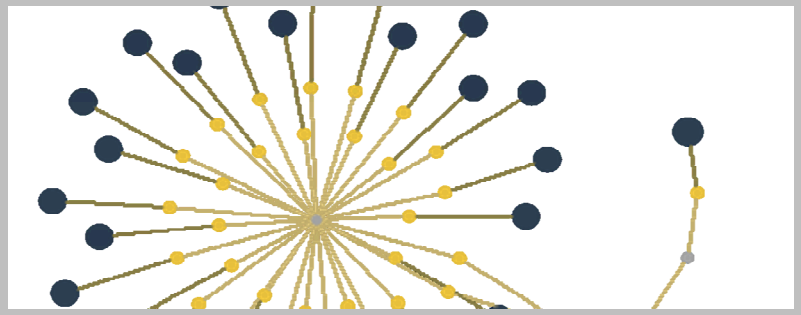
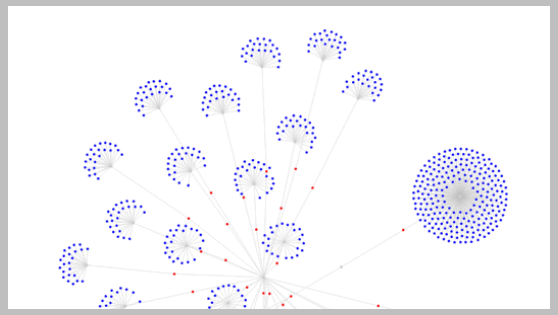
SAND2017-9164C





Emulytics and IoT at Sandia

David Fritz

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

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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.

The idea

- We're working to boot millions of IoT devices
- Coupled with millions of IT-infrastructure endpoints
- We want to study emerging threats at-scale
  - And scale seems to be "huge"

MIRAI botnet

- 600,000 IoT devices
  - 200,000-300,000 steady state population
- 65,000 in the first *20 hours* of infection spread
- 623Gbps at peak!
- Used for over 15,000 attacks
  - Krebs, DynDNS, …

Why model IoT at-scale?

Sandia National Laboratories

- IoT devices make great ~~botnets~~ peer-to-peer, decentralized networks
- We don't know what
  - DDoS
  - Botnets
- Look like at-scale
- Is my house on fire or is London on fire?

Emergent *mis*behavior

Sandia National Laboratories

- Autonomous vehicles
- P2P-IoT
  - Samsung and IBM – Autonomous Decentralized Peer-to-Peer Telemetry (ADEPT)
    - Overlay network based on a DHT (probably a variant of kademlia)
  - Filament
    - Uses a SHA-256 based DHT
- Lots of companies do under various terminology
  - Device-to-device
  - Machine-to-machine
  - Peer-to-peer

Emergent *mis*behavior

Sandia National Laboratories

- FOSCAM IP Cameras
  - Had an always-on P2P mode
    - With a non-functional option to disable it
  - Video stream is shared with Internet hosts
- FOSCAM uses ThroughTek's IOTC P2P platform
  - Also Maginon, Swann, QNAP (NAS!), and others
- ThroughTek claims to have 100 million IoT connections!

Well I don’t use IP cameras

- Maybe not, but you do occasionally get sick!
  - May 2017 – NHS hospitals crippled by ransomware
  - Apr 2017 – Erie Country Medical Center Level 1 trauma systems shut down for 6 weeks by ransomware
  - Nov 2016 – NHS Lincolnshire and Goole had to cancel surgeries and divert trauma patients due to malware
  - Mar 2016 – Methodist Hospital in Henderson KY shut down by ransomware
  - Feb 2016 – Hollywood Presbyterian shut down by malware

Healthcare

Sandia National Laboratories

- Hugely benefited by connected devices
- Deployment is moving at breakneck speed
  - Despite the amount of regulation and liability
- Ransomware that encrypts your hard drive is annoying, possibly causes financial loss
- Ransomware that threatens to administer a lethal dose of insulin on your Internet connected insulin pump (which exists!) is *terrifying*
- **A typical hospital bed has 5-15 connected devices**
- **A big hospital system has 5000 beds**

How is this still a problem?

Sandia National Laboratories

- IoT is too hard
  - Or engineers/programmers are too lazy…
- Surely standards will fix this
  - There are already LOTS of standards!
  - And everyone expects their standard to be supported
  - The degree of interoperability expectations render standards meaningless
- There's no market solution either
  - Because you don't care
  - Because the manufacturer doesn't care
  - Because government no longer has "purchasing power"

So what do we do?

Sandia National Laboratories

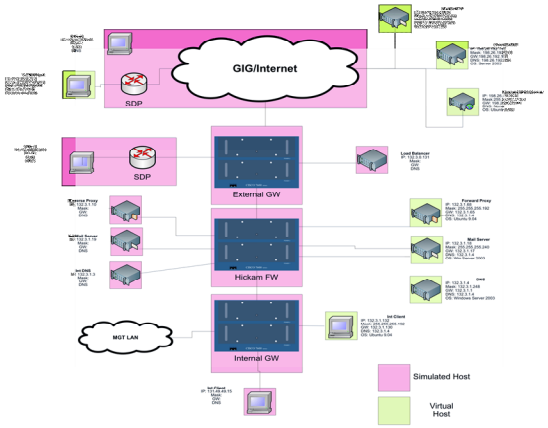
- IoT security research is a hot topic
- Lots of research on individual devices
  - Also lots of difficulties
  - Obtaining firmware is outrageously difficult
  - And possibly illegal (DMCA)
- Very little work on IoT cyber-security at-scale
- Enter Emulytics

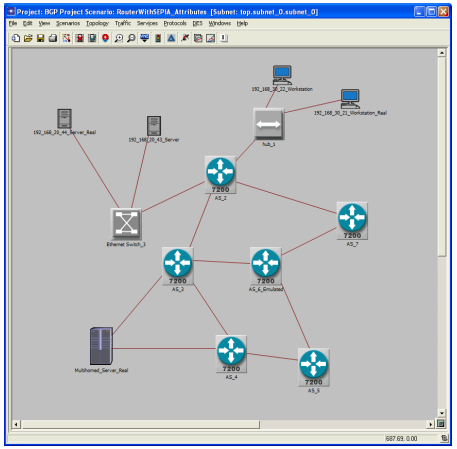
Sandia Emulytics

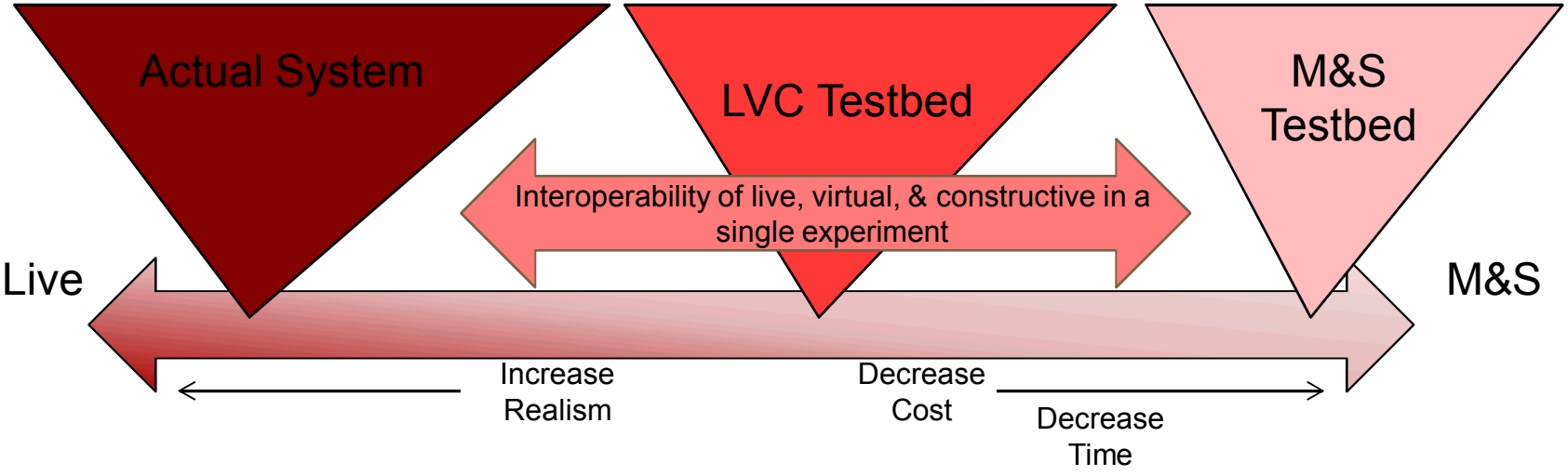
Sandia National Laboratories

- Emulytics = Emulation + Analytics
  - the art and science of modeling, simulating, instrumenting, and analyzing variable-scale networks that have significant dependencies upon networked systems
- How did it come about?
  - Outgrowth of Live, Virtual, Constructive concepts and traditional Modeling and Simulation
  - Confluence of HPC and cyber-security communities at SNL
- Our focus is on one platform - minimega

Live, Virtual, Constructive







Why bother with emulation?

