SAND2022-10154C

Sandia National Laboratories

Verification of Cyber Emulation Experiments Through Virtual Machine and Host Metrics

Presented by: Jamie Thorpe

Authors: Jamie Thorpe, Laura Swiler, Seth Hanson, Gerardo Cruz, Thomas Tarman, Trevor Rollins, Bert Debusschere

Cyber Security Experimentation and Test Workshop (CSET) 2022

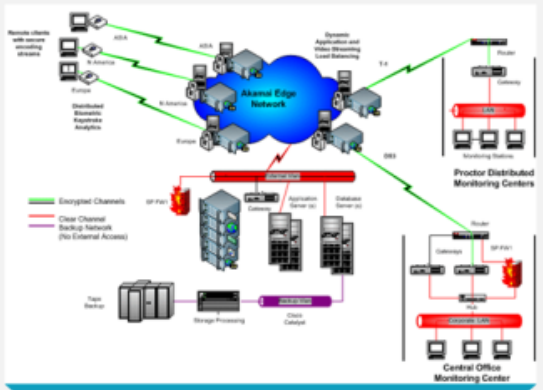
Session 3

August 8, 2022

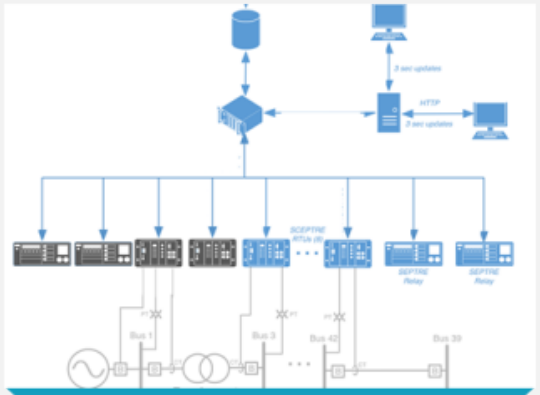
U.S. DEPARTMENT OF ENERGY | NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.0003525.

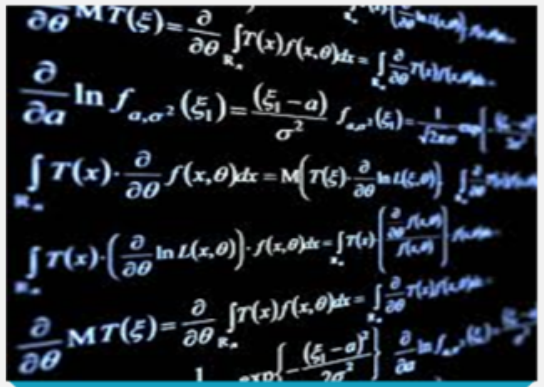
Cyber Experimentation



ACTUAL SYSTEM



VIRTUALIZED TESTBED



SIMULATION

"BAD DAY" BRAINSTORMING

Increasing Realism
Decreasing Flexibility
Increasing Cost
Increasing Time

Increasing Abstraction
Increasing Flexibility
Decreasing Cost
Decreasing Time

Verification

Is the experimental environment working as intended?

- If so, results can be used to better understand the system modeled
- If not, experiment results may not be reliable

Different Types of Verification

- Timing Realism – Processes and network traffic occur at expected rate
- Traffic Realism – Network traffic contains expected fields/data
- Resource Realism – Physical host has enough resources to support experiment

