System Architecture Support for Green Enterprise Computing

Maria Kazandjieva
Chinmayee Shah, Ewen Cheslack-Postava,
Behram Mistree, Philip Levis

Stanford University
Computing systems account for an estimated **13%** of the electricity use of office buildings. [DoE]

This amounts to about **2% of the total electricity consumption** in the US.
Office PCs spend the majority of their time at very low CPU utilization.
Office PCs spend the majority of their time at very low CPU utilization.

<table>
<thead>
<tr>
<th>Machine</th>
<th>5</th>
<th>50</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dell Optiplex 745</td>
<td>1%</td>
<td>9%</td>
<td>58%</td>
</tr>
<tr>
<td>High-end custom-built</td>
<td>0%</td>
<td>1%</td>
<td>57%</td>
</tr>
<tr>
<td>Dell Precision T3400</td>
<td>0%</td>
<td>4%</td>
<td>29%</td>
</tr>
<tr>
<td>HP Pavillion Elite m9250f</td>
<td>0%</td>
<td>0%</td>
<td>25%</td>
</tr>
<tr>
<td>Dell Precision T3400</td>
<td>0%</td>
<td>1%</td>
<td>13%</td>
</tr>
<tr>
<td>Dell Inspiron 530</td>
<td>1%</td>
<td>1%</td>
<td>8%</td>
</tr>
<tr>
<td>Dell Precision T4300</td>
<td>0%</td>
<td>1%</td>
<td>7%</td>
</tr>
</tbody>
</table>
two-thirds of office PCs have CPU<10% 75% of the time
1. Why is low desktop utilization a problem?
2. What about other, greener hardware?
3. A hybrid solution
4. Takeways
A graph showing the ideal relationship between power and utilization.
The graph shows the power consumption in watts for different levels of utilization, ranging from 0% to 100%. For the desktop, the power consumption increases linearly from 0 W at 0% utilization to 165 W at 100% utilization. For the ideal scenario, the power consumption is 110 W at 100% utilization, assuming a linear relationship.
Power Utilization

- **Ideal**
  - 0% utilization: 0 W
  - 100% utilization: 165 W

- **Desktop**
  - 0% utilization: 110 W
Hardware is not power-proportional so low utilization means a lot of waste.
Hardware: Thin Clients

No local compute resources
Displays the GUI of a remote machine.

15–20 watts for client itself
10–15 watts server share
Hardware: Laptops

<table>
<thead>
<tr>
<th>Power (watts)</th>
<th>Laptops</th>
<th>Desktops</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>100</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>150</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>200</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- Median: 17 W (Laptops), 100 W (Desktops)
- Mean: 16 W (Laptops), 109 W (Desktops)
- STD: 3 W (Laptops), 47 W (Desktops)

83 W
Laptops

performance \ll\ desktops

Thin Clients

not suitable for all workloads
Laptop power $\approx$ Thin Client power
1. Why is low desktop utilization a problem?
2. What about other, greener hardware?
3. A hybrid solution
4. Takeways
A hybrid compute architecture can save as much energy as a thin client without sacrificing performance.
Anyware combines low-power clients with a high-end shared server.
Anyware

low-power client

high-end shared server
1. Double-click to watch a video
1. Double-click to watch a **video**

2. Decide to use local resources to play the video
1. Double-click to watch a video

2. Decide to use local resources to play the video
1. Double-click to edit an image
1. Double-click to edit an image

2. Decide to offload the task
1. Double-click to edit an **image**

2. Decide to offload the task
How do we build Anyware so that

it is invisible to the user
and
it does not require application or OS changes
and
it is practical and easy to setup?
create a bare-bones VM that matches the client OS and architecture
create a bare-bones VM that matches the client OS and architecture

connect the VM and client via SSH and export the VM windowing system
create a bare-bones VM that matches the client OS and architecture

connect the VM and client via SSH and export the VM windowing system

identify files and folders that the client will export to the VM via a networked FS
Offload complete program execution.

Do so in user space by intercepting MIME type association.
But wait,

Who will decide where task are executed?

Will remote execution impact user experience?
User Study

Tasks

Application Features

- instructions executed
- IPC
- cache misses
- X drawing calls
- local: data read
- remote: network data in/out
Qualitative Results

From a user perspective, the majority of applications perform similarly, regardless of whether they run on a laptop or on a remote VM.

Tasks that are data—or graphics—heavy, have visibly worse performance when executed remotely.
A logistic regression model suggests a small set of workload features are indicative of where a task should be executed.

- number of instructions
- Remote Mb sent from client to VM
- Mb sent from VM to client
- Local number of subprocesses
Hardware Setup

- 4-core, 2.4 GHz
- 4 GB RAM
- 256 GB SSD

- 2-core, 1.6 GHz
- 4 GB RAM
- 256 GB SSD

- 12-core, 3.0 GHz
- 48 GB RAM
- 7200 RPM HDD

- Intel Xeon Server
- four cores
- 4 GB RAM
Image-processing Task

- **Open & Load**: 48.92s
  - **Laptop only**: 4.24s
  - **Desktop only**: 3.12s
  - **Anywhere (remote)**: 5.73s

- **Process Image**: 39.93s
  - **Laptop only**: 3.96s
  - **Desktop only**: 3.64s
  - **Anywhere (remote)**: 4.10s

- **Save & Exit**: 36.73s
  - **Laptop only**: 3.17s
  - **Desktop only**: 4.10s
  - **Anywhere (remote)**: 5.73s
A Video Workload

<table>
<thead>
<tr>
<th></th>
<th>desktop</th>
<th>Anyware</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>local</td>
</tr>
<tr>
<td>frames not</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>displayed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A Video Workload

<table>
<thead>
<tr>
<th>frames not displayed</th>
<th>desktop</th>
<th>Anyware local</th>
<th>Anyware remote</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>32%</td>
</tr>
</tbody>
</table>

Anyware will choose to run this locally
Energy Evaluation

110 - 165 W

14 - 24 W
(no screen)

130 - 270 W

5 - 11 W
(assuming 25 VMs)
Energy Evaluation

reduction of 68% – 77%

100% active

Desktop

165 W

110 W

35 W

19 W
Anyware

A practical system that uses established techniques to provide performance comparable to that of desktops while reducing energy costs by ~70%
Final Thoughts

The computing design space is large and the trade-off between power and performance is not linear.
Final Thoughts

Time to rethink the needs of enterprise computers:

local: graphics, I/O, network, memory
remote: cpu, memory