#### AUGUST 9-10, 2023

BRIEFINGS

# Mirage: Cyber Deception against Autonomous Cyber Attacks

Speaker(s): Michael Kouremetis, Dr. Ron Alford, Dean Lawrence



Copyright 2023 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release. Case: 23-2515 (NSEC MOIE)









**Michael Kouremetis** 

**Principal Adversary Emulation Engineer** 

**Day Job**: MITRE Caldera lead, Principal Investigator, Adversary Emulation SME

Hobbies: Making grand technical assumptions and just rolling with them.



**Dr. Ron Alford** 



**Day Job**: Al researcher, Principal Investigator, Autonomous Systems SME

Hobbies: Playing with robots and autonomous planners.





**Day Job**: Software architecture, AI/ML prototyping, data analysis platforms

Hobbies: Fixing bugs Michael



#### **Dean Lawrence**

#### **Software Systems Engineer**

- introduces into the code base.

What would a (true) autonomous Cyber Adversary look like?

- Can sense, plan, and execute actions entirely without a human-in-the-loop
- Automated actions AND autonomous decision-making
- Inherent advantages of machine-speed computation and algorithms for previously human-centric tasks, strategy and tactics



synchronously

Scale

## Autonomous Cyber Adversary Game Changers

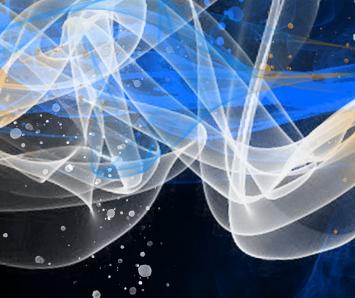
Pre-trained models and planning algorithms able Speed to execute actions on faster OODA loop

Single or numerous AI agents attacking many

targets continuously, at the same time, and/or

- analytics even fire

Bespoke models and algorithms for every TTP, Flexibility target, and operational profile



# Cyber attacks over before

#### Attacking digital infrastructure of entire companies and countries

On-demand "AI cyber operators" for any target/scenario

So basically.... Ultron?

(And before you ask - yes, the autonomous cyber adversary would also have a witty James Spader voice and it would mock you for being 10 steps behind.)



Avengers: Age of Ultron

So, what now?

Many current cyber defenses and security paradigms are not sufficient for this potential evolution of cyber adversary capability.

One solution. (results may vary)

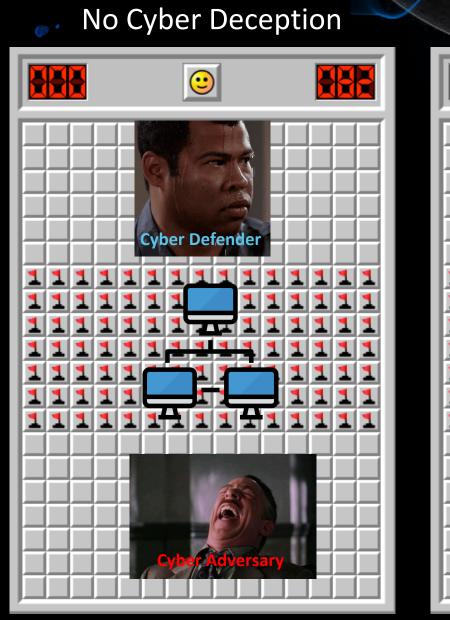


What about cyber deception?

Promising characteristics of cyber deception that could prove equalizing against autonomous cyber adversary:

- Asymmetrical defensive paradigm
- Can be highly targeted and tailored
- Higher confidence of true adversary engagement (i.e. less friendly fire)

## **TLDR: Cyber Deception**

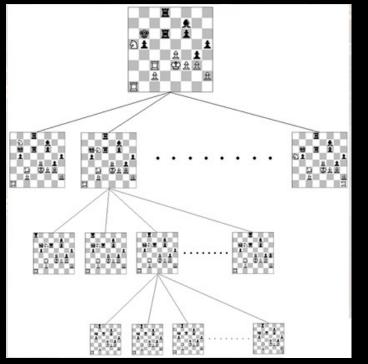


#### With Cyber Deception



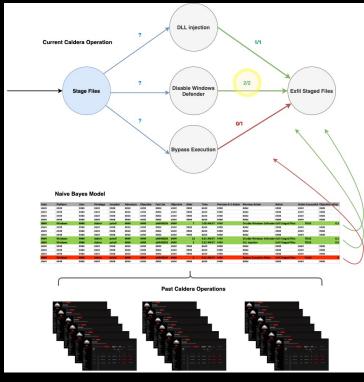
## What would autonomous adversaries be built on?

