CYSIV

White Paper

Faster Threat Detection and Response with Data Science and Next-gen SIEM

Cybersecurity is a Big Data Problem

Finding, triaging and investigating cyberthreats has never been more time-consuming, difficult or important for enterprises.

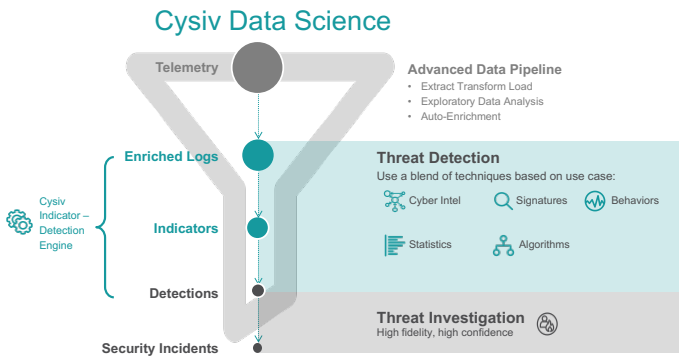
But cybersecurity teams are overwhelmed trying to deal with the massive amount of data generated by the security products they have deployed in their networks, on their endpoints and in their datacenters. With hundreds of new security incidents to deal with every week, and each one often taking days to investigate fully, SOC analysts can quickly fall behind and suffer alert fatigue as they manually try to add context to the data, often missing important signals in all the noise they must deal with. Adding to this frustration is the huge number of false positives that waste precious time. With more data traffic, an expanding security footprint, the adoption of new technologies like IIOT, analyst turnover, and a huge worldwide shortage of skilled cybersecurity professionals, the problem will only get worse.

In this environment, cybercriminals, cyberterrorists and hacktivists alike seemingly have the clear upper hand. Their attacks have become more sophisticated and difficult to detect and defend, and as a result, it's become more challenging to protect sensitive customer data and intellectual property, and prevent a breach. Profitability, share price, brand reputation, and even customer and staff safety, are on the line. Which is why cyber risk management, and solving this big data problem, have become an executive- and board-level priority.

This white paper provides a detailed description of Cysiv's modern, data science-driven approach to more quickly, effectively and efficiently detecting and investigating cyber threats, and the security operations and analytics platform, which many would consider a next-gen security information and event management (SIEM), that enables it.

Cysiv: Transforming Threat Detection Through Data Science

Cysiv was created to help enterprises better manage cyber risk, and overcome the big data problem and its associated issues (alert fatigue, skills shortage), through an advanced, managed security service model that's based heavily on automating critical daily SOC tasks related to threat detection, hunting, investigation and triage. Cysiv has developed and rigorously applied a broad range of data science techniques and technologies to automate, accelerate and improve the process of finding and prioritizing threats. At Cysiv, data science is used to more efficiently and effectively convert raw logs and data from other relevant sources, into actionable, high quality, high confidence detections and security incidents that warrant deeper human investigation.



Advanced Data Pipeline

Cysiv's advanced data pipeline ingests logs, data and other telemetry from as many relevant sources as possible. This ensures we have a comprehensive and clear picture of what's going on, and allows us to more confidently identify genuinely suspicious or malicious activity that warrants closer scrutiny. In general, the more data we have, the better we're able to make important correlations, reduce false positives and improve the confidence we have in the detections to be investigated.

Sources

Cysiv incorporates data from a broad range of sources:

Security controls

Data generated by security infrastructure and tools, including:

- Network IPS/IDS
- Firewall
- Endpoint protection platform
- Server / Workload / Container security
- Web proxy
- Email security

Infrastructure, monitoring and authentication

Data used to augment security control data sources; rich endpoint (server/desktop/laptop/workstation) and user activity data, including:

- Endpoint detection and response
- Windows security / Windows process launch / Sysmon / Linux system
- Active Directory (AD) authentication / Domain Controller / Linux auth
- IAM / SSO
- DHCP / Static IP
- DNS
- NAT / VPN / Proxy
- Cloud audit trail
- Network metadata

