elastic

Search. Observe. Protect.

Guide to high-volume data sources for SIEM

elastic.co

The need for more security data

The mission of solving security problems has always been complex and multifaceted, posing challenges with:

Visibility
Maintaining a high enough level of awareness across the environment

Focus
Identifying issues and knowing which ones are high priority

Speed
Addressing the highest-priority issues in a timely manner

Scale
Avoiding recurrences of known issues while identifying unknown issues

To meet these challenges, **security teams must have easy access to the right data**. However, "the right data" can be a fast-moving target due to global trends that further complicate these challenges. These global trends are accelerating and forcing security teams to collect more data, and more types of data, to keep up.

Why is there a need for more data? To understand this better, let's look at some trends in attack motives and attack methods.

Adversaries are more motivated

Cybercrime is driven by increasingly lucrative opportunities. According to a 2020 report from NIST, the total activity and assets at risk due to cybercrime in 2017 were valued at $11.9 trillion in the US alone. To the criminal syndicate, this represents an enormous opportunity.

The same report demonstrates total losses across all industries in the US alone to be between 0.9% and 4.1% of the total US gross domestic product (GDP), or between $167.9 billion and $770 billion. Consider that the analysis from 2020 used a Bureau of Justice statistics dataset from 2016, and also consider that the situation has since worsened: the author of the report asserts that "widely accepted estimates of cybercrime loss may severely underestimate the true value of losses."

In addition to financial gain, motivations to attack include nation-state politics/espionage and cyberwarfare, corporate espionage, or to gain control of proxy resources to facilitate a larger attack whose mission could be any of the above. In some cases, the adversary's mission is simply to be destructive.

Due to these high-stakes motivating factors, adversaries will execute more calculated and deliberate actions to accomplish their missions, by:

Leveraging more evasive methods: According to Mandiant's 2020 M-Trends Report, of all the malware families observed in the last year, 41% were previously unknown. This general trend is not specific to malware — investments in machine learning and threat hunting functions within the SOC are driven by the need to identify unknown threats.

Patiently executing an attack: Mandiant reports that overall median dwell times are down globally, which is good news overall, as it could mean organizations have better detection programs. But many investigations still show dwell times greater than nearly two years, and those may again be on the rise (see next section). Some disruptive and damaging attacks (ransomware, cryptominers) have traditionally shown shorter dwell times but some are now coupled with more patient attacks methods as well.

The challenge for security teams is to ensure they have all the context needed to verify threat activity early and to quickly investigate and contain it in an efficient manner. This translates directly to the need for more data:

To combat evasion: Security teams need a greater variety of data sources to gain more context, such as finding entry from uncommon vectors, identifying unusual activity as adversaries go out of their way to avoid triggering alerts, and performing analytics on a larger set of "typical" activities that alone might not be alarming, but that in combination might indicate malicious intent.

For long dwell times: Security teams need more historical data — again, from a wide variety of rich data sources — to obtain enough context to look back over not only months, but years, to find evidence of an initial breach and associated compromises of assets and user accounts, as well as to profile attacks and statistically analyze and identify patterns of activity in threat actor behavior.

Attack motivations will continue to evolve. Imagine a criminal syndicate working with an insider to short sell an organization's stock after hurting the brand through nontraditional means, such as social or media influence. Or imagine a malware campaign designed to phish developers committing changes to a project on GitHub. These are not as "classic" as some cybersecurity scenarios, but security teams will need to be prepared — new motivations mean new attack methods that coexist with, not replace, older methods.

Attacks range from basic to beyond sophisticated

Threat actors use a fast-growing barrage of new techniques. These techniques span a broad range of sophistication, from exploiting common and basic misconfigurations (especially in the cloud), to advanced persistent threats (APTs) that use highly nuanced reconnaissance methods and social engineering tactics to help actors gain a foothold, move laterally, and exfiltrate data.

Therefore, it is critical for security teams to protect against both basic and advanced methods:

- Basic methods often exploit configuration errors: This is a persistent and growing problem — according to Verizon's 2020 Data Breach Investigations Report, "*...the only action type that is consistently increasing year-to-year in frequency is Error.*"
- Advanced methods can exploit any potential available foothold: The Mandiant report mentioned previously asserts that "Traditional barriers to attacker success continue to lessen over time. Put simply, *more attackers can do more things in more diverse environments*."

