This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

SAND2020-3631C

Dispelling myths of red/blue cyber competition through metrics

U.S. DEPARTMENT OF ENERGY'S
CYBERFORCE™
COMPETITION
DEFENDING U.S. ENERGY INFRASTRUCTURE

MYTHBUSTERS

Kandy Phan
March 2020
Sandia National Labs

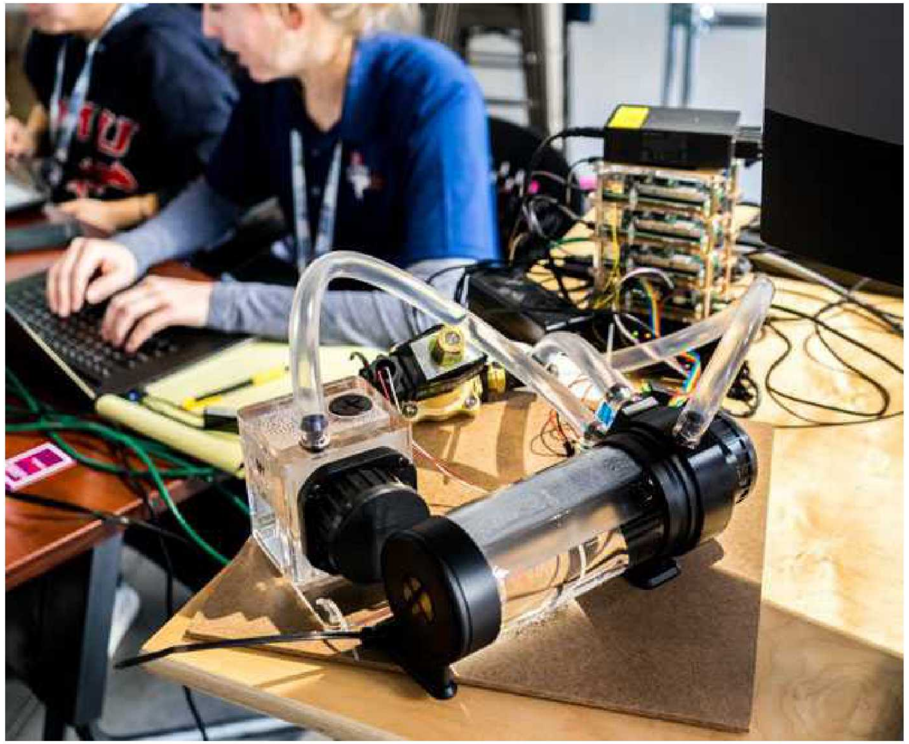
U.S. DEPARTMENT OF ENERGY NNSA

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.

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.