Enrichment

Identity, asset, vulnerability, and threat intelligence data that illuminates security context and impact during an investigation, including:

- Active Directory object properties / LDAP
- Asset inventory and classification / Configuration and patch management
- Indicators of Compromise (IOC)
- Vulnerability scan results

Applications

Data generated by mission critical applications running on servers, including

- Database
- ERP
- CRM
- APIs

Extract – Transform – Load (ETL)

All data sources are first extracted and normalized to the Apache Spot, Common Information Model for field naming and value normalization. This ensures elements like the source IP address are consistently represented the same way across all data, from every source. As well, empty and unnecessary fields are dropped, and the source and destination IPs are checked to ensure they only contain IPv4 or IPv6 addresses.

Exploratory Data Analysis

We use exploratory data analysis (EDA) to analyze each new potential data source using visual and mathematical models, to understand its key characteristics, and to test various hypotheses. EDA relies on modeling, ML, deep learning, statistical analysis and other techniques to determine how the data source can best contribute to the threat detection process. This may include, for example, corelating new data with existing data, to provide additional context and to help better understand the nature of potential attacks.

Auto-Enrichment

With all of the data now normalized, it's then enriched with critical contextual data to help understand the security context and impact of the detection. Auto-enrichment incorporates a range of contextual data sources including:

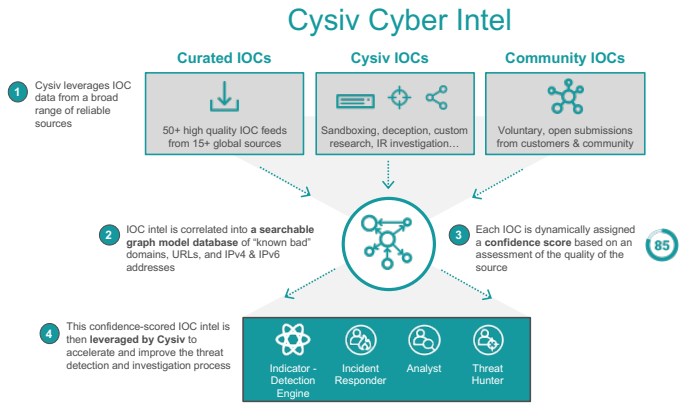
- **IP attribution** to determine a host's IP address at a point in time, to increase correlation of events generated by the different data sources
- **Active Directory** object properties to determine a user's role in the organization or the security context of a host
- **IP address geolocation** data to determine whether network communication is with a known cyber terror organization or nation state, for example
- **Configuration** data to confirm whether a version of the software/OS running on a system is vulnerable to a detected threat
- **Asset classification** data to help prioritize the investigation of suspicious detection on a production server, for example, over a development server
- **Vulnerability scan** results*, to ensure time isn't wasted investigating detections for systems that have already been patched (physical/virtually) to shield a vulnerability *(* Future capability)*
- **Cyber intel** with dynamically assigned confidence scores to assess the maliciousness of a detected threat, as well to identify the Tactics, Techniques and Procedures (TTPs) of a threat actor

Cysiv continues to explore and add new enrichments to help further enhance the ETL process.

Cyber Intel

One of the key elements of the auto-enrichment process is the incorporation of cyber intel. Cysiv leverages IOC data from a broad range of highly reliable sources including:

- **Curated IOCs:** 50+ high quality data feeds from over 15 global sources
- **Cysiv IOCs:** We draw on a range of sources from our customer deployments, including sandbox analyses from advanced breach detection systems and deception technologies, and from custom research and IR investigations
- **Community IOCs:** Customers and other organizations can voluntarily contribute anonymized IOC data.



All of this IOC intel is then correlated into a searchable graph model database of "known bad" domains, URLs, and IPv4 and IPv6 addresses. Each IOC is dynamically assigned a confidence score based on an assessment of the quality of the source. This confidence-scored IOC intel is then leveraged by the Cysiv Indicator-Detection Engine, as well as by Cysiv analysts, threat hunters and IR experts, to accelerate and improve the threat detection, triage and investigation process.

