Jacobs

Event Log Analytics using EXTREME SEARCH™

Accelerated Cybersecurity Incident Detection & Response Tool

Jacobs Engineering
Technical Solution Paper

May 2022

Contents

Executive Summary

The Jacobs EXTREME SEARCH cybersecurity solution enables customers to capture and search all event logs as they are generated, and to search multiple months of historic event logs in 25 minutes or less. Extreme Search is a revolutionary solution enabling functionality not available elsewhere in the market and offering organizations the opportunity to significantly enhance their cyber security. The solution delivers the event log capture and processing capabilities required by the May 2021 Cybersecurity Executive Order 14028. Extreme Search is designed to integrate with existing SIEM tools, such as Splunk, and can be transitioned seamlessly into an existing architecture with zero risk. Extreme Search is affordable and reduces costs for personnel, storage, and data ingest. The solution is enabled by two revolutionary technologies: 1) search capabilities without ETL or indexing, and 2) in-memory computing using neuromorphic techniques. Search rates on a single 2U server have exceeded 60 gigabytes/second, and by clustering multiple servers, petabytes of data can be search in 25 minutes or less. Extreme Search is currently deployed in two large enterprise customers, delivering timely cyber defense analytics results. CyberLynx, the underlying core SIEM interface software, has been deployed in secure infrastructure. Jacobs (formerly BlackLynx) has expertise and extensive background in deploying across unclassified and classified environments with defense, intelligence, and civilian government customers.

Solution Overview

The Need: Cyber Incident Response Teams Need Better Tools to Manage Threats

Large organizations are facing the multiple challenges of exponentially growing data, large and vulnerable enterprise networks, increasing numbers of cyber intrusions, and finite resources and personnel available for IT operations and cybersecurity.

Companies, government agencies, and other large organizations have seen an increasing number of cyberattacks with major impacts on finances, mission and reputation. Pandemic-driven changes in work patterns increased the attack surface and strained networks and IT admins, as workers quickly shifted to remote access and "bring-your-own" technology.

Cyber incident response (IR) teams are understaffed and overworked. 60% of IT security professionals say the biggest barrier to cyber resilience is a lack of investment in new cybersecurity technologies like artificial intelligence and machine learning, and the second biggest obstacle is hiring and retaining qualified personnel.[1] Incident response managers are facing lagging response times and security analysts must devote large amounts of time to investigations. On average, intrusion detection times range from 12 to 60 days,[2] and containment

[1] Sheridan, Kelly. 77% of Businesses Lack Proper Incident Response Plans. *Dark Reading*. March 14, 2018. https://www.darkreading.com/attacks-breaches/77-of-businesses-lack-proper-incident-response-plans
[2] Lee, Robert. 2021 SANS Cyber Threat Intelligence Survey Results. *SANS*. January 19, 2021. https://www.sans.org/webcasts/2021-cyber-threat-intelligence-cti-survey-results-116475/

takes an average of 287 days.[3] Security teams are dealing with a high number of alerts and false positives, which results in alert fatigue. According to a recent survey report by cybersecurity company, 70% of companies face over 100 security threat alerts per day,[4] and large enterprises can see thousands. Ransomware has been the most prevalent attack type in 2020 and 2021.[5] The good news is that IR teams with AI and automation built in have been shown to have much faster response times and lower costs; an IBM study found that breach costs were 80% lower for companies using automated security tools.[6]

Adding further challenges, the U.S. government cybersecurity Executive Order 14028 of May 2021 requires improved data retention, processing, and long term storage. Under this EO, Federal agencies must retain logs and cyber data for 12 months, resulting in multi-petabyte storage requirements that current indexing and search tools cannot meet within reasonable time and cost parameters. EO 14028 also requires agencies to process data in real time and have rapid search capability for over a year's worth of historical data, which in some cases is an extremely large volume of data. Therefore, government agencies are scrambling to add high-volume search and storage capability to meet these requirements and improve their cybersecurity incident response posture.

The Vision

Figure 1 depicts the incident response cycle and the need for cyber speed. Deploying automated cyber tools to find threats faster, improve detection and response times, focus resources where they are needed most, and provide a zero-risk seamless transition strategy are high priorities for organizations with the responsibility to protect their systems. With Extreme Search, Enterprise IT can be enhanced to analyze all IT infrastructure event logs more completely, including the application of artificial intelligence and machine learning (AI/ML) algorithms. The Extreme Search solution delivers a future where automated tools accelerate incident detection and response, and provide operators a comprehensive view of their IT operations. By processing data as it arrives and by making months or years of historical data immediately searchable, incidents can be detected, rapidly analyzed, and remediated. Cyber professionals now have the tools they need to make informed decisions quickly.

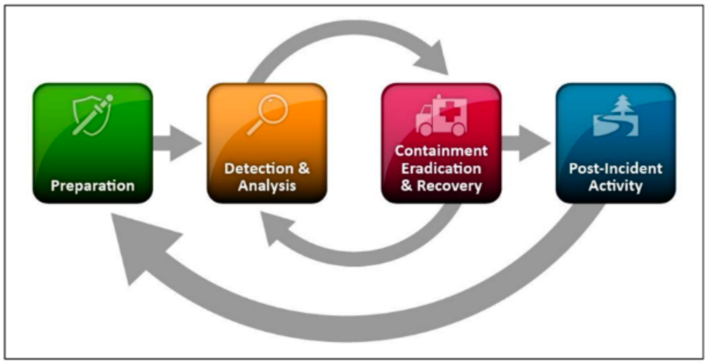


Figure 1: NIST Incident Response Life Cycle; Extreme Search supports all Phases of the Cycle.

[3] IBM, Cost of a Data Breach Report 2021, https://www.ibm.com/security/data-breach
[4] Red Canary, "The State of Incident Response, 2021." https://redcanary.com/resources/guides/the-state-of-incident-response-2021/
[5] IBM, X-Force Threat Intelligence Index 2021. https://www.ibm.com/downloads/cas/M1X3B7QG
[6] IBM, Cost of a Data Breach Report 2021, https://www.ibm.com/security/data-breach

