Operationalizing Threat Intelligence:

The Problems with Detecting at Relevance and Scale

ANOMALI

White Paper

According to recent reports, the average time to identify and contain a data breach in 2021 was 287 days.[1] This is why one key number that is generally accepted and that every CISO watches is “200-days.” While spending on security tools has increased, organizations are still often finding out they’ve suffered a breach after an attacker has infiltrated their network. Law enforcement, a business partner or independent researchers are often the ones to inform an organization that they’ve suffered a data breach.

In an effort understand possible indicators of compromise (IOCs), many organizations gather threat intelligence data. This data contains information about bad actors on the web, machine generated domain names, known bad IP addresses, lists of email addresses used for phishing campaigns and other data.

The threat analyst team will review the IOC data either manually or in a threat intelligence platform (TIP) to understand the relevance of the externally seen IOCs to their business in their vertical for their environment. To understand whether or not any of this data might be relevant for an organization, the organization’s security team will attempt to tie a relevant subset of the tens of millions of active IOCs collected from multiple sources to security information from inside their organization and contained in their security information and event management (SIEM) system log data.

> *“Traditional systems are good at monitoring what’s happening to your network. The good thing is, they detect breaches. The bad thing is that they do it after the fact.”*
>
> — Hugh Njemenze, President, Anomali

[1] Ponemon 2021 Cost of a Data Breach Report

© 2022 Anomali, Inc. All rights reserved.

SIEM Limitations

Today's modern SIEM is built on a concept from the late 1990s that data can be collected from a wide variety of security systems can be correlated to reduce false positives, highlight security issues, and support investigation and incident management. In the late 90s, no one could have predicted the large amounts of data streaming into the SIEM every second from a growing number of mission critical applications, security hardware and software solutions, and IT infrastructure. For the average enterprises, the math is obvious: 30,000 events per second (EPS) = 3 billion events per day = 600 GB per day = 120 TB over 200 days! The concept of threat intelligence is only about five years old. To address the need for additional methods of correlation, SIEM vendors added lookup and active list capabilities. These might scale to 500,000 items but certainly not consumption of tens of millions of indicators of compromise.

Pricing and scalability are often the two barriers to keeping enough data on-line for active breach detection over the 200-day window. The requisite computing power required for storage and correlation often isn't available to organizations. Many organizations only keep 30-90 days of data on-line and accessible to meet compliance regulations. In summary, the SIEM is not longer an adequate time machine for today's 200-day exposure window.

The Growing Volume of Threat Intelligence Data

Threat intelligence has evolved a great deal in a very short period of time. Many customers value having accurate and timely threat intelligence. There are hundreds of sources (free and paid), and many of them are of high quality. However, given the expertise of researchers at individual provider/vendors:

© 2022 Anomali, Inc. All rights reserved.

- One vendor will report an indicator of compromise before another
- Many threat feeds are offered in different formats making them hard to digest for downstream use in a SIEM
- None can tell you with 100% certainty that all their IOCs are relevant for your organization. This means that the threat intelligence vendors could be supplying millions of false-positives along with the few that are relevant.

Many organizations have subscribed to multiple feeds of threat intelligence. Some organizations subscribe to over 40. Threat Intelligence Platforms evolved to curate this data from multiple threat feeds and provide deduplication and post processing of the data to make it digestible for a SIEM.

Over the 200-days exposure window it is possible for an organization to have to review over 25 million active indicators of compromise from fee-based and open source threat intelligence providers. Since 2013, the number of active indicators has been growing at an average rate of 39% every month and shows no sign of abating.

Part of the growth in the number of IOCs comes from analysis of the almost one million never-before-seen threats being released into the wild on a daily basis and attackers that are using domain generation algorithms to create millions of throw-away domains. This technique called Domain Fluxing, is used for keeping a malicious botnet in operation by constantly changing the domain name of the botnet owner's Command and Control (C&C) server. It is also responsible for generating millions of random domains every hour. This reduces the value of threat data by reducing the overall timeliness.

The compute power required to correlate all the IOCs with 120TB of data simply isn't available to most organizations. Even if an organization is able to run a correlation search against the data, the search can take hours or days and in some cases never complete. The question of what IOCs collected are relevant for my environment over the last 200 days, goes unanswered.

© 2022 Anomali, Inc. All rights reserved.

