# ECEN 5623

## Reliability, Availability, Serviceability, Maintainability

Be Boulder.
University of Colorado Boulder

# ECEN 5623
# Real-Time Systems

*High Availability and Reliability for Mission Critical Systems*

© Sam Siewert

# RASM

- Reliability
  - High Quality Components (Unit Test)
  - Redundancy
    - Dual String
    - Cross Strapping
  - Stress Testing
- Availability
  - Five 9’s
  - Available for Service 99.999% of the time
  - Can be down 5 Minutes and 16 Secs Per Year
  - Recovery - Quick
- Serviceability
  - Field Debug
  - Field Component/Subsystem Replacement
- Maintainability
  - Field Upgrade
  - Remote Firmware/Software Upgrade

© Sam Siewert

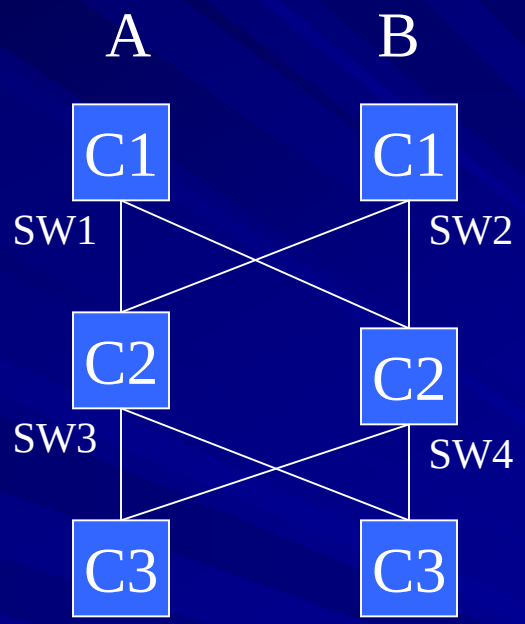
# Reliability

- Quality Units
  - Hardware
    - EMI/EMC, Thermal Cycling, ESD, Radiation
  - Firmware
    - Hardware/Software Integration
    - Hardware Fault Injection and Simulation
  - Software
    - Rigorous Test Criteria – Test Exit Requirements
    - White-box Unit Tests
      - Full path coverage testing
      - Statement coverage
      - Multiple Condition / Decision coverage
    - Black-box Unit and System Tests
      - Test Drivers and Test Harnesses
      - Feature Point and Function Testing
      - Stress Testing (Extremes, Corner Cases, Duration)
    - System Level Tests
    - Regression Testing of Units and System

© Sam Siewert

# Reliability and Recovery

- Redundancy
  - Dual String
    - Side A, Side B
    - Pilot, Co-Pilot
  - Fail-Over
    - Fault Detection, Protection, Recovery
  - Backup System
    - Independent Design
    - e.g. Backup Flight System
  - Cross Strapping of Sides
    - Dual String A & B
    - 3 Components C1, C2, C3
    - 8 Possible Configurations
    - 4 Component Switches
    - A|B Select Switch




| Configurations | C1 | C2 | C3 |
|---|---|---|---|
| 1 | A | A | A |
| 2 | A | A | B |
| 3 | A | B | A |
| 4 | A | B | B |
| 5 | B | A | A |
| 6 | B | A | B |
| 7 | B | B | A |
| 8 | B | B | B |

© Sam Siewert

# High Availability

- Service Up-time is Figure of Merit
  - Number of Times down?
  - How long down?
  - Quick recovery is key
- Hot or Warm Spare Equipment Protection
  - Fault Detection and Fail Over Without Service Outage
  - Excess Capacity
    - E.g. Diverse Routing in a Network
    - Overlapping Coverage in Cell Phone Systems
    - On-orbit spare satellites