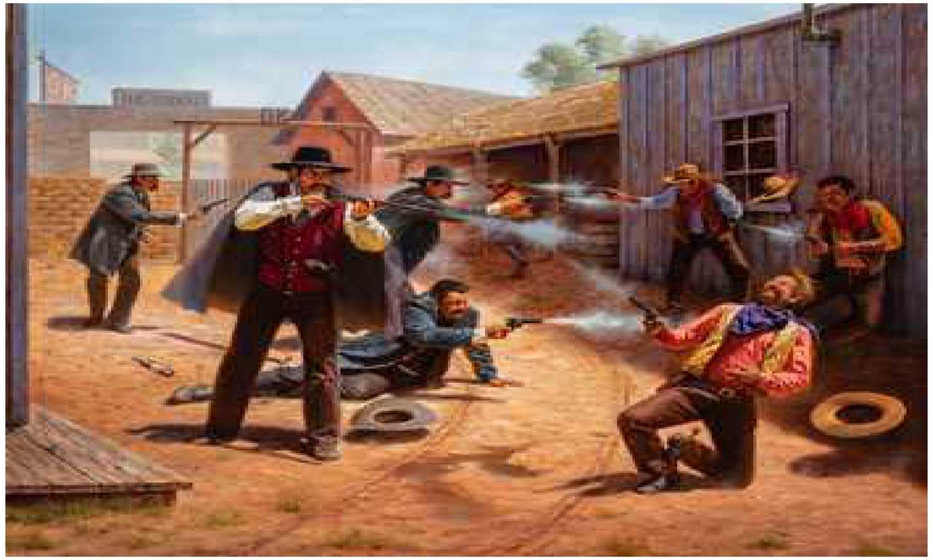
Outline

- DOE CyberForce Competition
    - Started in 2016
    - Headed by Argonne National labs
- College competition for cyber skills with an ICS flavor
- The red team portion:
- How we evolved
- Where we need to get better



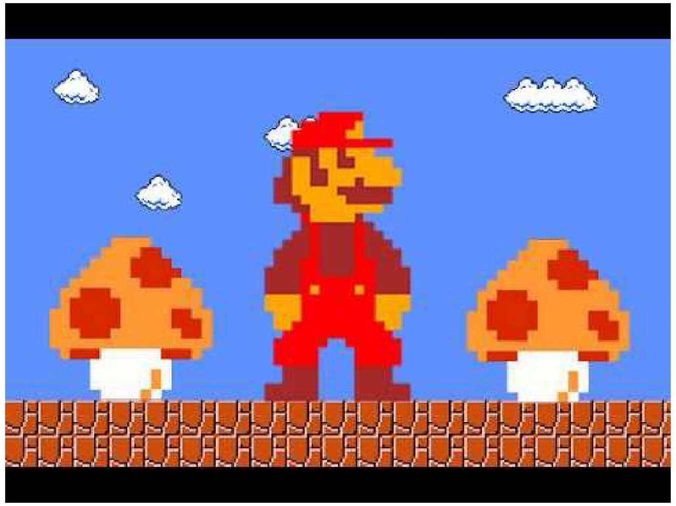
Pre-2018

- Throw in together a red and a blue team
  - Success, right?
- Wild west, adversarial but not necessarily in the good way
- More of a pentest-flavor instead of real red teaming
- Red focused on hackable teams:
  - Beating a dead horse



December 2018

- Pre-seed vulnerabilities
- Service uptime checks
- Mixed in some Active Directory action
- Getting closer to red teaming/threat emulation



November 2019

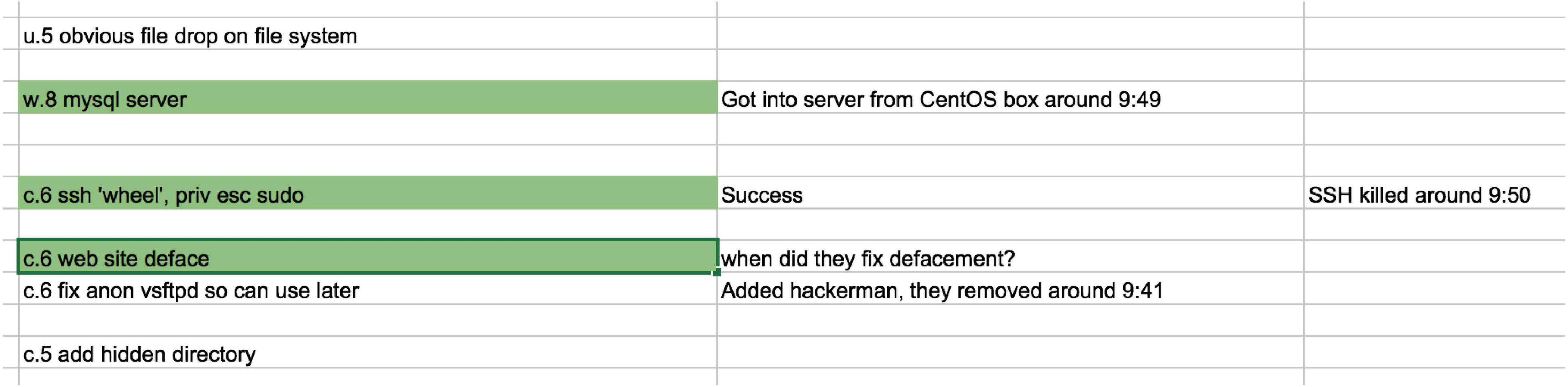
- Pre-seed vulnerabilities
- More prep-time (shared repo of exploits)
- Red teams with a more of professional approach