> *"To defend against cyber attacks, it is important for a defender to have timely access to relevant, actionable threat intelligence and the ability to act on that intelligence."*
>
> — NIST 800-150

The Next Big Data Problem – Threat Intelligence Data

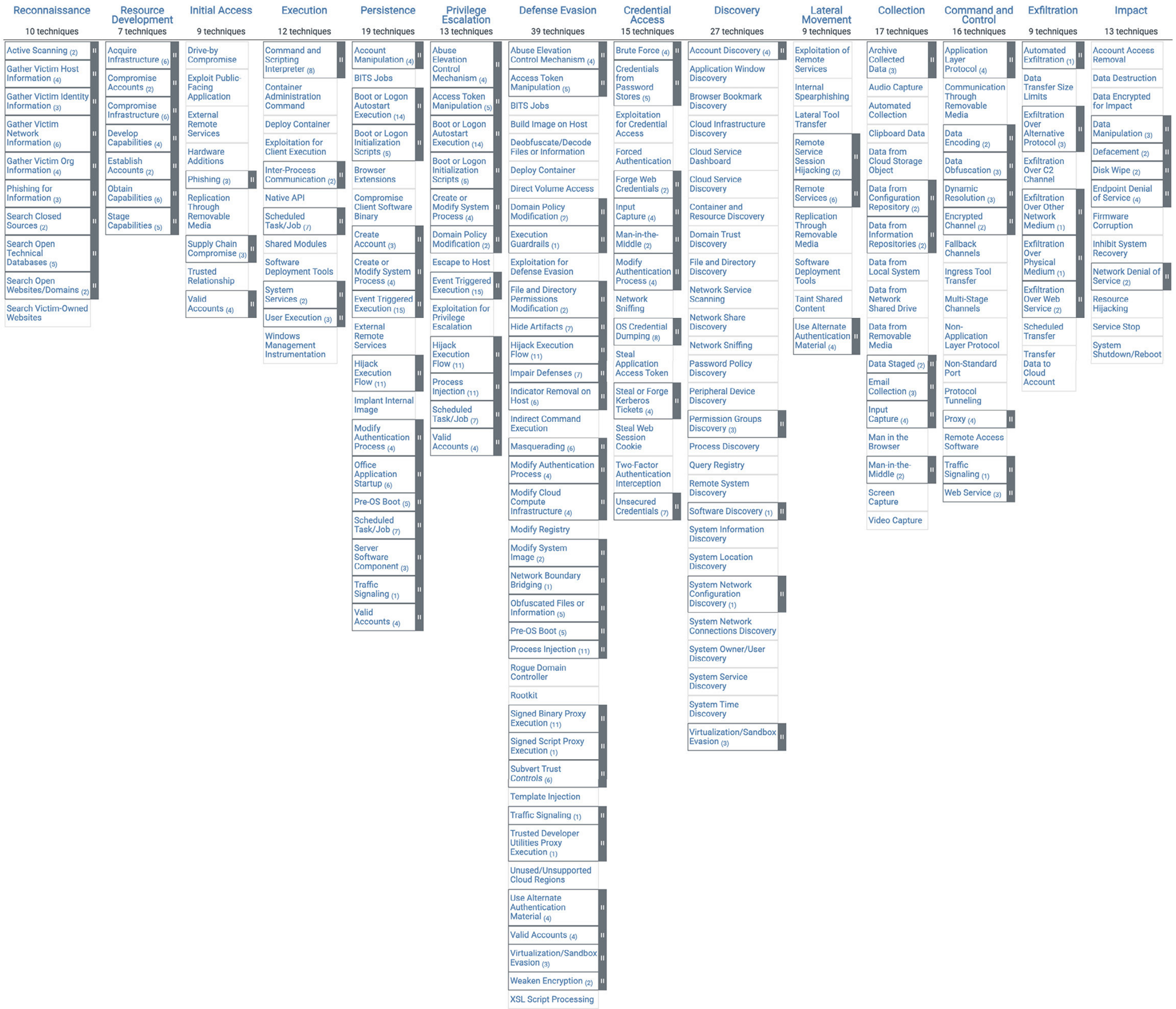
The SIEM has a role to play as the traditional single pane of glass for log data exploration and search, correlation for false-positive reduction, incident investigation, security metrics reporting (often for compliance) and facilitating incident management. Organizations have recognized that increasing amounts of relevant security data from traditional security and application data sources is already a scalablity problem. Creating matches between IOCs and security data is a new big data problem that requires a separate analytics engine that can address current and future scalability issues.

As security organizations mature, individual big data solutions that solve particular security problems will emerge that will make up a new layer of security ecosystem. These solutions will provide answers that will be fed back to the SIEM which will continue to facilitate data query, provide automated reports and perform security incident management and investigation.