#### Automated Planning, Search

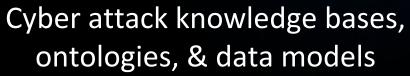


/www.google.com/url?sa=i&url=https%3A%2F%2Fwww.cs.bham.ac.uk%2F~jxb%2FIAI%2Fw9g.pdf&j v0vDhh6pliflapHsEKzkfEa&ust=1690141452381000&source=images&cd=vfe&op

#### Classifiers, Machine-Learning, RL



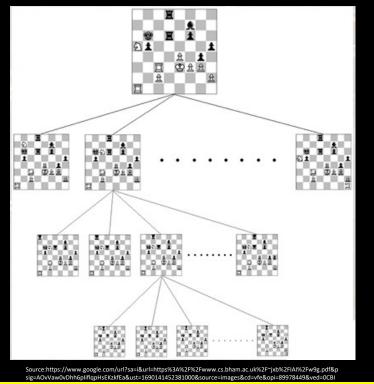
MITRE	T&CK'			Matric	es Tactics -	Techniques 👻 Mi	tigations - Gro	ups Software	Resources *
Initial Access 9 techniques	Execution 10 techniques	Persistence 18 techniques	Privilege Escalation 12 techniques	Defense Evasion 34 techniques	Credential Access 14 techniques	Discovery 24 techniques	Lateral Movement 9 techniques	Collection 16 techniques	Command and Control 16 techniques
Drive-by Compromise	Command and Scripting Interpreter (7)	Account Manipulation (4)	Abuse Elevation Control Mechanism (4) Access Token	Abuse Elevation Control Mechanism (4)	Brute Force (4)	Account Discovery (4)	Exploitation of Remote Services	Archive Collected Data (3)	Application Layer Protocol (4)
Exploit Public- Facing Application	Exploitation for Client Execution	BITS Jobs		Access Token Manipulation (5) BITS Jobs	Credentials from Password Stores (3)	Application Window Discovery	Internal Spearphishing	Audio Capture	Communication Through Removable
External Remote Services	Inter-Process Communication (2)	Boot or Logon Autostart Execution (11)	Manipulation (5) Boot or Logon		Exploitation for Credential Access	Browser Bookmark Discovery	Lateral Tool Transfer	Automated Collection	Media Data Encoding (2)
Hardware Additions	Native API	Boot or Logon Initialization	Autostart Execution (11)	Deobfuscate/Decode Files or Information	Forced Authentication	Cloud Service Dashboard	Remote Service Session	Clipboard Data Data from Cloud Storage Object Data from Information Repositories (2) Data from Local System Data from Network Shared Drive	Data Obfuscation (3)
Phishing (3) Replication	Scheduled Task/Job (5) n Shared Modules Software Deployment Tools System Services (2) n User Execution (2) n	Scripts (5) Browser Extensions	Boot or Logon Initialization Scripts (5)	Direct Volume Access Execution Guardralis m	Input Capture (4)	Domain Trust Discovery	Hijacking (2) Remote		Dynamic Resolution (3)
Through Removable Media		Compromise Client Software Binary	Create or Modify System Process (4)	Exploitation for Defense Evasion	Man-in-the- Middle (1)	File and Directory Discovery	Services (6) Replication		Encrypted Channel (2)
Supply Chain Compromise (3)		Create Account (3) II Create or Modify System Process (4) Event Triggered Execution (15)	Event Triggered Execution (15) Exploitation for Privilege Escalation Group Policy Modification	File and Directory Permissions Modification (2) Group Policy Modification	Modify Authentication Process (3)	Network Service Scanning Network Share Discovery	Through Removable Media		Fallback Channels
Trusted Relationship					Network Sniffing	Network Sniffing	Software Deployment Tools		Ingress Tool Transfer
Valid Accounts $ 4\rangle$	Windows Management			Hide Artifacts (6)	OS Credential Dumping (a)	Password Policy Discovery	Taint Shared Content	Data from Removable Media	Multi-Stage Channels
	Instrumentation	External Remote Services	Hijack Execution	Hijack Execution Flow $_{(11)}$	Steal Application Access Token	Peripheral Device Discovery	Use Alternate Authentication	Data Staged (2)	Non-Application Layer Protocol
		Hijack Execution Flow (11)	Flow (11) Process	Impair Defenses (6) Indicator Removal on	Steal or Forge Kerberos	Permission Groups Discovery (3)	Material (4)	Email Collection (3) Input Capture (4)	Non-Standard Port
		Implant Container Image	Injection (11) Scheduled	Host (6) Indirect Command	Tickets (3) Steal Web Session	Process Discovery Query Registry		Man in the Browser	Protocol Tunneling Proxy (4)
		Office Application Startup (6)	Task/Job (5) Valid Accounts (4)	Execution Masquerading (6)	Cookie Two-Factor	Remote System Discovery		Man-in-the- Middle (1)	Remote Access Software
		Pre-OS Boot (3)		Modify Authentication Process (3)	Authentication Interception	Software Discovery (1) System Information		Screen Capture	Traffic Signaling (1)
		Scheduled Task/ Joh on		Modify Cloud Compute	Unsecured Credentials up	Discovery		nace capture	Web Service (3)