Sandia National Laboratories

- The bad guys just use the Internet
- MIRAI was scaled up on the Internet
- The Internet has thousands (of millions) of nodes
  - That's the supercomputer
- As a legal entity, we would have some trouble with this approach…

minimega

Sandia National Laboratories

- Open source, publicly available
- Launch and manage VM-based experiments
- Setup complex virtual network topologies
- Integrate real hardware with virtual experiments
- Fast!
- Repeatable
- One tool in the toolset; the substrate for a number of programs

Core capabilities

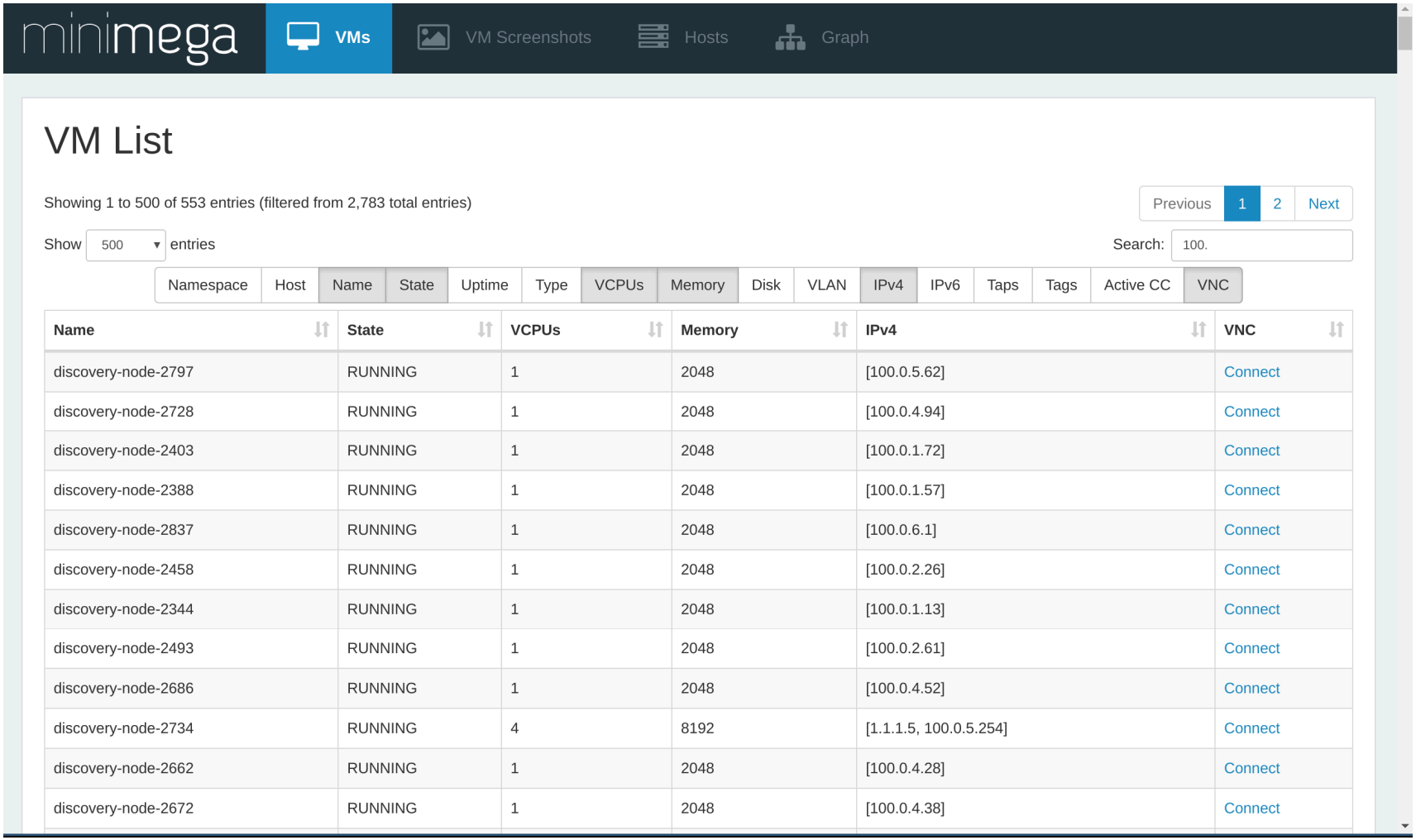
Sandia National Laboratories

- Programmable API
- Scripted/automated environment
- Fast
- Density/Scale
- KVM + container
- Cyber physical
- SCADA
- Command and control
  - Including networkless
- VNC record/replay
- Mobile
- Map-to-model
- Data aggregation
- Authentic layer 2

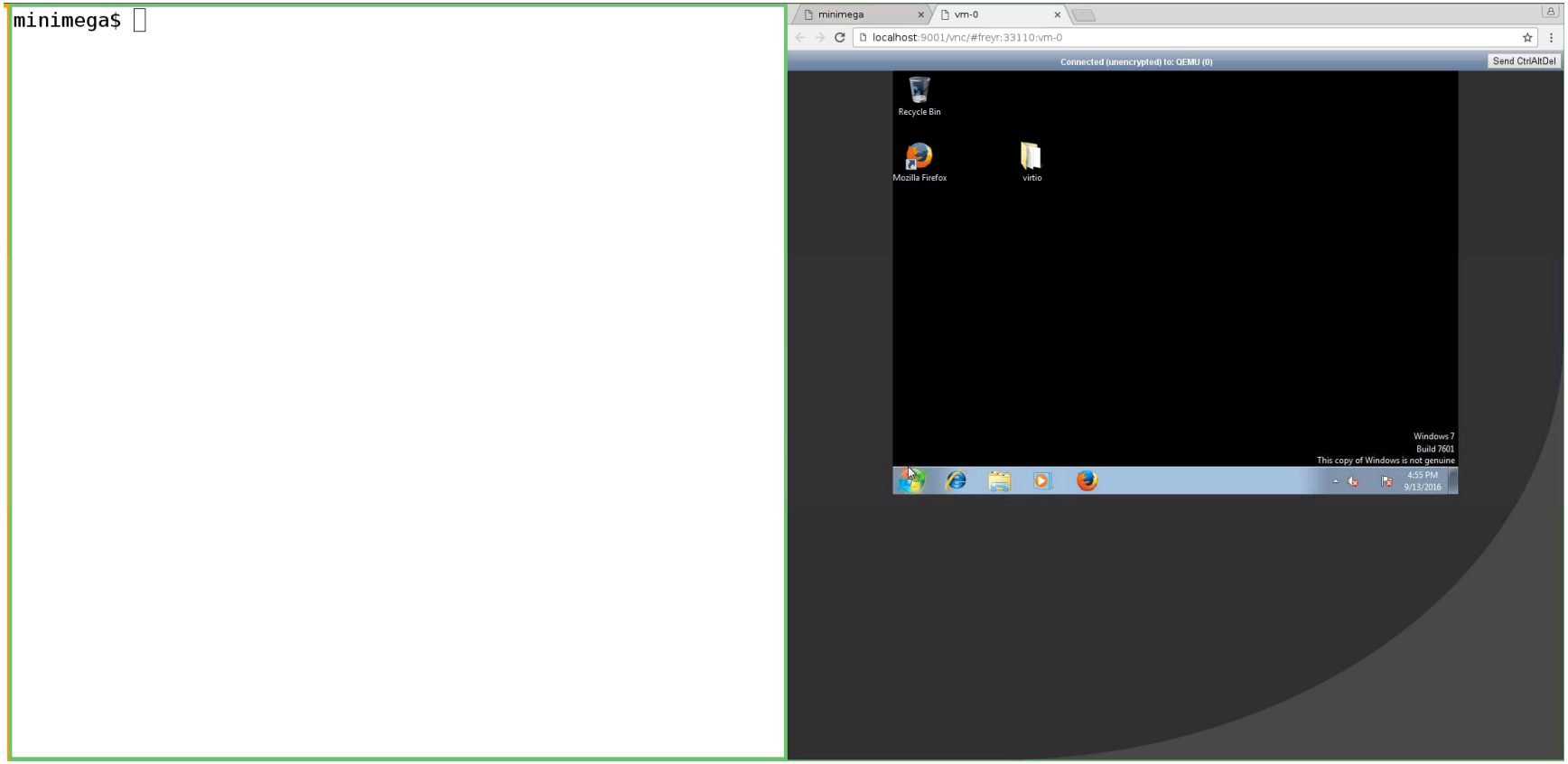
minimega scripting

```
# configure and launch some VMs
vm config disk foo.qc2
vm config net network-A network-B
vm launch kvm foo[1-1000]
vm config filesystem foo/
vm launch container 50000

# record user interactions on foo5
vnc record foo5 /tmp/foo.vnc
```