© 2022 Anomali, Inc. All rights reserved.

A View of IOCs Across the Threat Models

The scalable correlation of tens of millions of IOCs and over 200-days of customer data gives organizations a unique real-time view into attacker activity and the ability to align attacks with current threat reference models. The Mitre ATT&CK Framework is fast becoming one of the most widely used tools to help organizations profile their environment, conduct investigations and make informed decisions. Analyst users can quickly identify key areas of concern in their environment, and prioritize their response appropriately. Correlation of organization's log data against five key indicators of compromise yields visibility though the lens of each model.



Reconnaissance 10 techniques	Resource Development 7 techniques	Initial Access 9 techniques	Execution 12 techniques	Persistence 19 techniques	Privilege Escalation 13 techniques	Defense Evasion 39 techniques	Credential Access 15 techniques	Discovery 27 techniques	Lateral Movement 9 techniques	Collection 17 techniques	Command and Control 16 techniques	Exfiltration 9 techniques	Impact 13 techniques
Active Scanning (2)	Acquire Infrastructure (6)	Drive-by Compromise	Command and Scripting Interpreter (8)	Account Manipulation (4)	Abuse Elevation Control Mechanism (4)	Abuse Elevation Control Mechanism (4)	Brute Force (4)	Account Discovery (4)	Exploitation of Remote Services	Archive Collected Data (3)	Application Layer Protocol (4)	Automated Exfiltration (1)	Account Access Removal
Gather Victim Host Information (4)	Compromise Accounts (2)	Exploit Public-Facing Application	Container Administration Command	BITS Jobs	Access Token Manipulation (5)	Access Token Manipulation (5)	Credentials from Password Stores (5)	Application Window Discovery	Internal Spearphishing	Audio Capture	Communication Through Removable Media	Data Transfer Size Limits	Data Destruction
Gather Victim Identity Information (3)	Compromise Infrastructure (6)	External Remote Services	Deploy Container	Boot or Logon Autostart Execution (14)	Boot or Logon Autostart Execution (14)	BITS Jobs	Exploitation for Credential Access	Browser Bookmark Discovery	Lateral Tool Transfer	Automated Collection	Data Encoding (2)	Exfiltration Over Alternative Protocol (3)	Data Encrypted for Impact
Gather Victim Network Information (6)	Develop Capabilities (4)	Hardware Additions	Exploitation for Client Execution	Boot or Logon Initialization Scripts (5)	Boot or Logon Initialization Scripts (5)	Build Image on Host	Forced Authentication	Cloud Infrastructure Discovery	Remote Service Session Hijacking (2)	Clipboard Data	Data Obfuscation (3)	Exfiltration Over C2 Channel	Data Manipulation (3)
Gather Victim Org Information (4)	Establish Accounts (2)	Phishing (3)	Inter-Process Communication (2)	Browser Extensions	Create or Modify System Process (4)	Deobfuscate/Decode Files or Information	Forge Web Credentials (2)	Cloud Service Dashboard	Remote Services (6)	Data from Cloud Storage Object	Dynamic Resolution (3)	Exfiltration Over Other Network Medium (1)	Defacement (2)
Phishing for Information (3)	Obtain Capabilities (6)	Replication Through Removable Media	Native API	Compromise Client Software Binary	Domain Policy Modification (2)	Deploy Container	Input Capture (4)	Cloud Service Discovery	Replication Through Removable Media	Data from Configuration Repository (2)	Encrypted Channel (2)	Exfiltration Over Physical Medium (1)	Disk Wipe (2)
Search Closed Sources (2)	Stage Capabilities (5)	Supply Chain Compromise (3)	Scheduled Task/Job (7)	Create Account (3)	Escape to Host	Direct Volume Access	Man-in-the-Middle (2)	Container and Resource Discovery	Software Deployment Tools	Data from Information Repositories (2)	Fallback Channels	Exfiltration Over Web Service (2)	Endpoint Denial of Service (4)
Search Open Technical Databases (5)		Trusted Relationship	Shared Modules	Create or Modify System Process (4)	Event Triggered Execution (15)	Domain Policy Modification (2)	Modify Authentication Process (4)	Domain Trust Discovery	Taint Shared Content	Data from Local System	Ingress Tool Transfer	Scheduled Transfer	Firmware Corruption
Search Open Websites/Domains (2)		Valid Accounts (4)	Software Deployment Tools	Event Triggered Execution (15)	Exploitation for Privilege Escalation	Execution Guardrails (1)	Network Sniffing	File and Directory Discovery	Use Alternate Authentication Material (4)	Data from Network Shared Drive	Multi-Stage Channels	Transfer Data to Cloud Account	Inhibit System Recovery
Search Victim-Owned Websites			System Services (2)	External Remote Services	Hijack Execution Flow (11)	Exploitation for Defense Evasion	OS Credential Dumping (8)	Network Service Scanning		Data from Removable Media	Non-Application Layer Protocol		Network Denial of Service (2)
			User Execution (3)	Hijack Execution Flow (11)	Process Injection (11)	File and Directory Permissions Modification (2)	Steal Application Access Token	Network Share Discovery		Data Staged (2)	Non-Standard Port		Resource Hijacking
			Windows Management Instrumentation	Implant Internal Image	Scheduled Task/Job (7)	Hide Artifacts (7)	Steal or Forge Kerberos Tickets (4)	Network Sniffing		Email Collection (3)	Protocol Tunneling		Service Stop
				Modify Authentication Process (4)	Valid Accounts (4)	Hijack Execution Flow (11)	Steal Web Session Cookie	Password Policy Discovery		Input Capture (4)	Proxy (4)		System Shutdown/Reboot
				Office Application Startup (6)		Impair Defenses (7)	Two-Factor Authentication Interception	Peripheral Device Discovery		Man in the Browser	Remote Access Software		
				Pre-OS Boot (5)		Indicator Removal on Host (6)	Unsecured Credentials (7)	Permission Groups Discovery (3)		Man-in-the-Middle (2)	Traffic Signaling (1)		
				Scheduled Task/Job (7)		Indirect Command Execution		Process Discovery		Screen Capture	Web Service (3)		
				Server Software Component (3)		Masquerading (6)		Query Registry		Video Capture			
				Traffic Signaling (1)		Modify Authentication Process (4)		Remote System Discovery					
				Valid Accounts (4)		Modify Cloud Compute Infrastructure (4)		Software Discovery (1)					
						Modify Registry		System Information Discovery					
						Modify System Image (2)		System Location Discovery					
						Network Boundary Bridging (1)		System Network Configuration Discovery (1)					
						Obfuscated Files or Information (5)		System Network Connections Discovery					
						Pre-OS Boot (5)		System Owner/User Discovery					
						Process Injection (11)		System Service Discovery					
						Rogue Domain Controller		System Time Discovery					
						Rootkit		Virtualization/Sandbox Evasion (3)					
						Signed Binary Proxy Execution (11)							
						Signed Script Proxy Execution (1)							
						Subvert Trust Controls (6)							
						Template Injection							
						Traffic Signaling (1)							
						Trusted Developer Utilities Proxy Execution (1)							
						Unused/Unsupported Cloud Regions							
						Use Alternate Authentication Material (4)							
						Valid Accounts (4)							
						Virtualization/Sandbox Evasion (3)							
						Weaken Encryption (2)							
						XSL Script Processing							