© Sam Siewert

# What kind of Protocol is TCP?

A. High Reliability
B. High Availability
C. Both
D. Neither

# Availability vs. Reliability

- Are They the Same?
  - Are all Reliable Systems Highly Available?
  - Are all Highly Available Systems Reliable?
- Reliability = Long MTBF
  - Mean Time Between Failure
  - FDIR When Failures Do Occur
    - Fault Detection, Isolation and Recovery
    - Safing
    - MTTR (Mean Time to Recover)
- Availability = MTBF / (MTBF + MTTR) = % Uptime
  - MTBF = 8,766 hours (525,960 minutes)
  - MTTR = 5 minutes
  - Availability = 525,960 / (525,960 + 5) = 99.999% Uptime

© Sam Siewert

# Reliability Calcutions

From book, if you had 10 components with probability of survival of 99.999%, then system probability of survival is

.99999*10 = 99.999%

NO .99999^10 = 99.990%

What is Psys if Pcomp = 99.999% and 10000 components?

Psys = .99999^10000 = 90.48%

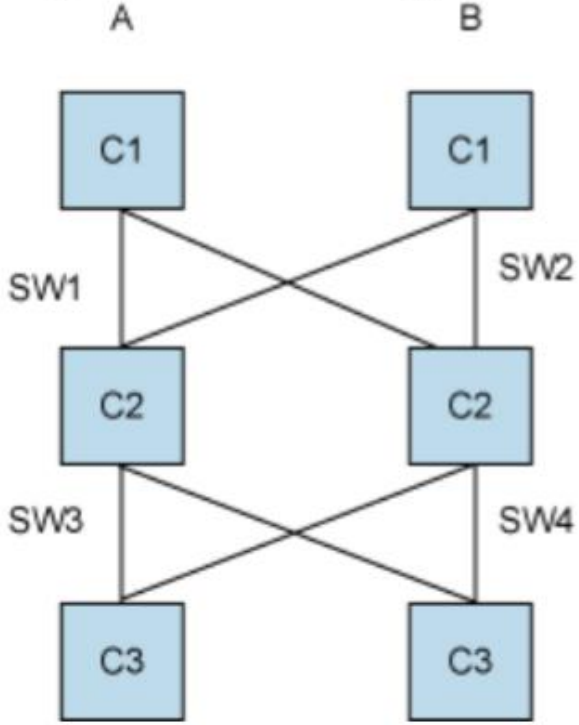
# ***Negative Tests*** - HA Design Criteria

- 5 9's – Out of Service no more than 5 Minutes per Year [Up 99.999% of the Time]
- Availability = MTTF / [MTTF + MTTR]
- MTTF ≈ MTBF
- E.g. 0.99999 ≈ (365.25x60x24 minutes) / [(365.25x60x24) + 5 minutes]
- MTBF=Mean Time Between Failures, MTTR=Mean Time to Recovery

Big iron lessons, Part 2: Reliability and availability: What's the difference?

Apply RAS architecture lessons to the autonomic Self-CHOP roadmap

**Figure 1. Dual-string, cross-strapped subsystem**




| Configuration | Component C1 | Component C2 | Component C3 |
|---|---|---|---|
| 1 | A | A | A |
| 2 | A | A | B |
| 3 | A | B | A |
| 4 | A | B | B |
| 5 | B | A | A |
| 6 | B | A | B |
| 7 | B | B | A |
| 8 | B | B | B |