minimega
VMs
VM Screenshots
Hosts
Graph
VM List
Showing 1 to 500 of 553 entries (filtered from 2,783 total entries)
Previous 1 2 Next
Show 500 entries
Search: 100.
Namespace Host Name State Uptime Type VCPUs Memory Disk VLAN IPv4 IPv6 Taps Tags Active CC VNC
Name State VCPUs Memory IPv4 VNC
discovery-node-2797 RUNNING 1 2048 [100.0.5.62] Connect
discovery-node-2728 RUNNING 1 2048 [100.0.4.94] Connect
discovery-node-2403 RUNNING 1 2048 [100.0.1.72] Connect
discovery-node-2388 RUNNING 1 2048 [100.0.1.57] Connect
discovery-node-2837 RUNNING 1 2048 [100.0.6.1] Connect
discovery-node-2458 RUNNING 1 2048 [100.0.2.26] Connect
discovery-node-2344 RUNNING 1 2048 [100.0.1.13] Connect
discovery-node-2493 RUNNING 1 2048 [100.0.2.61] Connect
discovery-node-2686 RUNNING 1 2048 [100.0.4.52] Connect
discovery-node-2734 RUNNING 4 8192 [1.1.1.5, 100.0.5.254] Connect
discovery-node-2662 RUNNING 1 2048 [100.0.4.28] Connect
discovery-node-2672 RUNNING 1 2048 [100.0.4.38] Connect

Interacting with VMs



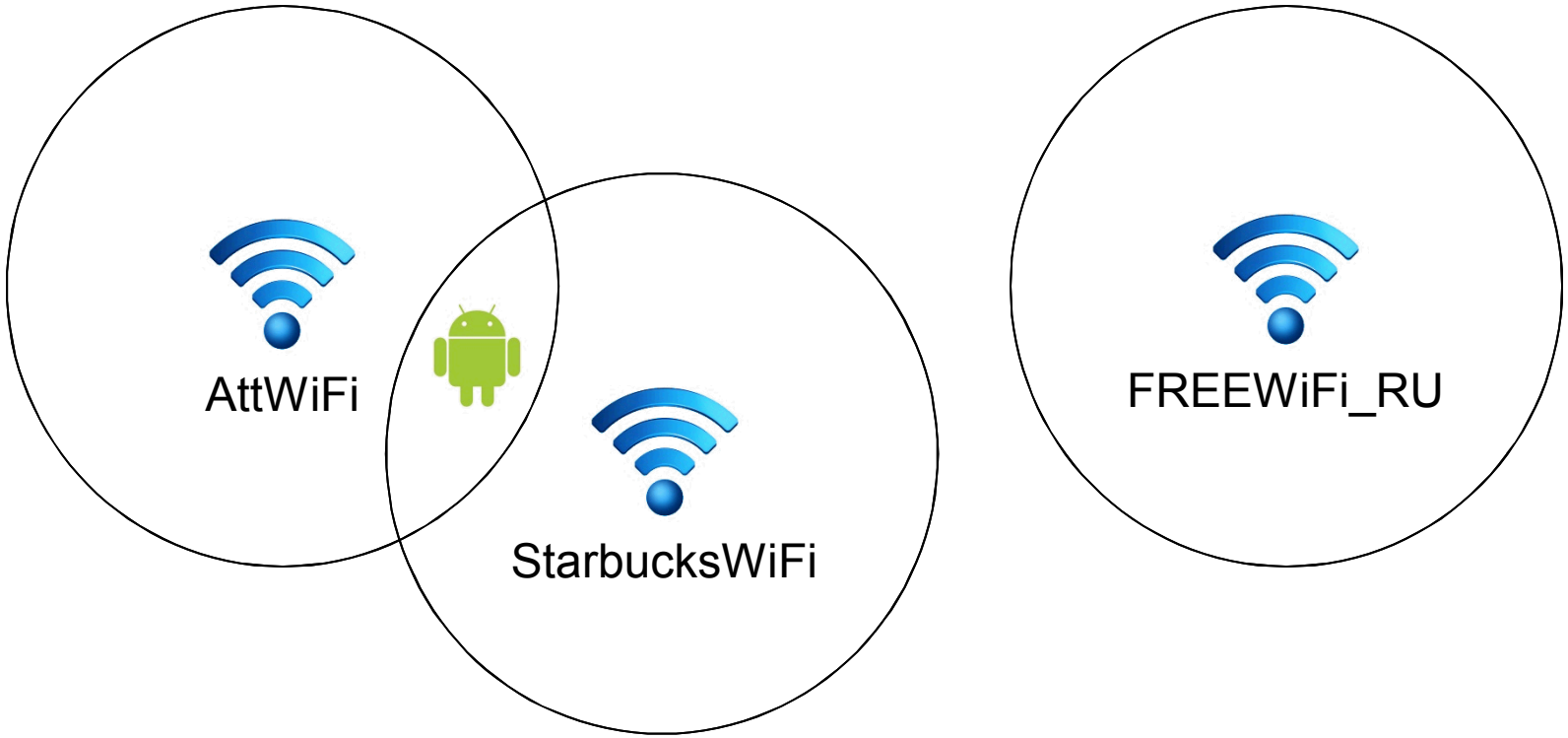
Android mobile support

Sandia National Laboratories

- Boot Android VMs
- Replace radios/sensors with shims to the host
  - GSM
  - GPS
  - 802.11
  - Accelerometer
- Provide the ability to inject data, attach simulators
- Provides simple way to blend IT and non-IT artifacts
  - 802.11 access point appear when “in range”, given by GPS