Looking at these together, **the challenge is, again, having enough data available**. Security teams need to maintain and improve:

- Consistent security posture: Ensure visibility and the ability to monitor and investigate in an efficient manner across both on-premises and cloud (IaaS and SaaS) environments.
- Easily accessible security-relevant context: Hunt with better-formed hypotheses, using as much diverse environmental data needed for higher-fidelity detections, more complete investigations, and better targeted responses.

In light of these ongoing and accelerating trends, how are security teams expected to keep up and even get ahead? In the security community we are all too aware of the "skills gap" — of course, we need to stay on top of training and stay abreast of the latest in attacker methodology and classes of threats, while we continually evaluate technologies and countermeasures that can help us. **Can our underlying technologies themselves help us in closing the skills gap?**

Below are a few examples of high-profile classes of attacks:

Cloud-based data breaches

Of the incidents that result in loss of data privacy, those that make headline news typically result from attackers finding and exploiting cloud service misconfiguration. According to a commissioned survey from IDC, nearly 80% of companies surveyed had experienced at least one cloud data breach in the past 18 months, and nearly half (43%) reported 10 or more breaches. Examples of high-profile attacks over the past several years include access through misconfigured web application security services, publicly accessible cloud compute instances, and unsecured databases deployed in public cloud environments. These attacks allowed threat actors to exfiltrate sensitive data and customer records.

New types of disruptive attacks

SUNBURST was massive, not only in terms of scale and scope, but in terms of forcing us to address a new level of threat: nation-state sophistication targeted at corporations. Just as NotPetya had lasting financial and technical repercussions, SUNBURST will likely have a long-term effect in many ways, not least of all how we collaborate as an industry and how we evaluate and address security from a supply chain perspective. The attack is extremely sophisticated, leveraging antidetection countermeasures and custom components per victim to reduce indicator of compromise (IoC) efficacy, and evades sandboxes and uses advanced techniques to hide beaconing and command and control (C2). Operationalizing IoCs is challenging because key context can come from sources such as domain name system (DNS), endpoint, certificate transparency logs, proxies, and many other sources. General guidance from security research teams also includes the recommendation for richer logging with much higher retention — investigations showed dwell times of over a year, providing plenty of opportunity for lateral movement and for actors to progress to later stages of their attack.

ICS/SCADA, OT, IoT attacks

These have continued evolving steadily and at a high rate since Stuxnet and the Mirai botnet. Brickerbot-inspired malware infects and destroys filesystems and device configurations, and other attacks access the environment through connected industrial control systems (ICS) and supervisory control and data acquisition (SCADA) systems, connected physical surveillance devices, and vulnerable services on industrial Internet of Things (IoT) and operational technology (OT) systems and devices may be insecure due to architecture, design, misconfiguration, or a lack of compensating controls. These can lead to access to IT systems, but the direct threat to disruption of services is also a top-priority issue. Manufacturers have suffered costly downtime, and the threat to critical infrastructure presents a risk to physical safety and national security.

In general, for these and many other classes of attacks, the commonality is that there is greater sophistication and patience — more attacks commonly leverage more of the environment to avoid detection, as well as longer timeframes to accomplish a much greater, more damaging mission with ongoing repercussions. Unlike in the past, these threats can and do target organizations directly, regardless of their motive or source.

As security teams mobilize further in addressing the larger security concerns resulting from the above trends, it is critical that they are not forced to be overly selective about what types and amounts of data to include or exclude in day-to-day operations.

High-volume data sources are critical

The above trends are causing many security teams to revisit the data layer. This is not a new phenomenon — security teams revisited the data fabric layer when relational databases, used to build earlier generations of security information and event management (SIEM) products, were struggling to support the scale and speed needed in the SOC. As a result, earlier SIEMs gave way to a "big data" approach to security analytics to handle multistage attacks.

Modern security teams are revisiting the data layer again, this time driven by the desire to avoid the pain of compromising their access to important data. These teams are too often forced to spend time and resources deciding which data to include and which data to leave out of day-to-day operations. They require more security-relevant data across a greater variety of environmental sources, including cloud infrastructure and cloud applications, richer endpoint data, DNS, NetFlow, wire data, IoT, and other data sources that are rich in context and typically higher in volume than some of the primary data sources ingested for daily operations by their SIEM.

