



Conclusion

- PERFUME provides holistic and self-adaptive performance and power control in a virtualized server cluster.
- Testbed implementation demonstrates
  - precise control of power consumption of virtualized blade servers
  - effective control of throughput and percentile-based response time of multi-tier applications.
  - flexible tradeoffs
  - control accuracy and system stability

ANY QUESTIONS?