Extreme Search Integration

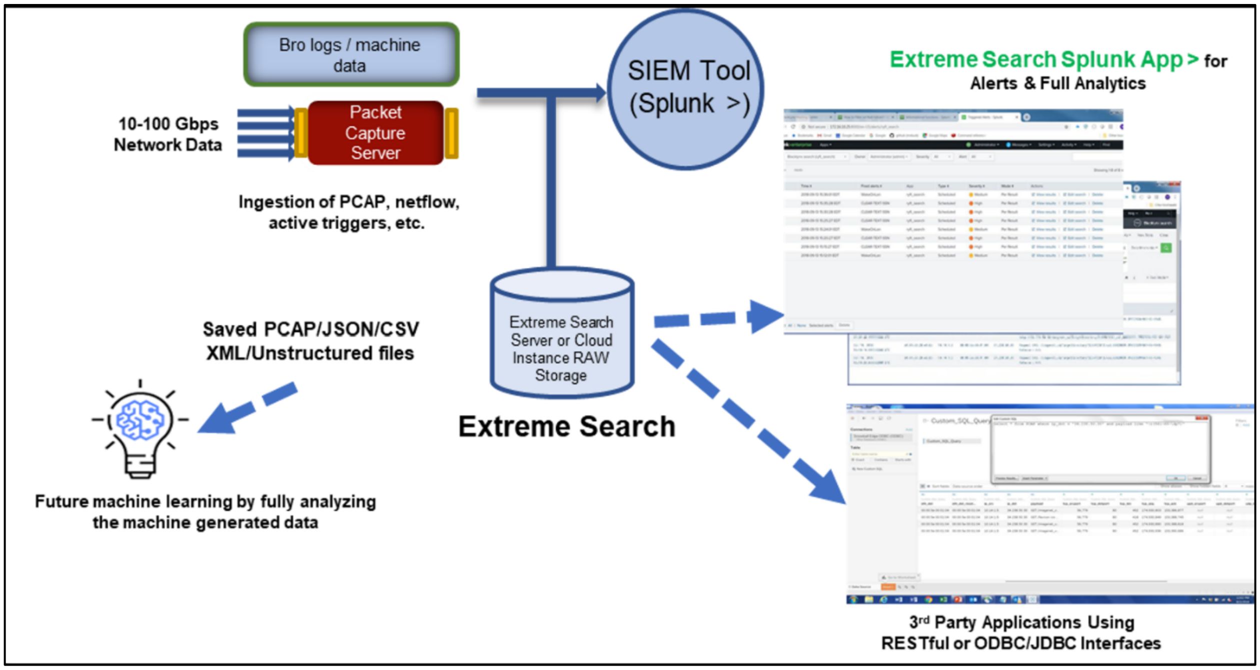

Bro logs / machine data
10-100 Gbps Network Data
Packet Capture Server
Ingestion of PCAP, netflow, active triggers, etc.
SIEM Tool (Splunk >)
Extreme Search Splunk App > for Alerts & Full Analytics
Saved PCAP/JSON/CSV XML/Unstructured files
Extreme Search Server or Cloud Instance RAW Storage
Extreme Search
Future machine learning by fully analyzing the machine generated data
3rd Party Applications Using RESTful or ODBC/JDBC Interfaces

Figure 2: Architecture integrating Extreme Search, CyberLynx, and Splunk.

Figure 2 shows the Extreme Search implementation architecture. Event logs are forwarded to the Extreme Search Server and the SIEM tool – Splunk in this case. CyberLynx is the underlying technology which provides the API to existing SIEM tools and to the Extreme Search server, enabling an operator to launch Extreme Search queries and receive results on the same UI as the SIEM tool. This approach enables a seamless transition and introduction of Extreme Search. The ongoing SIEM tool operations can remain constant and only be adjusted over time as the Extreme Search server demonstrates performance utility and begins to build the repository of historical data.

The Extreme Search tool provided by Jacobs is capable of using the CPU cores in the server to analyze the incoming data stream and provide alerting back to the SIEM tool dashboard. The Neuromorphic Processing Unit (NPU) search, provided by Lewis Rhodes Labs, enables the data to remain in Smart SSD and to be searched in place (see Figure 3). The Extreme Search solution is hosted on standard, high capacity servers with 24 Smart SSD drives integrated into the server. Each drive hosts a Xilinx FPGA capable of searching the 4 TB drive without moving the data. Currently, the total Extreme Search architecture is based on deploying an appliance. The CyberLynx tool can also be deployed in the cloud and used to analyze incoming data streams, and has been demonstrated on an AWS instance. In the future, additional cloud instances will be deployed utilizing in-memory computing solutions.

Figure 3: Extreme Search can be hosted in the cloud or on high capacity servers enhanced with NPU search.

Because Extreme Search integrates with existing enterprise tools, there is minimal need for operator training. The investment in existing licenses will remain valid and usage costs will likely decrease due to reduced need to ingest data. There will be a need to tailor and optimize queries based on searching all the data; however, the integration can occur in parallel with ongoing operations and immediately begin enhancing cyber security.

Distributed Queries – Search Across the Enterprise

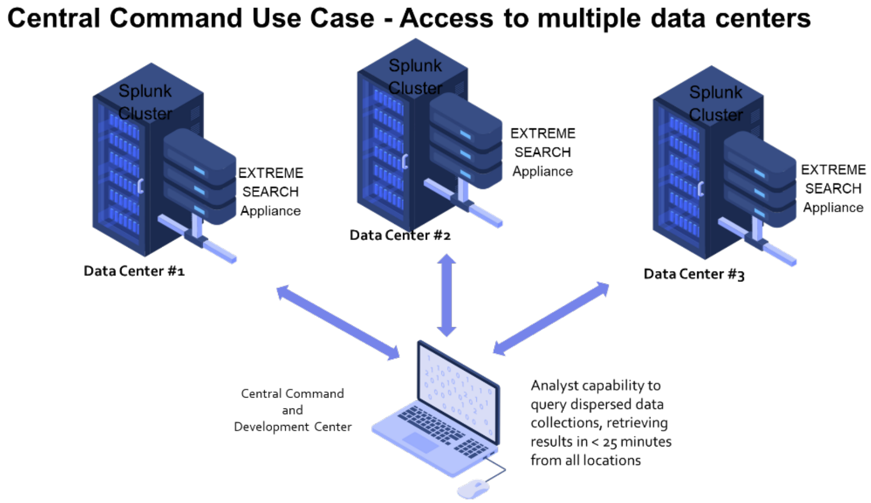
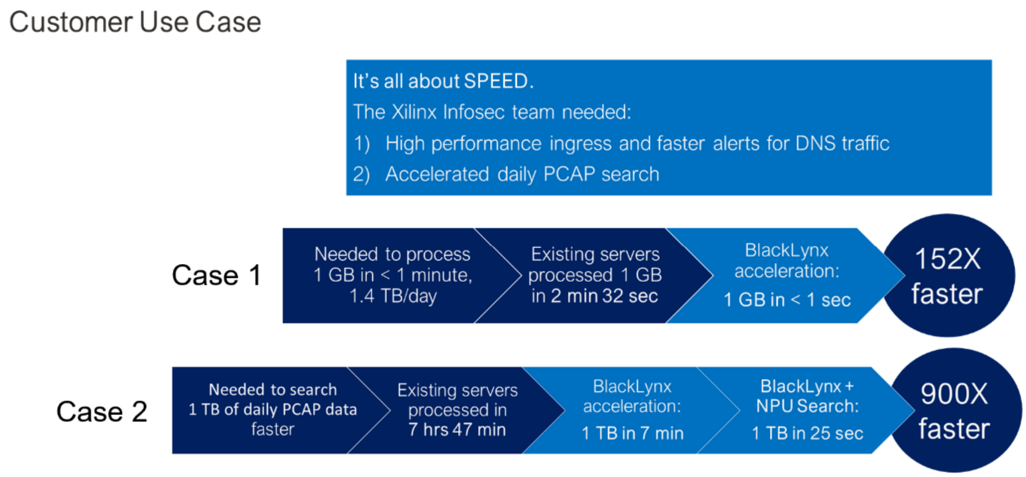