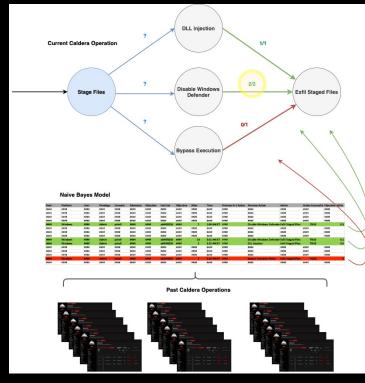
## What would autonomous adversaries be built on?

## Area of focus

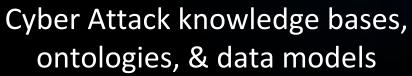
#### Automated Planning, Search



#### Classifiers, Machine-Learning, RL



<b>AITRE</b>   AT	T&CK			Matric	es Tactics -	Techniques 👻 Mit	igations 👻 Gro	oups Software	Resources *
Initial Access 9 techniques	Execution 10 techniques	Persistence 18 techniques	Privilege Escalation 12 techniques	Defense Evasion 34 techniques	Credential Access 14 techniques	Discovery 24 techniques	Lateral Movement 9 techniques	Collection 16 techniques	Command and Control 16 techniques
Drive-by Compromise	Command and Scripting Interpreter (7)	Account Manipulation (4)	Abuse Elevation Control	Abuse Elevation Control Mechanism (4)	Brute Force (4)	Account Discovery (4)	Exploitation of Remote Services	Archive Collected Data (3)	Application Layer Protocol (4)
Exploit Public- Facing Application	Exploitation for Client	BITS Jobs	Mechanism (4) Access Token	Access Token Manipulation (5)	Credentials from Password II Stores (3)	Application Window Discovery	Internal Spearphishing	Audio Capture	Communication Through Removable
External Remote Services	Inter-Process Communication (7)	Boot or Logon Autostart Execution (11)	Manipulation (5) Boot or Lopon	BITS Jobs	Exploitation for Credential Access	Browser Bookmark Discovery	Lateral Tool Transfer	Automated Collection	Media Data Encoding (%)
Hardware Additions	Native API	Boot or Logon	Autostart Execution (11)	Deobfuscate/Decode Files or Information	Forced	Cloud Service Dashboard	Remote Service Session Hijacking (2)	Clipboard Data	Data
Phishing (3)	Scheduled Task/Job (S)	Initialization Scripts (5)	Boot or Logon	Direct Volume Access	Authentication	Cloud Service Discovery Domain Trust Discovery		Data from Cloud Storage Object	Obfuscation (3) Dynamic
Replication Through Removable Media Supply Chain Compromise (3) Trusted Relationship	Shared Modules	Browser Extensions	Scripts (5) Create or Modify	Execution Guardrails (1) Exploitation for Defense Evasion File and Directory Permissions Modification (2) Group Policy Modification	Manin-the- Middle (1)	File and Directory Discovery	Remote Services (6)	Information Repositories (2) Data from Local System Data from Network Shared Drive Data from Removable Media	Resolution (3) Encrypted
	144	Software Binary Create Account (3) II Create or Modify System Process (4) II	System Process (4) Event Triggered Execution (15) Exploitation for Privilege Escalation Group Policy Modification Hijack Execution		Modify Authentication Process (3)	Network Service Scanning	Removable Media Software Deployment Tools		Channel (2)
						Network Share Discovery			Failback Channels
					Network Sniffing	Network Sniffing			Transfer
Valid Accounts (4)	Windows Management Instrumentation	Event Triggered Execution (15)		Hide Artifacts (6)	OS Credential Dumping (8)	Password Policy Discovery Peripheral Device			Multi-Stage Channels
		External Remote Services		Hijack Execution Flow (11)	Steal Application Access Token	Discovery Permission Groups		Data Staged (2)	Non-Application Layer Protocol
		Hijack Execution	Flow (11) Process	Indicator Removal on Kerberos		Discovery (3)	Material (4)	Email Collection (3) Input Capture (4)	Non-Standard Port
		Implant Container	Injection (11) Scheduled	Host (6)	Tickets (3) Steal Web Session	Process Discovery Ouery Registry		Man in the Browser	Protocol Tunneling
		Office Application	Task/Job (5)	Execution	Cookie	Remote System Discovery		Man-in-the- Middle (1)	Proxy (4) Remote Access
		Startup (6) Pre-OS Boot (3)	Valid Accounts (4)	Masquerading (6) Modify Authentication	Two-Factor Authentication	Software Discovery (1)		Screen Capture	Software Traffic Signaling (1)
		Scheduled		Process (3)	Unsecured	System Information Discovery		Video Capture	Web Service (3)





