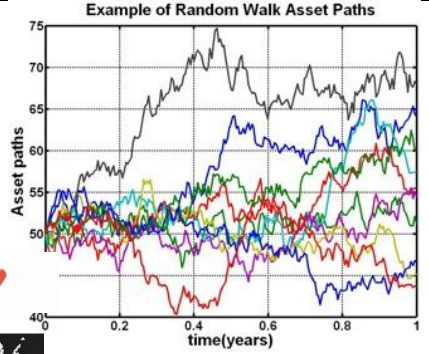
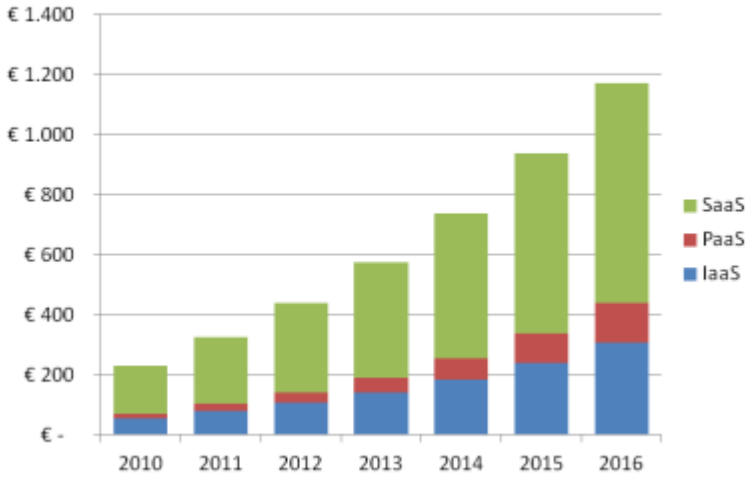
Statistical Characterization of Business-Critical Workloads Hosted in Cloud Datacenters