Central Command Use Case - Access to multiple data centers
Splunk Cluster
EXTREME SEARCH Appliance
Data Center #1
Splunk Cluster
EXTREME SEARCH Appliance
Data Center #2
Splunk Cluster
EXTREME SEARCH Appliance
Data Center #3
Central Command and Development Center
Analyst capability to query dispersed data collections, retrieving results in < 25 minutes from all locations

Figure 4: Distributed Queries Across Multiple Data Centers – Creating an Enterprise View

Figure 4 shows the capability to perform distributed queries across a data center infrastructure. Customers operate multiple data center sites and distributed IT instantiations. Because existing SIEM tools only analyze 20% or less of the total event logs and network traffic, it has been impossible to develop a comprehensive picture of what is happening across the enterprise. Anomalies detected in one data center should be alerted across the enterprise, and a comprehensive search for a malware signature should be possible. By adding the Extreme Search functionality, cyber operators can gain the upper hand on fighting cyber risk across distributed platforms.

Current Use and Deployments

The Extreme Search solution has been deployed in two locations – AMD/Xilinx infrastructure and at the Department of Energy's Sandia National Lab. CyberLynx has been used in multiple classified programs and has demonstrated performance searching large, unstructured or structured datasets without ETL or indexing. The following figures illustrate performance benefits seen in actual deployments.



Customer Use Case

It's all about SPEED.
The Xilinx Infosec team needed:
1) High performance ingress and faster alerts for DNS traffic
2) Accelerated daily PCAP search

Case 1

Needed to process 1 GB in < 1 minute, 1.4 TB/day

Existing servers processed 1 GB in 2 min 32 sec

BlackLynx acceleration: 1 GB in < 1 sec

152X faster

Case 2

Needed to search 1 TB of daily PCAP data faster

Existing servers processed in 7 hrs 47 min

BlackLynx acceleration: 1 TB in 7 min

BlackLynx + NPU Search: 1 TB in 25 sec

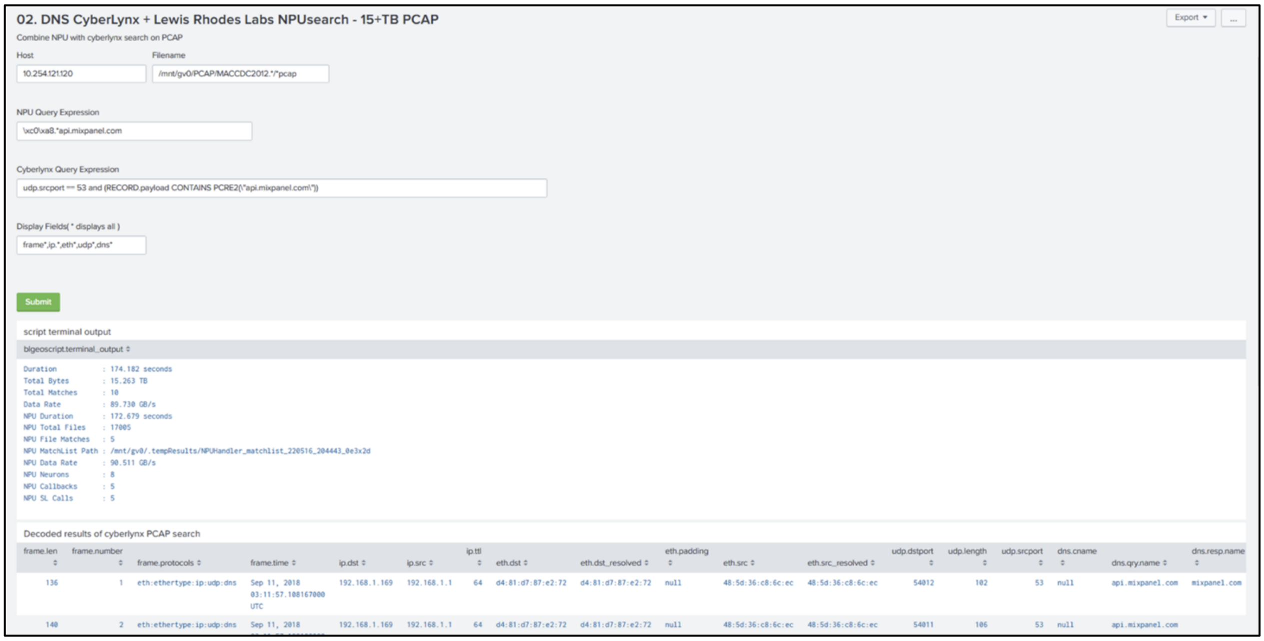
900X faster

Figure 5: Extreme Search Performance at Xilinx/AMD for Real-time and Historical Search Functions

Figure 5 illustrates the two main use cases. Case 1 shows how Extreme Search can process incoming data and provide alerts and detections at the speed of event log generation. Extreme Search is capable of searching over 20 gigabytes of incoming data every minute on a single server. If higher rates are desired, multiple servers can be clustered; however, the single server rate has been proven to be performant for deployments to date. Case 2 shows the benefit of searching in memory by going through 1 terabyte in 25 seconds, a 900x improvement over performance of the existing infrastructure. This technology will support searching petabytes of information in 25 minutes or less, and the solution scales horizontally as more information can be searched in parallel without increasing the search time.

Figure 6 shows a screen shot of measured results – showing a search rate of 90 Gigabytes/second on two Extreme Search appliances. Searching event logs at these rates enable cyber professionals to understand the performance and relationships in the data center, cloud, and IT infrastructure. Upon finding search results, other tools can be applied including graph analytics, entity mapping, and AI/ML. Existing SIEM or third party tools are still relevant and can be used, but now they can be supplied with relevant network traffic and can focus in on specific relationships – understand what is normal and what needs further investigation. Insider threats can be detected and analyzed when anomalous data flows are flagged by advanced analytics.

Jacobs Engineering, a leading cybersecurity solutions provider, is working on implementing Extreme Search within its infrastructure. Jacobs acquired BlackLynx in November 2021 in part because of the CyberLynx search technology and the potential to revolutionize cyber operations and analyze other large data flows. With the addition of the LRL Smart SSD neuromorphic search, the overall solution is increasingly attractive and responsive to the cyber needs of a modern IT infrastructure.



