Capacity Provisioning Problems in Geo-distributed Data Centers

Bhuvan Urgaonkar
Dept. of CSE
The Pennsylvania State University
Geo-distributed Data Centers

- Reasons for geo-distribution:
  - Latency
  - Availability
- What are the **cost** implications?
What's New?

• What is well-understood:
  - How to build single data centers cost-effectively?
  - How to create distributed applications using an existing pool of data centers (that were built separately)?

• What is (likely) less well-explored:
  - Building a fleet of distributed data centers from scratch for supporting large-scale distributed workloads

How do costs change when we build a geo-distributed version of a centralized DC?

• Approach: specific case studies -> general insights & challenges
A Simple Thought Experiment

- **Base IT cost**: Costs remain constant across different degrees of distribution.
- **Base + Spare IT cost**: Costs increase with higher degrees of distribution.
- **Costs of networking DCs**: Costs rise significantly with each step of distribution.

Graph showing IT cost vs. degree of distribution.
Costs: What have we made worse?

• Networking infrastructure to connect DCs
• Larger overall IT capacity
  - Redundancy for availability
    ▪ Higher for heterogeneous collection of DCs
  - Poorer statistical multiplexing
Costs: What have we made worse?

- Networking infrastructure to connect DCs
- Larger overall IT capacity
  - Redundancy for availability
    - Higher for heterogeneous collection of DCs
  - Poorer statistical multiplexing

How do we keep this “small”?
Costs: What have we made worse?

• Networking infrastructure to connect DCs
• Larger overall IT capacity
  - Redundancy for availability
    ▪ Higher for heterogeneous collection of DCs
  - Poorer statistical multiplexing
• Non-IT infrastructure (power+cooling) costs
  - To support higher IT capacity
Costs: What have we made worse?

• Networking infrastructure to connect DCs
• Larger overall IT capacity
  - Redundancy for availability
    ▪ Higher for heterogeneous collection of DCs
  - Poorer statistical multiplexing
• Non-IT infrastructure (power+cooling) costs
  - To support higher IT capacity

Can we keep non-IT Infra. “size” small?
Costs: What has improved?

- Revenue due to better latency improvements
Costs: What has improved?

- Revenue due to better latency improvements
- Aspects of availability
Costs: What has improved?

• Revenue due to better latency improvements
• Aspects of availability

Can we lower the availability of individual DC infra.?
An example of cost-effective IT provisioning
• Keeping non-IT infrastructure costs low
  • Lowering peak power related costs using batteries
• Conclusions
Problem Setting

• DC locations given
• Client demands known, time-varying
• Goal: determine total capacity at each DC
  - To meet latency constraints, and
  - To allow for one DC to fail
• Our optimizer: An LP
  - Generally, NP-hard facility location problems
Results

• DC locations
  - 6 MS data centers in the US

• Client demand model
  - Exhibits time zone specific variation
  - Proportional to population
Experiments using demand measured for one Microsoft cluster, and 6 MS DC locations within US. $L' = L$

TOTAL CAPACITY

- Optimized (no failure)
- Without time-of-day
- Optimized (support 1 failure)
- Nearest DC (no failure)
- Single DC capacity

Availability (against 1 failure) for free!

7/15/2014
Results

Experiments using demand measured for one Microsoft cluster, and 6 MS DC locations within US. $L=L'$

TOTAL CAPACITY

- Optimized (no failure)
- Without time-of-day
- Optimized (support 1 failure)
- Nearest DC (no failure)
- Single DC capacity

Excess capacity for high availability

An example of cost-effective IT provisioning

- Keeping non-IT infrastructure costs low
  - Lowering peak power related costs using batteries
- Conclusions
A Closer Look at Power Infrastructure

- Utility Substation
- Diesel Generator: $1/W
- UPS Battery: $0.6/W
- Power Distribution Unit (PDU): $0.3/W
- Server Racks: $0.2/W

Auto Transfer Switch (ATS)
Using Energy Storage

How to provision and harness ESDs in data centers?

Energy Storage Device (ESD) (No Performance Impact)
Why restrict ESDs to any one level of the datacenter power hierarchy (e.g., central or server)?
Specific Energy (Wh/kg) vs. Specific Power (W/kg) Ragone Plot

Compressed Air (CAES)
Flywheels (FW)
Lithium Ion
Fuel Cells
Batteries
Lead Acid
Ultracapacitors (UC)
Capacitors
Capital Cost (Energy and Power)

<table>
<thead>
<tr>
<th>Technology</th>
<th>Ultracapacitor</th>
<th>Flywheel</th>
<th>Lithium ion battery</th>
<th>Lead-acid battery</th>
<th>Compressed air</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost ($/kWh)</td>
<td>10,000</td>
<td>5,000</td>
<td>525</td>
<td>200</td>
<td>50</td>
</tr>
<tr>
<td>Power Cost ($/kW)</td>
<td>100</td>
<td>250</td>
<td>175</td>
<td>125</td>
<td>600</td>
</tr>
</tbody>
</table>

Why restrict to single ESD technology (e.g., Lead acid battery)?
Multi-level Multi-technology ESDs

- Diesel Generator
- Utility
- Battery
- Flywheel
- Compressed Air
- ESD
- PDU
- ESD
- Server H/W
- Rack
- Rack
- Rack

7/15/2014

Microsoft Faculty Summit 2014
Cost Savings for Google Workloads

(Savings, ESD cost) Total cost without ESD is $12k/day

✓ An example of cost-effective IT provisioning
✓ Keeping non-IT infrastructure costs low
  ✓ Lowering peak power related costs using batteries

• Conclusions
Related Work

• **IT capacity provisioning**
  - Capacity planning [Goiri et al. ICDCS’11]
    ▪ Showed that more DCs, where each is lower availability (lower cost) but extra geo-spares, better
    ▪ Computed optimal capacity placements

• **Lowering infrastructure availability/cost**
  - Reducing the “size” of power infrastructure
    ▪ Under-provisioning backup generators [Wang14]
    ▪ Reducing component redundancy [Govindan11,Kansal13]
  - Less aggressive cooling design
    ▪ Has similarity in offering an availability vs cost trade-off [Schroeder@Sigmetrics12]
    ▪ Related work in geo-distributed setting: [Wierman]
  - Lower availability IT
Conclusions

• Cost-effective capacity provisioning of geo-distributed data centers presents opportunities for novel problems in optimization and system design
  - Putting together lower availability data centers with appropriate fault tolerance mechanisms during subsequent operation
  - Key source of difficulty is uncertainty of subsequent workload evolution
    ▪ Typical facility location based formulations might be inadequate
    ▪ Stochastic optimization? Robust optimization?

• More information: http://www.cse.psu.edu/~bhuvan

• Joint work with: Anand Sivasubramaniam, Aman Kansal, Di Wang, Sriram Govindan, Hosam Fathy, Iyswarya Narayanan
Save the planet and return your name badge before you leave (on Tuesday)