Vincent van Beek
Siqi Shen
Alexandru Iosup

TU Delft

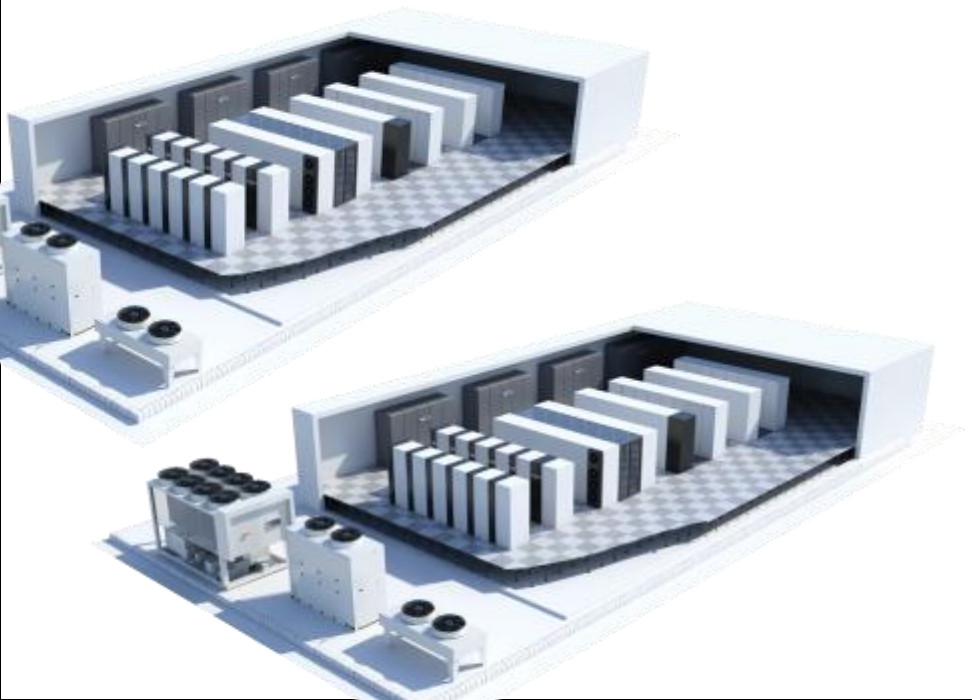


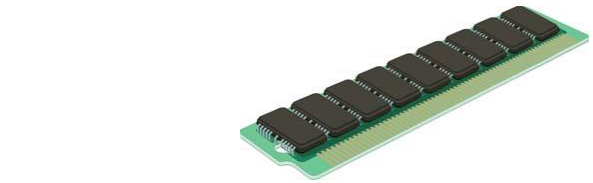
Monte Carlo simulation

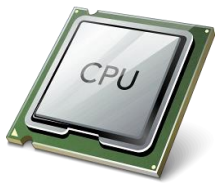


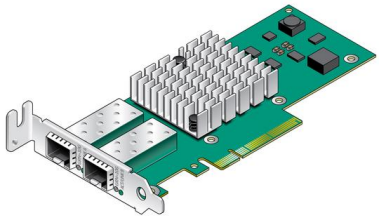
Source: http://www.themetisfiles.com

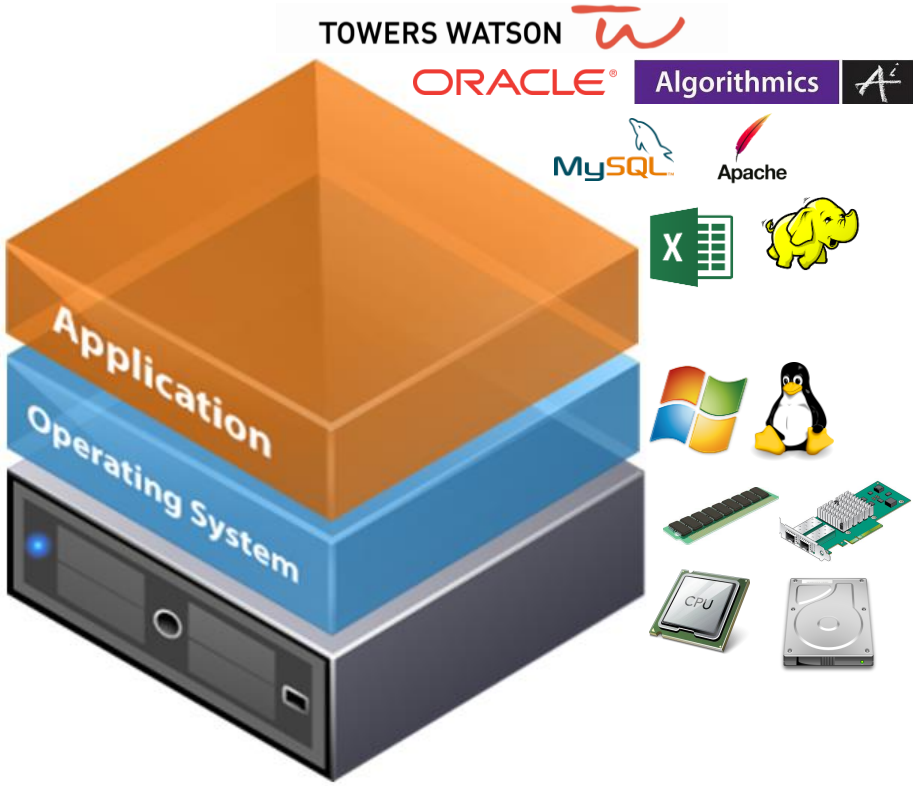
Business Critical Workloads



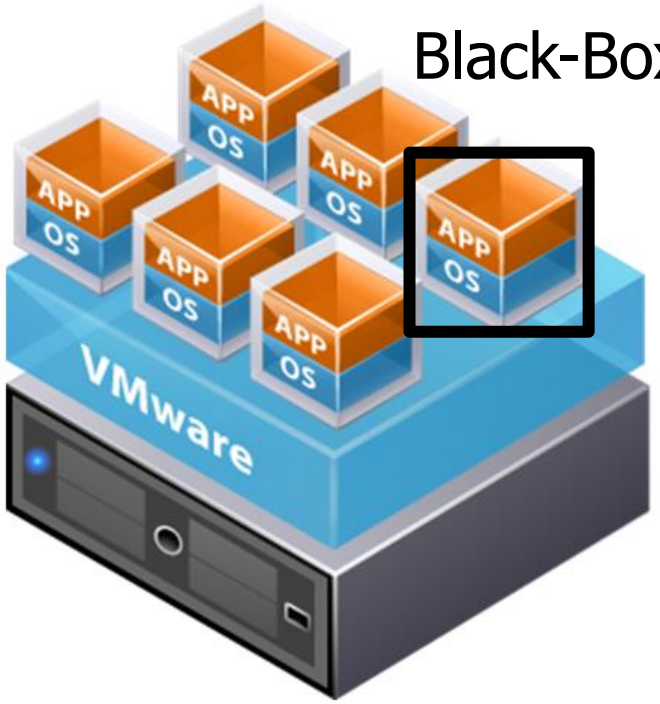








Traditional Architecture



Virtual Architecture

Requested resources

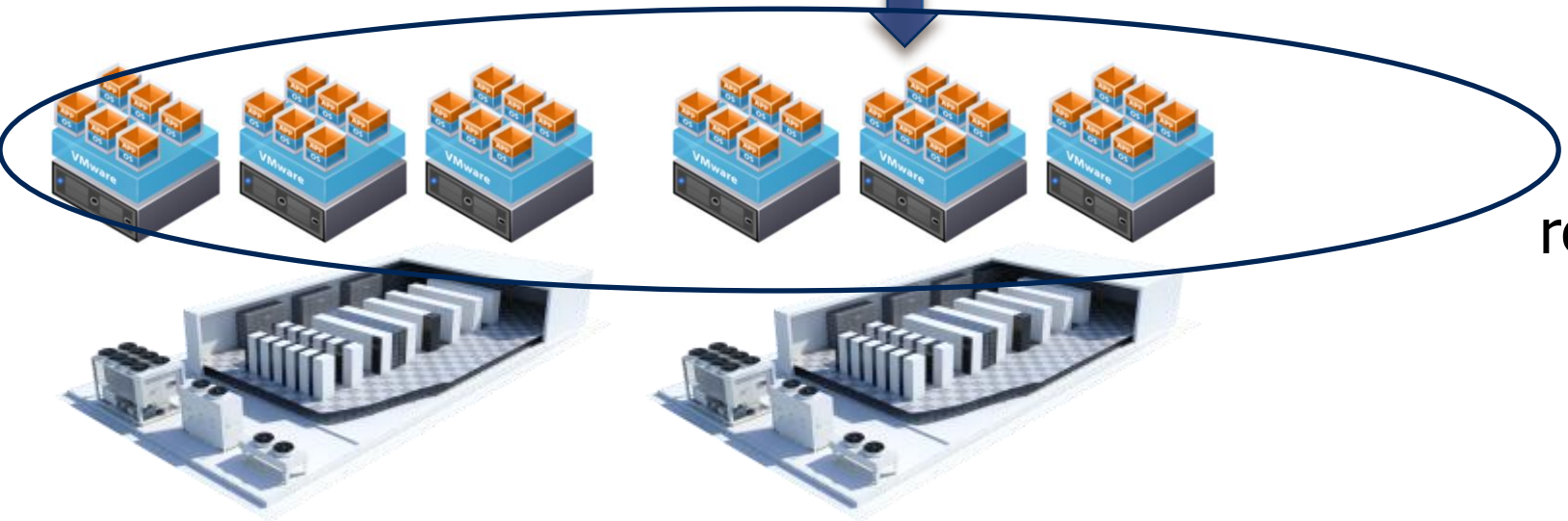
Per VM

- CPU: cores / speed (GHz)
- Memory: GB (100% reserved)
- Network: per NIC 10gbps
- Storage: GB / fast, slow tiers

Datacenter Manager

Ask about my other research

* Article in IEEE Computer this summer



Used resources

* V.S. Van Beek, J. Donkervliet, S. Hugtenburg, T. Hegeman, and A. Iosup: *Mnemos: Self-Expressive Management of Business-Critical Workloads in Virtualized Datacenters*, IEEE Computer, Special Issue on Self-Aware and Self-Expressive Computing Systems, 2015 (to appear)

Agenda

1. Introduction
2. **Collected Two Workload Traces**
3. Workload Characterization
    1. Requested resources per VM
    2. Used resources per VM
    3. Time patterns in resource usage
4. Conclusion