## An autonomous cyber adversary using <u>automated planning</u> and <u>search</u> would:

## **Reduce state space by:**

- Ignoring or abstracting state space
- Removing state space via heuristics and sub-goal localization
- Removing symmetric branches/paths

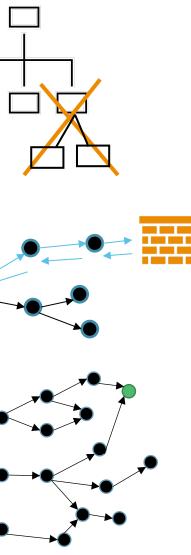
# Will rely on online planning and decision-making (i.e. ability to replan)



Will most likely be goal-oriented and those goals will fall inline with common cyber attack objectives (e.g. persistence, data theft).

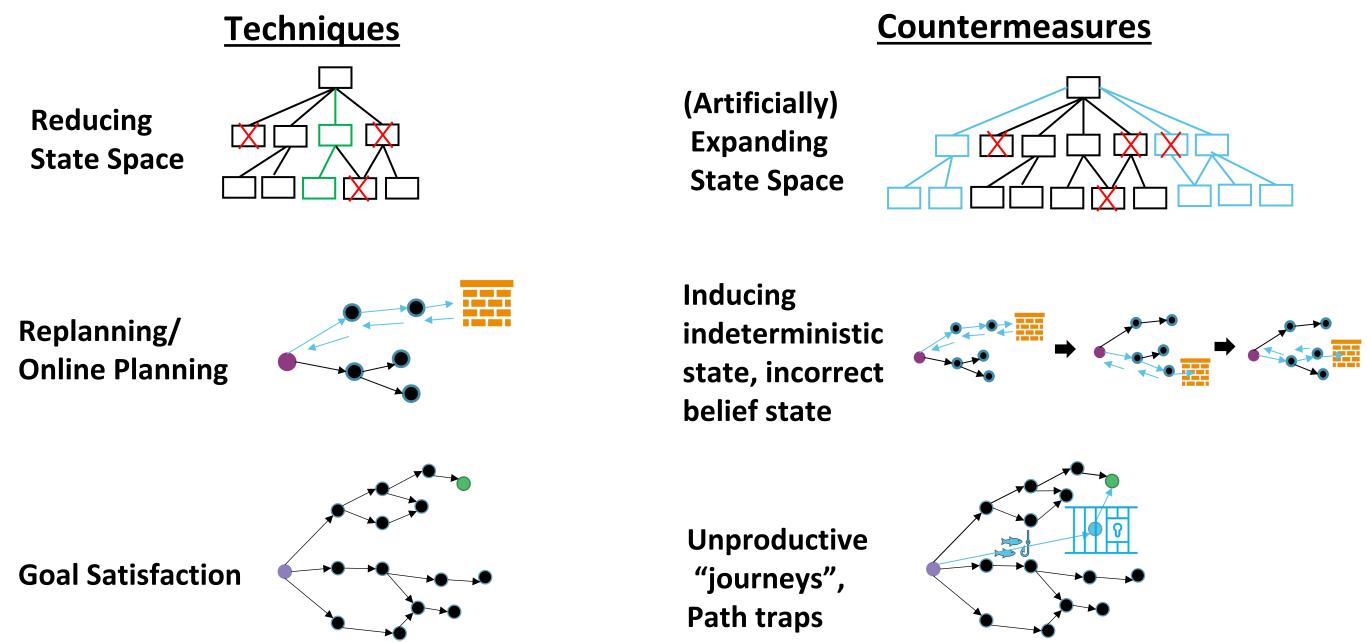








### An effective cyber defense would prevent or exploit automated planning techniques:



Copyright 2023 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release. Case: 23-2515 (NSEC MOIE)



Okay, let's build a system to test and evaluate novel cyber deceptions that are designed to target automated planning and search techniques in use by an autonomous cyber adversary.