In many cases, **high-volume data can present technical, operational, and business challenges** — either the sheer volume of data makes it difficult for analysts to perform responsive queries and analyses without a more robust architecture, or those data sources are deemed cost-prohibitive by architecture or engineering teams, and SOC managers don't have the budget to support the addition of more data sources.

Typical concessions these teams are forced to make can include:

- Shorter retention times for high-fidelity data used in SIEM rules/detections and machine learning
- Lower data verbosity/logging level for data used in incident investigation and response
- Altogether dropping/excluding “lower immediate value” data from use by the SOC
- Archival with limited or slow access, or even limited availability/accessibility (only a portion of the data)
- Archival but by other teams, e.g., for use by separate hunt teams, breach, or forensic analysts

High-volume data sources can help provide visibility into evasive activity, as well as the rich details needed to contextualize a threat. With the right archiving strategy, those data sources can provide the historical context needed to perform longer look-back analyses in response to a security incident or data breach, or for proactive threat analysis and adversary profiling.

Below is a table outlining the general security value of higher-volume data sources and the importance of including these data sources in day-to-day security operations. In light of the trends described above, these data sources are increasingly security-relevant and can provide critical context needed in detections, hunts, investigations, and incident response to help security teams verify, scope, and act more quickly — and, most importantly, with the right level of context needed so that those quick decisions have greater impact on containing threats effectively and minimizing downstream risk.

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
IaaS and SaaS	Cloud security posture monitoring, compliance, context-aware asset visibility, user monitoring, threat detection, incident investigation, detecting data exfiltration, identifying lateral movement	Cloud platform - As organizations increasingly deploy production workloads at scale in IaaS environments such as AWS, GCP, Azure, it is critical to maintain visibility across the hybrid environment.	Observability data, Cloudtrail, IAM, certs, infra config changes, unauthorized resource usage, permissions, security policy (FW-type - src/dst/port/protocol), o11y metrics, billing, WAF, VPC Flows	Cloud infrastructure and workload security can be a default priority instead of an afterthought, and "baked in" directly to the DevOps cycle, when IaaS visibility is a standard part of the security operations team's purview, and not treated as a separate silo of operations. In some cases, drops in performance can be indicative of a security issues.	Instance abuse (cryptomining etc), Billing changes/surges, Instance deletion, cloud discovery, DoS, Defacement, Application and system exploits, configuration changes, Data Theft
IaaS and SaaS	Cloud security posture monitoring, compliance, user monitoring, threat detection, incident investigation, anti-phishing, threat hunting, business context enrichment, detecting data exfiltration, behavior analysis	Cloud application logs - The acceleration to a remote workforce results in more business and sensitive data residing in cloud applications such as O365, Google Workspace, and Salesforce.com.	Observability data (APM), configuration changes, audit logs, application access, file access, user name, timestamps, service usage	Cloud application logs can provide business context needed to identify potential security issues, including insider threat. Cloud application logs can help with email tracing, behavioral baselining, and other investigative methods to help find activity associated with a threat.	Session, Token and Cookie theft, Account Discovery, Data Exfil, DoS, Account Manipulation, Phishing, Software Discovery