Bitbrains: A useful source of datacenter traces

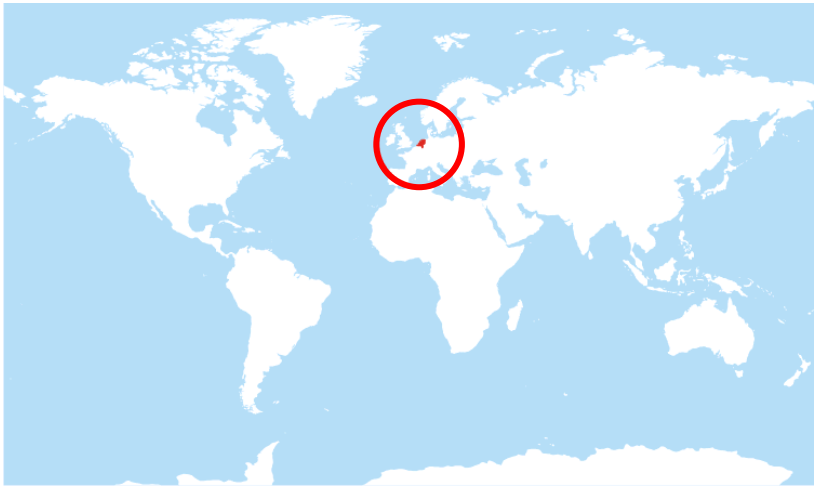
Leading HPC Cloud provider in the Netherlands

Customers

- ✓ ING
- ✓ NN
- ✓ Aegon
- ✓ ICS Cards
- ✓ Ahold
- ✓ …

2013

Gartner
Cool Vendor
Bitbrains



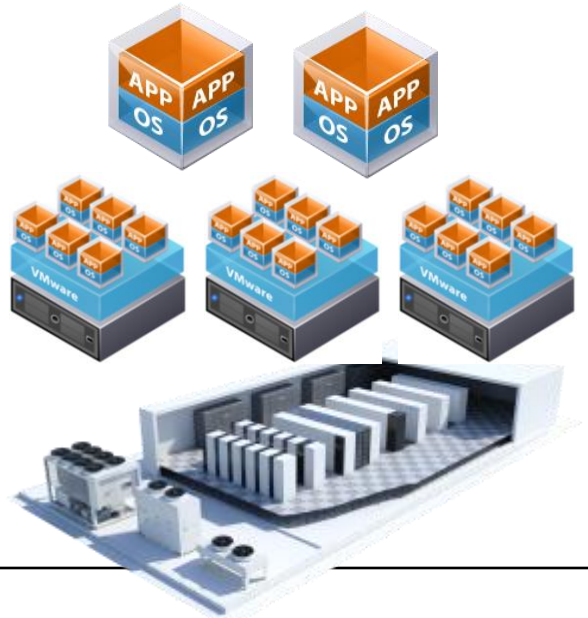
Bitbrains

designs, builds and supports leading edge Cloud Computing solutions and is specialized in High Performance Computing.

Collected Two Unique Workload Traces of Long-Running Black-Box VMs

Name of the trace	# VMs	Period of data collection	Storage technology	Total memory	Total cores
fastStorage	1,250	1 month	SAN	17,729 GB	4,057
Rnd	500	3 months	NAS and SAN	5,485 GB	1,444
Total	1,750	5,446,811 CPU hours		23,214 GB	5,501

- All resources:
  - CPU, Memory, Storage, and Network
- Large scale
- Long term



Collected Two Unique Workload Traces For Business-Critical workloads

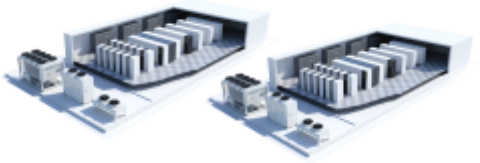
Prior work, per job:

- Google
- Facebook
- Taobao
- Scientific workloads
- Grids vs Google

Our study of, per VM:

- Requested and

- Used resources

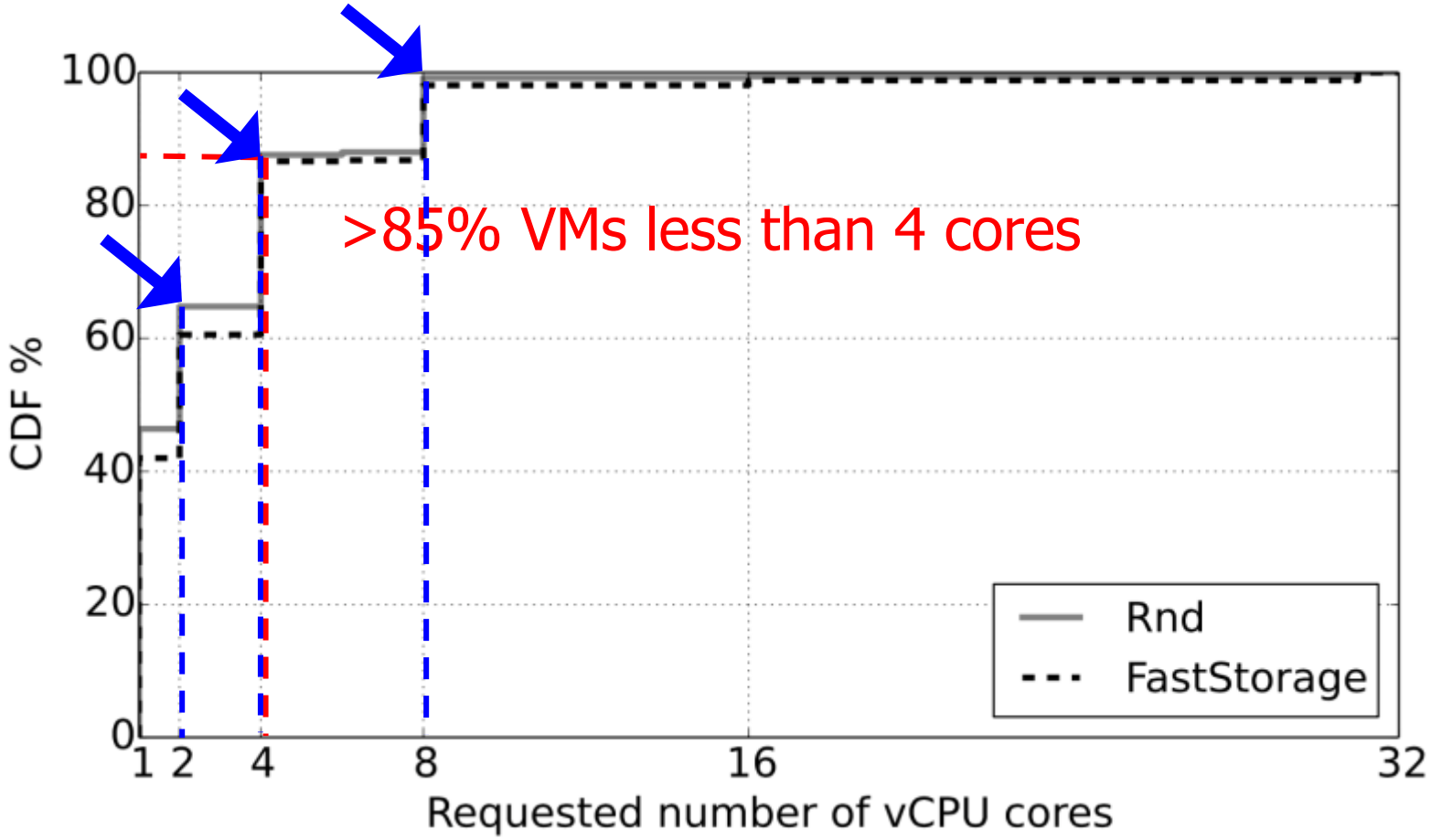


- For all resources
- Per VM

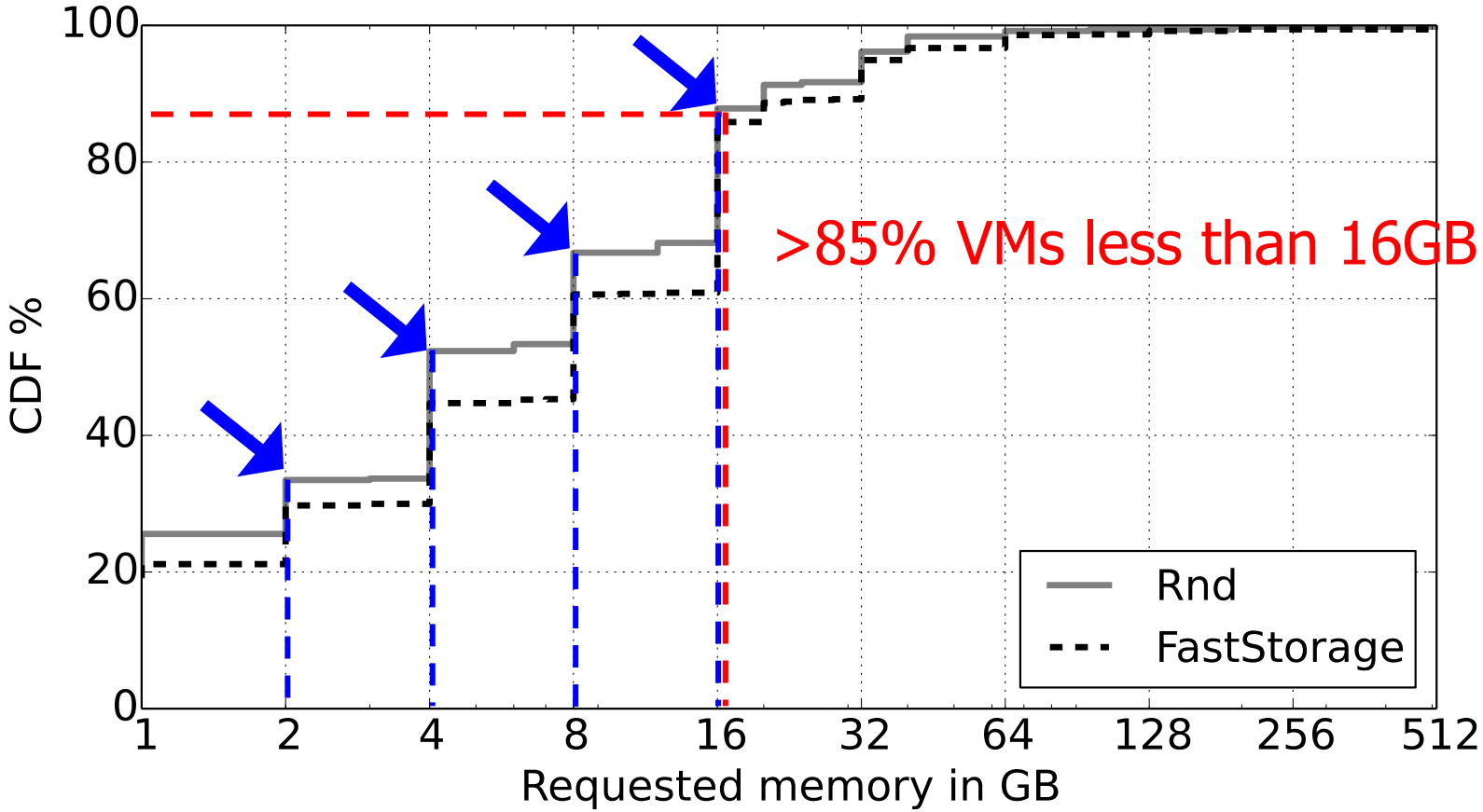
Agenda

1. Introduction
2. Collected Two Workload Traces
3. **Workload Characterization**
   1. Requested resources per VM
   2. Used resources per VM
   3. Time patterns in resource usage
4. Conclusion