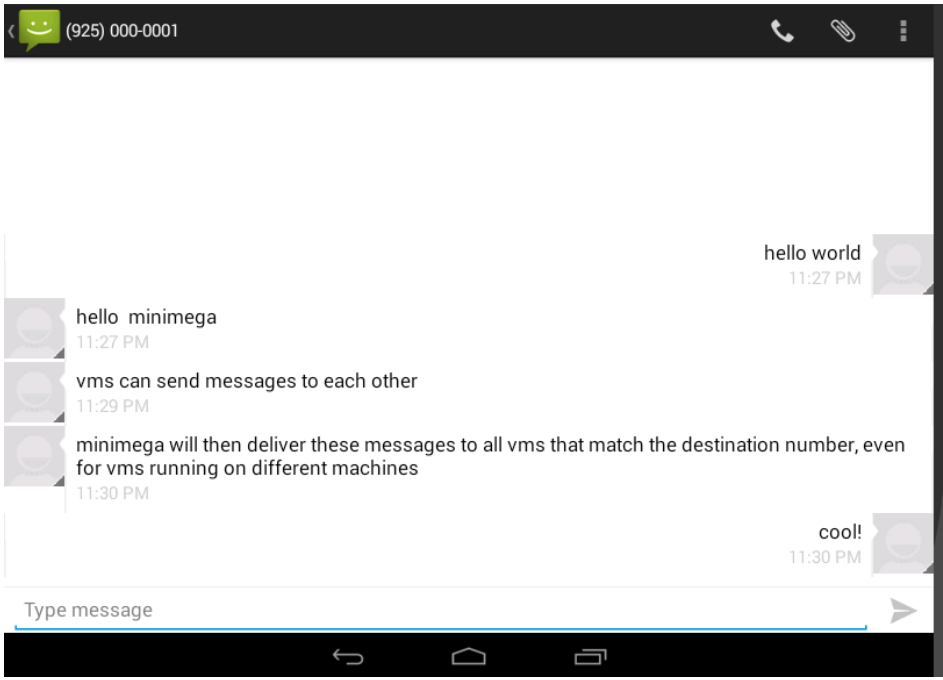
Mobile WiFi

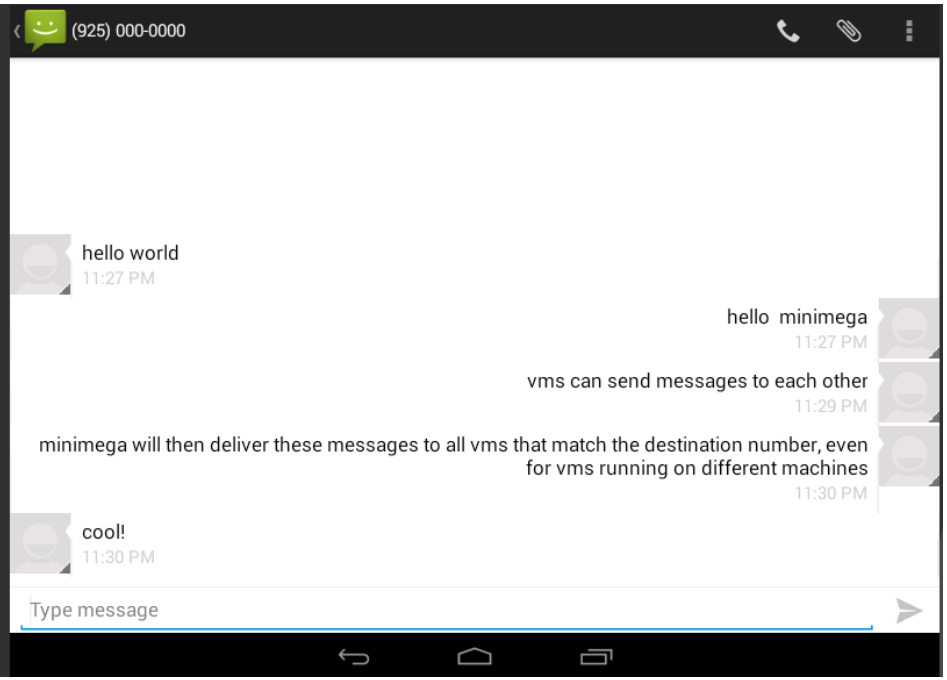
- Create WiFi AP with location, range, and VLAN
- Connect VM to APs in range, based on GPS location
  - Use minimega to update connected VLAN to that of connected AP



Mobile SMS

- Ability to inject, introspect SMS from the host
- Devices can text each other directly

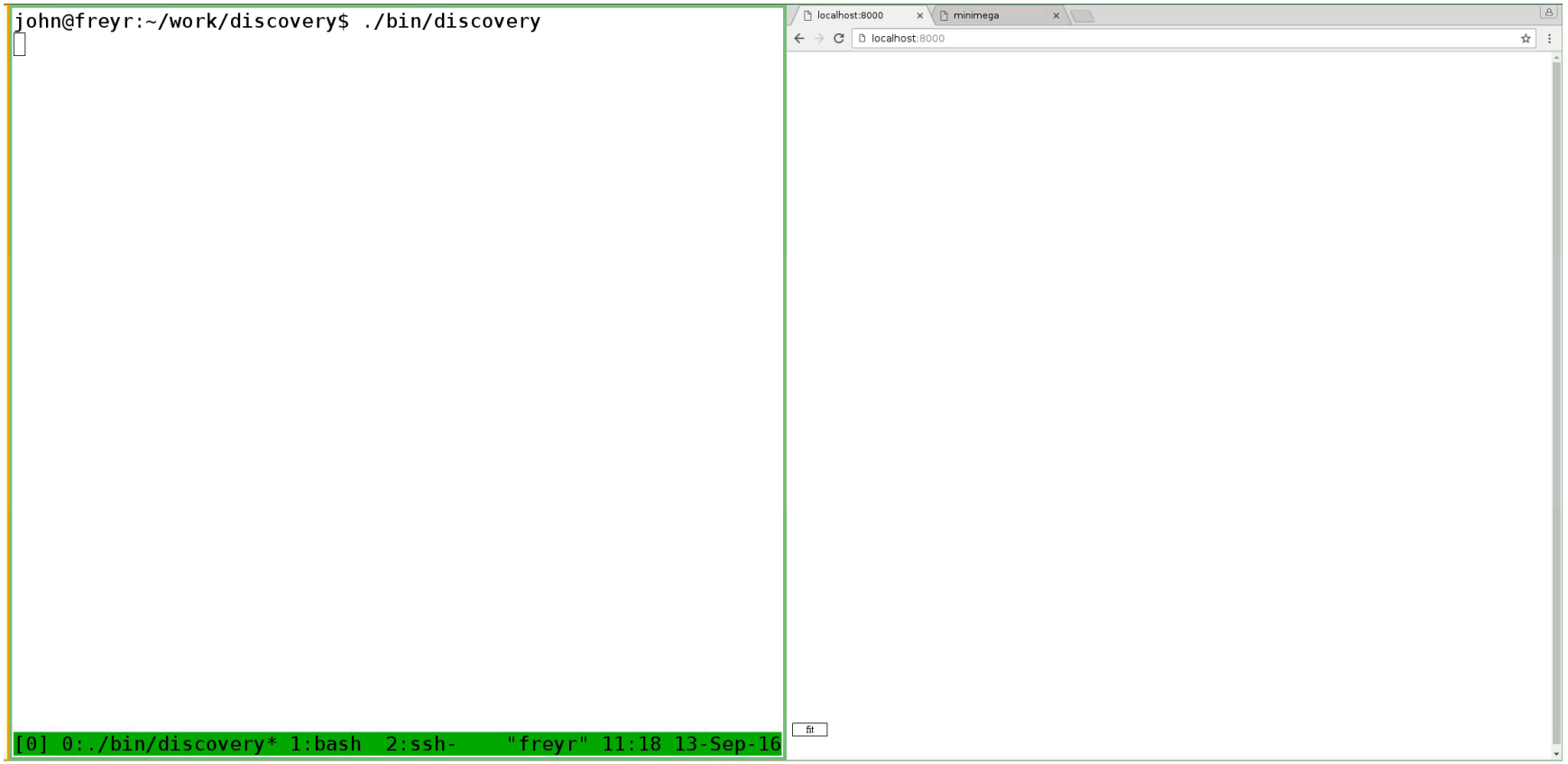




Automated map-to-model

- Ability to ingest a variety of data sources
  - PCAP
  - Active scans
  - Router configs
  - …
- Builds structural/behavioral graph representations
- Researcher can “put the magnifying glass” wherever needed