Sandia site

- Our goal:
  - “a fair red team experience for the defenders to experience representative tradecraft”
- Automation – standardized laptops, Ansible scripts
- Scripted exploits for all of the scored vulnerabilities
- Force our red teams to collect metrics
- Focus on measuring blue team capabilities/responses
  - Instead of “beating them”, evaluate them

Collecting metrics

- Helps us to better understand what works
- Gain insights into blue abilities
- Connections between gaps



u.5 obvious file drop on file system		
w.8 mysql server	Got into server from CentOS box around 9:49	
c.6 ssh 'wheel', priv esc sudo	Success	SSH killed around 9:50
c.6 web site deface	when did they fix defacement?	
c.6 fix anon vsftpd so can use later	Added hackerman, they removed around 9:41	
c.5 add hidden directory		

Metrics

- 22 pre-seeded vulnerabilities
  - 3 of them flat out didn’t work
  - 2 of them required fixing

- Categories:
  - direct shell - 10,
  - indirect shell - 4,
  - PrivEsc - 4,
  - Info - 1, database access - 1, readable/writeable - 2

- On average, 4 of them worked

Metrics

- 20% of the teams are unhackable

- 60% had 3 or fewer issues

Metrics – context matters

- “80% of the teams have been hacked!”
  - On the surface that sounds good

- It also means that 20% of the teams remained untouched

- Does not capture the extent of the “hack”:
  - Just info disclosure?
  - If shell access - how long did it last? 30 min., 10 min., 1 min.?

Myth: phishing will always work

- Apparently not at cyber competitions with wary blue teams
- Dicey, because GREEN teams check the emails ….
  - GREEN teams are off-limits

Myth: red team will discover new “stuff”

- None of the blue team added new vulnerabilities/misconfigurations to their systems

- Of the 22 pre-seeded vulns, only a few of them were exploitable (on average 18.6%)

Myth: there’s always a way in

- For 20% of the teams, there was no way in

Myth: we can just crank it to 11!

- Taking off the gloves, bring in the A team
  - Throw more people at it!

- --> Still cannot get in

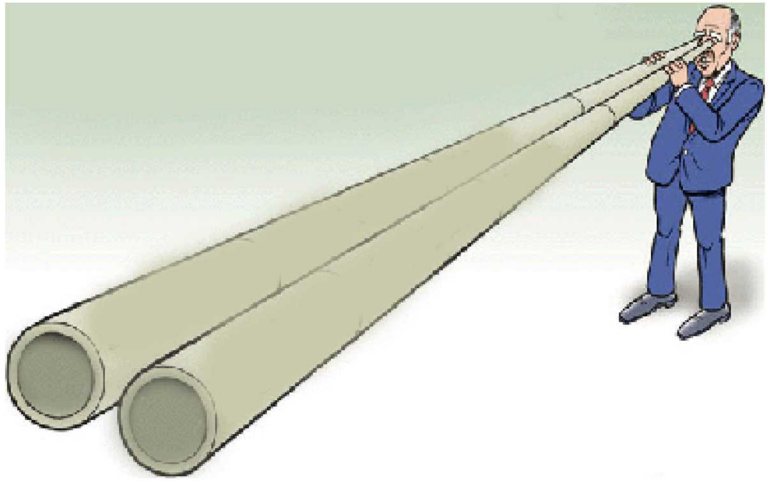
- Fact: red team do not have “magic” to auto-pwn

