Digital Institute

Analysis of Power-Saving Techniques over a large multi-use Cluster

Stephen McGough, Clive Gerrard & Jonathan Noble

Newcastle University

Paul Robinson, Stuart Wheater

Arjuna Technologies Limited

Newcastle University

Digital Institute

Overview

- Motivation and Background
- Power Management Policy and Simulation
- Conclusion

Digital Institute

Overview

- Motivation and Background
- Power Management Policy and Simulation
- Conclusion

Digital Institute

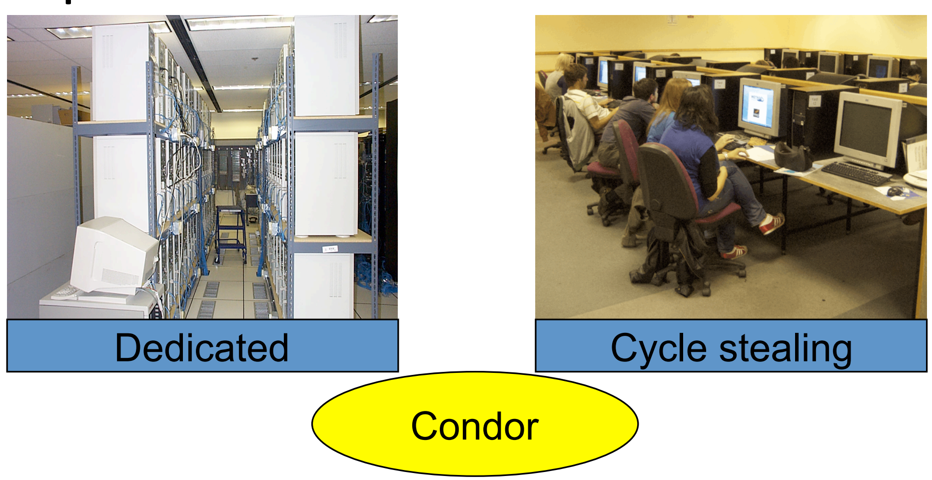
Motivation

- We have run a high-throughput cluster for ~6 years
  - Allowing many researchers to perform more work quicker
- University has strong desire to reduce energy consumption and reduce CO_2 production
  - Currently powering down computer & buying low power PCs
  - "If a computer is not 'working' it should be powered down"
- Can we go further to reduce wasted time?
  - Reduce computer idle time
  - Identify wasteful work sooner?
- Aims
  - Investigate policy for reducing energy consumption
  - Determine the impact on high-throughput users

Condor high-throughput cluster

Digital Institute

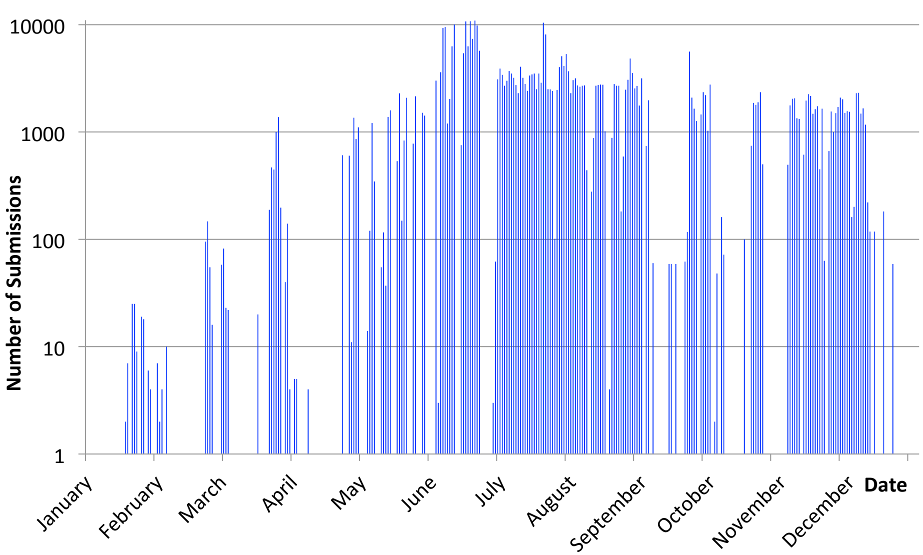
- Condor converts collections of distributed workstations and/or dedicated clusters into a distributed high-throughput computing (HTC) facility
- Condor manages both resources (machines) and resource requests (jobs)
- Established 1985
- Often used to exploit existing computing facilities
  - Though requires them to be turned on