02. DNS CyberLynx + Lewis Rhodes Labs NPUsearch - 15+TB PCAP

Combine NPU with cyberlynx search on PCAP

Host: 10.254.121.120

Filename: /mnt/gv0/PCAP/MACCDC2012.*/*pcap

NPU Query Expression: \xc0\xa8.*api.mixpanel.com

Cyberlynx Query Expression: udp.srcport == 53 and (RECORD.payload CONTAINS PCRE2(\"api.mixpanel.com\"))

Display Fields(* displays all): frame*,ip.*,eth*,udp*,dns*

script terminal output

blgeoscript.terminal_output ⇕

```
Duration          : 174.182 seconds
Total Bytes       : 15.263 TB
Total Matches     : 10
Data Rate         : 89.730 GB/s
NPU Duration      : 172.679 seconds
NPU Total Files   : 17005
NPU File Matches  : 5
NPU MatchList Path : /mnt/gv0/.tempResults/NPUHandler_matchlist_220516_204443_0e3x2d
NPU Data Rate     : 90.511 GB/s
NPU Neurons       : 8
NPU Callbacks     : 5
NPU SL Calls      : 5
```

Decoded results of cyberlynx PCAP search

frame.len ⇕	frame.number ⇕	frame.protocols ⇕	frame.time ⇕	ip.dst ⇕	ip.src ⇕	ip.ttl ⇕	eth.dst ⇕	eth.dst_resolved ⇕	eth.padding ⇕	eth.src ⇕	eth.src_resolved ⇕	udp.dstport ⇕	udp.length ⇕	udp.srcport ⇕	dns.cname ⇕	dns.qry.name ⇕	dns.resp.name ⇕
136	1	eth:ethertype:ip:udp:dns	Sep 11, 2018 03:11:57.108167000 UTC	192.168.1.169	192.168.1.1	64	d4:81:d7:87:e2:72	d4:81:d7:87:e2:72	null	48:5d:36:c8:6c:ec	48:5d:36:c8:6c:ec	54012	102	53	null	api.mixpanel.com	mixpanel.com
140	2	eth:ethertype:ip:udp:dns	Sep 11, 2018	192.168.1.169	192.168.1.1	64	d4:81:d7:87:e2:72	d4:81:d7:87:e2:72	null	48:5d:36:c8:6c:ec	48:5d:36:c8:6c:ec	54011	106	53	null	api.mixpanel.com	

Figure 6: Actual Results as Measured at BlackLynx Using Extreme Search on two 2U Servers

Accreditations

BlackLynx, now Jacobs Engineering, welcomes inquiries from customers regarding deployments and accreditations. BlackLynx/Jacobs is an industry leader in deploying secure and automated IT infrastructure, having gained multiple accreditations in the highest security environments. CyberLynx is currently running on classified infrastructure.

Conclusions

This document demonstrates specific applications of the Extreme Search tool and Splunk app for cybersecurity and incident response. As enterprise IT and information security teams face increasing threats, automation can maximize utility of available resources, improve time to discovery and reduce cost impact of cyber incidents. Extreme Search offers a suite of customizable tools that can easily integrate with existing business processes. These tools can be used for real-time network traffic analysis, and forensic searches of current and historic logs across a range of structured or unstructured data types. Extreme Search can help incident response teams improve and accelerate incident detection, analysis, containment and recovery to better protect critical networks, data and systems.

Appendix 1: Use Cases for Extreme Search

Our analysis is grounded in current industry best practices and is based on real-life experience using Extreme Search's CyberLynx and neuromorphic compute tools, embedded in the Splunk dashboard, to accelerate incident detection and response.

This playbook is based on the NIST cyber framework and incident response guidance,[7] and the open-source Incident Response Consortium model playbooks.[8] The playbook leverages the MITRE ATT&CK framework, as well as published reports and analysis by commercial threat hunters, incident response, cybersecurity and data analytics firms.

Appendix 1 will be useful for CISO's, CTO's, IR teams and managers, cybersecurity analysts, threat hunters, and enterprise IT and network operations teams. This playbook will be most useful to those responsible for network traffic monitoring for large enterprise scale operations, e.g. handling cyber data on the scale of gigabytes to terabytes to petabytes. Potential end users might include corporate cybersecurity and network operations, internet service providers, data centers, utilities, hospitals, and manufacturing facilities; all of which are frequent targets for ransomware and other cyber attacks.

Extreme Search Accelerated Analytics Solutions

Extreme Search has developed high performance search tools that integrate with commonly used data analytics software to significantly improve speed and performance for cybersecurity applications. CyberLynx is a specialized cyber security analytics tool that is designed to rapidly search large amounts of packet data directly without requiring any preprocessing or indexing of the data. CyberLynx is geared towards network forensic analysis on raw files with PCAP (Packet Capture Data). It can also be used to search a range of file types, such as raw unstructured text, CSV, JSON, or XML. The Extreme Search application stack can be integrated with existing tools such as Splunk, Wireshark, Tableau, Power BI, and Google Earth, using open API connectors. Extreme Search typically achieves several orders of magnitude performance improvements over commonly used search tools while providing the capability for exact, fuzzy (Hamming or Levenshtein), PCRE2 regular expressions, date, time, IPV4/6, PCAP and geospatial searches. We use standards-based interfaces such as C, C++, Java, Python, ODBC, JDBC, RESTful JSON, and command line/scripting interfaces. Our heterogeneous computing algorithms make the best use of available hardware (CPU, GPU or FPGA) to dynamically shifts computing loads, and are accelerated with Xilinx FPGAs.

Extreme Search has developed a customized Splunk app and dashboard interface which can easily be set up to run alongside and accelerate an organization's existing Splunk instance. The diagram below demonstrates a sample architecture using the Splunk app integration, and the Splunk app will be used for discussion throughout this paper. However, the CyberLynx tool can

[7] Getting Started with the NIST Cybersecurity Framework, https://doi.org/10.6028/NIST.SP.1271 and NIST Computer Security Incident Handling Guide, https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf
[8] Incident Response Consortium, used with permission. https://www.incidentresponse.com/playbooks/

also be set up and integrated with other tools, including custom architectures. When integrated with Splunk, CyberLynx acts as a filter on large data sets to avoid lengthy and costly indexing or extract, transform and load (ETL) processes. CyberLynx filters the raw data to find the most relevant files, which are then funneled and stored in Splunk to leverage its native capabilities.

Incident Response Life Cycle

Although there are unique aspects of different types of cyber incidents, these stages generally apply across a wide range of potential cyber activity that an organization may need to respond to. For the purposes of this paper, we will focus on detection and response to some common IR scenarios, including:

- Malware outbreak,
- Virus outbreak,
- Phishing,
- Data exfiltration/theft,
- Unauthorized access, and
- Improper usage.

Network traffic analysis and forensic data queries are key tools that can be used across different attack types to detect, analyze, contain and eradicate malicious actors or applications. For further discussion and detailed incident response playbooks for different attack types, see the Incident Response Consortium playbooks.[9]

Preparation

The first stage of incident response occurs before an incident is detected. Standard best practices for preparation include defining roles and responsibilities and establishing playbooks for analysts and incident response teams. Preparation also includes acquiring and establishing appropriate hardware and software capabilities to automate detection and handle incidents when they arise. These tools can establish a robust detection and response capability that will improve response times and reduce the cost of breaches.

Steps an organization can take to prepare for cybersecurity incidents using Extreme Search tools include: acquire CyberLynx tool; if using Splunk, install app in dashboard; or use any of available APIs to connect with custom or third party applications. Preparation can also include searching historical data to identify sources of precursors and indicators;[10] set alerts using known or blacklisted DNS; and other such proactive measures specific to your organization. Graph Analytics and other AI/ML solutions are increasingly useful in anticipating and finding incidents. Extreme Search enables these tools to be more comprehensive and insightful by assessing the entire set of event logs in context.