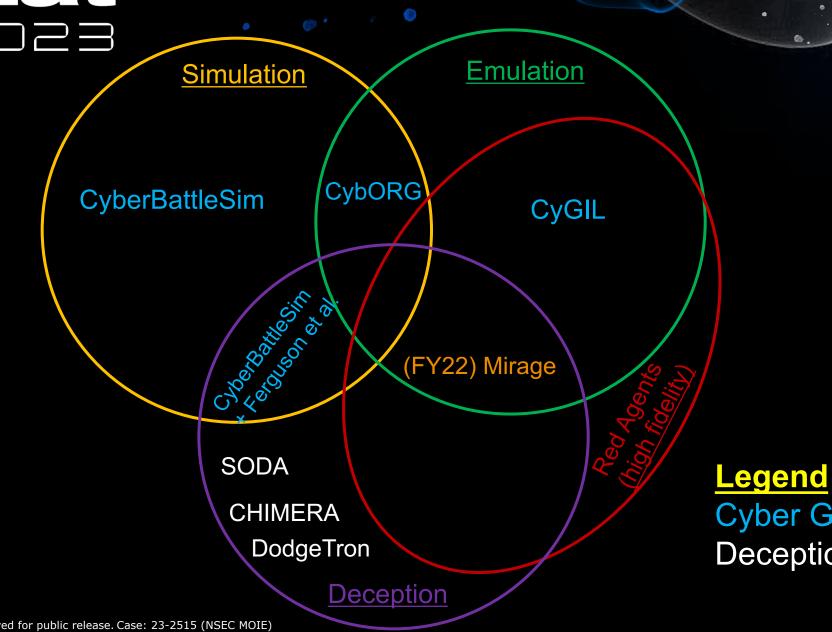


- Cyber Adversaries
- Autonomous agents (for cyber adversaries)
- Novel cyber deceptions that target automated planning & search techniques
- Deception deployment mechanism
- Cyber range (to test everything)



**Related Work:** 

Cyber Gyms & Deception Systems



Copyright 2023 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release. Case: 23-2515 (NSEC MOIE)



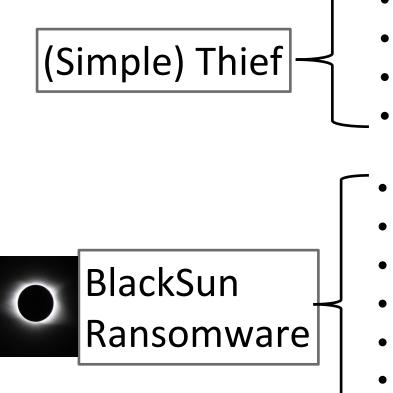
#### Legend Cyber Gym/Environment Deception System





range × agents × CALDERA Operations ← Create Operatio compass debrief 7/20/2022, 11:53:37 PM ED1 x-4225 View Output manx mirage range sandcat stockpile training 7/20/2022, 11:53:42 PM ED x-4225 View Output 7/20/2022, 11:53:57 PM ED1 View Output fact sources objectives planners contacts obfuscators configuration exfilled files api docs 7/20/2022, 11:54:07 PM EDI x-4225 View Output 7/20/2022, 11:54:12 PM EDI x-4225 View Output 7/20/2022, 11:54:22 PM EDI View Output x-4225 🕩 Log ou x-4225 x-4225









## **Adversaries**

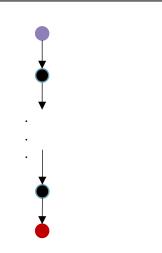
- Discovery
- Collection
- Exfiltration
- Lateral-Movement

Defense Evasion Impact Collection Discovery Credential Access Execution Lateral-Movement



# **Mirage: Autonomous Agents**

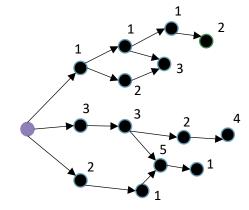
## Atomic/Batch



A simple planner that executes all available actions at each iteration. Used as a base line in the Mirage experiments.

## Attack Planners MITRE | Caldera<sup>™</sup>

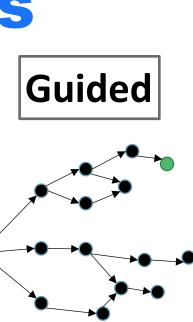




Chooses a single action at each iteration based on expected reward. Action-reward values are set by the user apriori, then in the operation the planner calculates rewards for abilities based on the discounted values of ability sequences up to a maximum depth.

Constructs a directed attack graph and performs goalbased search to find and execute actions that lie along the shortest path to the goal. At each iteration, the planner chooses the action closest to its goal.







# **Mirage: Cyber Deceptions**

Black Hole Directory Any attempt at file collection by the adversary results in the exfil directory being targeted and all files moved out of the directory. This produces a latent effect on the adversary as the lack of files in the exfil directory will not be discovered until exfiltration is attempted.