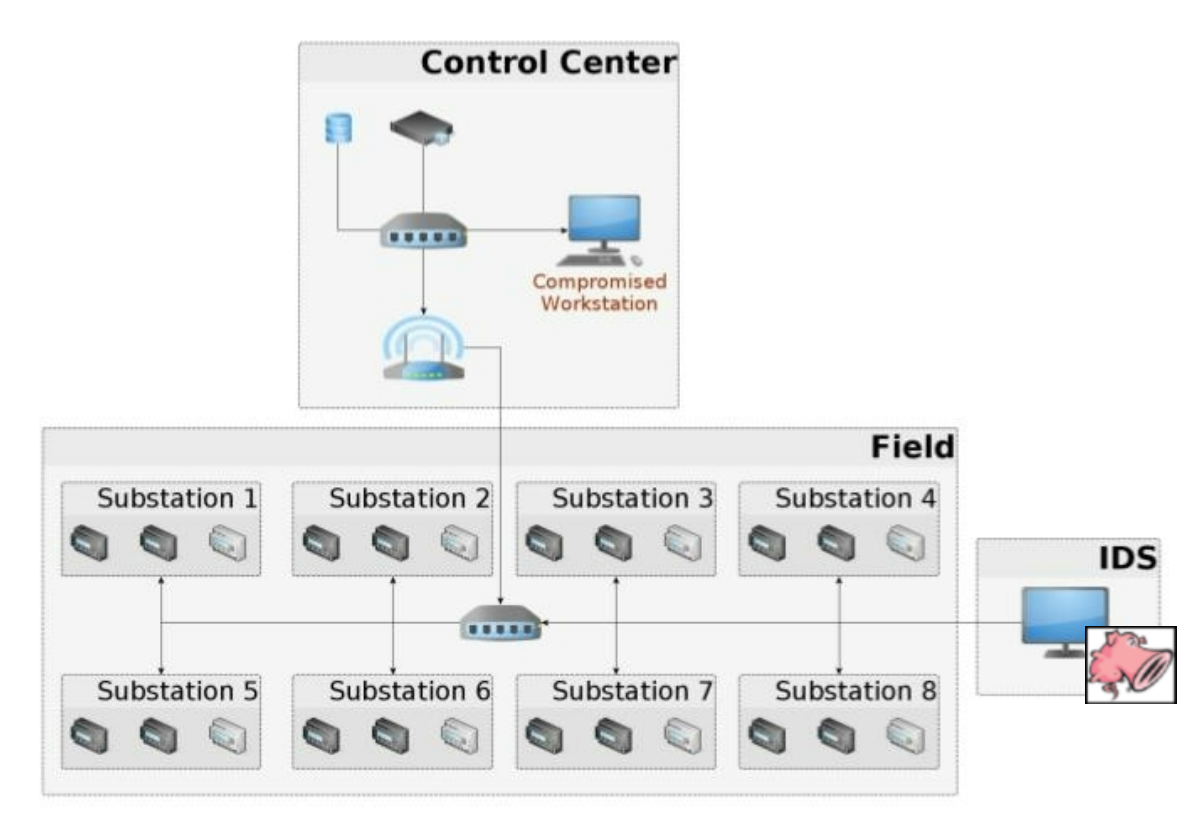
Approach

1. Devise mechanism for increasingly stressing physical host resources
   - Run more experiments (replicates) in parallel
2. Run multiple replicates in each resource setting
3. Collect key telemetry and results data from each replicate
   - Physical host load (telemetry)
   - In-experiment virtual machine functionality (telemetry)
   - In-experiment results
4. Compare telemetry from replicates under different resource settings with experiment results

Can a Telemetry-Based Metric be Used to Determine if the Results of a Replicate are Unreliable?