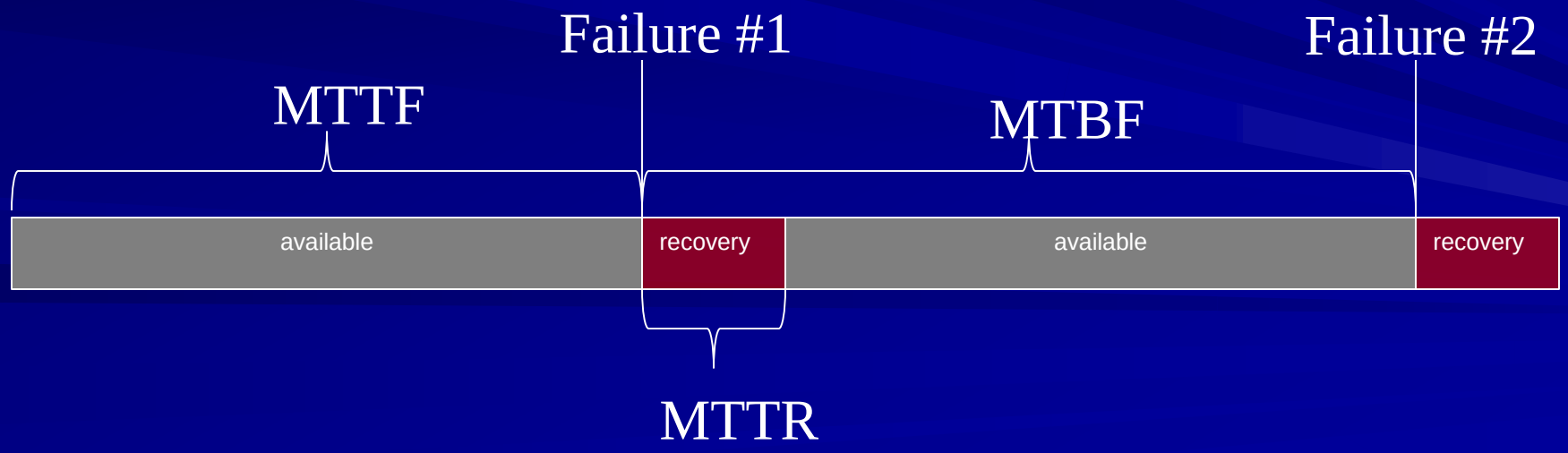
© Sam Siewert

# Availability

- Time System is Available for Use
- Time System is Unavailable for Use
- Ratio of Time Available over Total Time Both Available and Not Available

- Some Variation in Definitions of MTTF, MTBF, MTTR
  - MTBF = MTTF + MTTR [In Book]
  - MTTF is Simply Time Between Failures (Observed in Accelerated Lifetime / Stress Tests)
  - MTBF Often Stated – Does it Really Include MTTR or Not?
  - Typically Recovery Time is Small, So Negligible, But if Long, Difference is More Important
  - Just Be Clear on Definitions [Can Vary]

- % of Time System is Available for Use Each Year (Or Other Arbitrary Period of Time)




© Sam Siewert

# Serviceability

- Field Anomaly Diagnosis
  - Trace capabilities
  - Debug capabilities
  - Error detection and prediction
  - Performance monitoring
  - Ease of Component, Sub-system Replacement
  - Diagnostics and Isolation

© Sam Siewert

# Maintainability

- Field Upgrade Capabilities and Limitations
- Make Before Break
  - Can I upgrade while Unit remains in service?
  - Can I test upgrade prior to service switch over?
  - Can I fall back to old configuration on failure of new?
  - E.g. – Upload new Boot Code, it Fails, Rolling Reboot Causes Fail Over to Old Image

© Sam Siewert

# Component, Unit, Sub-Sytem Reliability

- Stress Testing
  - Hardware
    - Thermal Cycling
    - Vacuum Chamber
    - Radiation Dosing
    - Mechanical Shock and Vibe Tests
  - Firmware/Software
    - Hardware sub-system or component failure
    - Repeated Testing for Long Duration
      - Memory Leaks
      - Race Conditions
      - Priority Inversions
      - Deadlock
    - Monte Carlo Analysis

© Sam Siewert

# Monte Carlo Analysis

- Repeatable Pseudo Random Sequence
  - Can Re-run A Failure Case with Same Seed
  - Nominal Value +/- Dispersion
    - Dispersion is a Random Variation
    - Uniform or Normalized Distribution
  - Drive Timing Variations
  - Drive Input Variations
  - Drive Simulation Variations

© Sam Siewert

# Embedded Autonomic?

- Autonomic Systems appear to be Self-Healing Through the Management of Redundancy with Automatic Recovery using Excess Capacity and Distributed Hot Spareing of Resources and Components.
- CHOP – Configuring, Healing, Optimizing and Protecting. Healing means replacing damaged portion. Automate replacement of a FRU. Used to be considered science fiction, but new technology may make cost-effective.