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
IaaS and SaaS	Monitoring and cyberhygiene, threat detection, incident investigation	Communication Apps (Zoom, Slack, Teams, etc) Digital transformation and remote work are pushing teams to collaborate more extensively in real-time, and organizations are leveraging tools such as Zoom, Slack, and Microsoft Teams to help them stay connected.	timestamps, source ips, users names, actions taken, login information, attachments, message interactions, recordings, deletions	With the increase in remote workforce, abuse and malicious behavior such as user impersonation or eavesdropping on meetings with corporate-sensitive discussions is on the rise. Protecting against these by monitoring usage on meeting collaboration tools has become more important than ever.	Data Exfil, Account Takeover, Phishing, Zoom Bombing
User	User monitoring, privileged user monitoring, compliance, context-aware asset visibility, threat detection, incident investigation, threat hunting, identifying lateral movement, behavior analysis, hunt augmentation, resolution management/ containment	Authentication - Authentication systems verify a user is who they say they are by challenging them to provide a predetermined piece of information (something they know or have) to validate their identity. Organizations use premises-based tools as well as cloud-based services such as Okta.	Credential information and user activity, including what devices, services and applications were used, when.	Authentication logs can be used to identify users that have been compromised and potentially uncover what data may have been exfiltrated. Authentication events can help identify unusual user activity by looking at time (day/night), frequency, regularity, and pattern of access (what devices, what services/applications) to identify trends that help security teams distinguish between “normal” vs. “suspicious”. For users that are known to be compromised, the logs can help identify which assets and systems, services and applications were accessed, which can help during investigations to determine the type of data that may have been stolen or compromised.	Brute Force Attacks, Golden Ticket, Credential Theft, Privilege Escalation, Credential Stuffing, Password Spraying, MITM

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Network activity	Network activity monitoring, compliance, context-aware asset visibility, threat detection and prevention, malware prevention, ransomware protection, threat hunting, incident investigation, identifying lateral movement, detecting data exfiltration, behavior analysis, resolution management/ containment	Next-Gen Firewall (NGFW) - Firewall functionality can be delivered as a stand-alone device (physical or virtual) or as part of an integrated gateway solution. It controls access to the network and performs multi-function inspections to try to identify attacks within the traffic.	Source IP, source port, destination IP, destination port, service, application, user, traffic flow information. NGFW logs will also include threat context from threat intelligence (signature match, payload analysis, threat type, threat source, threat intelligence source)	Firewall logs provide insights into communications to the IP space, applications and users. They can be used to: Uncover resources, services and applications being targeted. Identify malicious traffic trends (command and control traffic, geo-location spikes, etc.). Traffic to the same URL at the same interval every day (which could be malware beaconing (phoning home) to notify attacker of successful installation and get further instructions. Flag unusual activity to a controlled asset. Highlight traffic reversals that represent traffic going in a direction that is not normal to a device, which indicates that device has been compromised.	DoS, Application Exploits, OS Exploits, Protocol based attacks, VLAN Hopping, malware delivery

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Network activity	Network activity monitoring, compliance, context-aware asset visibility, threat detection, threat hunting, incident investigation, identifying lateral movement, detecting data exfiltration, behavior analysis	VPN - Virtual Private Networks (VPNs) provide remote access capabilities into corporate resources, for employees and sometimes partners. Depending on the vendor, VPN capabilities may be integrated directly into a perimeter firewall, or they may be deployed separately as a VPN concentrator.	IP addresses, user names, login times, session duration, auth failures, policy check failure, frequency of session timeouts, key exchange issues, tunnel metrics including location	A compromised remote access VPN account can allow an attacker full access to corporate resources that reside on-premises. VPN logs can provide insights into unusual activity such as unusual login times, login from multiple physically distant locations at once (landspeed violation), host policy violations, and other activity that should be flagged for investigation.	MITM, session hijacking, brute force attempts, DoS/SYN flood, spoofing
Network activity	Network activity monitoring, compliance, context-aware asset visibility, threat detection and prevention, malware prevention, ransomware protection, incident investigation, threat hunting, identifying lateral movement, detecting data exfiltration, behavior analysis	IDS/IPS - Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) look for and blocks attacks within network traffic. They use a multi-pronged approach to detection to identify potentially malicious behavior and generate alarms. High confidence alerts may be automatically blocked by an IPS to prevent propagation into other areas of the network.	Attack definition, CVE information, vulnerability data, as well as general traffic info: source IP, source port, destination IP, destination port, time, services and applications.	IDS/IPS logs provide insights into network attack vectors. They often provide the first clue, generating the alert that leads to an investigation. They can be used to identify the specific exploit and vulnerability, which can help identify other potential victims throughout the network; verify reconnaissance activity, investigate the path of an attack, follow additional clues that paint a more complete picture; flag unusual activity (e.g. new lateral traffic through the network that looks suspicious); and gain protocol-level insights.	DDos, Smurf Attack, Ping of Death, Fragmentation, Probing, ARP Spoofing, Port Scanning, Fingerprinting, TLS Evasion