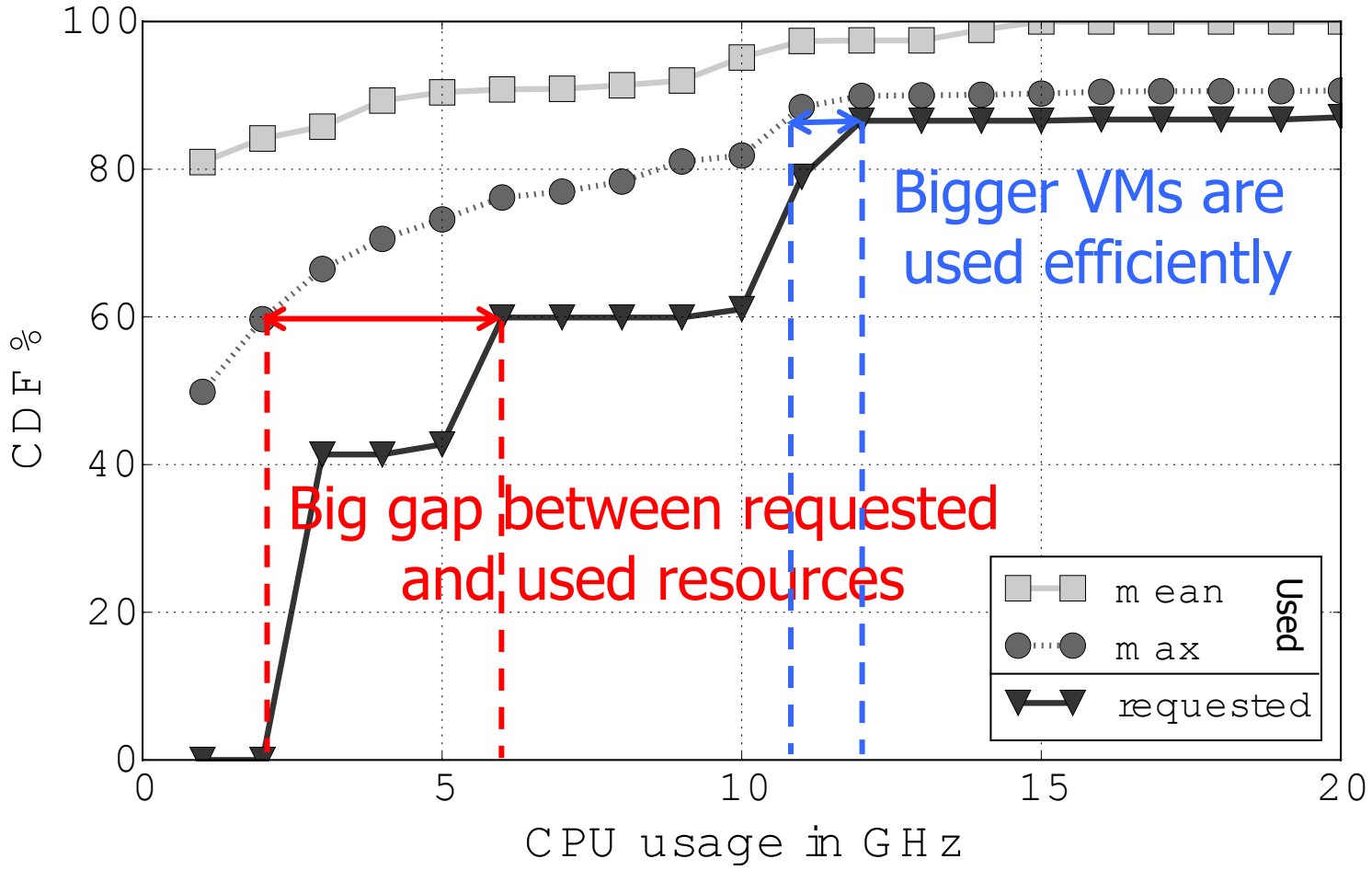
Requested Number of CPU Cores is Low, and Power of Two Scaling



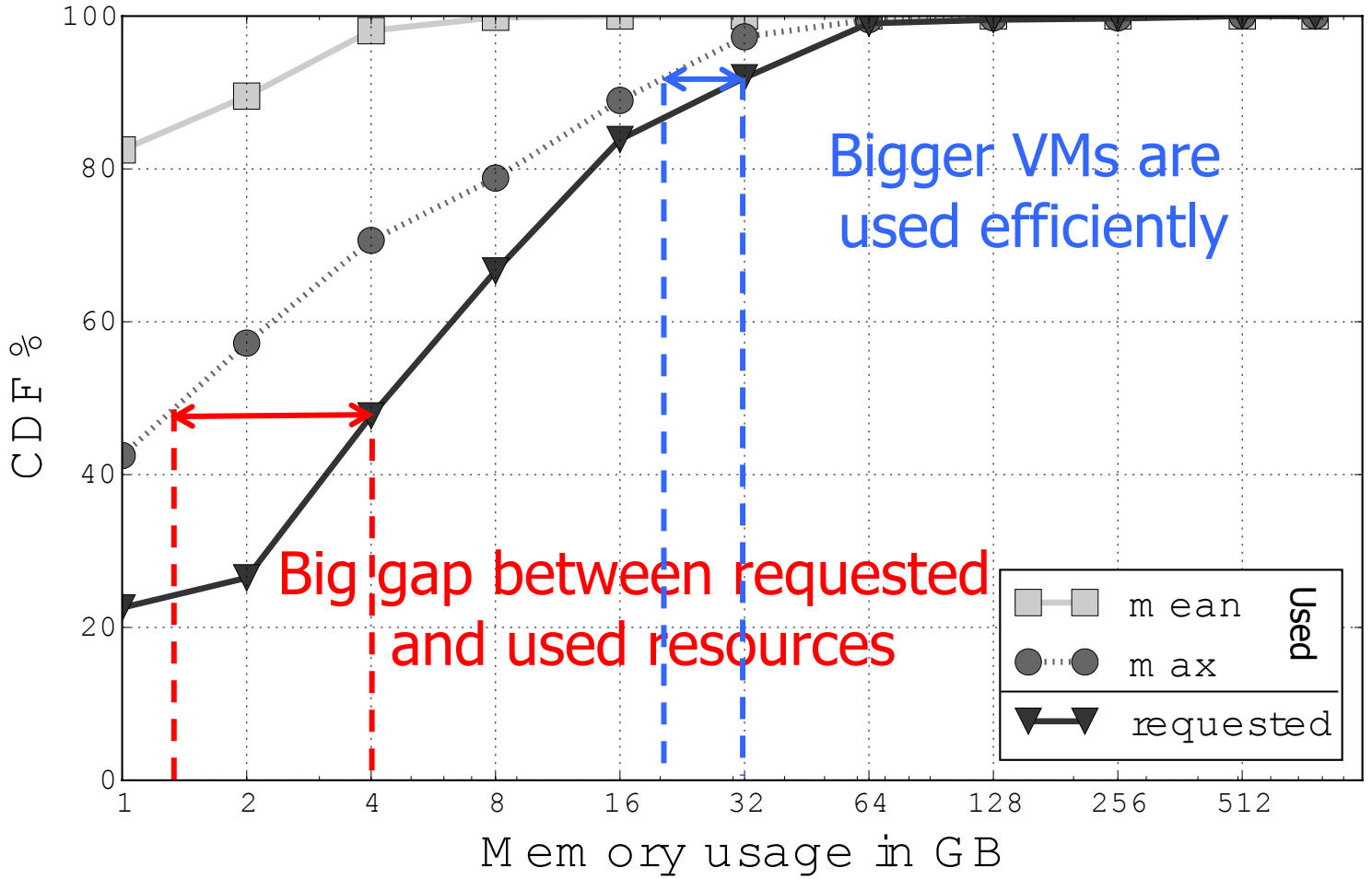
Requested Memory is Low, and Power of Two Scaling