Autonomic – requires little intervention. USB – self configuring.

- Are Autonomic Principles System Level Only? Internet multi-path Routing
- Do they Apply to Embedded Systems?
- To SoC?
- To Enterprise Systems? Storage – Raids
- Project to backup the entire net.

© Sam Siewert

# Beyond RASM?

- Autonomic Architectures
- 4 Goals (self-CHOP)
  - Self-Configuring
  - Self-Healing
  - Self-Optimizing
  - Self-Protecting
- Eight Features
  1. "Self Knowing" – *Status at All Levels, BIST, Used by Peers and in Hierarchy*
  2. "Configurable/Reconfigurable" – *FPGA, Fusing, Cross Strapping*
  3. "Self Optimizing" – *Built In PMU, Trace, SPA*
  4. "Self Healing" – *Recovery, Fusing, Excess Capacity, Redundancy, Autonomous Repair*
  5. "Self Protecting" – *Virus, Security, Data Integrity*
  6. "Discovery" – *Finding and Identifying Resources*
  7. "Interoperable" – *Component, Subsystem, System*
  8. "Self Managing" – *Low or No Admin (Zero Uplink)*

© Sam Siewert

# Motivation for Autonomic

- Moore's Law Reaching Limit?
  - GHz of CPU Cycles, Multiple CPUs
  - Gbits/sec of Bandwidth
  - GBytes of Memory
  - Terra Bytes of Storage
- Ability to Manage These Resources Reaching Limit?
- Maintenance Bigger Issue than Capacity, Throughput, and Bandwidth
- Comparable to Evolution of Telecomm
  - Party Lines (Users Manage and Arbitrate for Resources Directly)
  - Operators for Exchanges Configure Connections (Centralized Management)
  - Switching Systems (Automated Operations, Service Provisioning)
- Number of Administrators/Operators Per System?
  - N to 1 (Mainframes and Supercomputers)
  - 1 to 1 (Minicomputers, File Servers, Workstations, PCs)
  - 1 to N (Clusters, GRID Computing, RAID)

© Sam Siewert

# Autonomic Characteristics

- Examples of Autonomic Features Today
  - CPU
    - Watch-dog timer for Software Sanity
    - Software recovery for missed service deadline
    - eFuse in circuit enable/disable for post manufacture reconfiguration
    - PMU for on-chip performance monitoring
  - IO
    - Sensor Fusion and Sensor/Actuator Cross strapping
    - Diverse routing through networks
    - CRC, timeouts, re-transmission
  - Memory
    - EDAC memory
  - Storage
    - RAID5 volume rebuild on drive failure
    - RAID6 dual failure handling

© Sam Siewert

# Some Web Links

- RASM
  - http://www.sun.com/service/support/ras/
  - http://www-1.ibm.com/servers/eserver/pseries/hardware/whitepapers/p690_ras.html
  - http://www-1.ibm.com/servers/eserver/pseries/hardware/whitepapers/s80ras_2.html
  - http://www.microsoft.com/serviceproviders/support/isp_topic_reliablity.asp
  - http://www.microsoft.com/windows2000/server/evaluation/business/relavail.asp
  - http://www.cisco.com/en/US/products/hw/switches/ps1925/products_technical_reference_chapter09186a00800eac71.html
- Autonomic
  - http://www.autonomic-conference.org/
  - http://researchweb.watson.ibm.com/journal/sj42-1.html
  - http://www-106.ibm.com/developerworks/autonomic/newto/
  - http://www-106.ibm.com/developerworks/autonomic/library/technical/autonomic.html

© Sam Siewert