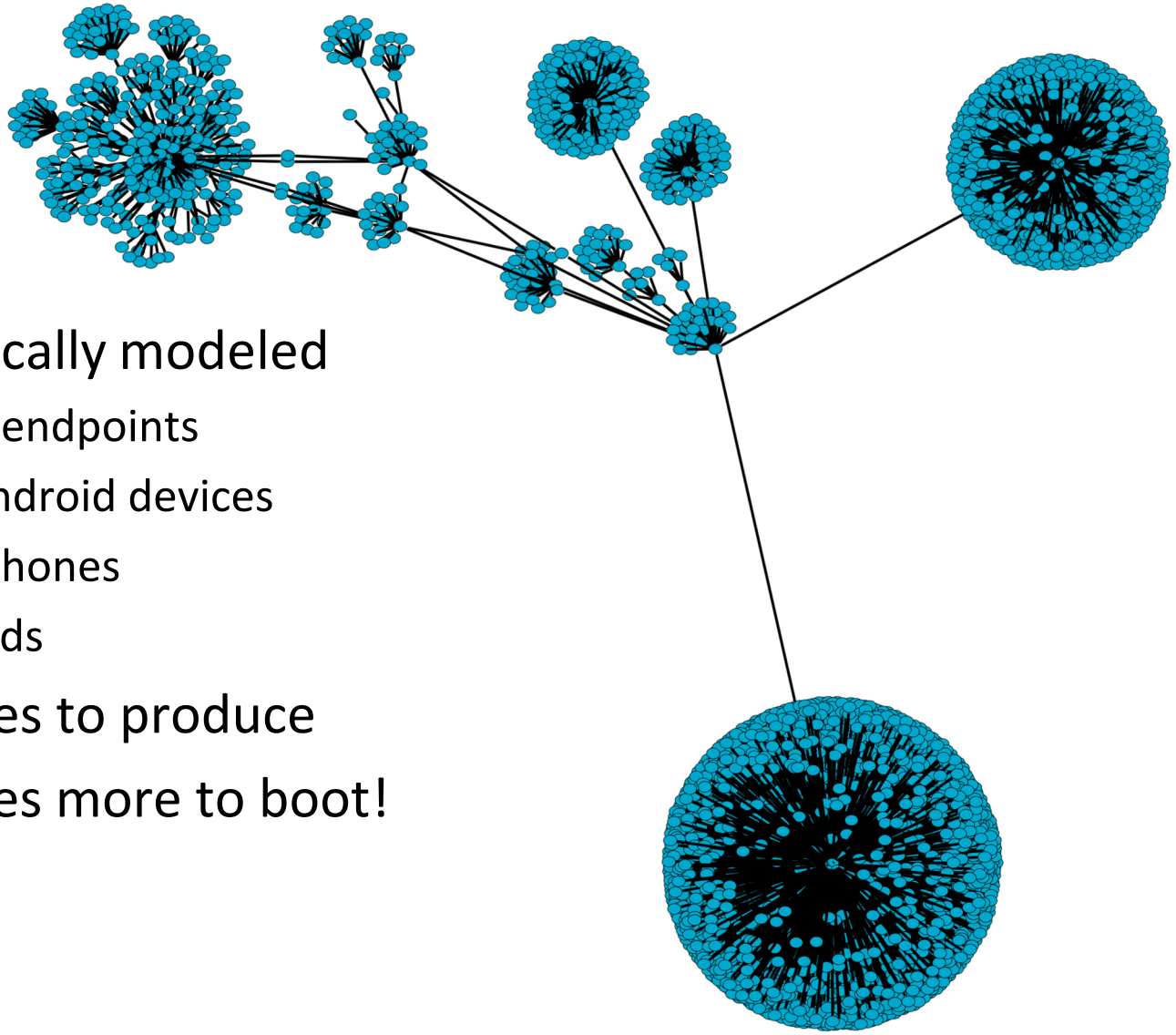
Automated map-to-model



Scale

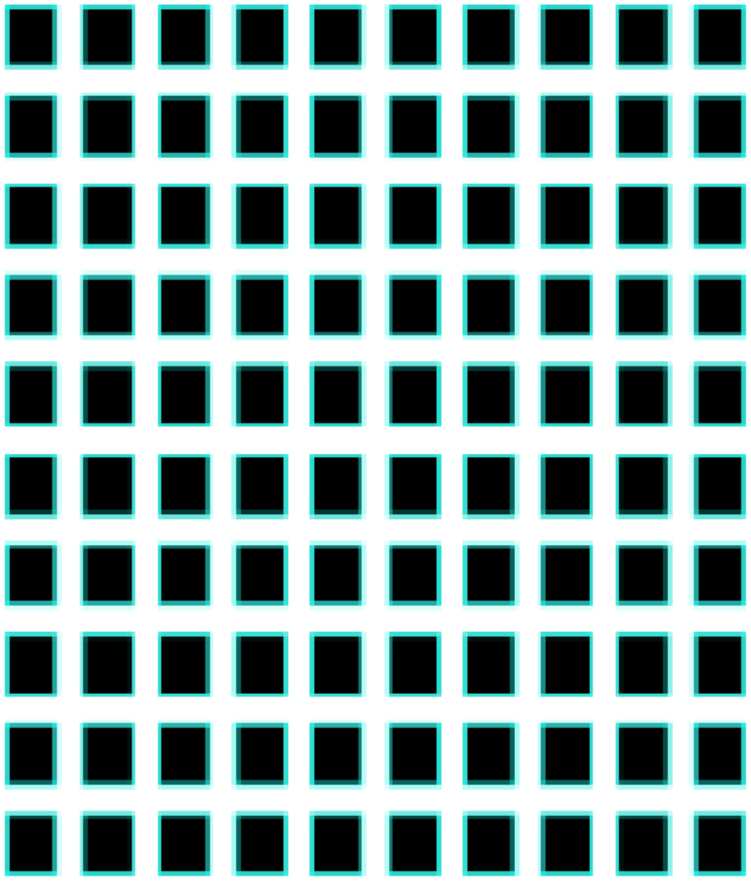
- Example: SC16 SCinet
- High capacity network
  - 9x 100Gbps links
- Exists for 2 weeks a year
- Supports SC conference
  - 10k attendees