Scenario 1 – Scanning and Detection

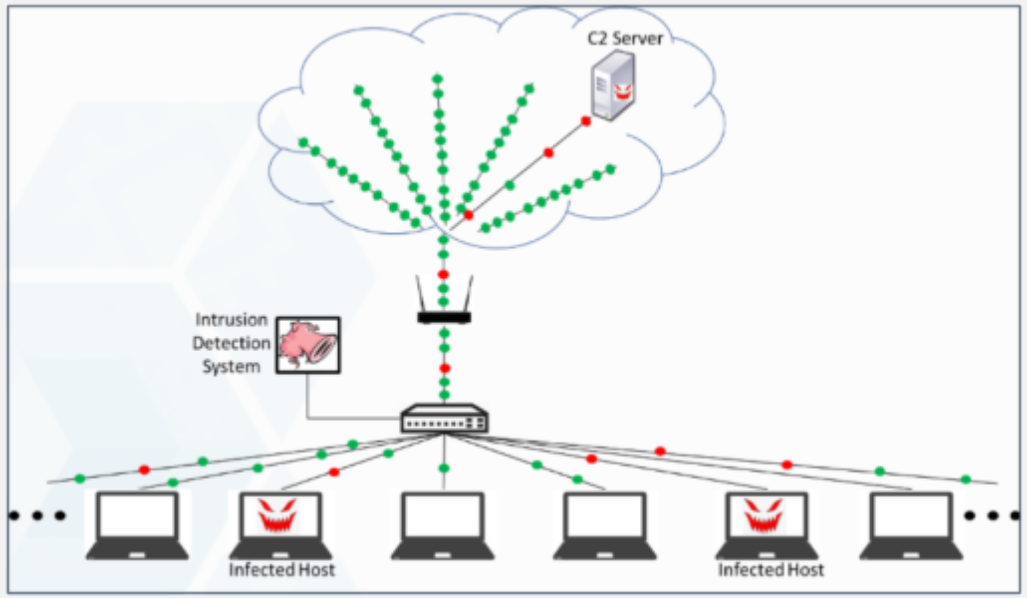
Detect adversary running port scan on 24 nodes



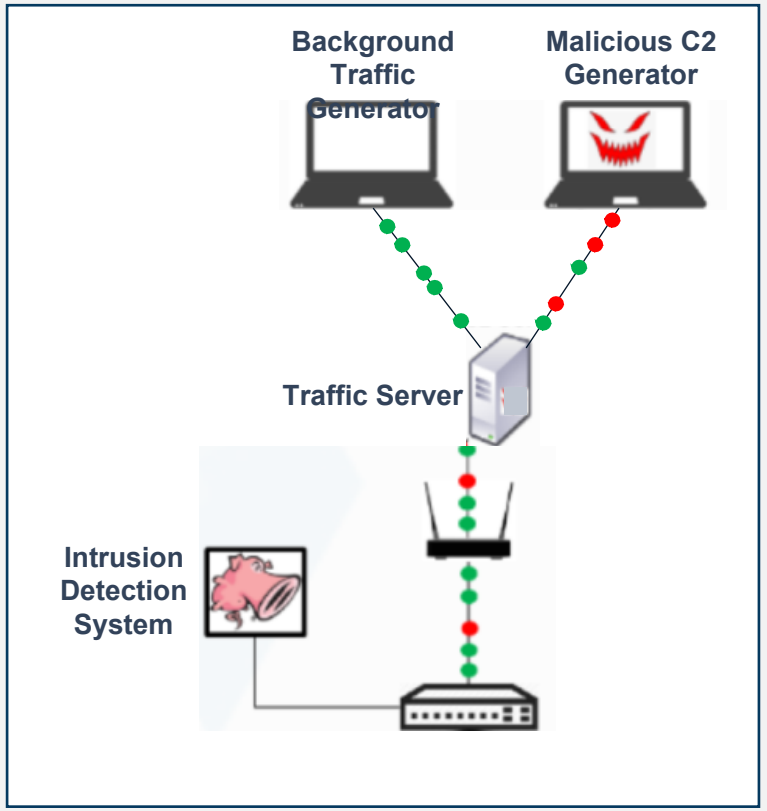
- Quantity of Interest: Detection Time
- Deterministic Scan Order
- No Packet Loss Assumed

Scenario 2 – Command and Control (C2)