© 2022 Anomali, Inc. All rights reserved.

Evolving to Face Today's Threats

Becoming a Modern SOC must include detection capabilities that do not rely solely on alerts, but one that can scale to collect over 200 days of data from logs and match millions of externally collected IOCs in real-time and:

- Operationalizes threat data for immediate use for incident responders and SOC operations personnel
- Provides for SIEM scalability and makes the SIEM smarter
- Means that indicators provided are relevant to your specific company or agency
- Allows you to identify and measure which streams of threat intelligence are the most relevant for your organization
- Moves from security operations center (SOC) to Intelligence-driven security operations center (ISOC[2]).
- Enables security teams to leverage advanced analytics across a large and wide range of telemetry to detect threats.

Summary

To address growing scalability issues, organizations will need to embrace a distributed computing approach to security. In addition, Threat Intelligence Platforms will need to create organizational relevance for threat intelligence data. The SIEM will continue to collect log data from traditional security sources and the next generation of threat intelligence platforms will have to answer the question of what's relevant across the window of exposure. These answers to the relevance question will be pushed into existing workflows for analysis and automated response.

[2] How to Build and Operate a Modern Security Operations Center

About Anomali

Anomali offers intelligence-driven extended detection and response solutions that help organizations quickly identify and respond to threats in real-time by automatically correlating ALL security telemetry against active threat intelligence to expose "(un)known" threats. Anomali Match uses automation to correlate tens of millions of threat indicators against your real time network activity logs and up to 5 years of forensic log data. Anomali's approach enables detection at every point along the kill chain, making it possible to mitigate threats before material damage to your organization has occurred.

© 2022 Anomali, Inc. All rights reserved.