

Large-Scale Systems Group

Dennis Abts (Google), Nathan DeBardeleben (LANL), Greg Bronevetsky (LLNL), John Daly (DoD), Armando Fox (Berkeley), Sarah Michalak (LANL), Jon Stearley (SNL), David Walker (Princeton)

***DRAFT* Challenge Problems**

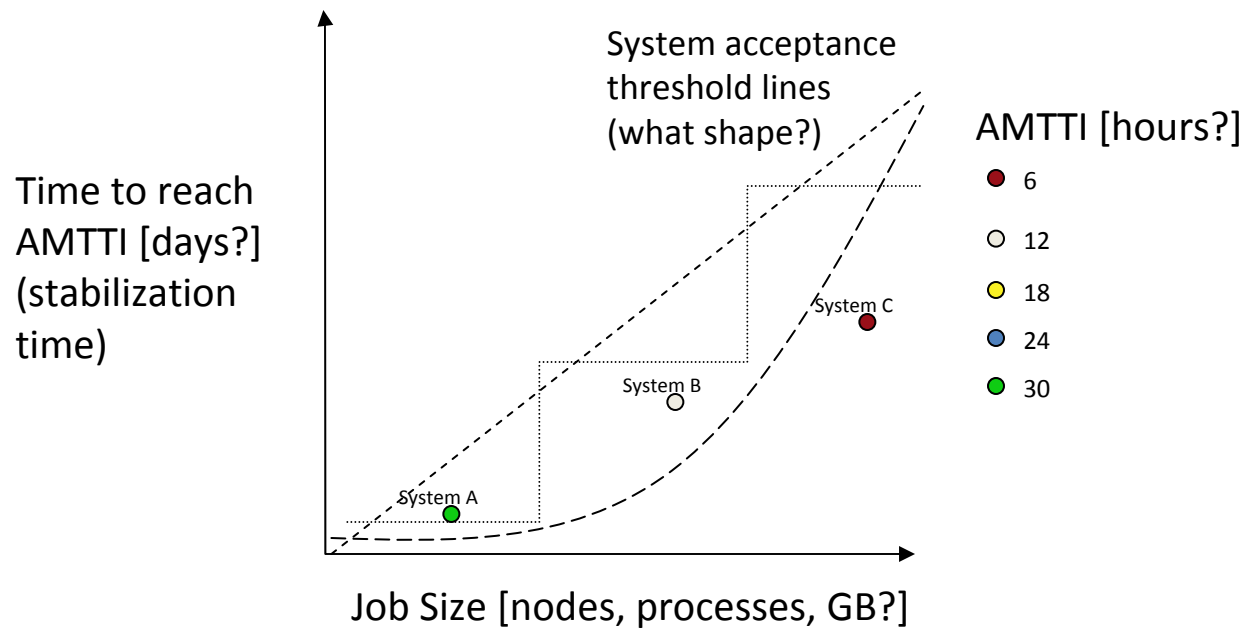
- Challenge Goal 1: Decrease system stabilization time (the time from the system being turned on to when it may be turned over to users) from its current time, which is sometimes measured in months, to X days.
 - This goal addresses bringing the system to a required level of reliability & robustness.
- Challenge Goal 2: Develop a new paradigm for fault tolerant applications.
 - This goal addresses resilience strategies.
- Note: Any solution(s) would need to impose less than 10% overhead in power, performance, space, and equipment.

But...How Do We Define Concepts Like Reliability, Resilience, and Robustness?

- MTBF: Mean time Between Failures
- AMTTI: Application Mean Time to Interrupt
 - System Reliability + Application Resilience
- Reliability = $AMTTI \approx MTBF \gg 1$
- Resilience = $AMTTI \gg MTBF \approx 1$
- Expansion Factor in Time To Completion (TTC)
 - Additional increment in TTC due to failures
- Utility/\$ for a required workload
 - Workload throughput; purchase/power/maintenance costs

Stabilization Time Acceptance Criteria

- What would a plot of stabilization time (days to reach AMTTI=XXX) for historical systems look like?
- What do we want a plot of acceptance criteria to look like for future systems?
- A plot of AMTTI values a system has attained during stand-up would be one way to track its stabilization progress.



There is currently only mediocre visibility into how different sites calculate reliability metrics (eg AMTTI) for acceptance criteria and ongoing operation, so a clear standard and adding the time-to-attain-this-value (eg stabilization time) is a meaningful challenge goal.

Plot should also show number of samples used to calculate each mean (1?).

Note: This slide contributed by Jon Stearley.