Requested vs Used CPU Resources Leave Big Gap for Optimization

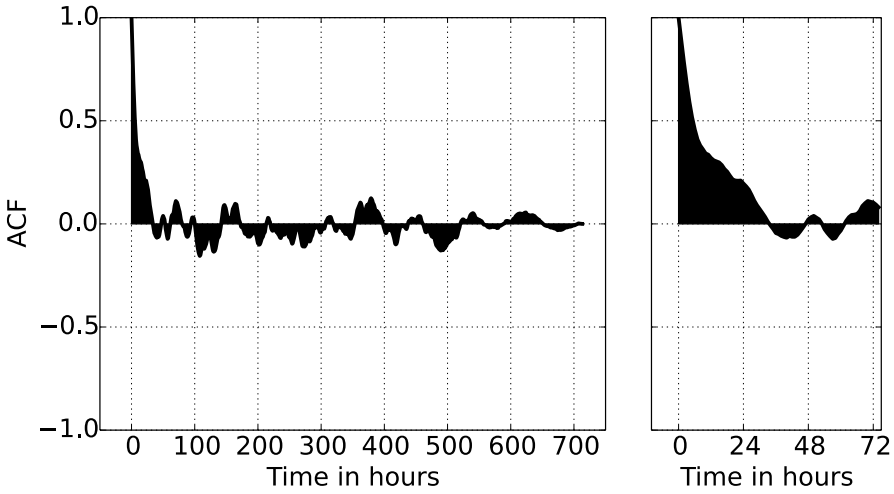


Requested vs Used Memory Resources Leave Big Gap for Optimization

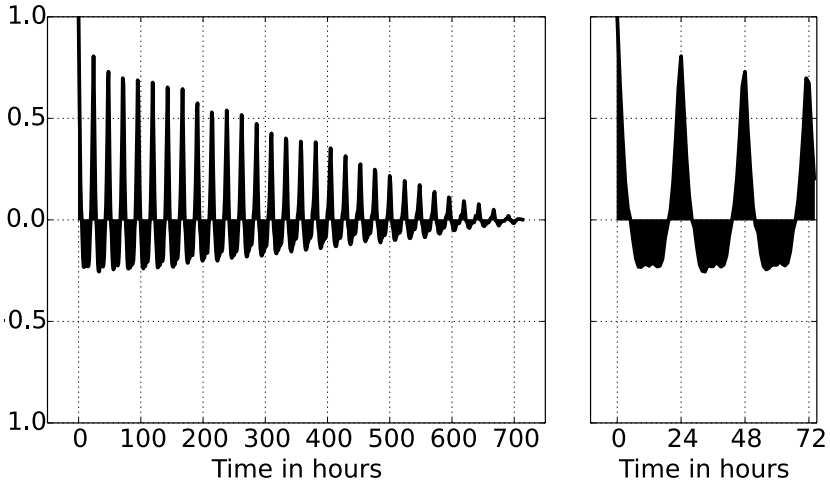


Short and long term Auto-correlation

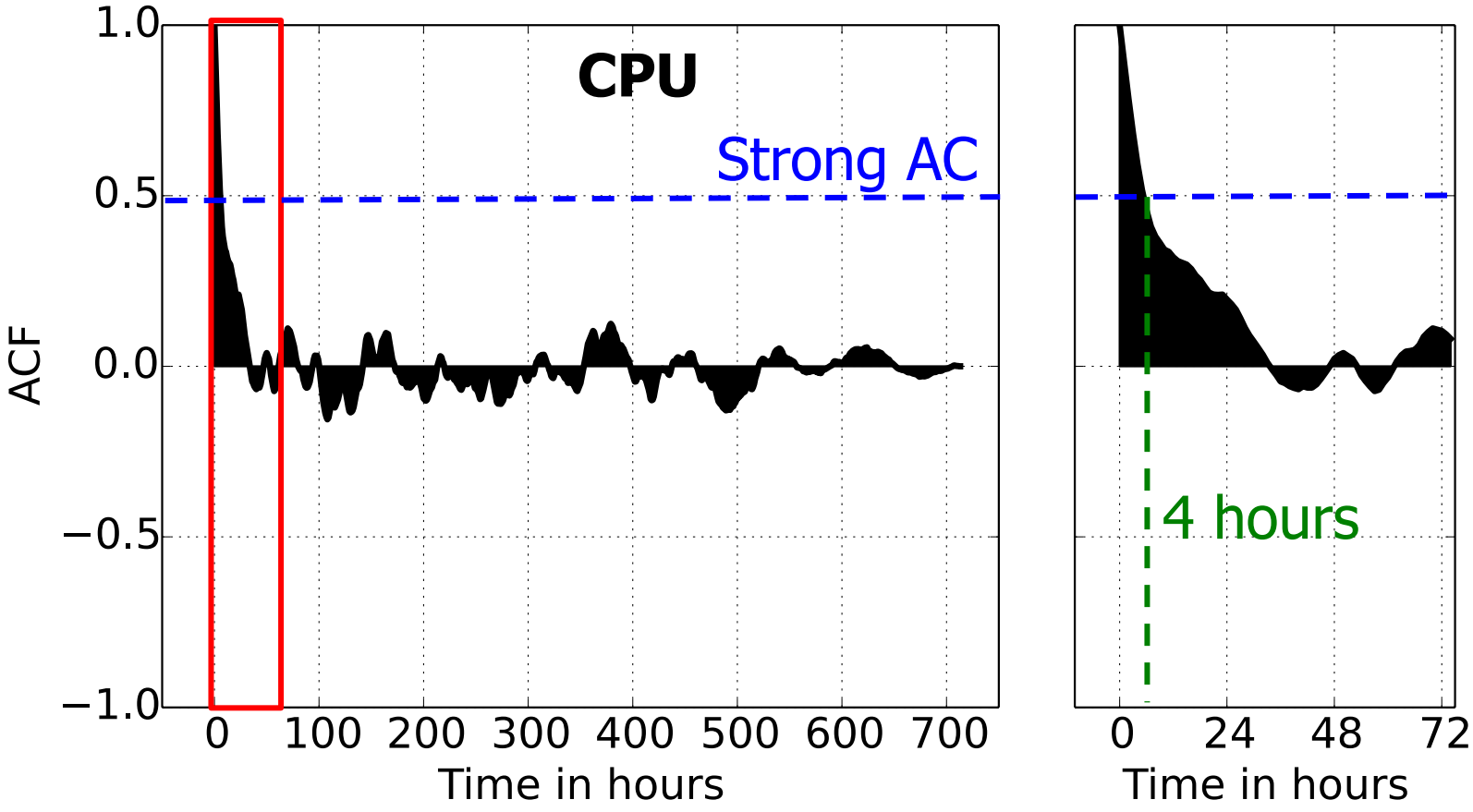