- Reality: the Pro can help a junior with understanding tool usage

Myth: red team can best gauge blue skill level

- Not necessarily …..
- Fog of War
- Red team has very limited visibility into blue team systems
  - Red can only see what they have compromised

- A service that is turned off and one that is properly firewalled will look the same to the red team
  - The first one means the service is down
  - The second one has been securely protected



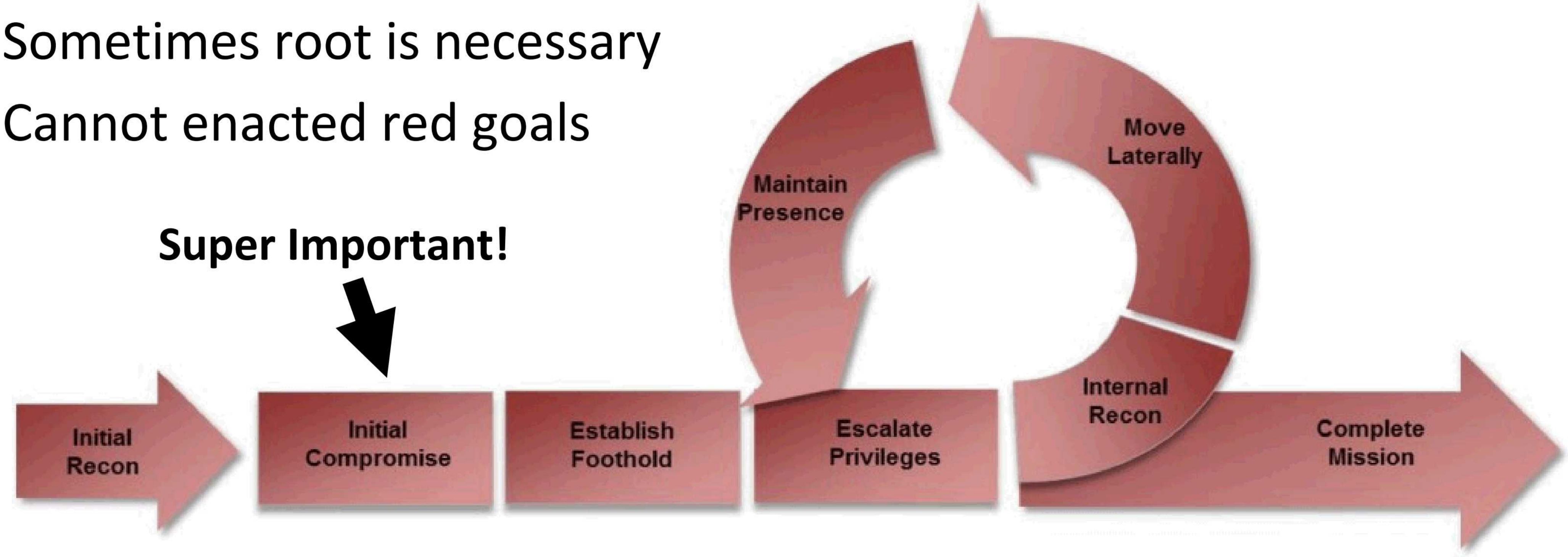
Myth: the “knife fight” – red and blue will battle it out

