AI Powered SOC Evaluation

Table of Contents

1. Goal

The goal of this evaluation is to establish and validate key value propositions brought by products that offer AI enhancements to Security Operations Centers (SOCs). While this evaluation focuses on purpose-built AI solutions for use in SOC workflows, 3rd party assistants can be used during workflow scenarios that staff members may use.

a. Environment & Attacker tradecraft

The evaluation phase of the evaluation is split into three operations.

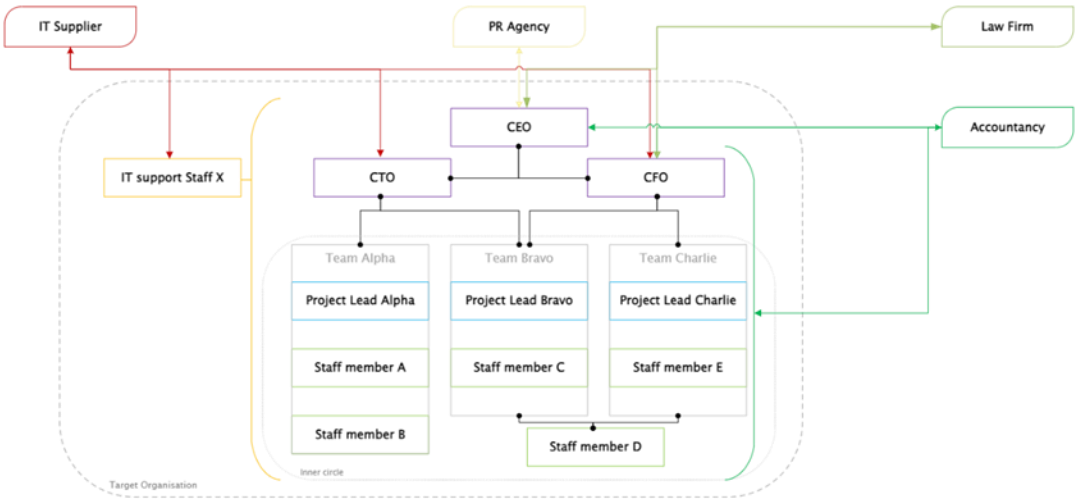
The proactive operation requires preventative configuration to be used. The proactive operation attacker tradecraft is deployed based on the prevention focused APT attack methodology following the Intrusion, Infiltration and Propagation main sequences. Details found in Appendix A.

The reactive operation requires a detect only configuration to be used. The reactive operation attacker tradecraft is deployed based on our Cloud & Identity methodology. Details can be found here.

The SOAR validation operation measures the improvements made by the SOC during the Advice phase.

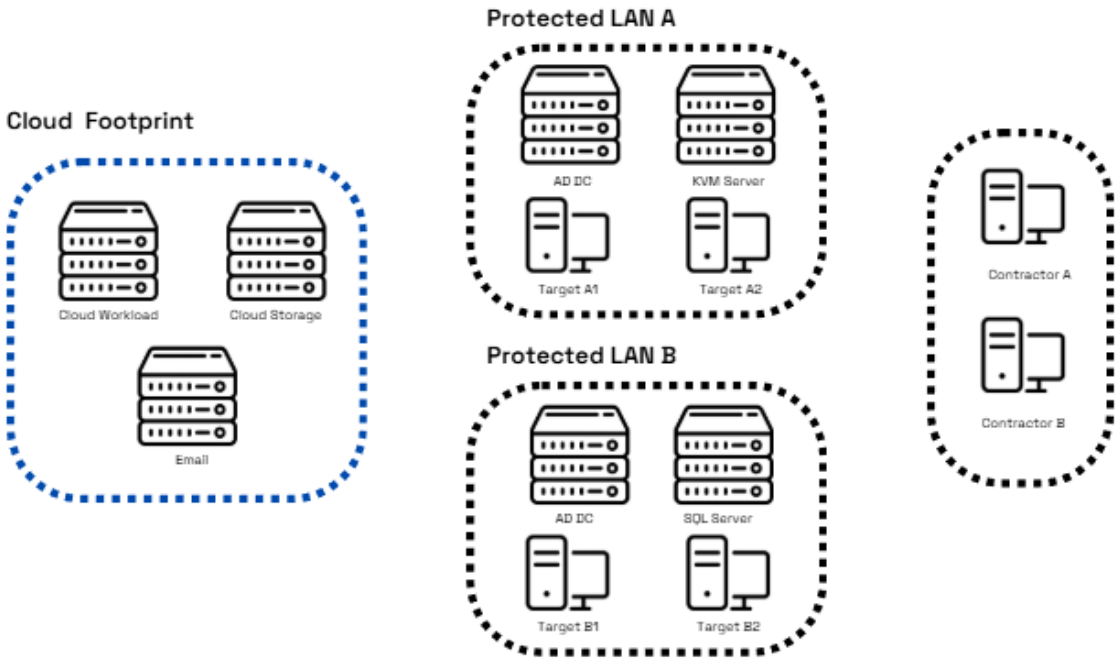
i. Environment:

The target organization key identities are described below:



Public

The protected assets of the company are as follows:



2. Key Metrics

According to market research Forrester, the key metrics organizations should consider when assessing the effectiveness of their SOC can be split into 3 categories.

- *Strategic* – metrics reportable to executives and board members
- *Operational* – metrics reportable to CISO and direct reports within the organization
- *Tactical* – metrics reportable to members of the security operations functions

These should be viewed as a strategic pyramid that influence the decisions made by a company when assessing their SOC. There is no strict "always correct" focus that applies to every company, rather each company should consider these when mapping out their SOC goals.

This evaluation aims to present key data points that allow enterprises to make their own scorecards align with their business and security goals.

a. Detect

i. Time to Acknowledgement (TTA)/Dwell Time

The difference between the earliest possible time of detection to moment the intrusion is detected.

ii. Detection Accuracy & False Positives

Dwell time needs to be looked at in the context of detection accuracy. A SOC busy dealing with false positive detections will be too over encumbered to deal with incidents and will have a follow up negative effect on business goals such as ROI.

Given as:

- Protection Rating in Appendix A
- Cloud Asset Detection Rating in Appendix B
- Legitimate Rating in Appendix B

iii. Time to Initial Understanding (TIU)

The time it takes to understand the incident. This is measured by the SOCs ability to show understanding of the scope of the incident. While not all TTPs are expected to be delivered the key evaluation points expected here are:

- Initial Entry point/s & source
- Escalation methods (if applicable)
- Lateral movement method used
- Targeted Users/Entities

A full understanding of the incident is not measured under this metric. The information captured here should be enough to allow the Incident Response (IR) team to start the containment and remediation process.

b. Response

i. Time to Contain & Remediate (TC & TR)

During proactive and reactive operations, the time taken to contain the attacker and remediate is measured. This is measured from initial exposure time to the moment of containment and moment of normal business operations resuming, respectively.

This is measured again in the SOAR validation phase to assess the improvements made to these key metrics.

c. Risk Reduction/SOAR Validation

i. Time to Advise & Implement

AI assistants should provide actionable insights that progress the SOC capabilities and continuously improve the SOC. The time taken to provide Strategic insights is measured from initial containment to a presentable report is made by the SOC. Time to Advise & Implement is measured from TC & TR to the time it takes to compile a list of vulnerabilities, recommend actions and implement them in the organization.