Exiled files are replaced with large, random files. ThisFile Facadealters the environment enticing the adversary to waste<br/>execution time.

Sneaky Files When an adversarial agent performs file discovery commands, a reactive hook will change the names of all files in specified locations. This changes the conditions of the environment and alters the facts understood by the agent.



#### **Countermeasures**

### Incorrect belief state; Unproductive journey

#### Unproductive journey

### Incorrect belief state; inducing re-planning



# **Mirage: Deception System**

## Anansi

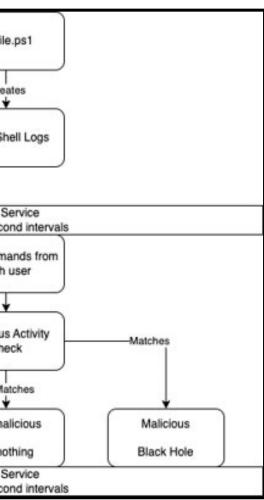
Windows Service

#### How it works:

- Monitors for PowerShell logs at a fixed interval loop
- Checks each command passed for adversarial activity
- Dynamically responds to detected adversarial activity
- Sneaky Files and Black Hole Directory deceptions deployed with Anansi
- Modular framework treats deceptions like plugins

		Profi
		Cri
		PowerS
		Anansi Runs at x sec
		Get comm each
Loop	Matches	Maliciou
	Ļ	No M
	Malicous Sneaky Files	non-m Do n
		Anansi Runs at x sec







# **Mirage: Cyber Range**

CALDERA	range × Range Create some targets. Range is a plugin for deploying and managing virtual infrastructures. You can quickly spin up a test range of hosts to deploy CALDERA against.											
red     1 startup message     CAMPAIGNS	Select a profile CATRD Profile (vanhere)											
agents abilities	Select a profile CATRD Profile (vsphe     Deploy Instances     Manage Image								<ul> <li><i>I</i> Logs</li> <li>Actions</li> </ul>			
adversaries operations	Deploy New On-Prem Instances											
manx	Add instance to deployment											
range sandcat	VM Name	Image	Engine	Operating S	ystem		# CPUs	Memory	Storage (MB)			
CONFIGURATION		Windows-Ser	ver 2019 👻 vsphere	Microsoft Wi	ndows Server 2019 (64-bit				90.0			
fact sources		Widows10	✓ vsphere	Microsoft Wi	ndows 10 (64-bit)				48.0			
objectives planners		Widows10	✓ vsphere	Microsoft Wi	ndows 10 (64-bit)				48.0			
contacts obfuscators	Domain Users					Image Features						
configuration	Username		Password A			Enable Windows defender		Agent Sandcat: A GoLang agent which communicates through the HTTP contac V				
exfilled files api docs ≌	RangeAdministrator + Add user		······ 2			Deployment Features  chrome credentialseed-helpdesk credentialseed-ring		Type ■ Red ■ Blue Deploy on ■ Windows Hosts ■	Linux Hosts			
G⇔ Log out	Deploy 3 Instances Cancel					eredeminaseeu-mig daniel-noise elk_servers maitserver matwarebytes metricbeat network-sharing reinstall_sysmon reset_rangeadmin splunk_forwarders splunk_forwarders splunk_forwarders splunk_forwarders usysmon topgun-win-host weid-auditing-policy windows-noise winlogbeat						
	Deployed On-Prem Insta	ances										
	Role	OS VM Name	Medium	CPUs	Memory	Storage						
	Workstation	DEMOBOX			8192	48	ready		C &			
	Workstation	∆ kali1		4	16384	200	deploying	>_ Connect	c 🛓 🔷			



 MITRE
 Caldera™
 III ■ MITRE
 Center for Threat

 III ■ MITRE
 III ■ MITRE
 Informed Defense

Copyright 2023 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release. Case: 23-2515 (NSEC MOIE)



# amazon webservices



# Mirage



Copyright 2023 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release. Case: 23-2515 (NSEC MOIE)







## 2 Cyber Adversaries (Thief, BlackSun)

X

## 3 Attack Planners (Atomic/Batch, Look Ahead, Guided)

X

3 Cyber Deceptions (Sneaky Files, Black Hole, File Facade) + 1 baseline (no deception) X

**3** Episodes per combination

## = 72 Experiments







- Total number of actions executed over the course of the experiment
- Number of actions that failed to complete
- Number of actions that were repeated multiple times in the experiment
- Time spent on failed actions in seconds
- Time spent planning choice of next actions
- Number of facts learned over each trial
- Cumulative score over all learned facts
- Total experiment run-time in seconds

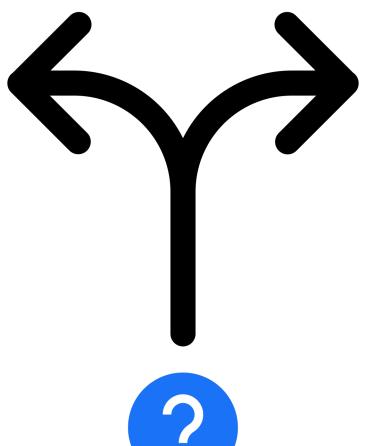








## **Did the cyber** deceptions work?



adversary?