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Network activity	Network activity monitoring, compliance, context-aware asset visibility, threat detection and prevention, malware prevention, ransomware protection, threat hunting, incident investigation, detecting data exfiltration, behavior analysis, resolution management/ containment	Web Proxy Servers - Protects web traffic, blocking malicious URLs (matched to a threat/reputation database), scanning web content and performing malware analysis to provide insights into the threats you are facing.	Inbound/Outbound Web traffic, including information on URLs, domains and portable executables (files, exe, DLL, docs, JavaScript, etc.).	Web proxy logs can be used to tie users/ systems involved in an attack to web traffic to: Identify initial infections that originated from new, malicious sites or legitimate Web sites that were been compromised. Uncover data exfiltration by uncovering unusual outbound activity (e.g. looking for frequency and a pattern of “automated” transmissions.) E.g. Beaconing (phone home) traffic of malware trying to establish communication with C2 to get instructions (traffic to the same URL at the same interval every day) Confirm infections by correlating to payload analysis.	Data Exfil, C2 Communication, Phishing attempts, Droppers

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Network activity	Network activity monitoring, compliance, anti-phishing, malware prevention, ransomware protection, incident investigation, detecting data exfiltration, behavior analysis, threat hunting, enrichment and threat intelligence, hunt augmentation	DNS - A Domain Name Server (DNS) associates IP addresses with a unique name that identifies a particular computer, service or resource connected to the Internet or a private network.	IP addresses, assets and their correlating domain name, users/systems that conducted DNS lookups, time and location.	DNS logs can be used to look for post infection activity. They can be used to: Identify an attacker looking for ways to communicate back to the command and control server (if using domain names (not IP) they will need to do a DNS lookup); can correlate that information with outbound connections. Uncover: Failed DNS lookups Suspicious domains Transmissions that used the DNS protocol Anomalous DNS activity – e.g. high number of DNS requests coming from a particular client, compared to a baseline	Cache poisoning, domain hijacking, dns flood attack, DRDoS, DNS Tunneling, Subdomain attacks
Network activity	Network activity monitoring, compliance, context-aware asset visibility, incident investigation, identifying lateral movement, behavior analysis, threat hunting, hunt augmentation	DHCP - A Dynamic Host Configuration Protocol (DHCP) allows an IP address to automatically be assigned to a computer from a range of numbers that have been configured for a particular network.	IP addresses, MAC addresses interfaces, services, DHCP requests, time and location	DHCP logs can be used to map user and device activity on the network. If there is suspicious connection, DHCP can be used to identify the specific machine that initiated the connection, providing vital details on which devices have been compromised and are being used as part of the attack.	DHCP Poisoning, DHCP Starvation, MITM, DHCP Spoofing, Theft of Service

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Network activity	Network activity monitoring, cyberhygiene, behavior analysis, identifying lateral movement, incident investigation	Wifi Access Points Wifi Access Points are used throughout the enterprise to allow for secure internal wireless access to corporate-connected resources, and also to provide connectivity to an organization's Internet gateway for both employees and guests. Wifi access points are typically configured to allow free roaming throughout a campus, therefore collecting association/event logs is critical to maintaining security visibility.	Timestamp, error/event message, mac address, hostname, etc	Client connection to access points are tracked. It is important to know which access point a client used to connect to a network — it's a way of discovering unauthorized wireless supplicants attempting to connect to the network, rogue/ fake access points configured to force association from unsuspecting users, or an attempt to connect point-to-point to legitimate access points to sniff traffic.	Unauthorised access, Physical device compromise, MITM, Evil Twin, Rogue Access Points, Wardriving, Warshipping, Packet Sniffing
Network activity	Network activity monitoring, cloud security posture monitoring, context-aware asset visibility, threat detection and prevention, detecting data exfiltration, behavior analysis, incident investigation, resolution management/ containment, hunt augmentation	DLP systems - Data Loss Prevention (DLP) systems build a digital security perimeter around the enterprise and analyze all outgoing (and sometimes incoming) data.	Timestamp, DLP Event type (print, usb, email, etc), policy breached, block/permitted, user name, ip, data size, document name, etc	DLP Systems can monitor many different forms of data leakage. Anything from copy/ pasting from a document, to including specific text in an email. Crucial for data exfiltration use cases.	Data Exfil