Detection and Analysis

[9] Incident Response Consortium. https://www.incidentresponse.com/

[10] NIST Computer Security Incident Handling Guide, Table 3-1, pg 36. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf

Incident detection and analysis is the primary area where Extreme Search tools can add value for incident response teams. Standard indicators of cyber incidents often include unknown or unexpected network traffic, which can be detected using packet capture (PCAP) and analysis. In the case of a malware or virus outbreak, indicators might include unusual traffic between nodes or locations in an enterprise network. In cases of data theft, indicators can include data leaving a network in unusual patterns, in unusually large amounts, or at unusual times. Unauthorized access or improper usage can be indicated by anomalous user behavior or web traffic, large data transfers, or access to unauthorized or blacklisted websites. For further discussion of tactics and techniques, see MITRE's ATT&CK Network Traffic taxonomy.[11]

The Extreme Search tools described below can be applied to define or detect threat indicators in cyber incident scenarios. These queries can be used to conduct analysis of PCAP data, and the results can be used to set alerts in Splunk, or other real-time network traffic monitoring tools. These tools can profile networks and systems to detect unusual or unauthorized changes, or correlate different types of data from multiple sources to validate indicators of an attack.

Respond: Contain, Eradicate & Recover

Once an incident has been detected and analyzed, the next steps are to contain, eradicate and recover from the threat. A key part of containment and forensic analysis is examining databases, vulnerability logs, system logs, and other sources. Extreme Search can help incident response teams identify how far an attack has spread, identify the IT Services being impacted, identify the systems that have been affected, or identify the vulnerabilities being exploited.

Post-Incident Activity

After a cyber incident occurs is the time to collect valuable information in attack indicators, vectors, and vulnerabilities. Data analysis tools can assist in documenting lessons learned and establishing processes and alerts to prevent future attacks. Extreme Search can be applied to evaluate all available historical data for indicators or clues that lead to the eventual incident.

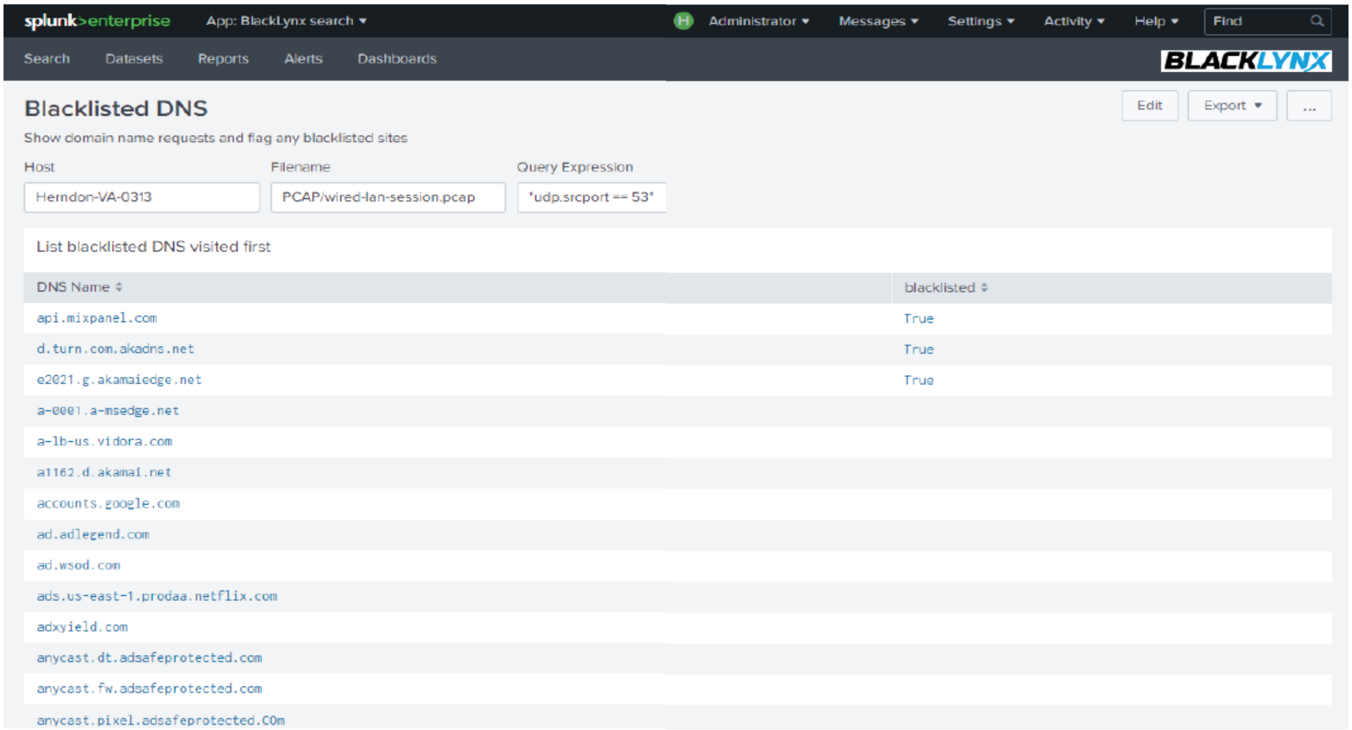
Extreme Search Solutions

Malware and Data Theft: Blacklisted DNS Query & PCAP Search

In many cyber incident scenarios, a blacklisted DNS query can help detect threat indicators using an accelerated PCAP search. The "Blacklisted DNS Query" performs a search on a PCAP file on DNS requests, extracts the requested sites, and flags any requests for known blacklisted sites. This tool can be used to run routine or forensic searches on periodic batches of PCAP data, and can be conducted in near real time. The known blacklisted sites are maintained in a file by the system administrator, and can be informed by third-party cybersecurity services. Blacklisted DNS data can also be used to set up automated alerts during the Prepare phase.

[11] MITRE ATT&CK Framework, Network Traffic. https://attack.mitre.org/datasources/DS0029/

The Extreme Search dashboard provides a list of all DNS requests and flags those that appear in the blacklist file. The dataset shown in the figure below is a demonstration of PCAP capture run on Extreme Search office networks; in this example case, the search returned three results matching blacklisted sites. The administrator can then act on this event.



Blacklisted DNS

Show domain name requests and flag any blacklisted sites

Host: Herndon-VA-0313

Filename: PCAP/wired-lan-session.pcap

Query Expression: 'udp.srcport == 53'

List blacklisted DNS visited first

DNS Name ⇕	blacklisted ⇕
api.mixpanel.com	True
d.turn.com.akadns.net	True
e2021.g.akamaiedge.net	True
a-0001.a-msedge.net	
a-lb-us.vidora.com	
a1162.d.akamai.net	
accounts.google.com	
ad.adlegend.com	
ad.wsod.com	
ads.us-east-1.prodaa.netflix.com	
adxyield.com	
anycast.dt.adsafeprotected.com	
anycast.fw.adsafeprotected.com	
anycast.pixel.adsafeprotected.COm	

> Query Inputs:
- Host –Extreme Search server hostname or IP Address
- Filename –PCAP file(s) to be searched
- Query Expression –search criteria, in this case UDP source port 53 (DNS server)

> Operation and Output:
- Search of native PCAP files extracting those packets to port 53
- Decode of selected packets to extract the requested domain names
- Return of the results to Splunk in a JSON format

The Blacklisted DNS query can be used for a wide variety of incident detection and response scenarios, in particular for malware or virus detection, phishing, data theft, unauthorized access, and improper usage.