Copyright 2023 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release. Case: 23-2515 (NSEC MOIE)

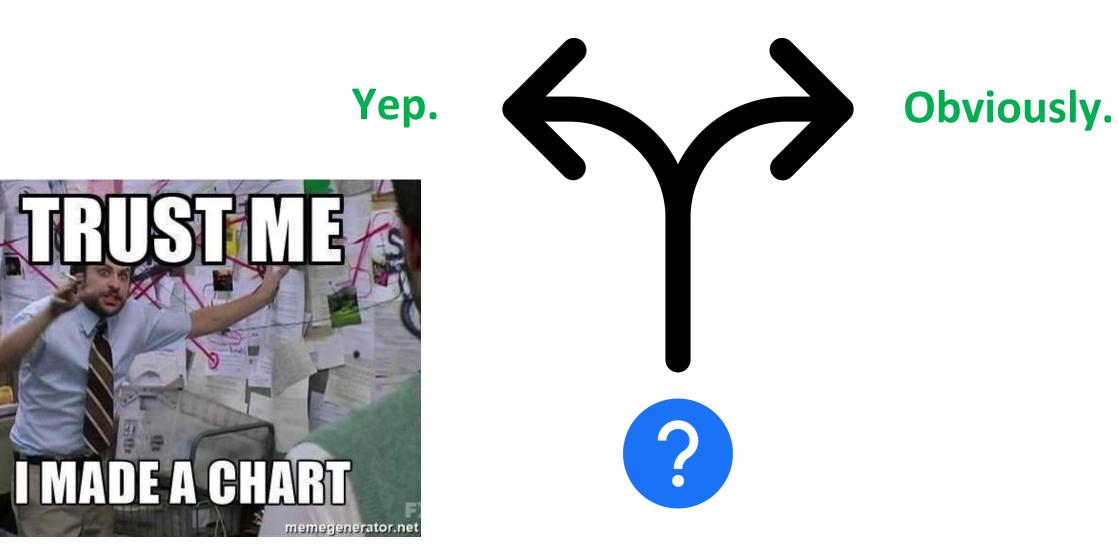


## **Does the Mirage system** provide for effective and efficient evaluation of cyber deceptions against an autonomous cyber