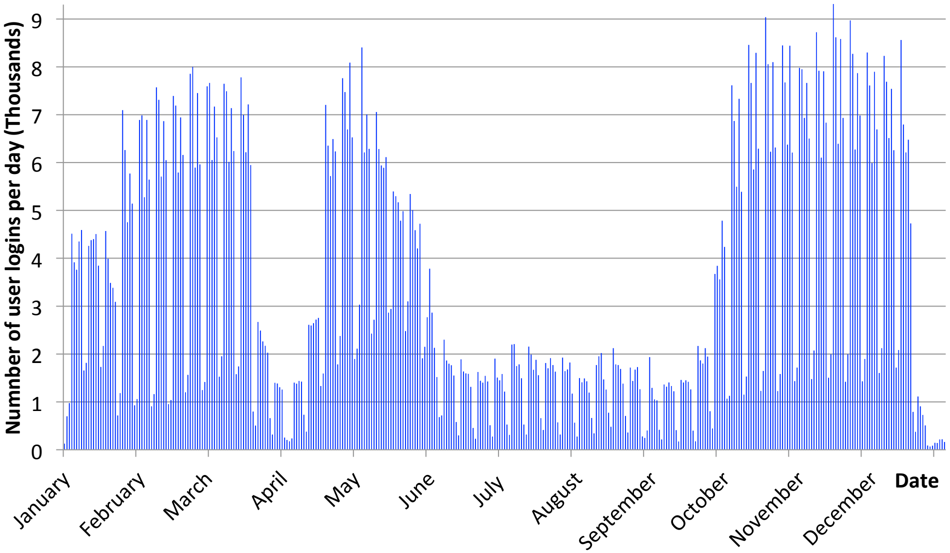
Condor At Newcastle

- Comprises of ~1300 open-access computers based around campus in 35 ‘clusters’
- All computers at least dual core, moving to quad

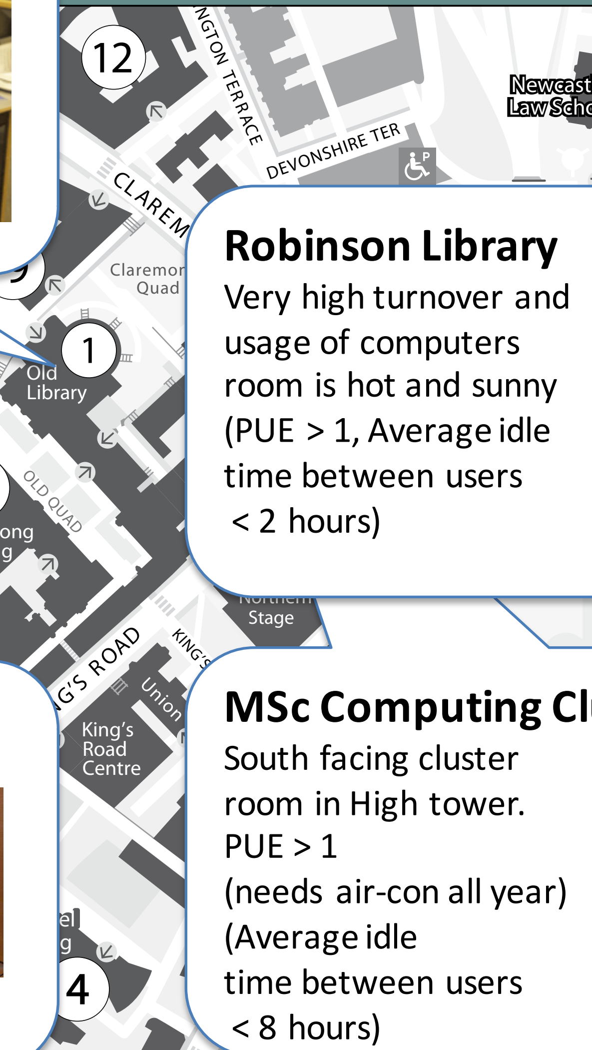
Job Submissions



User Logins



ter Locations

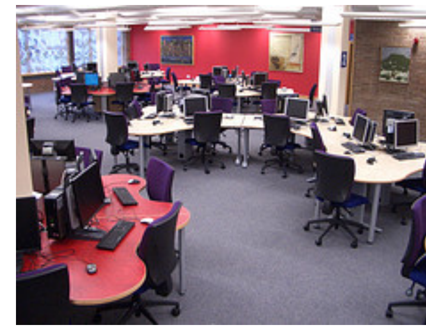


Old Library

Basement Cluster room
Needs heating all year
PUE < 1 (offset heat from computers against room heating) (Average idle time between users < 5 hours)