SCinet 16

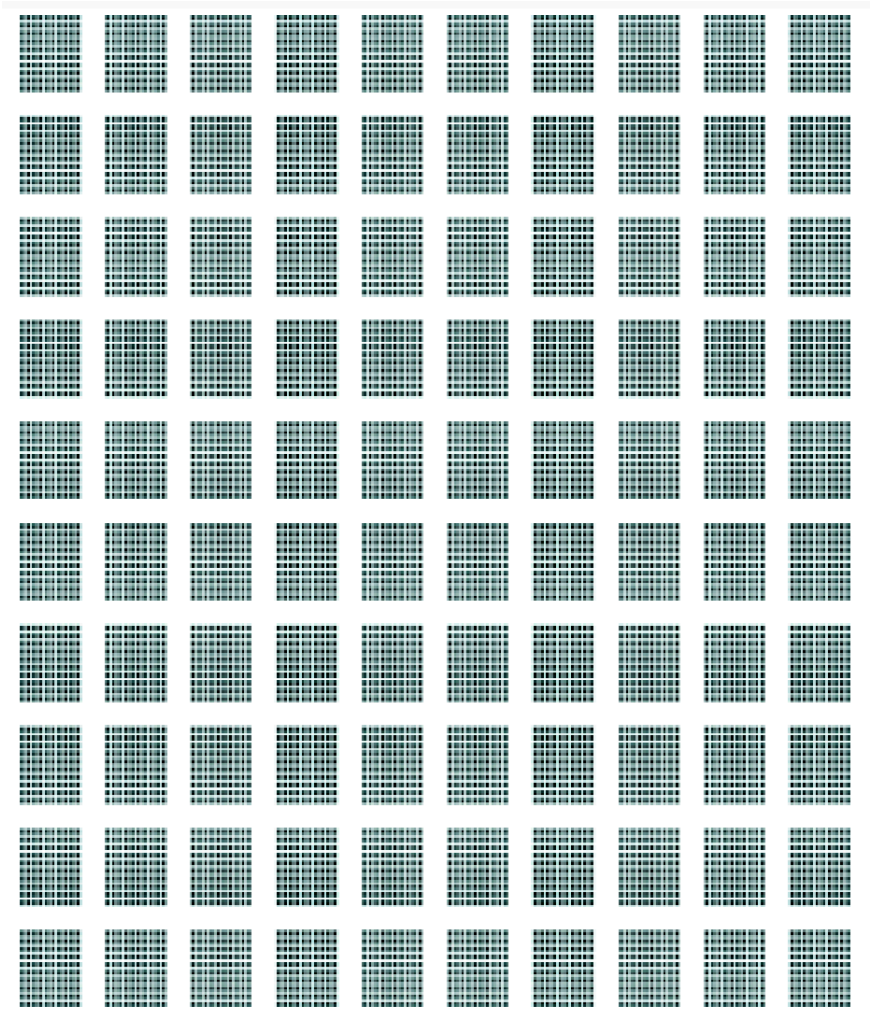


- Automatically modeled
  - 10,000 endpoints
  - 2099 android devices
  - 1933 iphones
  - 485 ipads
- ~2 minutes to produce
- ~2 minutes more to boot!

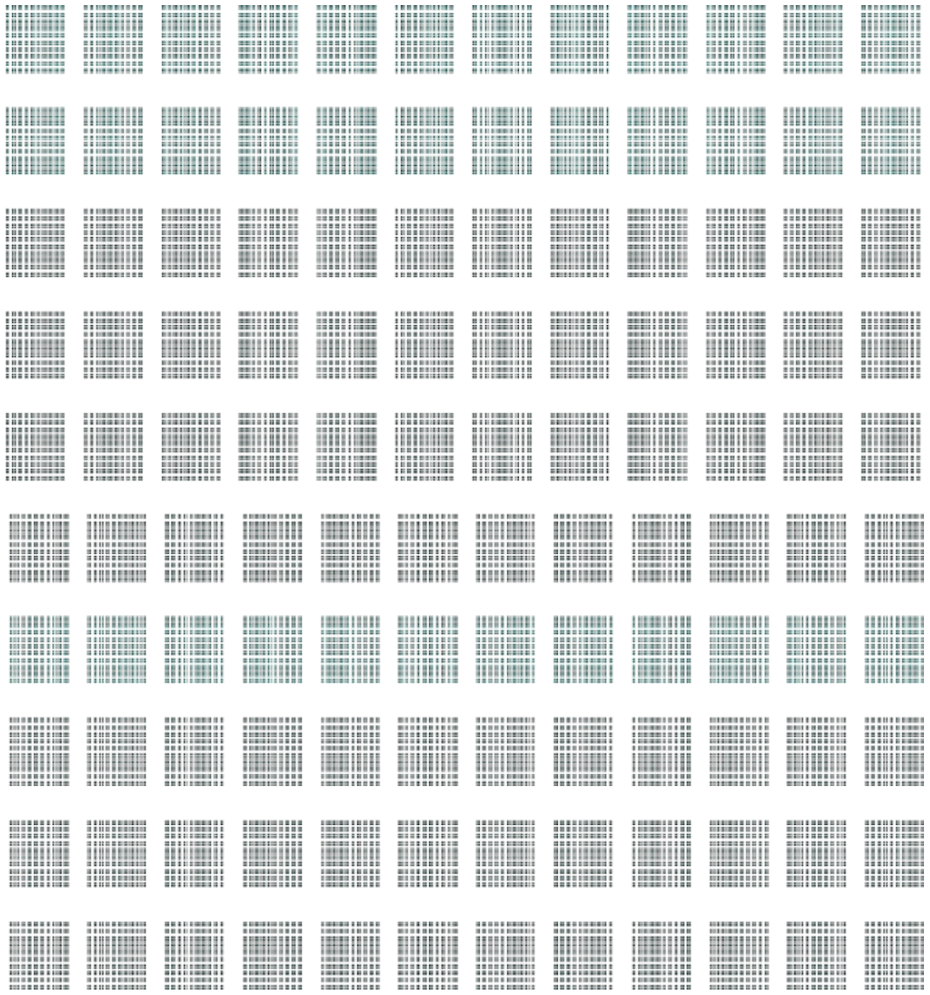
A note on scale: 100 nodes



10,000 nodes (supercomputer)

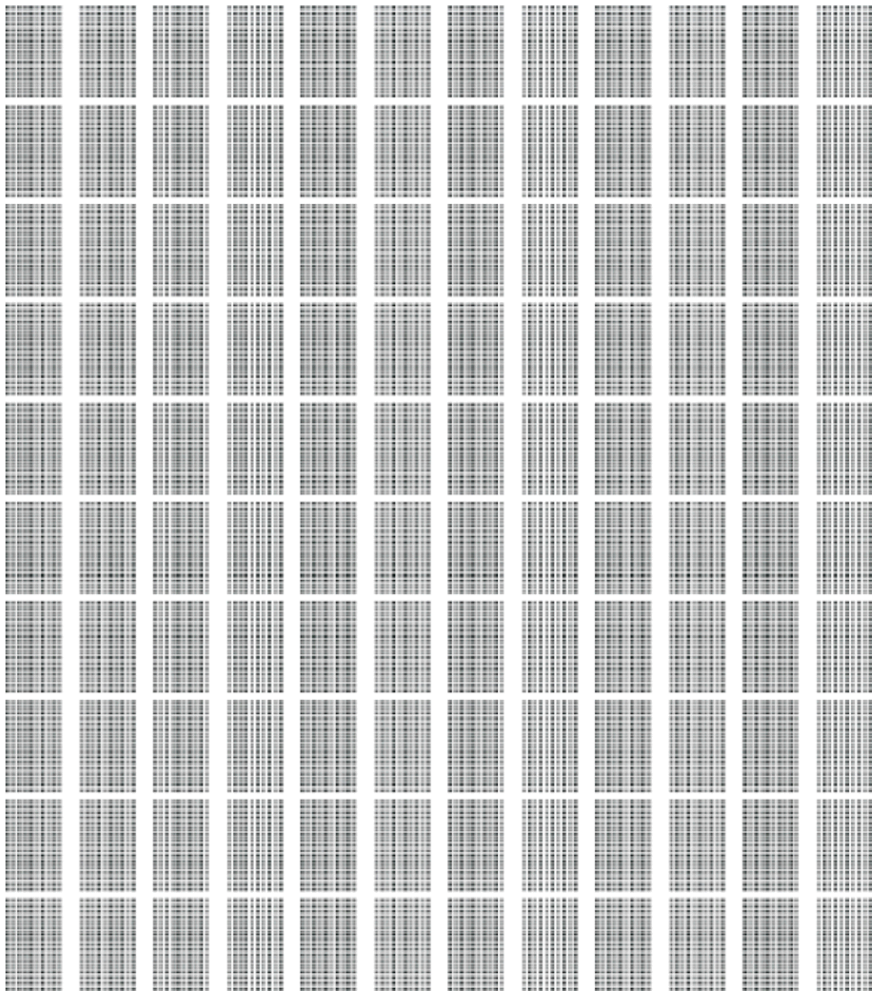


1M nodes (we've run out of pixels)



10M nodes (diffraction pattern)