Threat Detection Methodology

False positives are a well-known security analytics problem. Dealing with them can be further complicated when multiple security tools create disparate alerts that can easily be taken out of context and can't be readily correlated. Cysiv addresses this problem with:

- Target-centric correlation of events generated by data sources
- Pre- and post-analytic contextual enrichment
- The combination of multiple threat detection methods that use signatures, behaviors, statistics and algorithms.

Indicator-Detection Model

The Cysiv Indicator-Detection Model is based on strict pre-processing rules that ensure each data source event is stored with as much identity attribution as possible. This ensures any post-ingestion analytics is run on highly correlated data, which leads to significantly fewer false positives and saves a security analyst several hours of identity attribution investigation. It also allows for more straightforward rule processing and better ML results.

Typically, a small percentage of raw logs generated by the data sources will indicate "high-confidence" known bad activity that needs to be investigated: there's a simple and direct relationship between one event, one indicator, and one detection. However, that leaves a large percentage of raw logs that by themselves are not sufficient to trigger a detection. But by combining these raw logs with other data sources, or enriching them with context or intel, we can generate higher-fidelity, "high confidence" detections. This model of enriching, combining and aggregating, describes the Cysiv Indicator-Detection model.

Entity Attribution Model

In the Cysiv Indicator-Detection model, all triggered indicators and detections are attributed to an entity, which can be a user or a device/host. This approach provides the analyst with a clear, simple timeline of what is occurring and what is related to the entity under investigation. This threat-centric viewpoint can show the propagation of a threat through a network.

Indicator-Detection Engine

Our proprietary Indicator-Detection Engine first converts enriched logs to indicators, and then converts these indicators to detections.

Step 1: Convert Enriched Logs to Indicators

In this step, the engine applies rules-driven use cases, to generate "indicators", in batches. An indicator is a single enriched log, a sequence or aggregation of enriched logs, or an analytics model result based on many enriched logs that indicate activity which may be malicious, but may also be legitimate. An indicator, by itself, is not normally enough to raise an alert or require a response, but it can contribute to a detection. This drastically reduces false positives.

Each use case incorporates enriched logs from a defined time period and applies a set of indicator rules to the batch of data. The time period for each use case is defined by the data science team, and may range from "continuously", to "last five minutes", "last 60 minutes", "last 24 hours", or even "last thirty days" to look for "low and slow" activity.

Step 2: Convert Indicators to Detections

Anything flagged as an indicator will then trigger the Engine to do further analysis. Any single indicator, or group of indicators, that matches one or more pre-defined patterns is flagged as a "detection". A detection is a high-confidence, high-fidelity set of logically-grouped indicators, generated by Cysiv's proprietary Indicator-Detection Engine, enriched with contextual data, correlated to cyber intelligence, and attributed to a host or user that indicates a potential threat.

For example, the following pattern would be flagged: suspicious email received + suspicious PDF attachment opened + communication port opened.

Cysiv leverages user and entity behavior analytics (UEBA) to look at patterns of human behavior, and then applies algorithms and statistical analysis to detect meaningful anomalies from those patterns—anomalies that indicate potential threats.

Detection Techniques

The Cysiv Indicator-Detection Engine applies four different types of detection techniques:

- **Signature-based** detection techniques match all or some attributes of an object to a known bad object.
- **Behavior-based** detection techniques match some type of digital pattern, footprint, human activity, or network behavior to known bad behavior.
- **Statistics-based** detection techniques use clustering, grouping, stack counting, baseline and variation, outlier detection, logistic regression and other methods to detect anomalous activity.
- **Algorithm-based** detection uses ML techniques such as supervised/unsupervised learning or deep learning to detect malicious/anomalous activity or to predict attacks.

Use Cases: Applying the Methodology

While the following table provides a small snapshot of the many use cases leveraged by the Indicator-Detection Engine, the data science team is constantly developing new use cases, and refining existing ones.