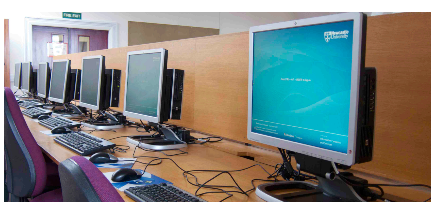
Robinson Library

Very high turnover and usage of computers room is hot and sunny (PUE > 1, Average idle time between users < 2 hours)



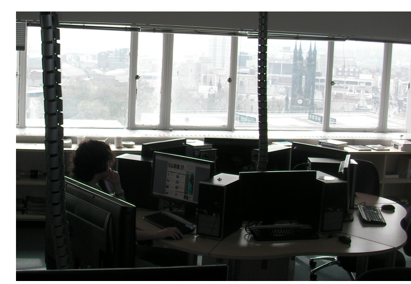
School of Chemistry (Chart)

Very low usage of Computers (PUE ~ 1, Average idle time between users ~23 hours)



MSc Computing Cluster

South facing cluster room in High tower.
PUE > 1
(needs air-con all year)
(Average idle time between users < 8 hours)

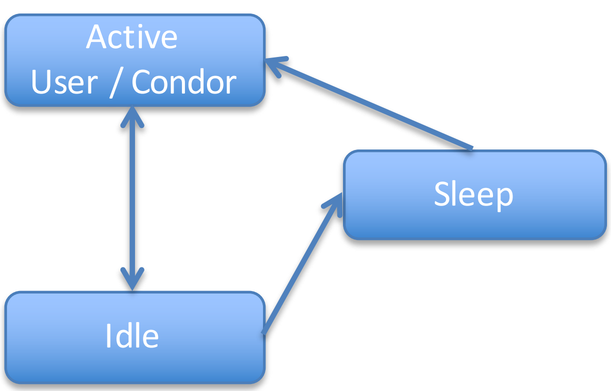


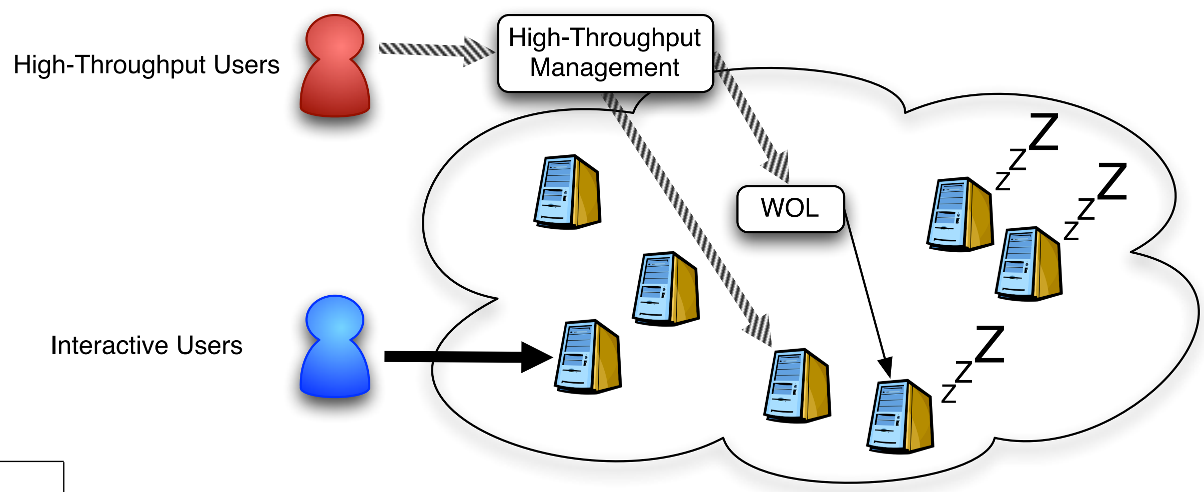
Power Usage Effectiveness (PUE) – depends on location of computer (and time)