Detect malicious traffic between host(s) and C2 server



Scenario as Described



Scenario as Modeled

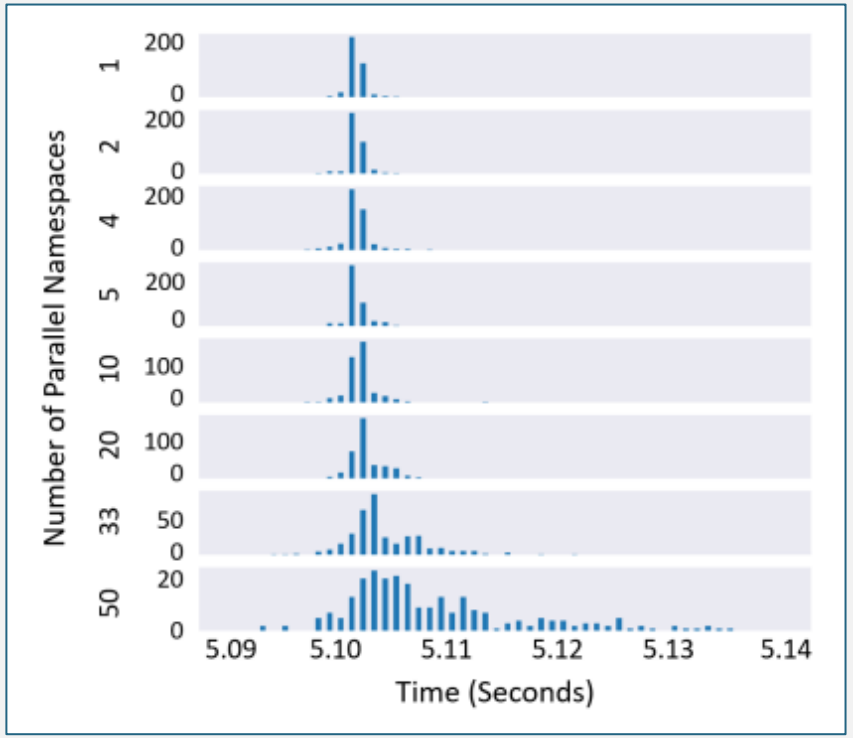
- Quantity of Interest: Number of Alerts at Certain Timestamps
- No Packet Loss Assumed

Results – Scenario 1 (Scanning and Detection)

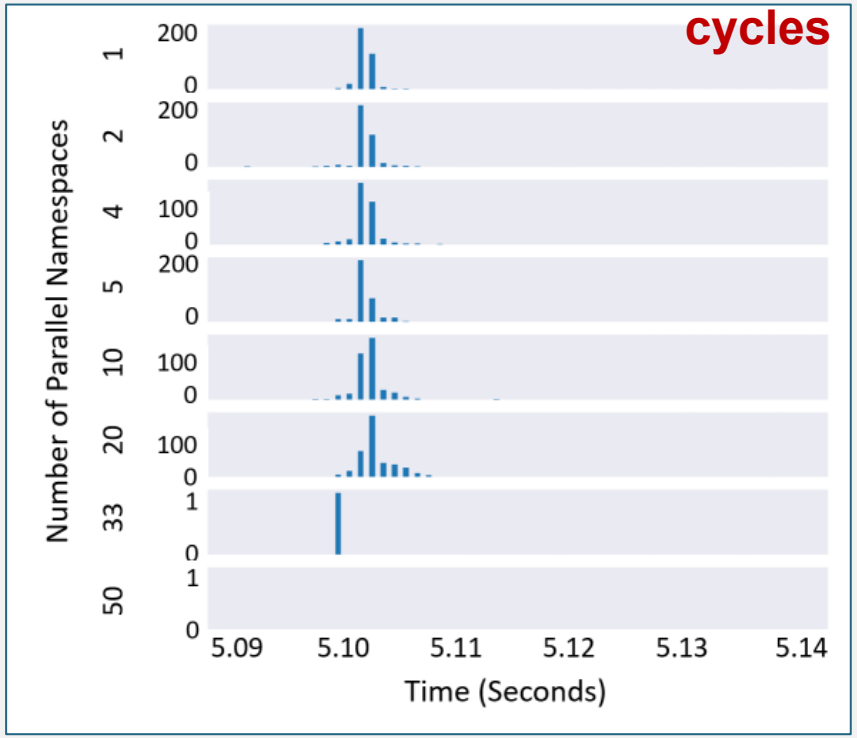
Example Metrics:

- Stolen Cycles = 0
- Load ≤ 64 Processes
- Throughput ≥ 250k bps

All replicates

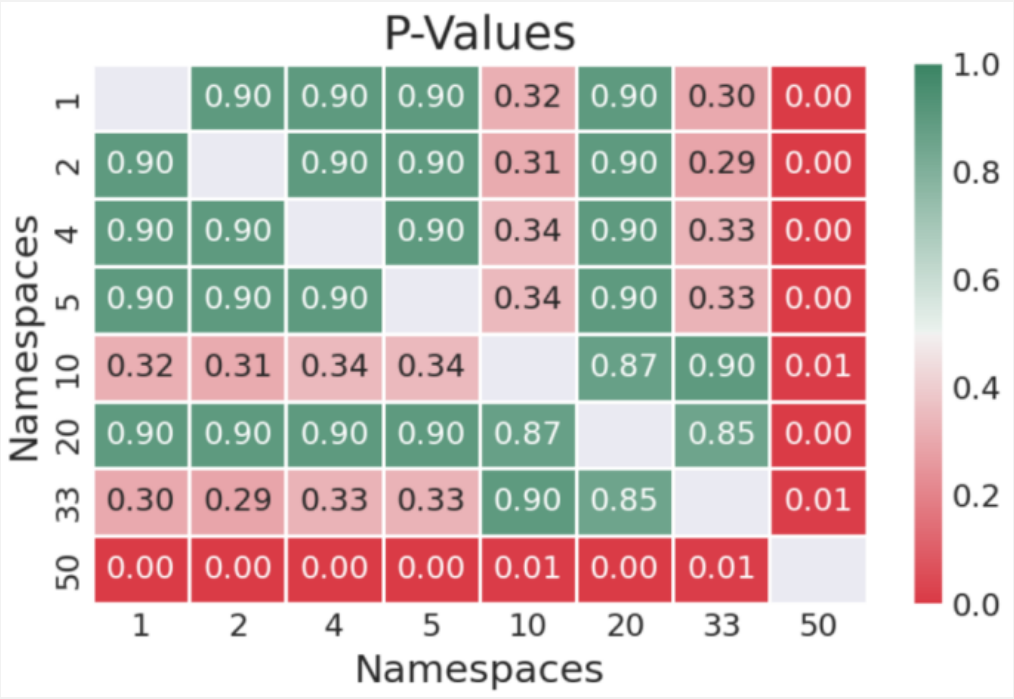


No stolen cycles

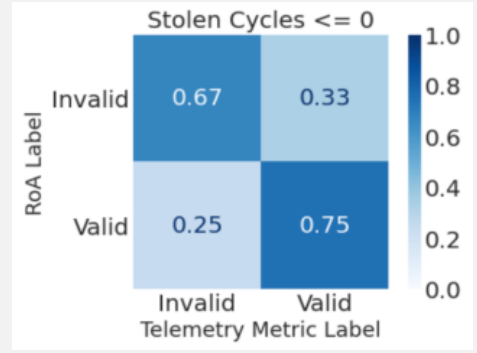


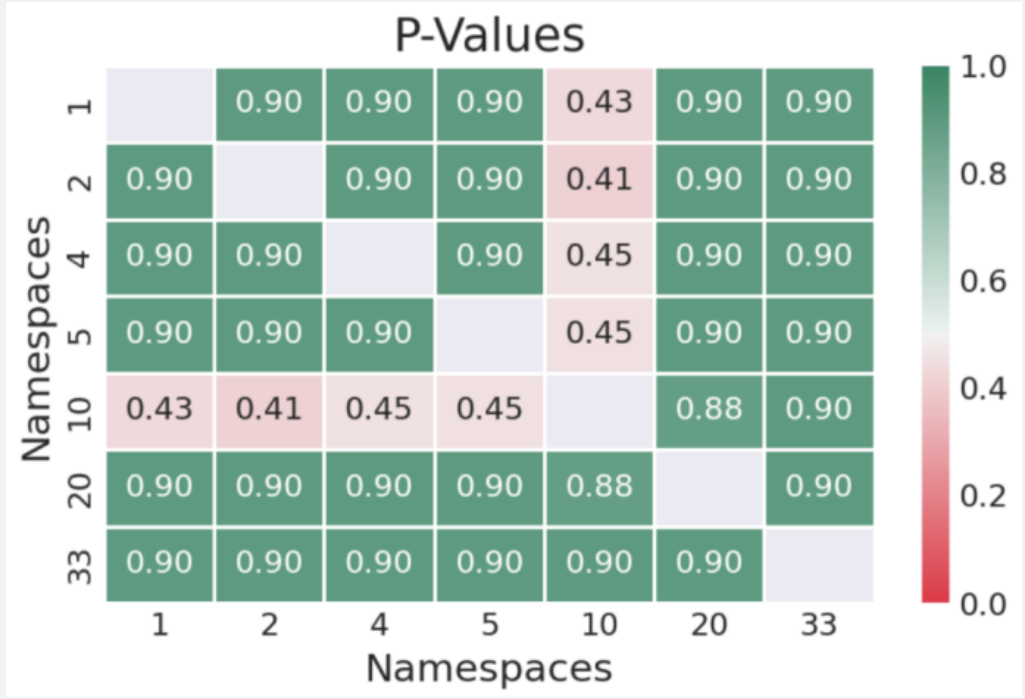