Use Case Examples	Description	ATT&CK Tactics	Detection Technique
High Risk User/Host			
Detection of High Risk User/Host	Checks which user/host has a greater risk of compromise by checking for the usage of P2P applications, Tor, and cracked software	• Initial Access • Execution • Privilege Escalation • C&C	• Signature-based • Behavior-based
Endpoint Compromise			
Malware Outbreak	Determines if more than 3 systems become infected with the same malware within a 24- hour period	• Initial Access • Execution • Privilege Escalation	• Signature-based • Statistics-based • Algorithm-based
Server Threat			
Suspicious Login to File Server/Database	Monitors all the logins and access for non-working hours. Checks the geo location to find unusual behaviour	• Initial Access	• Signature-based • Behavior-based

Use Case Examples	Description	ATT&CK Tactics	Detection Technique
Bad Web Application Configuration	Checks for successful common web application attacks such as directory traversal, XSS, SQL injection, file redirect, binary file upload, and plain text password usage over the network	• Initial Access	• Signature-based
Network Threat			
Brute Force Detection	Detects brute force attempts made by external / internal users	• Initial Access	• Signature-based • Behavior-based
Suspicious User / Host			
Insider Threat Detection	Looks for unusual or unexpected access attempts by an employee, to a network, application or database	• Execution • Privilege escalation • C&C	• Signature-based • Statistics-based • Algorithm-based

Human Investigation

Cysiv analysts and incident response experts then investigate these high fidelity / high confidence detections via Cysiv Command, tag their severity level, and determine an appropriate course of action.

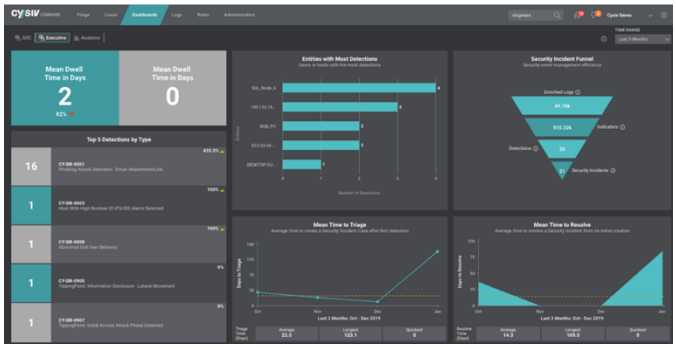
This triage process often includes customer interaction. In the event of a true positive incident, Cysiv IR experts will assist with:

- Data acquisition and analysis
- Identifying the source of the incident
- Developing a strategy to isolate the threat and affected systems
- Determining an appropriate removal and remediation strategy

Cysiv Security Operations and Analytics Platform

Cysiv has developed a powerful security operations and analytics platform that is the foundation for all Cysiv services, and the foundation for enabling and conducting all of the data science outlined in this white paper.

Cysiv Command



Cysiv Command is a cloud-native, data-science driven, security operations and analytics platform that empowers you to co-monitor or co-manage your security alongside Cysiv experts. All Cysiv customers have access to Command, and it is the focal point for a broad range of security-related activity, including accessing all the enriched logs, managing indicators, detections, security incidents and cases, and generating reports.

Key functionality includes:

- **Advanced** Data Pipeline: Eliminates manual correlation and enrichment of logs, accelerating incident response times and increasing analyst productivity. Ingested logs are parsed, cleansed, normalized, enriched, and indexed. During enrichment, logs are attributed with IP. Active Directory and critical asset data provide important context during an investigation. Following indexing, the Indicator-Detection Engine correlates logs to integrated cyber intel to trigger high-fidelity indicators and detections.
- **Security Information and Event Management:** Massively scalable, purpose-built, indexed data lake, with tiered data storage (hot, warm, and cold) enables costs to be better managed, and supports compliance requirements. It can complement, or for some, replace an existing SIEM investment.
- **User and Entity Behavior Analytics:** Leverages machine learning methods to augment and enhance rule-based threat detection with behavior analytics.
- **Integrated Cyber Intel:** Enables collecting, classifying, corroborating, and scoring large volumes of unstructured and highly related IOCs, resulting in all-source, finished intelligence that is included and integrated with the security operations and analytics platform. Incorporates a broad range of high-quality curated IOC feeds, Cysiv IOCs and community-sourced IOCs shared by customers that have opted-in to share them.
- **Cyber Intel Sharing:** Leverages anonymized IOCs that can be auto-extracted and shared amongst a network of industry peers to better identify targeted attacks and alert contributing members of the key threats to prioritize. Customers can opt-in to participate in Cysiv's global intelligence sharing community.
- **Automated Threat Detection and Triage:** Leverages a blend of signature-based, behavior-based, statistically-based and algorithmic (supervised and unsupervised machine learning), detection methods to automatically identify potential threats. Cysiv aggregates logs into indicators, correlates them with integrated cyber intel and then aggregates indicators into detections through its proprietary Indicator-Detection Engine. Automatically prioritizes security incidents based on the highest severity detections, focusing attention on the investigation of the most critical detections first, thus streamlining the analyst workload.
- **Case Management:** Provides workflow capabilities, tight integration, transparency, and seamless communication and collaboration during detection handling and incident management. Supports integrations with 3rd party case management systems.
- **Dashboard and Reporting:** Persona-based dashboards and the ability to create your own dashboards for monitoring and reporting on key operational and cyber security risk metrics, which can readily be shared as PDFs.

Cysiv Connector

This software component enhances our ability to extract log data from your environment, by normalizing it before it's then encrypted for secure delivery to Cysiv Command. It ensures continuous detection capabilities can be delivered with minimal disruption to local security and administrative teams, while optimizing the quality and reliability of local telemetry aggregation. The Connector is installed locally, and is available for VMware, Amazon Web Services, Microsoft Azure and Google Cloud Platform. Alternatively, it can be hosted for acquisition of data from SaaS products.

Benefits

Cysiv's data science-driven approach automates, simplifies and accelerates the critical tasks of threat detection, threat investigation and threat hunting, to reduce the probability of a successful attack, breach, theft or disruption. Key benefits to your organization:

- Reduce the Mean-Time-To-Detect: Shortens the length of time required to detect suspicious activity, providing more time to respond and take preventative or other actions.
- Reduce Alert Fatigue: Dramatically reduces the number of alerts and false positives that analysts must investigate, giving them more time to investigate higher quality alerts, more thoroughly.
- Avoid Costs and Damages: Reduces the probability of incurring costs (legal fees, regulatory fines, customer service costs, etc.) and the brand damage associated with a successful attack.
- Improve Agility / Scalability: By addressing this big data problem, you'll be better positioned to scale your business – whether it's adding employees, cloud workloads, new business units or acquisitions, or investing in IIoT – all without the cost and difficulty of adding security analysts, or materially changing your cyber risk profile.

Cysiv SOC-as-a-Service

Cysiv's data science-driven approach to detecting and investigating threats, and its cloud-native security operations and analystic platform (or next-gen SIEM) are available to customers through Cysiv Security Operations Center (SOC)-as-a-Service.

Cysiv provides SOC-as-a-Service to enterprises that need to better manage cyber risk by accelerating and improving the process of detecting, investigating and remediating threats, and of managing critical security controls. Cysiv uniquely combines all the elements of a proactive, 24/7 threat hunting SOC–including security and threat experts, global intel, processes and playbooks, and technology–with a managed security product stack, and delivers them as a service, with consumption-based, monthly billing. Cysiv's cloud-native, next-gen SIEM overcomes the limitations, deployment challenges, and frustrations associated with traditional solutions. It combines essential technologies and functions into a single co-managed platform, and leverages advanced data science techniques to automate the time-consuming, complex but critical activities and processes for truly effective threat detection, hunting, investigation, and remediation.

To learn more, visit www.cysiv.com.