CPU workload



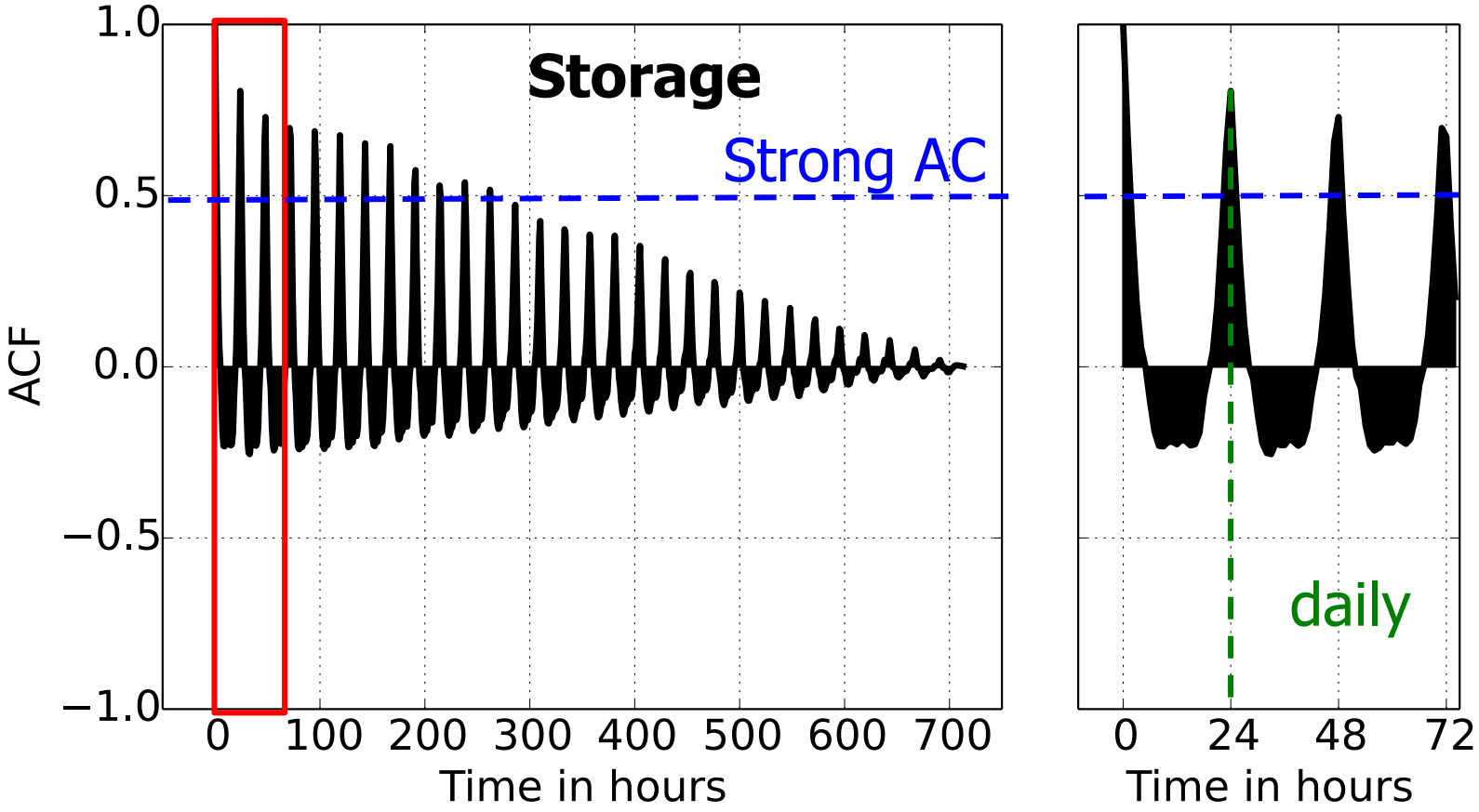
Storage Read workload



Strong hourly Auto-Correlation for CPU

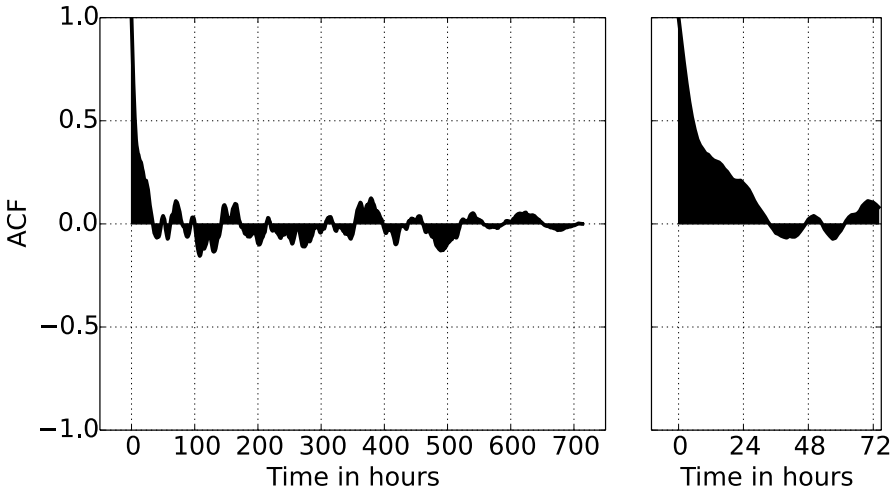


Strong daily Auto-Correlation for Storage



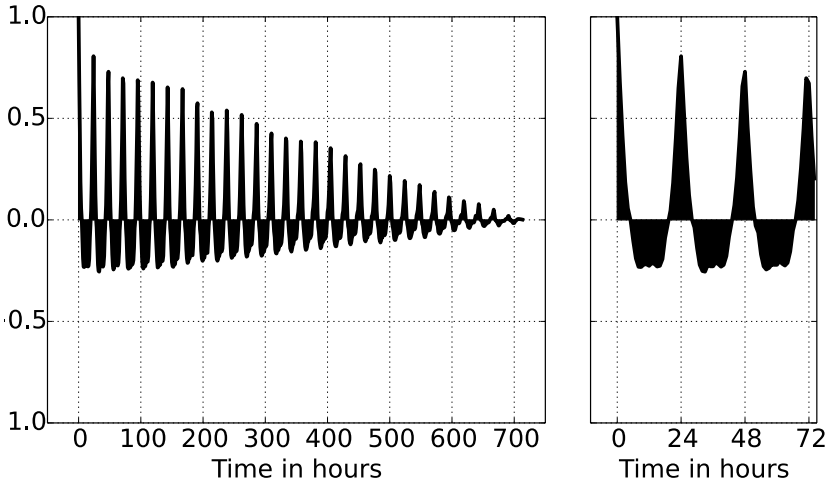
Strong Auto-Correlation only for Storage