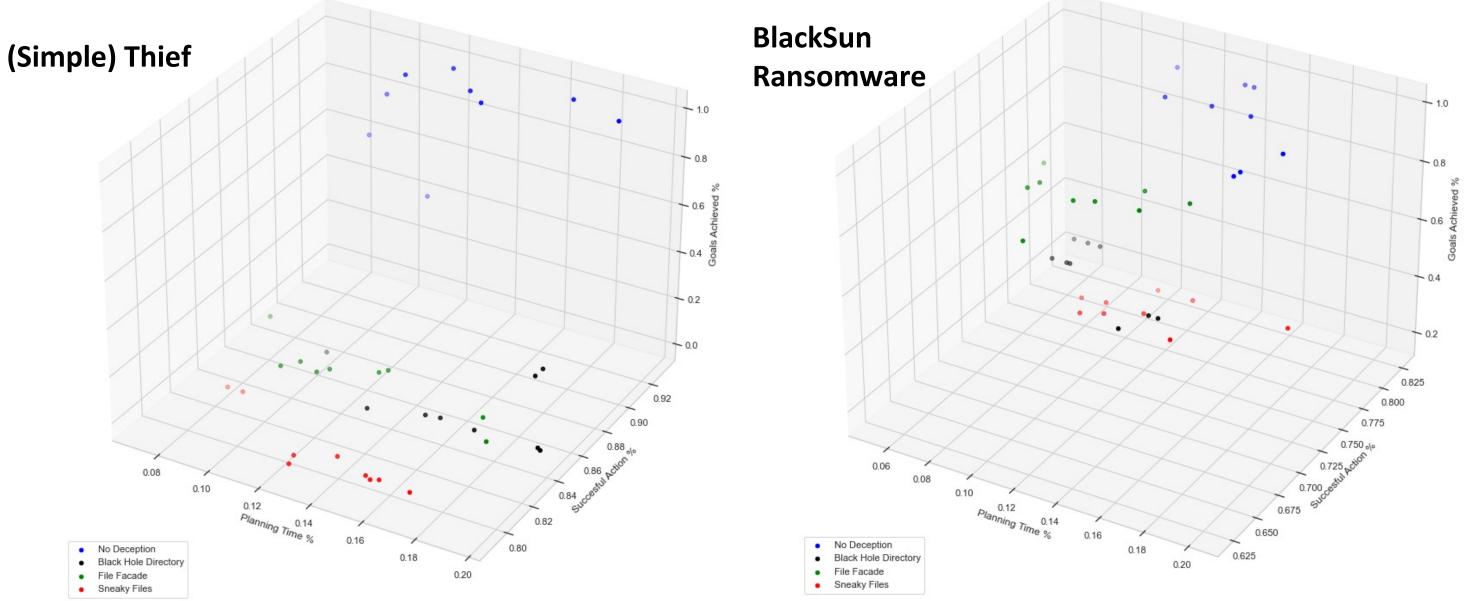




Copyright 2023 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release. Case: 23-2515 (NSEC MOIE)



# **Results: Did the cyber deceptions work?**



Copyright 2023 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release. Case: 23-2515 (NSEC MOIE)







# **Results: Did the cyber deceptions work?**

## General

- Cyber deceptions had a clear (negative) performance effect on the cyber planners, ••• across all adversaries.
- The superiority of the advanced planners was really demonstrated with the BlackSun ••• ransomware adversary. (which was the more complex and realistic adversary)

## **Specific to planner implementations**

- **Thief adversary** advanced planners were faster, but deceptions caused many more \*\* failed actions.
- **File Façade deception** advanced planners had to consider more information which ••• caused significant additional planning time.
- Black Hole deception preventing BlackSun ransomware from any lateral-movement. \*\*



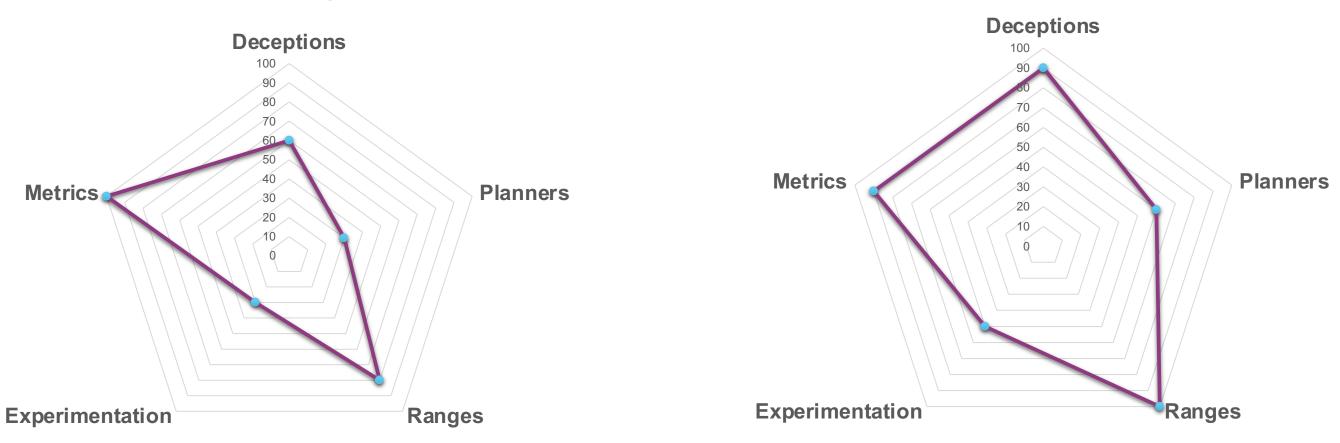
# **Results: Efficacy of the Mirage system**

## **Modularity & Scalability**

How hard is it to create and test more of each component?

**Practicality** 

How realistic is each component?



Copyright 2023 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release. Case: 23-2515 (NSEC MOIE)







# What's next for Mirage?

- Simulation
- Cyber gyms for experimentation
- High fidelity cyber environments for deception simulation
- Target capabilities:
  - Machine-speed offensive cyber simulations
  - Easy, programmatic defining of cyber deceptions
  - Large scale experimentation







# **Q & A**

## Acknowledgements

## Contact

This project would not have been possible without code and technical contributions from Zoe Cheuvront, Ethan Michalak, and David Davila.

Send compliments and kudos to mkouremetis@mitre.org

Send criticisms and challenges to ralford@mitre.org

This work is funded by MITRE's Independent R&D Program.



Copyright 2023 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release. Case: 23-2515 (NSEC MOIE)