Results – Scenario 1 (Scanning and Detection)

Tukey - All replicates



Tukey – No stolen cycles



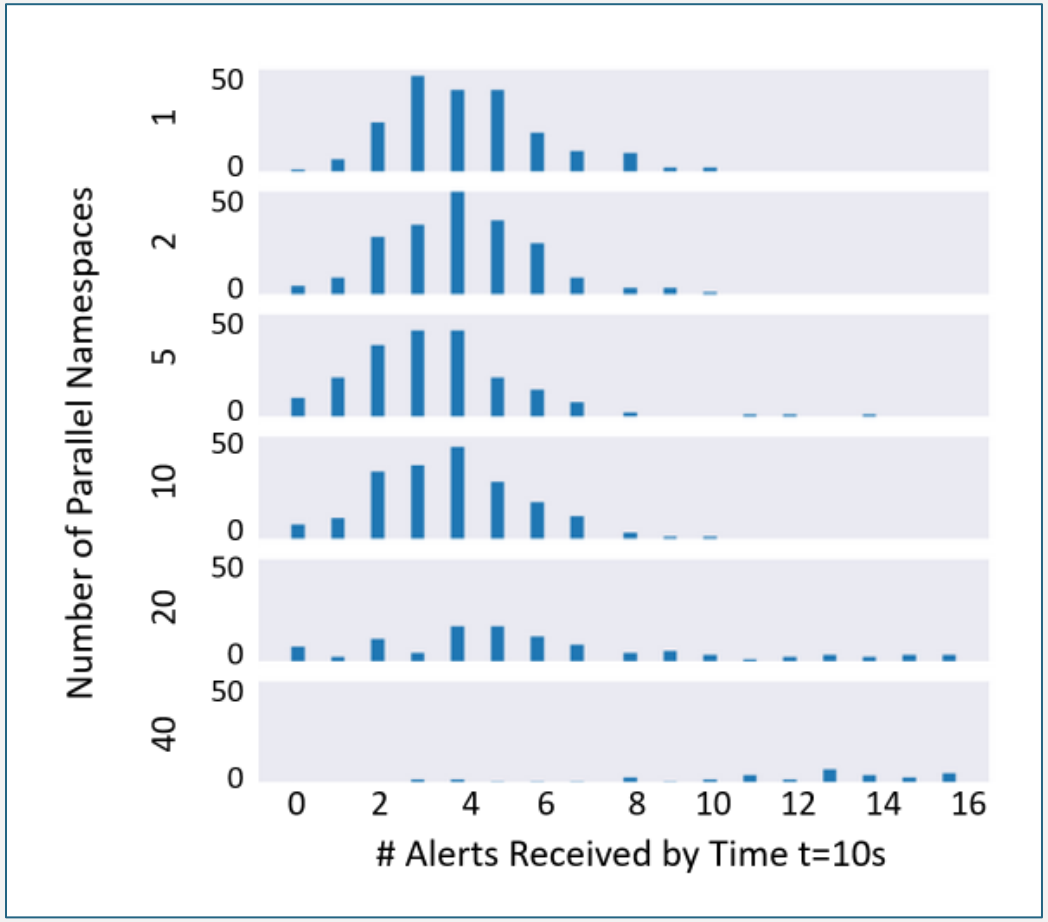


Results – Scenario 2 (Command and Control)