Power Efficiency: efficiency = flops/(PUE * watts)

Cluster Simulation

- High Level Simulation of Condor
  - Trace logs from the last year are used as input
    - User Logins / Logouts (computer used)
    - Condor Job Submission times (and duration)





Type	Cores	Speed	Power Consumption		
			Active	Idle	Sleep
Normal	2	~3Ghz	57W	40W	2W
High End	4	~3Ghz	114W	67W	3W
Legacy	2	~2Ghz	100-180W	50-80W	4W

Digital Institute

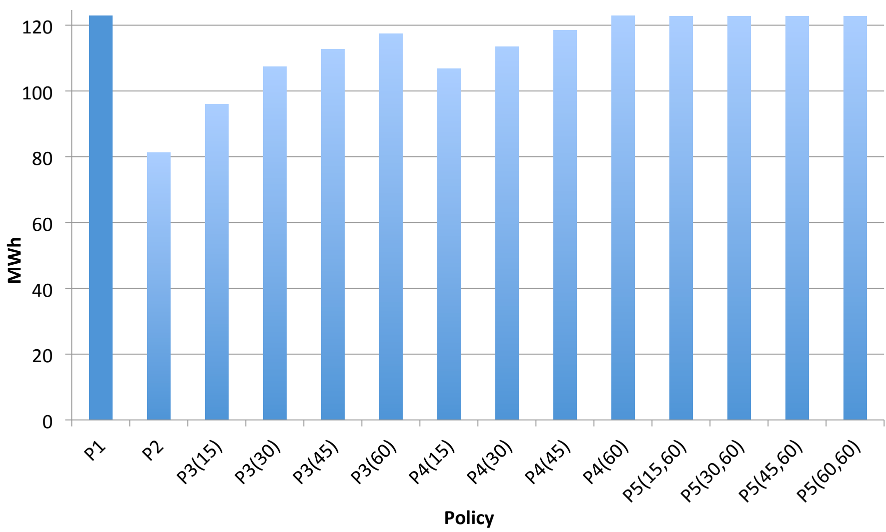
Overview

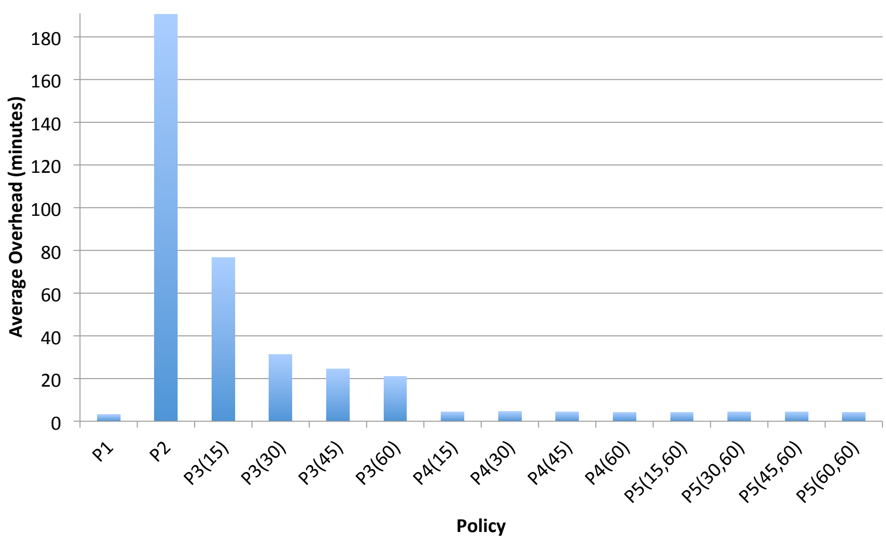
- Motivation and Background
- Power Management Policy and Simulation
- Conclusion

Digital Institute

Power State Policy

- P1: Computers are always on
- P2: On during cluster open hours and off otherwise, no mechanism to wake up
- P3: Computers sleep after n minutes of inactivity with no wake up
- P4: Sleep after n minutes of inactivity but can be woken up
- P5: Sleep after n mins of inactivity but Condor is only informed every m mins

