Multi-agent Approaches to Data Center Energy Management: A Research Agenda

Jeff Kephart (kephart@us.ibm.com)
IBM Thomas J Watson Research Center
Hawthorne, NY
Multi-agent systems and autonomic data centers

- I envision data centers of the future as a complex ecosystem of interacting semi-autonomous entities – an autonomic, multi-agent system

- Autonomic computing definition

- Software agent definition
  - “An encapsulated computer system, situated in some environment, and capable of flexible, autonomous action in that environment in order to meet its design objectives.” Jennings, Sycara and Wooldridge, A Road Map of Agent Research and Development, JAAMAS 1998
  - Multi-agent systems: collections of agents that interact with one another to achieve individual and/or system goals

- Agents will
  - represent, or be embedded in, different products from different vendors
  - reside at many levels of the management stack
  - manipulate control knobs at all levels of the stack (from hardware/firmware up through middleware and facilities)
  - collectively manage the data center to specified objectives and constraints (some relating to power)
  - interact in intended and unintended ways with one another, and with other types of automated management processes directed towards maintaining high levels of performance, availability, reliability, security, etc.

- This vision is a natural extrapolation of present-day facts and trends
  - Industry and academia are developing a multitude of control knobs and automated techniques to save energy
  - These will be incorporated into a multitude of management products from different vendors
  - They will operate simultaneously within and across multiple levels of the stack
  - Somehow, these products will need to work together effectively, requiring some cooperative interactions

- Data centers are a “killer app” for multi-agent systems
  - Conversely, MAS architectural and algorithmic concepts are essential to energy-efficient, autonomic data centers
Example: Interaction between power and performance agents

- How might semi-autonomous power and performance agents interact?
  - Mediated through coordinator agent, or
  - Direct bi/multi-lateral interactions

Scenario (with mediation)
- Performance manager observes subset $s_{\text{perf}}$ of system state, and controls application placement and load balancing weights
- Power manager observes subset of $s_{\text{pwr}}$ of system state, and controls on/off state of servers
- Coordinator understands overall power-performance tradeoffs as expressed in a joint utility function, and queries performance and power agents for likely impact when $n$ servers are turned on, finding optimal number $n^*$

Another example: OS and Middleware agents can interact to coordinate appropriate control actions to satisfy power-performance tradeoffs

Kephart, Chan, Das, Levine, Tesauro, Rawson, Lefurgy. Coordinating Multiple Autonomic Managers to Achieve Specified Power-Performance Tradeoffs. ICAC 2007. (Emergent phenomena can occur when autonomic managers don’t communicate effectively.)
Research Challenges

- **Marketplace realities** dictate de-centralized MAS solutions to energy management
  - Interaction among agents responsible for different dimensions of management
  - Interaction across layers of the stack

- **Architectural questions**
  - What is a best (minimal) set of interfaces among agents?
  - Can a multi-agent approach work, using negotiating agents and mediators to manage performance, power, availability, reliability …?
  - Are markets and auctions effective coordination mechanisms when there are numerous agents and “goods”?
    - What are the goods in this case (e.g. one core in a multi-core system running in turbo mode)?
    - We may need hierarchical markets that extend across multiple data centers
    - What happens when data center markets are coupled to the global economy?

- **Algorithmic (and other) challenges**
  - Building/tuning deterministic and statistical what-if models on the fly
  - Avoiding undesirable emergent phenomena (IBM and HP Research have observed this!)
  - Eliciting preferences (tradeoffs between power, performance, …)

- **Beyond IT**
  - There is much to be gained by coordinating workload placement, load balancing, etc. with facilities management, e.g. co-managing cooling and workload migration
  - Agents will represent PDU’s, CRAC’s, chillers, etc., vastly increasing the size and variety of the MAS