Extreme Search researchers and Xilinx security teams have used the Blacklisted DNS query tool in real-world cyber incidents, which has led to improvements in threat detection and incident response. A sample use case of this tool would be detection of a network user visiting blacklisted domain names, which could alert security personnel to potential phishing, malware, or improper usage incidents. This tool can also be used to detect data movement to a blacklisted DNS- which could indicate data exfiltration incidents such as IP or other sensitive data theft.

USE CASE: Data Exfiltration

IP Theft Prevention and Export Controlled Data Monitoring

Some types of information must be carefully controlled, like sensitive national security data and critical intellectual property. The CyberLynx app in the Splunk dashboard can be used to monitor for data movement to IP addresses that are prohibited.

Imagine you are a company CISO with sensitive technical designs that are highly valuable and vulnerable to IP theft, or that fall under export control laws and can not be shared with other nations. To protect your intellectual property from theft, and avoid potential export-control fines, your security team sets a list of prohibited IP addresses associated with known Advanced Persistent Threat (APT) actors that will alert security teams to take action if data exfiltration to these addresses is detected.

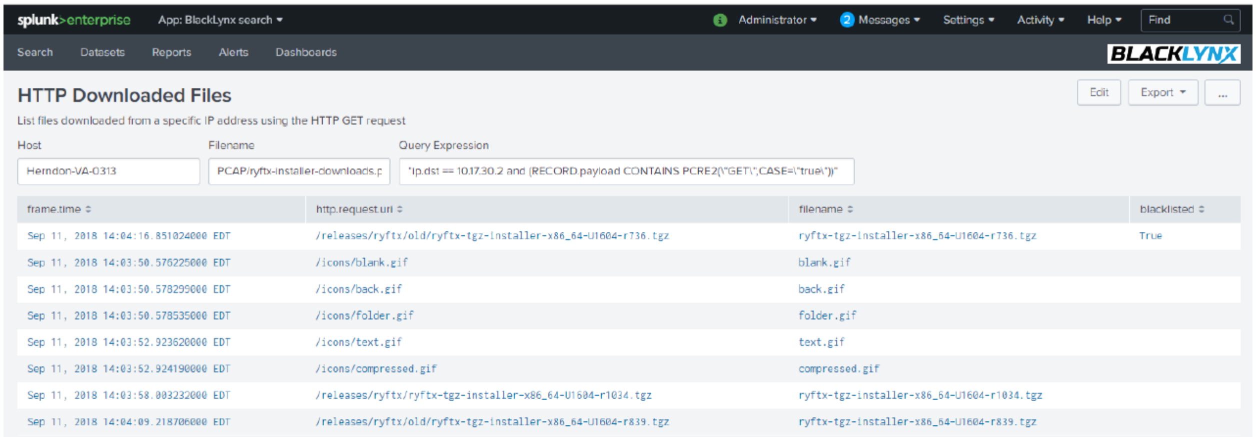
Malware Detection: HTTP Downloaded Files Query

The "HTTP Downloaded Files Query" can be used for incident detection, analysis, and containment. This query can identify the vulnerability being exploited by searching vulnerability logs and system logs and comparing them against incident or threat databases.

This query performs a search on a PCAP data to identify files downloaded from a specific IP address. It is an example of quickly extracting data from a native PCAP file. In the "HTTP Downloaded Files" query the list of files is compared against a known list of compromised files and any matches are flagged. The known compromised sites are maintained in a file by the system administrator and can be informed by historical experience or alerts from third-party cybersecurity services.

> Query Inputs:
- Host –Extreme Search server hostname or IP Address
- Filename –PCAP file(s) to be searched
- Query Expression –search criteria, in this example the IP address (10.17.30.2) and a payload search for an HTTP GET statement

> Operation and Output:
- Search of native PCAP files extracting those packets with a specific IP address and a payload search for HTTP GET requests
- Decode of selected packets to extract the requested download file names
- Return of the results to Splunk in a JSON format



HTTP Downloaded Files

List files downloaded from a specific IP address using the HTTP GET request

Host: Herndon-VA-0313

Filename: PCAP/ryftx-installer-downloads.p

Query Expression: "ip.dst == 10.17.30.2 and (RECORD.payload CONTAINS PCRE2(\"GET\",CASE=\"true\"))"

frame.time	http.request.uri	filename	blacklisted
Sep 11, 2018 14:04:16.851024000 EDT	/releases/ryftx/old/ryftx-tgz-installer-x86_64-U1604-r736.tgz	ryftx-tgz-installer-x86_64-U1604-r736.tgz	True
Sep 11, 2018 14:03:50.576225000 EDT	/icons/blank.gif	blank.gif	
Sep 11, 2018 14:03:50.578299000 EDT	/icons/back.gif	back.gif	
Sep 11, 2018 14:03:50.578535000 EDT	/icons/folder.gif	folder.gif	
Sep 11, 2018 14:03:52.923620000 EDT	/icons/text.gif	text.gif	
Sep 11, 2018 14:03:52.924190000 EDT	/icons/compressed.gif	compressed.gif	
Sep 11, 2018 14:03:58.003232000 EDT	/releases/ryftx/ryftx-tgz-installer-x86_64-U1604-r1034.tgz	ryftx-tgz-installer-x86_64-U1604-r1034.tgz	
Sep 11, 2018 14:04:09.218706000 EDT	/releases/ryftx/old/ryftx-tgz-installer-x86_64-U1604-r839.tgz	ryftx-tgz-installer-x86_64-U1604-r839.tgz	

HTTP Downloaded Files query can be used for incident detection and response, in particular for malware or virus detection and improper usage scenarios. For example, a security analyst might detect a malicious file that has been downloaded by a user. The analyst could then examine log files to determine whether any other files had been downloaded from the same IP address, and if any are found, isolate them to prevent further damage.