Example Metrics:

- Stolen Cycles ≤ 1
- Load ≤ 14 Processes
- Interrupts ≤ 2250/s

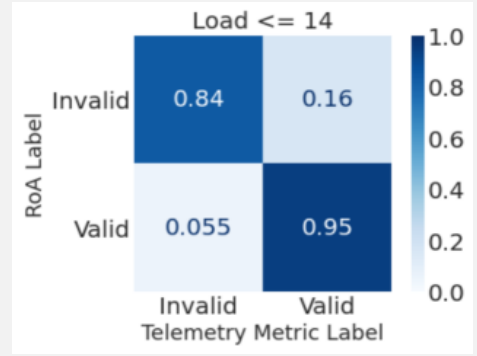
All replicates

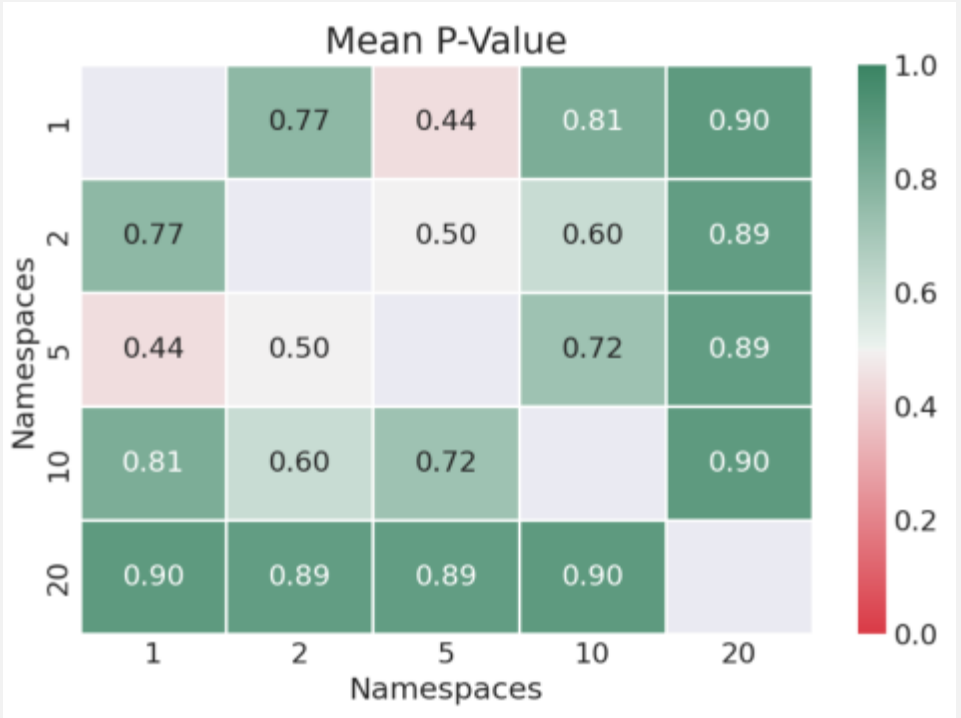


Results – Scenario 2 (Command and Control)

Tukey - All replicates

Tukey – Load ≤ 14 Processes





Outcome

Verification helps ensure cyber experiment results can be used to accurately understand real cyber systems

Failure to reproduce cyber experiment results could be due to emulation environment rather than faulty experiment design – the **emulation environment should be verified**

This work successfully demonstrates a generalizable process for resource verification


Thank You!