Sandia National Laboratories



At note on scale

- Cannot have global knowledge
- Even configuration has to be figured out computationally
- Most tools are designed for a small world - we live in a large world
  - At 10M nodes a DHCP file is 350MB
- Simple monitoring
  - 1 bit at 1Hz
  - 1.2Mbytes/sec
  - And we want more than 1 bit!
- Remember that 256-bit hash?
  - Node “distance” is a function of the hash
  - Geography isn’t the only measure of the structure of the network
  - You need to populate the space or you just get fragments
  - Only choice is to make the scale “huge”

Using minimega for IoT experiments

Sandia National Laboratories

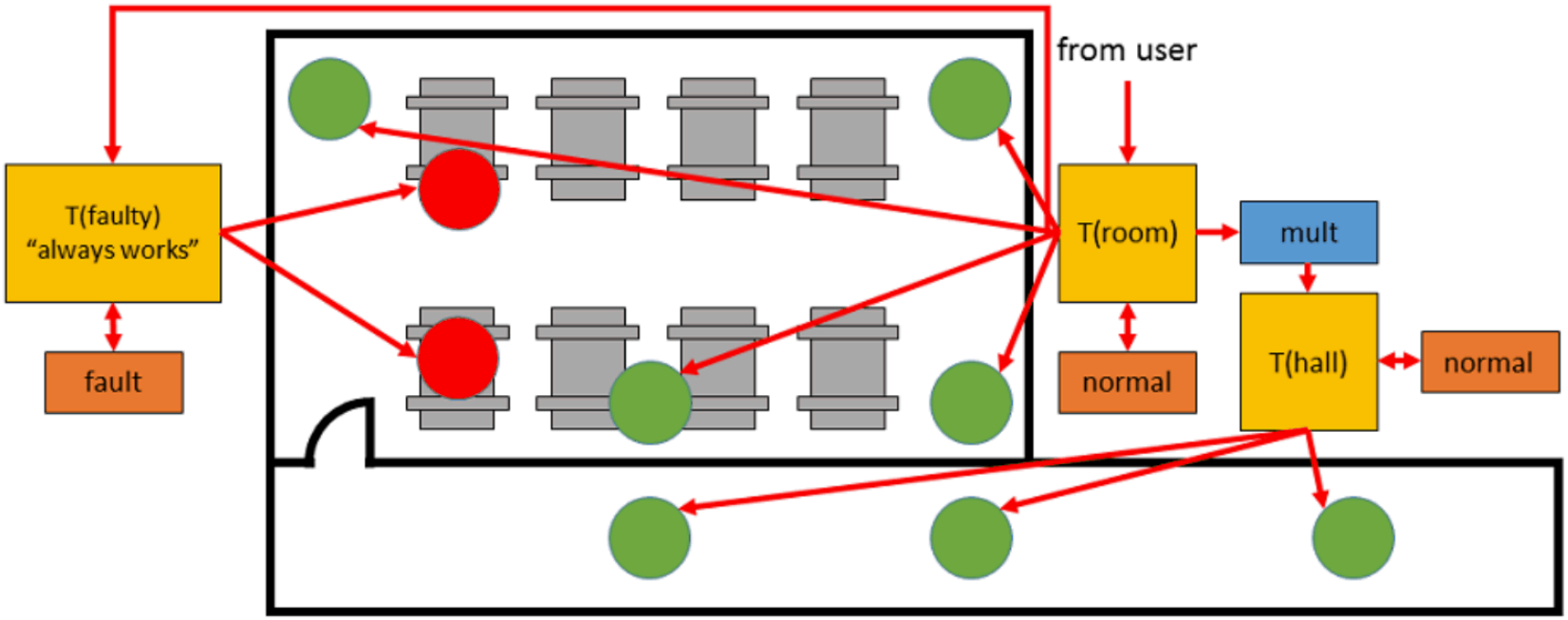
- Bruce Schneier - "Give the Internet hands and feet, and it will have the ability to punch and kick.“
- How do we plumb non-network based connectivity?
  - Buttons
  - Serial connections
  - Vehicle radars
  - Sound, light, heat…
  - Cyber-physical interactions of all kinds
- How do we support the variety of IoT devices on the market?

miniplumber

Sandia National Laboratories

- A networkless, out-of-band, interprocess communication layer
- Quick specification of communication pathways (pipelines)
- Uni- and multi-cast
- Scales with minimega
- Borrows concepts from the Plan 9 Plumber
- Works on host, in minimega, and on any guest endpoint
- Allows bridging simulation layers to emulation layers to real hardware
  - Across your cluster

miniplumber



Using minimega with SCADA

Sandia National Laboratories

- Sandia SCEPTRE
- ICS modeling and simulation with bindings to minimega/IT systems

SCEPTRE Components

Sandia National Laboratories

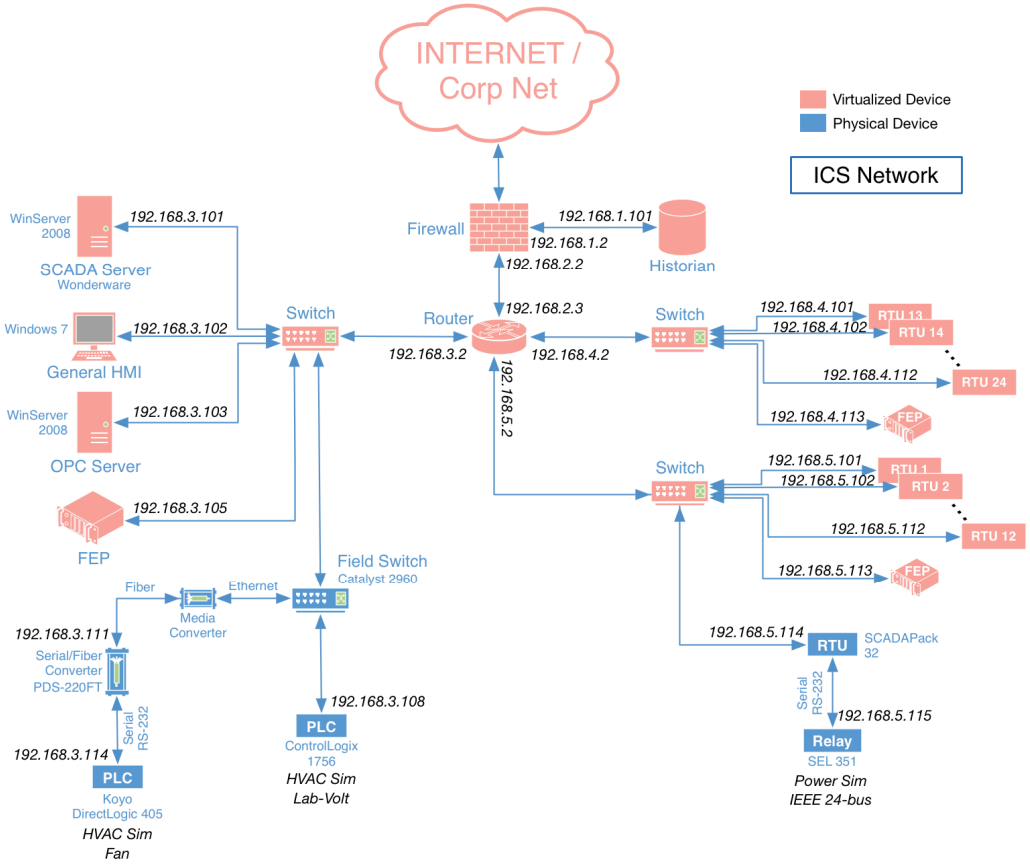
- Control Systems devices
  - Low fidelity simulated ICS devices
    - RTUs, PLCs, protection relays, FEPs
  - Emulated PLCs, HMI services
  - Hardware-in-the-loop (HITL) devices such as relays, PLCs, RTUs
- High fidelity SCADA protocols
  - ModbusTCP, DNP3, IEC 61850 and 60870
  - Written to specification
  - Enabling technology that allows communication between HITL and simulated devices
- Process simulation
  - Leverage industry standard software where possible (V&V)
    - PowerWorld, PyPower, PSS/E
  - Develop our own simulated process when needed
    - Water treatment, refinery, natural gas pipeline, railroad signaling