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Network activity	Network activity monitoring, compliance, context-aware asset visibility, threat detection and prevention, threat hunting, incident investigation, identifying lateral movement, behavior analysis, resolution management/ containment	Web Application Firewalls - A specific form of application firewall that filters, monitors, and blocks HTTP traffic to and from a web service.	HTTP headers, source/ client IPs, HTTP request payloads, policy hit and matches and scores, timestamp, user agents, geo, etc	Web application firewalls are design to spot layer 7 attack on web server/sites. Tremendously vital information, as the entire payload is checked for potentially malicious behaviour based on rules, with the ability to block if necessary.	OWASP Top 10 and other Layer 7 attacks (including DoS), Account Takeover, Session Hijacking, Heartbleed, Shellshock, Poodle.
Endpoint	User and application monitoring, compliance, context-aware asset visibility, threat detection and prevention, anti-phishing, file integrity monitoring, malware prevention, ransomware protection, threat hunting, incident investigation, identifying lateral movement, detecting data exfiltration, behavior analysis, resolution management/ containment, hunt augmentation	Endpoint Security Solutions (AV, EDR, EPP, anti-spyware, host-based firewalls / IPS) Endpoint security solutions look for and may block attacks on the endpoint. They use a variety of detection mechanisms to identify potentially malicious behavior and generate alarms. High confidence alerts may be automatically blocked to prevent propagation.	Endpoint security logs may contain specific file names, attack definitions, vulnerability information, executables, new applications,	Endpoint security solutions could provide insights into suspicious/attack activity on the device. They often provide the first clue, generating the alert that leads to an investigation. They can be used to: Identify the specific exploit and vulnerability, which can help identify other potential victims throughout the network. Investigate the path of an attack, following the bread crumbs that lead to a more complete picture.	Malware and Ransomware, Fileless attacks, Memory Modification, Data Exfiltration, Credential Based attacks, Insider Threat, Registry Manipulation

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Endpoint	User and OS monitoring, compliance, context-aware asset visibility, threat detection, anti-phishing, file integrity monitoring, malware prevention, ransomware protection, threat hunting, incident investigation, identifying lateral movement, detecting data exfiltration, behavior analysis, hunt augmentation	Endpoints (OS, event logs) Verbose endpoint (OSquery, sysmon, other beats) All endpoints record all their activity within their event and OS (Windows, Apple, Android, etc.) logs.	Endpoint logs contain system information, such as names of running processes, changes to credentials, privilege escalations, time, frequency and location of activities. Event and process IDs, system metrics (CPU, memory utilization, temp), FIM, anomaly detection (user access, file access, process launch/kill/etc)	Endpoint logs could: Record an attacker installing and running malware. Identify automatic processes and programs that might be used by the attacker to conduct malicious activity (e.g. automatically send data to a server as soon as it is saved to the endpoint) Flag unusual administrative activity (e.g. deleting event logs, which could indicate and attacker trying to delete evidence)	Malware and Ransomware, Fileless attacks, Memory Modification, Data Exfiltration, Credential Based attacks, Insider Threat, Registry Manipulation
Server	User monitoring, anti-phishing, malware prevention, ransomware protection, detecting data exfiltration, behavior analysis, threat hunting, incident investigation, enrichment, resolution management/ containment, hunt augmentation	Mail Servers - Mail servers handle all email transmissions in and out of the organization.	Email logs contain sender/recipient information, as well as the email payload, including any links or attachments.	Mail logs can be used to scan attachments for malware/viruses and compare email domains to databases of known spam/ phishing attacks. They can be used to: Identify the source of an attack, pinpointing when and how an endpoint that is behaving strangely was infected by a malicious link/ attachment. Correlate the click through or installation of a malicious link/ attachment with outbound activity on that endpoint to learn more about the exploit and identify the progression and spread of the attack.	Phishing Campaigns, Malware/Ransomware Delivery, Data Exfiltration, Extortion, Insider Threat,