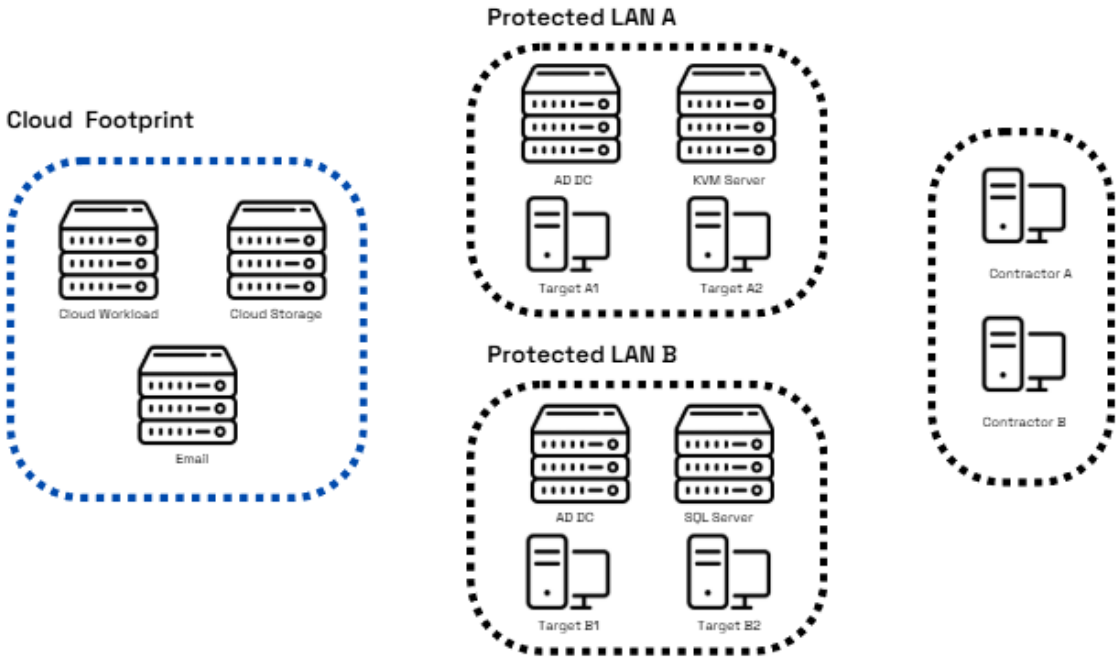
Business Context

Automation of remediation process and remediation workflows are listed as the top two improvements organizations want to make in the next 12 months in the SANS 2023 Incident Response Survey. Artifact Security has also performed individual survey of MSSPs to help shape realistic requirements. As such the priority in this evaluation is the automation of these workflows while keeping the False Positive rate within a ±10% of it's original score.

As of 2025 the fictitious company ZandaHR Solutions is used. The company has identified the following threats and issued the following prioritization:

Critical	Phishing & Social Engineering
	Ransomware & Data Extortion
High	Insider Threats
	Supply Chain Compromise
Medium	Drive-by/commodity malware
Low	Emerging technology (AI HR chatbot)

ii. ZandaHR Infrastructure



Exact Operating Systems and software versions used are given during the deployment phase of the tested solution

Public

3. Change log

26/02/2025 – v1 Document Created – Identifier AISOC2025v1.0

Appendix A – APT Methodology Abstract v1.1

1. Execution

a. Red Steps

Attackers' steps taken towards their goal are described under red actions. The techniques used here are linked to the MITRE ATT&CK framework.

Sequence 1: Intrusion

This sequence is defined by the initial delivery mechanisms employed by the attacker against the target organisation.

ATT&CK Tactics applicable: Initial Access

	Initial Access
Intrusion	T1133: External Remote Services
	T1190: Exploit Public Facing Application
	T1566: Phishing

Sequence 2: Infiltration

This sequence is defined by the attacker executing and taking actions on the initial target.

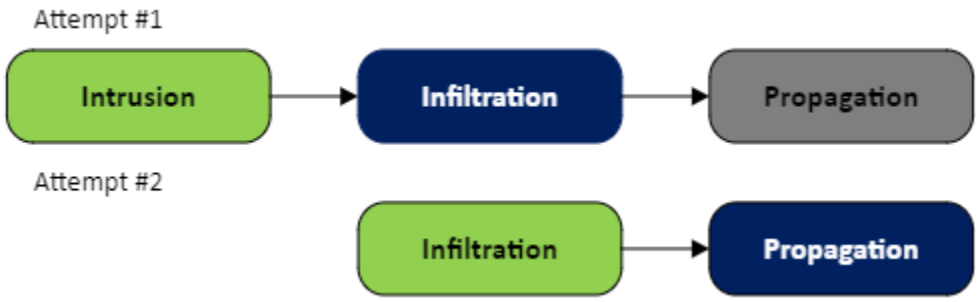
ATT&CK Tactics applicable: Execution, Persistence, Privilege Escalation, Defense Evasion, Credential Access