SCEPTRE Operational Overview

Sandia National Laboratories

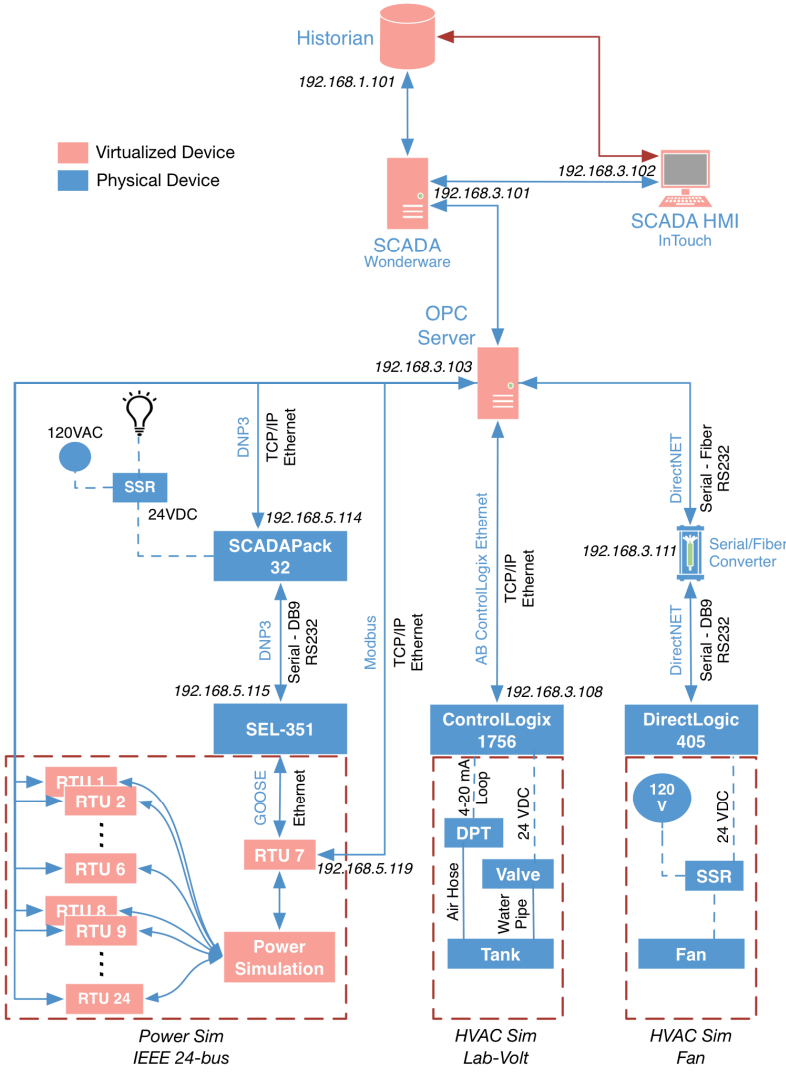
- SCEPTRE is an application that uses an underlying network (like Sandia's Emulytics™ Platform technology) to run
- ICS devices (simulated, emulated, real) communicate and interact via high fidelity SCADA protocols
- Process simulation data is provided to all the ICS devices
- All ICS devices are able to interact with the simulation, providing both updates and subscribing to the current state of the simulation
- When the simulation state updates, all devices receive the current state so there is a common view of the simulation
- Overall simulation is able to bridge multiple infrastructures into the same experiment to show interdependencies.
- Real-time vs discrete event simulations

Example ICS Environment

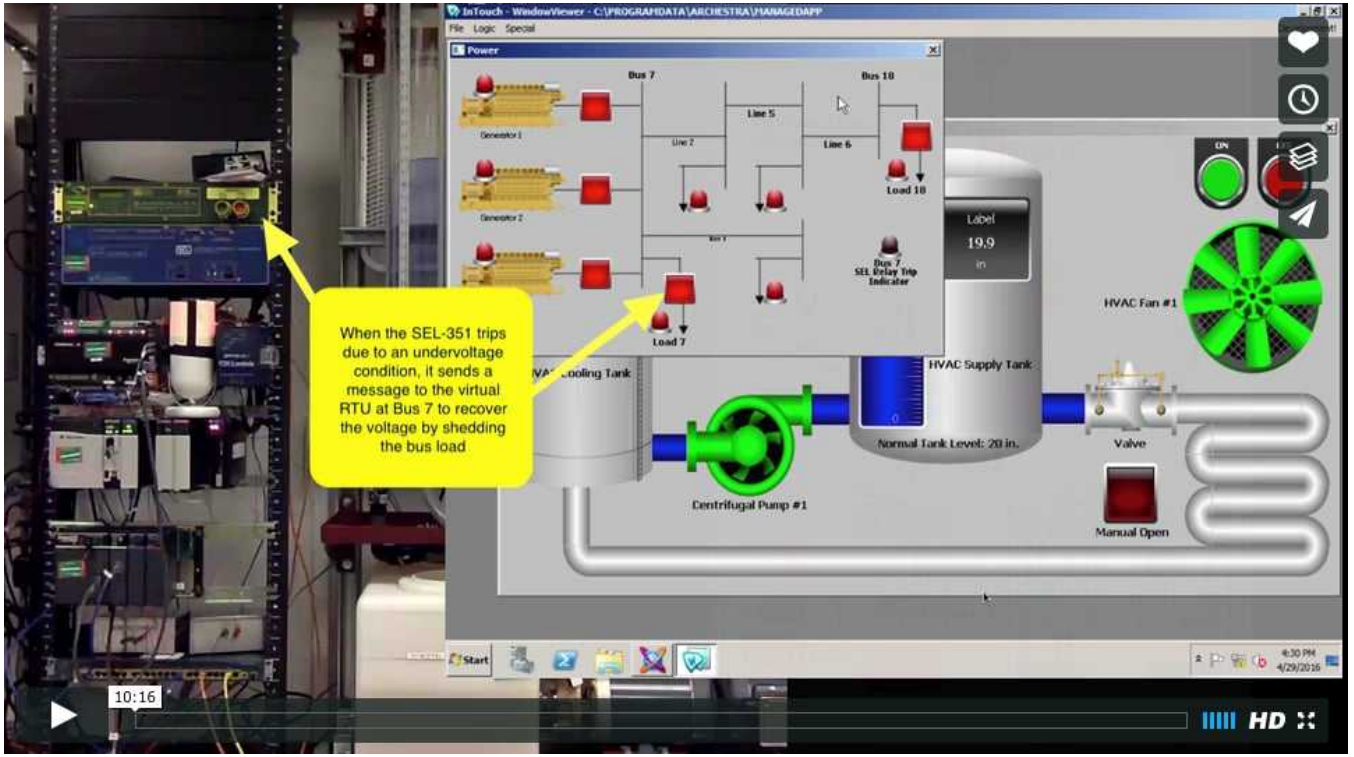


ICS Network Diagram

ICS Logical Diagram



SCEPTRE Demonstration



https://vimeo.com/178492617

Lessons Learned

Sandia National Laboratories

- Scale is bigger than you think
- It’s not surprising that everything breaks
- We’ve upset every applecart
- Analytics at-scale isn’t as mature as you may think
- IoT is a big space


Homeland security
Buildings
Energy
Home
Healthcare
Transportation
Public safety
Industrial
Air pollution
Environmental
Industrial control
Emergency comms
Mass transit
Structural
Perimeter access control
Traffic congestion
Smart meters
Alarm systems
Digital signage
Building automation
Smartphone detection
Radiation detection
CCTV
Lighting control
Health monitoring
Water quality

But it's all important!

- It's the emergent properties of interconnecting everything that is interesting
- If you focus on a sector, you miss the big picture
- It used to be a washing machine with a computer in it
- Now it's a computer with a washing machine attached to it
  - Or a swarm of vehicles
  - And they're connected to the Internet

Thank you

David Fritz, PhD

Sandia National Laboratories

djfritz@sandia.gov

minimega.org