Energy-Efficient Video-Sharing Servers
Hang Yuan, C.-C. Jay Kuo

Motivation
- Data centers consume 30 billion watts of electricity
  - 90% of the power is consumed by idle devices
- Ubiquity of large-scale video-sharing services
  - 72 hours of videos are being uploaded to YouTube every minute
  - Internet video will account for 57% of the Internet traffic in 2014
- Storage systems with a large number of disks
  - Consume 25%–35% of the total energy
  - Video-sharing services rely on such storage systems

Background
- Disk Energy Management
  - Use multi-speed disks
- Video-Sharing Services
  - Huge number of videos and users
  - Diverse video repository

Modeling
- Disk Energy Management
  - Use multi-speed disks

Approach
\[ T_s = \min(T_r, t_D); \quad t_D \in (k_{n-1}, k_n) \]
\[ T_r; \text{ sleep time;} \]
\[ T_s; \text{ time until the first unknown deadline;} \]
\[ p(t); \text{ pdf of } T_s \]
\[ p(t) = \begin{cases} 
\lambda_0 e^{-\lambda_0 t} & \text{if } t \leq \min(k_0, t_D) \\
\alpha_1 \lambda_1 e^{-\lambda_1 t} & \text{if } t \in (k_{i-1}, \min(k_i, t_D)] \\
\alpha_i \delta(t - t_D) e^{-\alpha_i t} & \text{if } t \geq t_D 
\end{cases} \]

Energy Cost:
\[ \frac{P_1T_s + O_1}{T_s + R_I} \]

Delay Cost:
\[ D_I = \frac{\lambda_i R_I^2}{2} + \frac{(\lambda_i R_I + \eta_i)\lambda_i R_I T}{2n_I} \]

We want to minimize energy with a constraint on service delay
\[ \min(E), \text{ subject to } D < D_C \]

It can be solved using Lagrangian relaxation
\[ \min(C), \text{ where } C = E + \mu D \]
Each \( \mu \) will solve the optimization for some particular \( D_C \)

Optimal mode:
\[ M = \arg\min_{t} C_t \]

Results

Conclusions
- Conducted the first study of the energy issues for large-scale VSS
- Jointly optimized energy and service delay
- Proposed a model for disk idle time in VSS
- Proposed a prediction-based approach to make optimized power mode decisions
- Future Work
  - Energy-aware caching
  - Energy-aware placement

Contact: hangyuan@usc.edu