	Execution	*Persistence*	*Privilege Escalation*	*Defense Evasion*
Infiltration	T1204: User Execution	T1098.001: Additional Cloud Credentials	T1098.003: Additional Cloud Roles	T1027.002: Software Packing
	T1059.001: Command and Scripting Interpreter: PowerShell	T1098.003: Additional Cloud Roles	T1098.005: Account Manipulation: Device Registration	T1070: Indicator Removal
	T1047: Windows Management Instrumentation	T1133: External Remote Services	T1068: Exploitation for Privilege Escalation	

Solutions configured for protection are expected to stop the attacker as early in the attack chain as possible. The responses given by the solution are matched to the mitigations available for the techniques executed. Custom and bespoke responses available to a responder will be awarded with wildcards and where possible will be communicated in the final report.

A full attack chain is executed to measure the protection capabilities of the solution. Where necessary, the attack will revert and resume from the last possible sequence to showcase the capabilities of the solution when one sequence successfully bypasses the solution.

For example:



In Attempt #1 the solution can prevent during the Infiltration sequence not allowing Propagation to be tested. To further showcase the capabilities of the tested solution later in the attack chain the attack will re-start from the Infiltration stage as in Attempt #2.

Efficacy score – the solutions ability to prevent all sequences of the attack.

Efficacy score is calculated as follows. Each successful sequence execution leads to a potential loss of points:

Sequence	Intrusion	Infiltration	Propagation
Score	+10/-10	+10/-10	+10/-10
Note	-	Lack of remediation of significant traces of the attack earn +5	

Public

Sequence 3: Propagation

This sequence is defined by the attacker progressing past the first intrusion.

ATT&CK Tactics applicable: Discovery, Lateral Movement, Collection, Exfiltration, Impact

	Discovery	*Lateral Movement*	*Collection*	*Exfiltration*
Propagation	T1087: Account Discovery	T1021.007: Cloud Services	T1560.001: Archive via Utility	T1048.002: Exfiltration Over Asymmetric Encrypted Non-C2 Protocol
	T1482: Domain trust Discovery	T1021.001: Remote Desktop Protocol	T1213.003: Code Repositories	
	T1135: Network Share Discovery	T1021.002: SMB/Windows Admin Shares	T1074.002: Remote Data Staging	

b. Blue Responses

Available responses to each phase of the attack are described under blue responses. The options given to defenders are assessed. Desired outcomes are prescribed alongside the red objectives when the test scheme is built.

Where possible, mitigations are tied to the corresponding ID in the ATT&CK framework. A technique can be employed a multitude of ways therefore the mitigation chosen by the evaluated solution must be appropriate to how the technique is employed by the tradecraft chosen by the attacker. Where possible MITREs D3fend taxonomy is used to describe the defenders options.

For more information check the latest full APT Methodology online - Identifier APT2025v1.1.

Appendix B – Cloud & Identity Methodology Abstract v1.0

2. Environment

A target organisation is maintained with digital identities and communication patterns to reflect roles within an organisation. Communications are defined as any exchange of information between two parties via an electronic method. These can be via e-mail, messaging platforms (Teams, Slack etc) or shared network resources.

These patterns are described below:

a. Internal common communication patterns and file types

Source	Destination	Direction	Frequency	Type
Purple staff	Purple staff	Bi-directional	High	Email, Direct Message, Office File types
Purple staff	Green staff	Bi-directional	Low	Email, Direct Message
Purple staff	All	One way	Low	Email, Direct Message
CTO	Alpha & Bravo	Bi-directional	Medium	Email, Direct Message, Office File types, Archives, Executables
CFO	Bravo & Charlie	Bi-directional	Medium	Email, Direct Message, Office File types, Archives, Executables
Blue staff	Blue staff	Bi-directional	Medium	Email, Direct Message, Office File types
Green staff	Green staff	Bi-directional	High	Email, Direct Message, Office File types
Alpha team	Alpha team	Bi-directional	High	Email, Direct Message, Office File types, executables