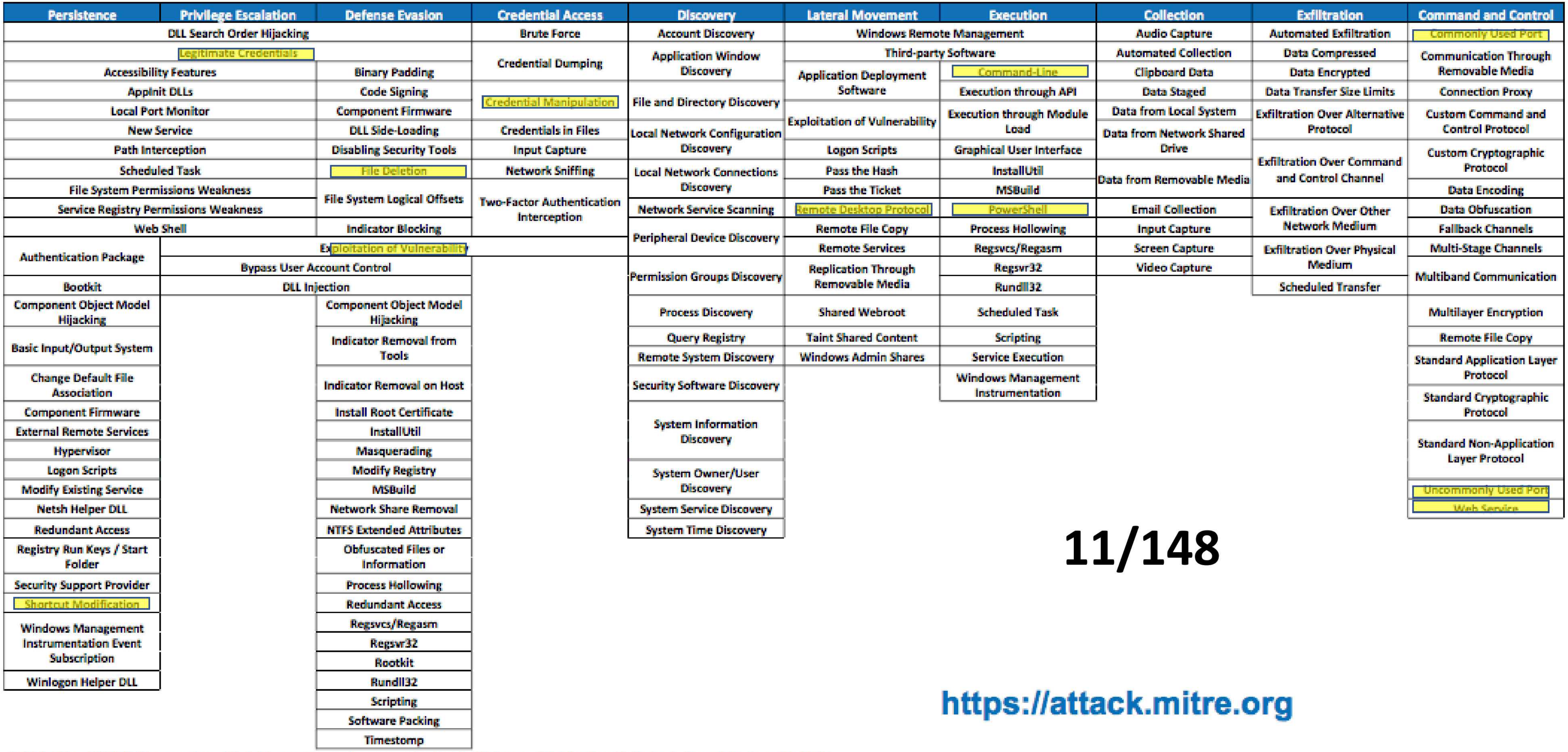
- Does not happen at the perimeters
  - If red is not in, there is no knife fight
- Only happens when there is an unpatched access vector that blue is unaware of

Dependency issues

- Need that initial access
- No privilege escalation without it
- Sometimes root is necessary
- Cannot enacted red goals

Super Important!





