Characterizing and Evaluating Desktop Grids: An Empirical Study

Derrick Kondo[†], Michela Taufer[†§], Charles Brooks III[§], Henri Casanova[†‡], Andrew A. Chien[†]

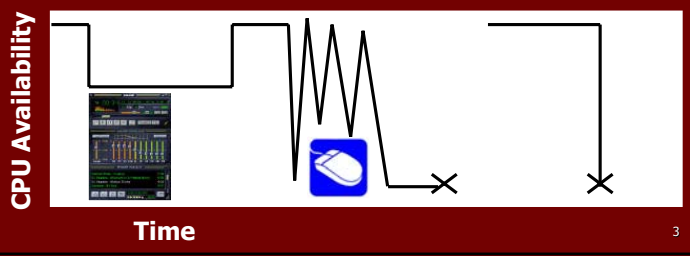
[†] University of California at San Diego
[‡] San Diego Supercomputer Center
[§] The Scripps Research Institute

Desktop Grid Background

- Set of (shared) network-connected resources
- High computational power at low cost
  - Reuse existing infrastructure of resources
- Successful deployment of compute-intensive applications
  - E.g. SETI@home, folding@home, fightaids@home
- Computing platform
  - Internet
  - Enterprise

Desktop Grid Resources

- Resources are extremely heterogeneous
  - E.g. in terms of CPU, memory, disk space, network connectivity, OS
- Resources are volatile



Goal & Approach

- Determine the utility of desktop grids for high throughput, task parallel applications
  - Develop performance model
  - Quantify utility in terms of cluster equivalence
- Measurements of resource availability

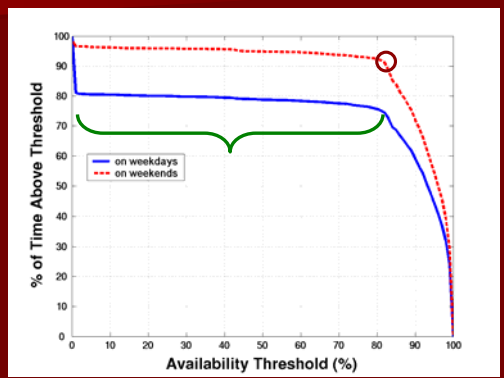
Related Work

- Monitored CPU availability [Wolski99, Wolski99+, Dinda98, Bolosky00, Arpaci95]
  - Difficult to determine effect on desktop grid behavior
    - OS idiosyncrasies
    - Ignores keyboard/mouse activity
      - E.g. hard to infer task failures

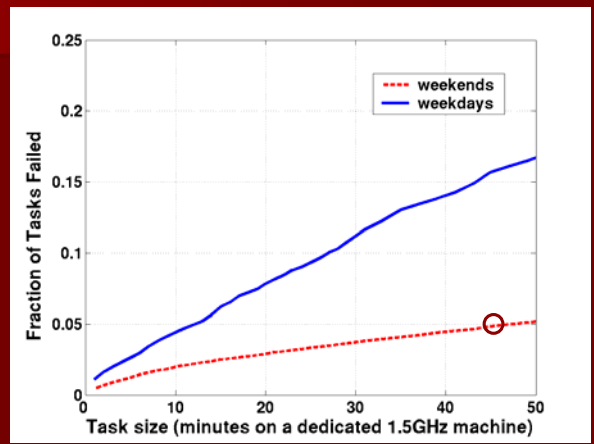
Method

- ***Intrusive*** measurements on ***Entropia*** desktop grid system
  - Fixed time-length tasks
  - Every 10 seconds the program writes the number operations completed to file
  - Output files assembled to produce a CPU availability trace
    - Interpolated gaps due to system overhead
  - ~220 machines at SDSC
  - Cumulative measurement period: 1 month

CPU Availability



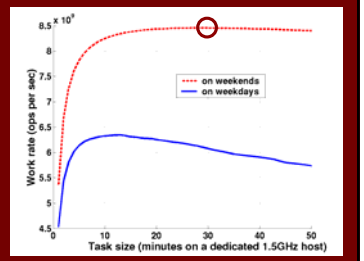
Task Failure Rate



Performance Model

- **N**: number of hosts
- **s**: operations per task
- **f(s)**: failure rate
- **r**: average ops per sec for a host
- **g**: average system overhead per task
- **W(s):** aggregate ops per sec

Optimal Task Size

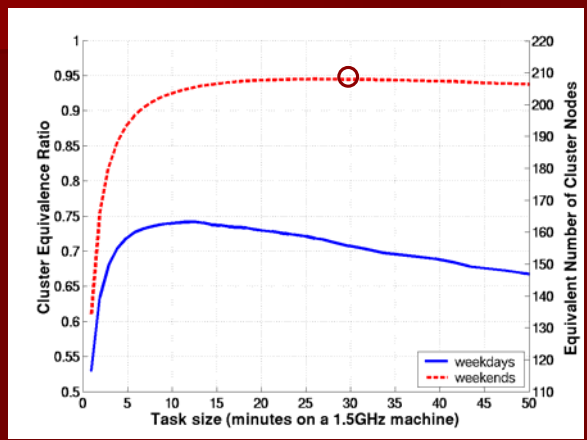


$$W(s) = N * \frac{r\,(1-f(s))}{1+(r/s)*g}$$

Cluster Equivalence

- Compare utility of desktop grid with that of a dedicated cluster
  - High throughput, task parallel applications
- Determine M/N cluster equivalence ratio
  - Given N-host desktop grid, what is equivalent M-node dedicated cluster

Cluster Equivalence



Contributions

- Measurement data
  - Captures temporal structure of resource availability
- Model of desktop grid work rate
- Quantify desktop grid utility for high throughput, task parallel applications using cluster equivalence metric

Current and Future Work

- Traces of other desktop grids
  - Xtremweb, BOINC
- More detailed characterization
  - E.g. at host level
- Resource selection for rapid application turnaround