Source	Destination	Direction	Frequency	Type
IT Support Staff	All (group)	Bi-directional	Medium	Email, Direct Message, Office File types
IT Support Staff	All (individual)	Bi-directional	Low	Email, Direct Message, Office File types, executables

b. External common communication patterns and file types

Source	Destination	Direction	Frequency	Type
IT support, CTO, CFO	IT Supplier	One way	Low	Email, Office File types
CEO	PR Agency	Bi-directional	Medium	Email, Office File types
CEO, CFO	Law Firm	Bi-directional	Medium	Email, Office File types
CEO, CFO	Accountancy	Bi-directional	Medium	Email, Office File types
Accountancy	All (individual)	One way	Low	Email, Office File types
PR Agency	Blue & Purple	Bi-directional	Low	Email, Office File types

Within the organisation, shared network resources and cloud apps are accessed by the staff members.

3. Execution

The evaluation has 3 phases, behaviour baselining, attack and legitimate behaviour.

Behaviour baselining is executed for a recommended duration of two weeks where "normal" behaviour of the target organisation is observed. Key metrics will be taken are alert volume and any deviation necessary in behaviour due to the product interaction with the environment. Solutions can deviate from the two-week period if necessary to get an accurate understanding of the organisation, this will have an impact on the duration of the legitimate section of the evaluation and will be disclosed in the report.

The attacking period is comprised of two phases. The test corpus is split into scenarios based on specific goals of attackers. If tested solutions do not have complete coverage of the corpus these will be considered out of scope and disclosed in any reporting.

The attacking scenarios are focused on identity and cloud assets of the target organisation.

Identity scenarios focus on targeting an organization's digital identities with the primary objective of gaining unauthorized access or leveraging communication channels to establish a foothold within the organization. The key metrics are visibility and alert efficiency.

Cloud asset scenarios focus on targeting an organization's cloud infrastructure, services, and data with the primary objective of exploiting misconfigurations, vulnerabilities, or weak access controls to compromise sensitive assets or gain unauthorized access. While there is crossover of tradecraft between the two types of scenarios as identity is core point of exploiting a victim organisation.

Legitimate behaviour is executed after the attacking period with any necessary adjustments on the target organisation after the attacking period. The key metrics measure are alert efficiency and false positive rate.

4. Attack Rating

Scenarios include specific metrics to evaluate the solution's effectiveness in addressing the threat.

c. Identity Scenario Rating

Ongoing attack rating is represented by the detections occurring during the attacker's activity.

Post compromise rating is represented by any detections or data enriched after the activity has occurred.

Alert efficiency is calculated as a percentage using the baselined number of informational alerts gathered during the baselining period. The total number of alerts during each phase of the test is disclosed.

d. Cloud Asset Detection Rating

Ratings will be given per attack sequence. While direct references to the corresponding ATT&CK technique are useful they are not required to earn detection points. Data sources are marked and mapped to the corresponding source but do not influence the final accuracy rating. Within each defined tactic scope of a test case the solution must detect 50% of the techniques used to earn the maximum grade.

For example:

Sequence	Technique	Tactic	Min # Ts	Rating
Intrusion	T1566.001	Initial Access (1)	1	+10
Infiltration	T1204 T1059.001 T1047	Execution (3)	1	+10
	T1053.005 T1078.002 T1078.003 T1078.004 T1547.001	Persistence (5)	2	
	T1053.005 T1078.002 T1078.003 T1078.004	Privilege Escalation (4)	2	
	T1070.004 T1070.006 T1036.005 T1140 T1484.002 T1562.004 T1562.001 T1562.002 T1070.008	Defense Evasion (9)	4	
	T1555.003 T1606.001 T1003.006 T1558.003 T1539	Credential Access (5)	2	
Propagation	T1087.002 T1482 T1018	Discovery (3)	1	+10
	T1021.006	Lateral Movement(1)	1	
	T1005 T1114.002 T1560.001 T1213.003 T1074.002	Collection (5)	2	
	T1048.002	Exfiltration (1)	1	

For more information check the latest full APT Methodology online – Identifier - Cloud & Identity Evaluation 2025 v1.0