URL Hijacking & Malware Detection: Cybersquatting Query

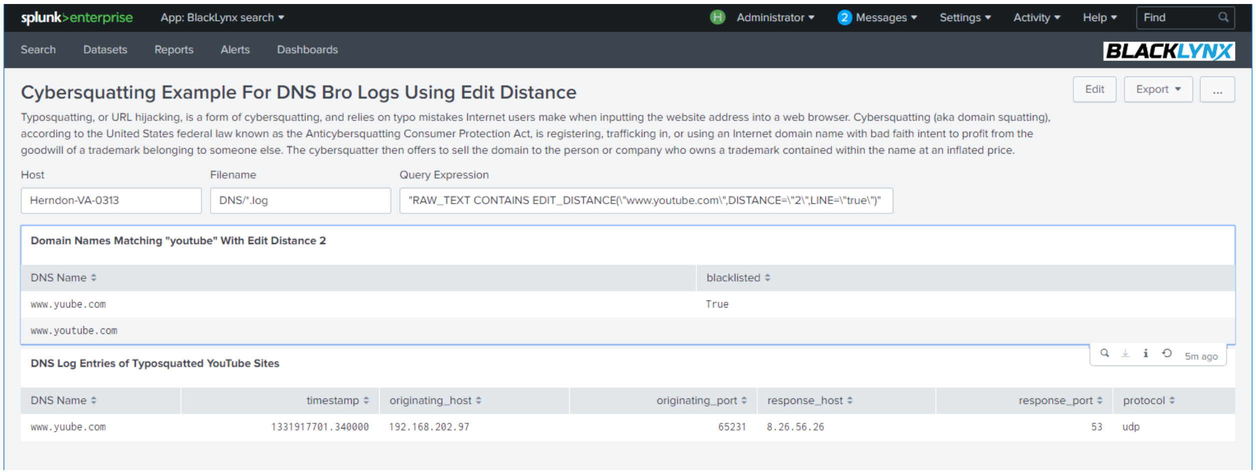
Cyber attackers often use malicious URLs, either sent via phishing emails or “cybersquatting” on URLs that are similar to legitimate websites. Typosquatting, or URL hijacking, is a form of cybersquatting and relies on typo mistakes Internet users make when inputting the website address into a web browser. Cybersquatting (aka domain squatting), according to the United States federal law known as the Anti-cybersquatting Consumer Protection Act, is registering, trafficking in, or using an Internet domain name with bad faith intent to profit from the goodwill of a trademark belonging to someone else. CyberLynx includes a customized cyberquatting query that can be used to examine PCAP logs to detect whether a network user has visited one of these malicious URLs.

> Query Inputs:
  - Host – Extreme Search server hostname or IP Address
  - Filename – PCAP file(s) to be searched relative to configured directory
  - Query Expression – search criteria, in this case sites similar to www.youtube.com

> Operation and Output:
  - Search of native CSV files (BRO Logs) using a Levenshtein (fuzzy edit distance) search (distance = 2) to find DNS requests that are one or two characters different from the expected site
  - Return of the results to Splunk in a JSON format

Sample use cases for this tool might include a security analyst running a search on logs for cybersquatting to detect malicious URLs masquerading as legitimate websites. In this case the analyst might compare a published source of known indicators of compromise with malicious

URLs (for example from MITRE or DHS-CISA)[12] against their enterprise log files. The CyberLynx cybersquatting edit distance search performs a high speed search for these malicious sites. The search returns results in Splunk, and the analyst can set up alerts to detect vulnerabilities in the future.



Cybersquatting Example For DNS Bro Logs Using Edit Distance

Typosquatting, or URL hijacking, is a form of cybersquatting, and relies on typo mistakes Internet users make when inputting the website address into a web browser. Cybersquatting (aka domain squatting), according to the United States federal law known as the Anticybersquatting Consumer Protection Act, is registering, trafficking in, or using an Internet domain name with bad faith intent to profit from the goodwill of a trademark belonging to someone else. The cybersquatter then offers to sell the domain to the person or company who owns a trademark contained within the name at an inflated price.

Host: Herndon-VA-0313

Filename: DNS/*.log

Query Expression: "RAW_TEXT CONTAINS EDIT_DISTANCE(\"www.youtube.com\",DISTANCE=\"2\",LINE=\"true\")"

Domain Names Matching "youtube" With Edit Distance 2

DNS Name	blacklisted
www.yuube.com	True
www.youtube.com	

DNS Log Entries of Typosquatted YouTube Sites

DNS Name	timestamp	originating_host	originating_port	response_host	response_port	protocol
www.yuube.com	1331917701.340000	192.168.202.97	65231	8.26.56.26	53	udp

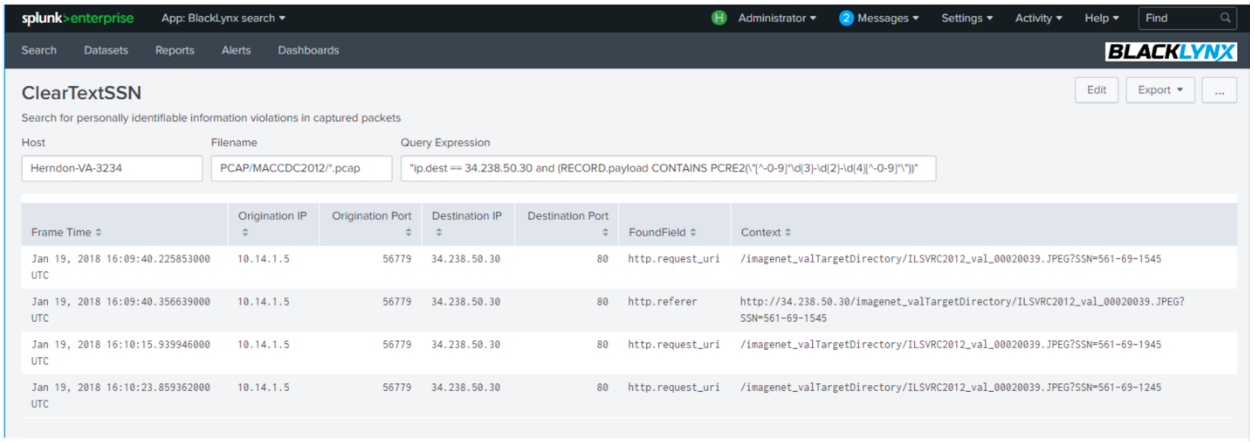
Prevent Data Theft: ClearTextSSN Search

Another example of a CyberLynx regular expression search that can be applied for cybersecurity detection and analysis is the ClearTextSSN search. Companies across many industry sectors have faced data theft; while in some cases data theft may result in financial or reputational harm, there are some data types such as personally identifiable information (social security numbers, addresses, etc.) that are protected by State and Federal regulations, with penalties for security failures. ClearText SSN search is just one of many custom queries that can be set up to monitor for specific data entering or leaving a network. The "ClearTextSSN" performs a search on a PCAP file(s) looking for any case where a Social Security Number is transmitted in clear text. In the demonstration project shown below using a dataset from the 2012 Mid-Atlantic Collegiate Cyber Defense Competition (MACCDC2012), 16GB of PCAP data was searched in less than 5 seconds.

> Query Inputs:
- Host – Extreme Search server hostname or IP Address
- Filename – PCAP file(s) to be searched relative to "/ryftone" directory
- Query Expression – search criteria, in this case a specific IP destination and a PCRE2 search of the payload data.