CPU workload



Strong short correlation

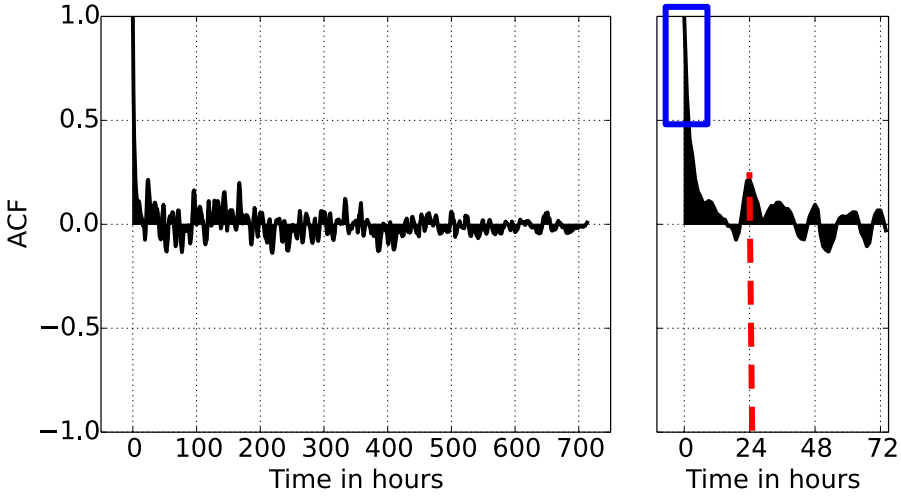
Storage Read workload



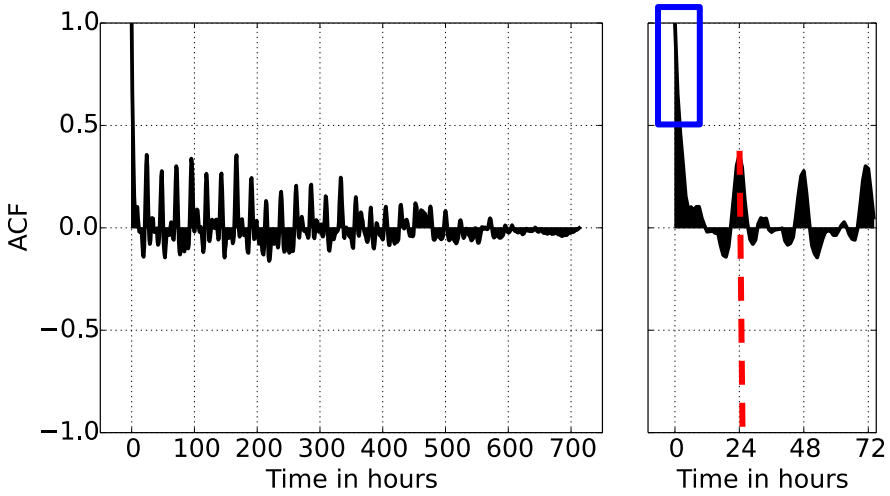
Strong daily correlation

Auto-Correlation for Network traffic

Network receive



Network transmit



Strong short auto correlation for all VMs

Daily auto correlation for some VMs

Business-Critical vs known workloads

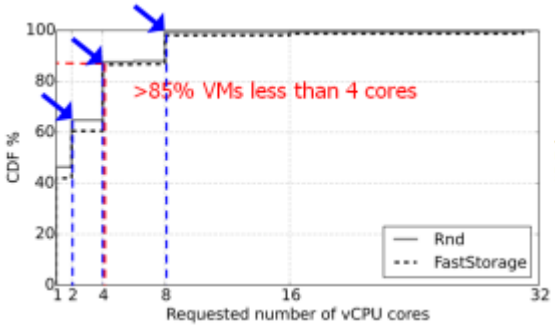
- Long running VMs vs short running jobs
- Compared to parallel workloads, small in size (cpu and memory)
- Much more diverse in nature compared to data analysis workloads from Facebook, Google, and Tabao
  - Monte Carlo Simulation in the financial domain
  - Data analysis of business data
  - Office automation (e.g. web, mail)
  - High available web-services for complex applications (e.g. finance, retail, credit card systems)

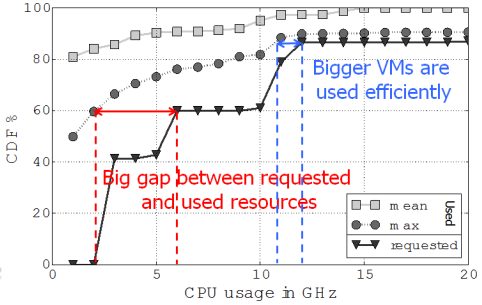
Agenda

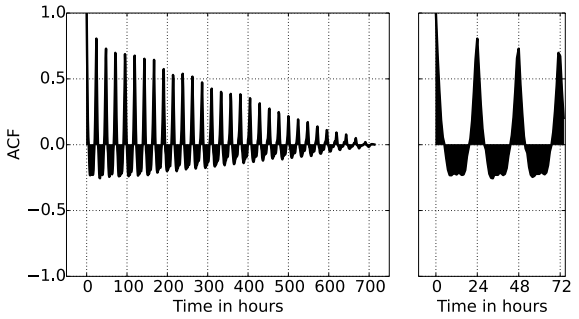
1. Introduction
2. Collected Two Workload Traces
3. Workload Characterization
    1. Requested resources per VM
    2. Used resources per VM
    3. Time patterns in resource usage
4. **Conclusion**

First Study of Business-Critical workloads

- Different from other workloads
- Big gap between requested and used resources
- Different behavior between resource types
- Real-world datacenter optimization at Bitbrains







Business-Critical workloads

- ✓How are we helping the community
  - ✓Method for characterizing workloads
  - ✓Publicly available workload traces
    - ✓The Grid Workload Archive GWA (http://gwa.ewi.tudelft.nl)
  - ✓We are planning on making our analysis tools open source

✓ You can HELP!

✓ Let us analyze YOUR workload traces!