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Server	Monitoring and compliance, context-aware asset visibility, privileged user monitoring, threat detection, detecting data exfiltration, behavior analysis, threat hunting, incident investigation, hunt augmentation	Database Audit Logs - Databases are some of the most crucial production hosts in an environment, as they house IP, PII, CC info, and more.	timestamp, database user, database name, database table, query, response time, byte size, authentication method, etc	Knowing the who, what, and when of database queries is absolutely vital — and is also a compliance requirement. Elastic Security can model queries and detect usual query patterns between applications and their associated database.	Data Exfil, Insider Threat, Data Manipulation, Data Defacement, Financial Fraud
Server	Monitoring and compliance, threat detection, cloud security posture monitoring, incident investigation, threat hunting	Certificate Transparency Logs Certificate authorities (CAs) are required to publish a public log of every single certificate they sign, which is millions every day.	Timestamp, Log entry/ store, entry stage, common names, subject alternative names, creation date, expiry, fingerprint	Extremely valuable to detect rogue certificate generation, certificates generated for phishing and more.	Phishing, CA compromise, Insider Threat
Server	Monitoring and compliance, threat detection, cloud security posture monitoring, incident investigation, threat hunting	CDNs (Cloudflare, Akamai, etc) CDNs host all static web assets for an organization's website, and many times combine WAF.	Very similar to WAF/ HTTP but can also include a combination of other fields/values depending on the services the CDN offers	These will be very useful for determining web based attacks, just like web server logs.	OWASP Top 10 and other Layer 7 attacks (including DoS), Account Takeover, Session Hijacking,etc
Wire and flow data	Monitoring and compliance, context-aware asset visibility, threat detection, incident investigation, threat hunting, hunt augmentation, behavior analysis, identifying lateral movement, detecting data exfiltration, anti-phishing, malware prevention, ransomware protection	PCAP - Short for packet capture, serves as an API for capturing network traffic.	protocol, source and destination ips and ports (depending on protocol), frames and sequences, client/ server ciphers for TLS handshakes, payloads, etc	Similar to netflow and other network based logs, without the dependency of relying on OS collection. Being able to collect network traffic as it appears "on the wire" allows a team to spot attack traffic that spans 1000's of hosts by using a network tap, and also allows to see if malware/ users have interfered with TCP/UDP streams as they leave a host.	Remote Access, C2 Communication, Malware Download, TLS Injection, Session hijacking, Data Exfiltration, DGA attacks, Watering hole, Brute Force

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Wire and flow data	Monitoring and compliance, context-aware asset visibility, threat detection, incident investigation, threat hunting, hunt augmentation, behavior analysis, identifying lateral movement, detecting data exfiltration	Flow data - Analyzed to monitor network performance, optimize infrastructure, and detect malicious traffic. NetFlow - Introduced on Cisco routers (1996) to provide the ability to collect IP network traffic as it enters or exits an interface. IPFIX - Internet Protocol Flow Information eXport (IPFIX) has since superseded the NetFlow protocol.	source and destination IPs and ports and bytes sizes, combined flow size, protocol, packet count, flow/community ids, flow duration, etc	Determining irregular data transfers and significant outbound connections, with the ability to model this behaviour using ML	Data Exfil, DoS
Connected devices and physical security	Monitoring and compliance, threat detection, incident investigation, threat hunting, hunt augmentation, behavior analysis, detecting data exfiltration	ICS - An industrial control system (ICS) consists of integrated hardware and software that monitors and controls the operation of industrial equipment. IoT - Internet of Things (IoT) includes devices for measuring, monitoring, and controlling physical devices through cloud-based processes.	Device name/ID/IP, message, message type/producer, severity, location, unit	Just like any other device with a network connection, IOT and ICS generate traffic, which can, of course, be the avenue of attack delivery, data theft, rogue manipulation and more (DDoS etc).	DoS, Account Takeover, Device Takeover, Module Discovery, Data Exfil, DoC, File Injection