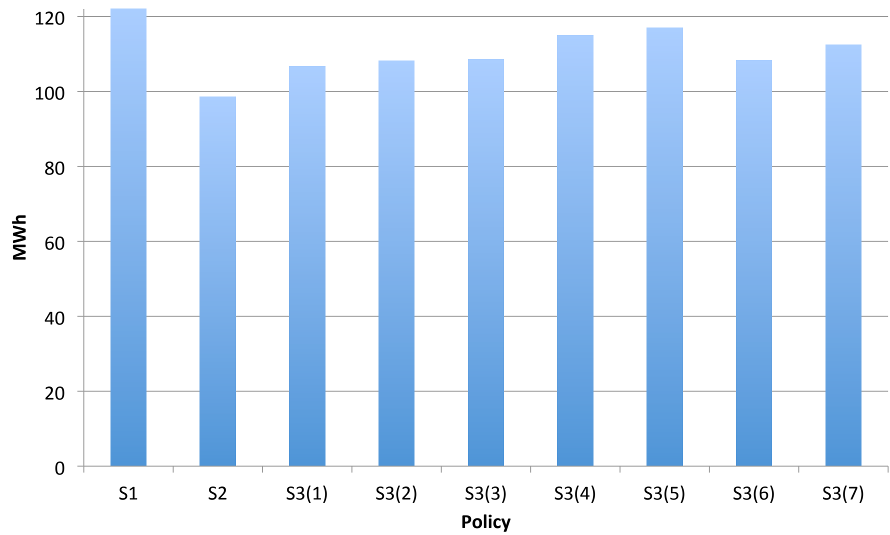


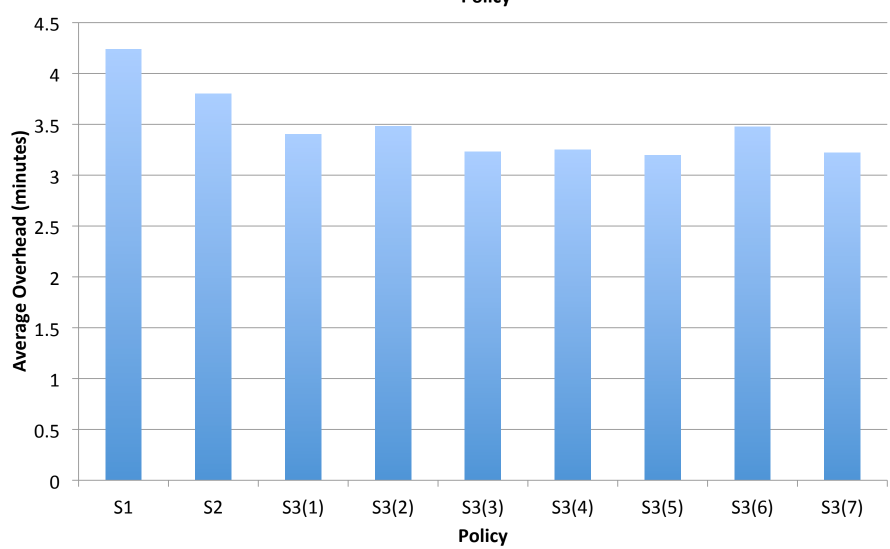


Computer Selection Policy

Digital Institute

- S1: Condor Default (random)
- S2: Target most energy efficient computers
- S3: Target least used computers
  - Least number of interactive logins
  - Largest intervals between logouts and logins