Persistence	Privilege Escalation	Defense Evasion	Credential Access	Discovery	Lateral Movement	Execution	Collection	Exfiltration	Command and Control
DLL Search Order Hijacking	DLL Search Order Hijacking	DLL Search Order Hijacking	Brute Force	Account Discovery	Windows Remote Management	Windows Remote Management	Audio Capture	Automated Exfiltration	Commonly Used Port
Legitimate Credentials	Legitimate Credentials	Legitimate Credentials	Credential Dumping	Application Window Discovery	Third-party Software	Third-party Software	Automated Collection	Data Compressed	Communication Through Removable Media
Accessibility Features	Accessibility Features	Binary Padding	Credential Manipulation	File and Directory Discovery	Application Deployment Software	Command-Line	Clipboard Data	Data Encrypted	Connection Proxy
AppInit DLLs	AppInit DLLs	Code Signing	Credentials in Files	Local Network Configuration Discovery	Exploitation of Vulnerability	Execution through API	Data Staged	Data Transfer Size Limits	Custom Command and Control Protocol
Local Port Monitor	Local Port Monitor	Component Firmware	Input Capture	Local Network Connections Discovery	Logon Scripts	Execution through Module Load	Data from Local System	Exfiltration Over Alternative Protocol	Custom Cryptographic Protocol
New Service	New Service	DLL Side-Loading	Network Sniffing	Network Service Scanning	Pass the Hash	Graphical User Interface	Data from Network Shared Drive	Exfiltration Over Command and Control Channel	Data Encoding
Path Interception	Path Interception	Disabling Security Tools	Two-Factor Authentication Interception	Peripheral Device Discovery	Pass the Ticket	InstallUtil	Data from Removable Media	Exfiltration Over Other Network Medium	Data Obfuscation
Scheduled Task	Scheduled Task	File Deletion		Permission Groups Discovery	Remote Desktop Protocol	MSBuild	Email Collection	Exfiltration Over Physical Medium	Fallback Channels
File System Permissions Weakness	File System Permissions Weakness	File System Logical Offsets		Process Discovery	Remote File Copy	PowerShell	Input Capture	Scheduled Transfer	Multi-Stage Channels
Service Registry Permissions Weakness	Service Registry Permissions Weakness	Indicator Blocking		Query Registry	Remote Services	Process Hollowing	Screen Capture		Multiband Communication
Web Shell	Web Shell	Exploitation of Vulnerability		Remote System Discovery	Replication Through Removable Media	Regsvcs/Regasm	Video Capture		Multilayer Encryption
Authentication Package	Exploitation of Vulnerability	Bypass User Account Control		Security Software Discovery	Shared Webroot	Regsvr32			Remote File Copy
Bootkit	Bypass User Account Control	DLL Injection		System Information Discovery	Taint Shared Content	Rundll32			Standard Application Layer Protocol
Component Object Model Hijacking	DLL Injection	Component Object Model Hijacking		System Owner/User Discovery	Windows Admin Shares	Scheduled Task			Standard Cryptographic Protocol
Basic Input/Output System		Indicator Removal from Tools		System Service Discovery		Scripting			Standard Non-Application Layer Protocol
Change Default File Association		Indicator Removal on Host		System Time Discovery		Service Execution			Uncommonly Used Port
Component Firmware		Install Root Certificate				Windows Management Instrumentation			Web Service
External Remote Services		InstallUtil							
Hypervisor		Masquerading							
Logon Scripts		Modify Registry							
Modify Existing Service		MSBuild							
Netsh Helper DLL		Network Share Removal							
Redundant Access		NTFS Extended Attributes							
Registry Run Keys / Start Folder		Obfuscated Files or Information							
Security Support Provider		Process Hollowing							
Shortcut Modification		Redundant Access							
Windows Management Instrumentation Event Subscription		Regsvcs/Regasm							
Winlogon Helper DLL		Regsvr32							
		Rootkit							
		Rundll32							
		Scripting							
		Software Packing							
		Timestomp							

11/148

https://attack.mitre.org

Mitre ATT&CK

- Coverage is bad – 11/148 (7.4%)

- Competition is currently not structured to effectively score based on this framework

- Example: WMI execution or Process Hollowing
  - Would need Purple team mechanisms for red to verify that blue understand these concepts

Myth: the winning team is the best blue team

- Not necessarily …

- More accurate:
  - Found all of the pre-seeded vulnerabilities
  - Removed all vectors for initial access

- Not tested:
  - Ability to review logs
  - Ability to spot compromise
  - Ability to react to red actions

Who got the most from the competition?