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Connected devices and physical security	Monitoring and compliance, threat detection, incident investigation, threat hunting, hunt augmentation, behavior analysis, detecting data exfiltration	MDM Logs - Mobile device management (MDM) logs help enhance corporate data security by enabling the monitoring, managing, and securing of employees' various portable/mobile devices.	Device name, device action, policy, user name, IP address, owner, MDM action etc	Any organisation that has any form or corporate mobile devices would (and should) be monitoring and managing these devices through an MDM, particularly corporate phones, iPads, etc. Changes to policies, unusual app installs, login locations, lost devices and more can all be logged and monitored	Physical Theft, Data Exfil/Theft, Mobile Supply Chain attacks, Mobile Malware, Credential Misuse
Connected devices and physical security	Monitoring and compliance, threat detection, incident investigation, threat hunting, hunt augmentation, behavior analysis, detecting data exfiltration	CCTV Motion Events - Closed circuit television-detected activity at a given location.	Endpoint security solutions could provide insights into suspicious/attack activity on the device. They often provide the first clue, generating the alert that leads to an investigation. They can be used to: Identify the specific exploit and vulnerability, which can help identify other potential victims throughout the network.	Similar to physical access control, we can determine irregular motion events, and, if face recognition is included, we can identify and model user irregularities	Social Engineering, Data Exfil, Physical Theft, Insider Threat
Connected devices and physical security	Monitoring and compliance, threat detection, incident investigation, threat hunting, hunt augmentation, behavior analysis, detecting data exfiltration	Printer Logs - Enable security teams to keep track of print jobs on their business' printers.	Printer name/ip, source ip/user, sheets printed	Using non-traditional data sources, data exfil can also be detected in the physical world. If a user has never printed 300 pages before, at 6am, we probably want to take action on it	Data Exfil

High-volume data sources for SIEM

Category	Primary Use Cases	Data source (name & description)	Fields within the data / classes of logs / things to look for	Security significance / value	Examples of attacks / attack methods you can uncover with this data source
Code repository	Monitoring and compliance, threat detection, incident investigation, threat hunting, hunt augmentation, behavior analysis, detecting data exfiltration	Source Control Logs - Allow for monitoring of actions your users perform on source-controlled objects (GitHub, etc.) across your organization.	timestamp, repository names, user names, organisation names, action taken, source ips, etc	As organisations start doing everything "as code," it is vital to monitor irregular activities on repositories, as well as things like password/ secret access, api calls, logins and more.	Supply Chain Attacks, Insider Threat, Data Exfil, Code manipulation, Secret discovery and theft

Let your data help you close the skills gap

As mentioned on page 4 above, in cybersecurity, the pain of the skills gap has been all too real for many organizations. This critical issue has been compounded as digitization accelerates — more attack surface, unfamiliar environments, new data formats, and the need to reverse engineer emerging methods and new attack tactics, techniques, and procedures seen in the wild.

For example, security teams can get bogged down in learning the new "language" of cloud services — which cloud functionality is equivalent to devices deployed on-premises, architecture and deployment scenarios, vendor-specific logging formats, security-relevant fields within the data they need to focus on, and so on — for each cloud provider's specific environment. Add wire data, NetFlow, IoT, and a few other less commonly logged data sources into the mix, and teams can get overwhelmed and struggle to keep up.

Streamlining access to the right data sources is the first step to helping your organization to close this gap. With Elastic, security teams can stop worrying about whether they have the right data, or whether the data they need is readily accessible and usable, and instead prioritize what matters most — doing their jobs better.

Use Elastic in conjunction with your existing SIEM to handle higher-volume data sources, or if you're new to SIEM, start using Elastic as your SIEM today and collect as much data as you need for security use cases, regardless of scale or deployment type. Elastic enables your security team to adapt effectively, evolve more quickly, and helps you focus on becoming a better security practitioner.

About Elastic

Elastic makes data usable in real time and at scale for enterprise search, observability, and security. Elastic solutions are built on a single free and open technology stack that can be deployed anywhere to instantly find actionable insights from any type of data — from finding documents, to monitoring infrastructure, to hunting for threats. Thousands of organizations worldwide, including Cisco, Goldman Sachs, Microsoft, The Mayo Clinic, NASA, The New York Times, Wikipedia, and Verizon, use Elastic to power mission-critical systems. Founded in 2012, Elastic is publicly traded on the NYSE under the symbol ESTC. Learn more at elastic.co.

Want to check out Elastic Security for yourself?

Try Elastic Security on Elastic Cloud (14 days free, no credit card required). Or, deploy it on-prem, where it's always free.

Start Elastic Security free