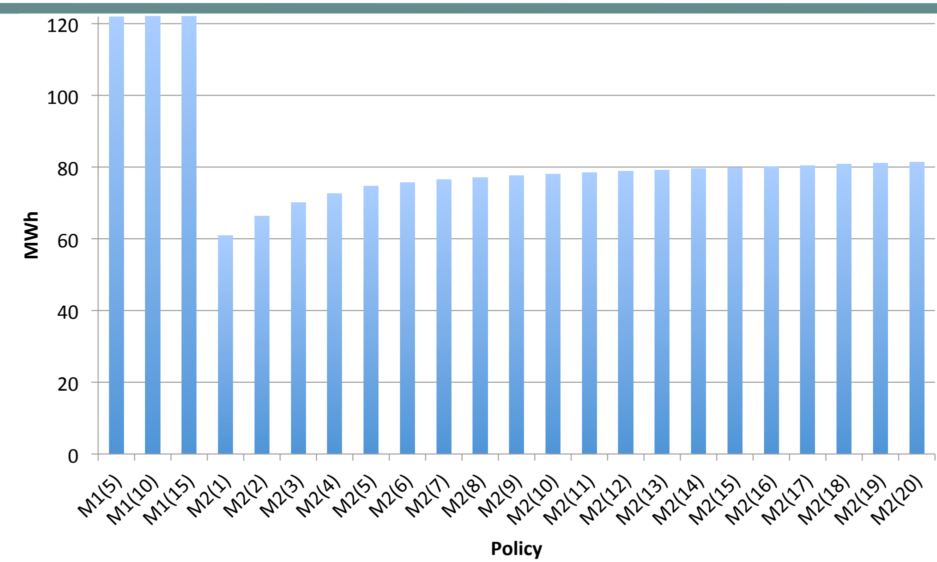


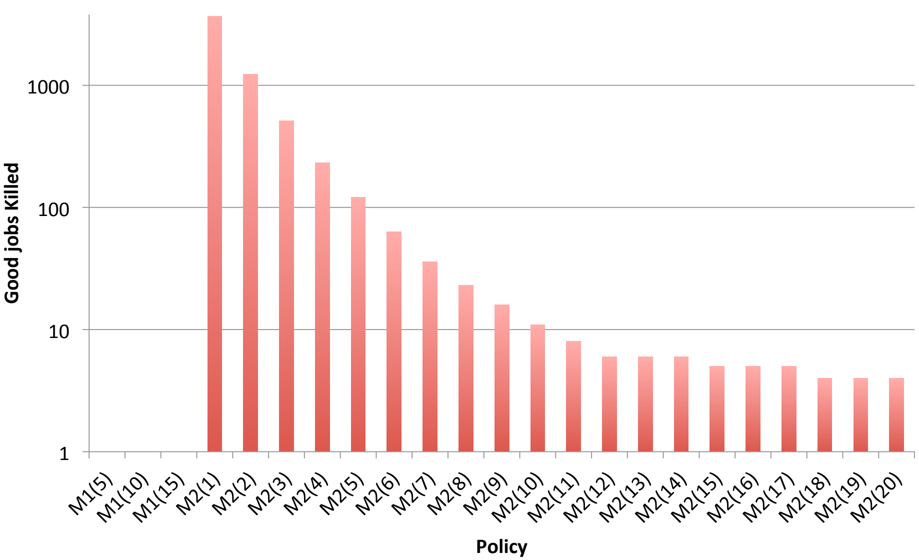


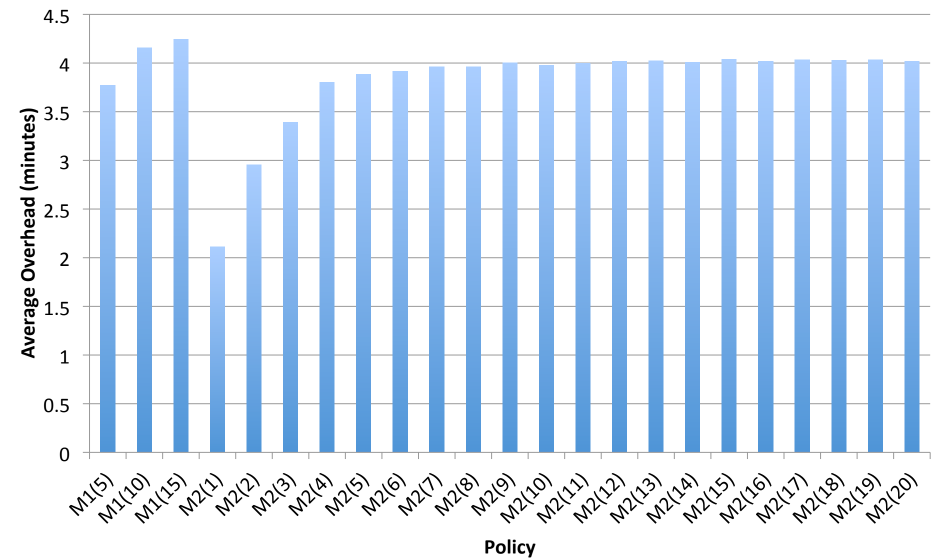
Digital Institute

Management Policy

- M1: Computer is idle for at least n minutes before a Condor job can run on it
- M2: If a job is started more than n times mark it as ‘miscreant’ and don’t re-start