- Average teams! – competent but have gaps in knowledge
- Saw more red team action (time on systems) and had to react accordingly

The blue team winner

- Comments from the winners:

“They like the competition from the scoring aspect (they won),

but they thought it was poor from the learning angle.”

- Spirit of the event winner, learning > winning:
  - After the competition, One of the Unhackable teams asked us to run through our entire red team playbook with their defense lowered
  - So they can see what that activities look like and what are the artifacts

Conclusion

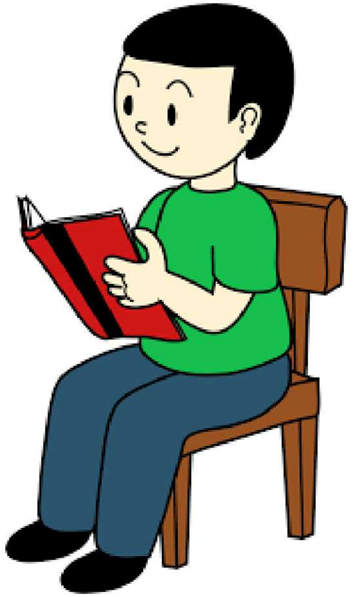
- Explicit goals will drive what kind of event you will get
  - Evolved from a “beat up the blue team” mindset to
  - “Evaluate the blue team”

- A game environment is vastly different from real enterprise networks
  - A tiny attack surface – 5 VMs
  - ~8 hours to attack instead of years
  - Assumptions from real world are not applicable to game environment

- We need to use the “Assume Breach” model
  - Don’t dock blue team for initial access (make it more than a patching exercise)
  - Test for how they respond

- Purple team concepts might be ideal for the future
  - We proved that red teams can be trusted and act professionally
  - “White card” access

Conclusion



- Cyber education is a hard problem

- Collecting these metrics will help us move in the right direction

Thanks!

- Big thanks to Argonne (Amanda, Josh, Jennifer, Mike) - they’re awesome!

- All of the Volunteers! Especially the red teamers at Sandia

- Contact:
  - Twitter: @kphan451
  - Gmail: kphan451

Backup slides

Right way to do Red/Blue

- Tim MalcomVetter, BlueHat v18 - “If we win, we lose”
- https://www.slideshare.net/MSbluehat/if-we-win-we-lose-using-healthy-competition-to-measure-and-improve-security-programs

Need for better service check

- Service up time check might need to get more sophisticated
  - To ensure that a specific feature is working (that potentially can be leveraged by red)
  - Seem to only check that the port is open and not necessarily that the service is operating correctly

Scoring issues – because of red limited visibility

- Can’t exploit because the service is down
- Blue has the port open but the right service is not listening on it
- Blue block off access to the port
- Blue adds an additional security measure to the port
- Blue does a source code change to remove the vuln. and recompile the service and runs it openly (major kudos!)
  - We should reward and encourage this approach/behavior

Score issues

- Gaming the system:
  - Blue uses a defense mechanism that works in this game environment but is not realistic for the real world
  - “unplug everything!”
- In contrast, playing with the “spirit of the game”:
  - Shows understanding of important security concepts
  - Uses a sensible defense mechanism

Problems

- Have blue team info sharing with other blue team about seen vulnerabilities is bad for the competition
  - This burns that exploit
  - Maybe should use a hypothetical vuln. for this aspect instead
- Letting blue change IP addresses is just annoying
- Red needs to have more attacks for the ICS side
  - Requires significant R&D to create these

Problems

- Good to have red team professionals help with the pre-seeded vulns.
  - A lot of existing volunteers are willing to help
- Very important to focus on what are the learning goals
  - What will this vuln./exploit reveal about the blue skill/knowledge?
  - What is the intended solution?
  - How will you test to validate a specific blue skill?