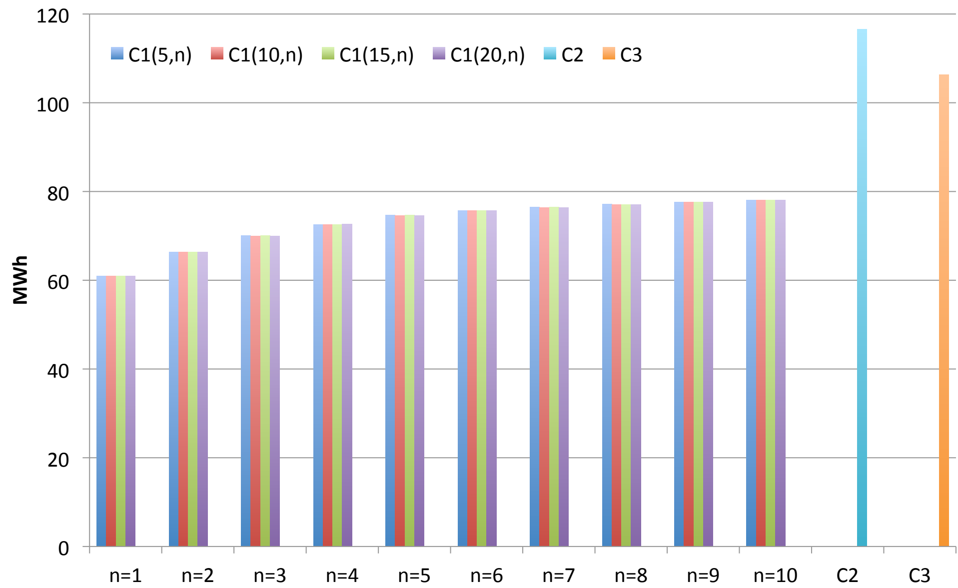


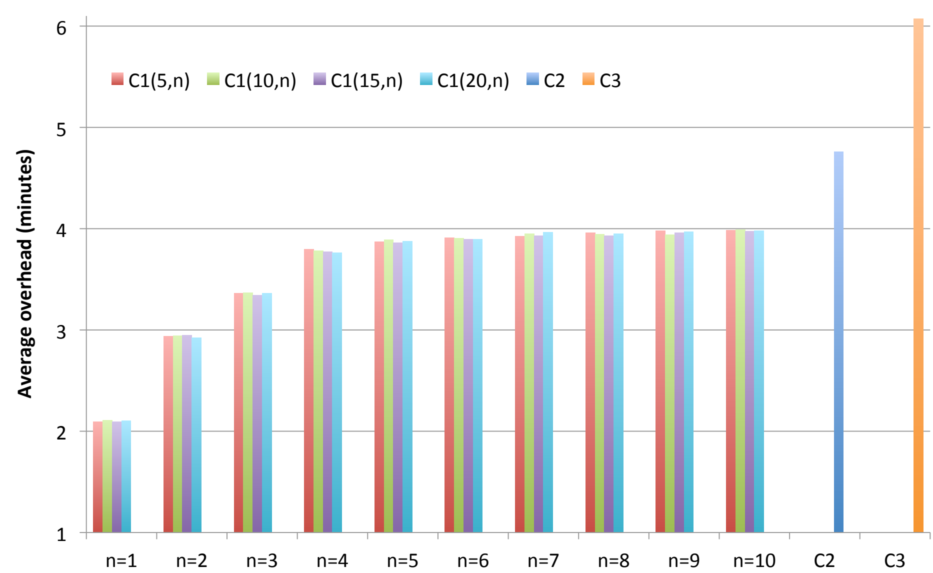




Cluster Change Policy

- C1: Dedicated computers for ‘miscreant’ jobs
  - Run these jobs on computers where they can’t be evicted
- C2: High-throughput jobs defer nightly reboots
- C3: High-throughput jobs use computers at the same time as interactive users





Overview

- Motivation and Background
- Power Management Policy and Simulation
- Conclusion

Digital Institute

Conclusion

- We can save energy (with minimal user impact)
- P4 is the most optimal policy
- S3 – greater impact on overhead
- S2 – greater impact on power consumption
  - These could be merged
- M2 can kill off lots of good jobs
  - Fix this by using C1
- Benefits of C2 and C3 lost due to number of miscreant jobs
  - Need a better way to identify these
- Policies are not mutually exclusive
  - could save ~70MWh (~60% of current usage) without significant impact on high-throughput user
- Powering down cluster saves the most energy

Digital Institute

Questions?

stephen.mcgough@ncl.ac.uk

Newcastle University