> Operation and Output:
- Search of native PCAP files extracting those packets to a specific IP Address
- Native PCAP search of the packet payload data using any of the Extreme Search search primitives, in this case, a regular expression search for a string formatted "nnn-nn-nnnn" without a leading or trailing digit.
- Return of the results to Splunk in a JSON format

[12] MITRE: https://attack.mitre.org/, DHS-CISA: https://www.cisa.gov/uscert/ncas/alerts



ClearTextSSN

Search for personally identifiable information violations in captured packets

Host: Herndon-VA-3234

Filename: PCAP/MACCDC2012/*.pcap

Query Expression: "ip.dest == 34.238.50.30 and (RECORD.payload CONTAINS PCRE2(\"[^-0-9]"\d{3}-\d{2}-\d{4}[^-0-9]"\"))"

Frame Time ⇕	Origination IP ⇕	Origination Port ⇕	Destination IP ⇕	Destination Port ⇕	FoundField ⇕	Context ⇕
Jan 19, 2018 16:09:40.225853000 UTC	10.14.1.5	56779	34.238.50.30	80	http.request_uri	/imagenet_valTargetDirectory/ILSVRC2012_val_00020039.JPEG?SSN=561-69-1545
Jan 19, 2018 16:09:40.356639000 UTC	10.14.1.5	56779	34.238.50.30	80	http.referer	http://34.238.50.30/imagenet_valTargetDirectory/ILSVRC2012_val_00020039.JPEG?SSN=561-69-1545
Jan 19, 2018 16:10:15.939946000 UTC	10.14.1.5	56779	34.238.50.30	80	http.request_uri	/imagenet_valTargetDirectory/ILSVRC2012_val_00020039.JPEG?SSN=561-69-1945
Jan 19, 2018 16:10:23.859362000 UTC	10.14.1.5	56779	34.238.50.30	80	http.request_uri	/imagenet_valTargetDirectory/ILSVRC2012_val_00020039.JPEG?SSN=561-69-1245

Use Case: Detecting Theft of Customer Data

Loss of customers' personal information is a public relations nightmare for many companies. Proactive monitoring using custom ClearTextSSN search on stored data can help ensure personal information is secured properly. Analysts can also set alerts to detect personal information moving out of the network.

Imagine you are the CIO for a major bank or health insurance company, and suddenly receive an alert that large data transfers are occurring out of your networks. Your network security team informs you that the data includes clear text Social Security numbers- and that the attack was detected because the CyberLynx Cleartext SSN alert was triggered in Splunk. The security team rushes to block the attack, conducts forensics, and discovers other data stored as clear text. You team puts new security measures in place to encrypt stored personal data of your customers and protect against a future attack.

Detect Unauthorized Access: Wake on LAN Alert

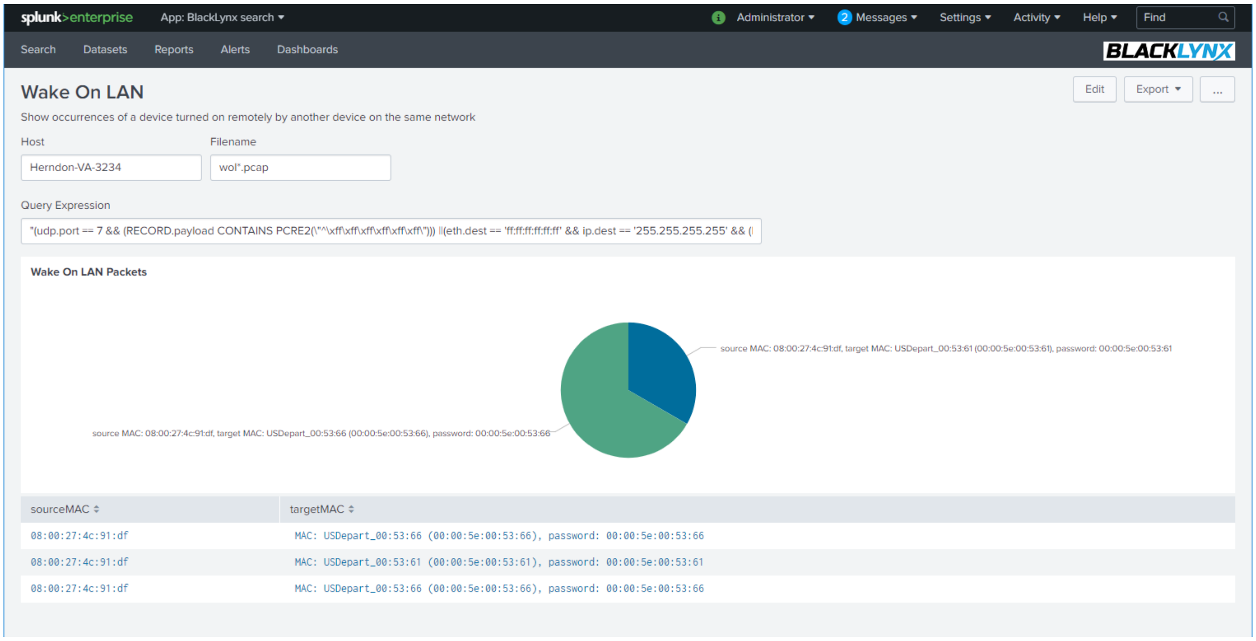
Wake on LAN is an ethernet networking standard that allows a computer to be turned on or awakened by a network message. This type of unauthorized access is a particular security concern for utilities, manufacturing facilities, internet of things, and edge devices. The CyberLynx dashboard searches a PCAP file or files for occurrences of the Wake-on-LAN message and returns the source and destinations for the message.

> Query Inputs
- Host – Extreme Search server hostname or IP Address
- Filename – PCAP file(s) to be searched relative to "/ryftone" directory
- Query Expression – search criteria, in this case the components of the Wake-On-Lan Message

> Operation and Output:
- Search of native PCAP files extracting those packets that match one of the two forms of the Wake On LAN message.

- Use of non-ascii patterns in the payload query
- Return of the results to Splunk in a JSON format

Use cases for this tool include unauthorized access or improper usage of a network; setting an alert for a Wake on LAN message would assist a cyber analyst in detecting unauthorized access.


splunk>enterprise
App: BlackLynx search ▾
Administrator ▾
2 Messages ▾
Settings ▾
Activity ▾
Help ▾
Find
Search
Datasets
Reports
Alerts
Dashboards
BLACKLYNX
Edit
Export ▾
...
source MAC: 08:00:27:4c:91:df, target MAC: USDepart_00:53:61 (00:00:5e:00:53:61), password: 00:00:5e:00:53:61
source MAC: 08:00:27:4c:91:df, target MAC: USDepart_00:53:66 (00:00:5e:00:53:66), password: 00:00:5e:00:53:66

Wake On LAN

Show occurrences of a device turned on remotely by another device on the same network

Host: Herndon-VA-3234

Filename: wol*.pcap

Query Expression:

```
"(udp.port == 7 && (RECORD.payload CONTAINS PCRE2(\"^\xff\xff\xff\xff\xff\xff\"))) ||(eth.dest == 'ff:ff:ff:ff:ff:ff' && ip.dest == '255.255.255.255' && (|
```

Wake On LAN Packets

sourceMAC ⇕	targetMAC ⇕
08:00:27:4c:91:df	MAC: USDepart_00:53:66 (00:00:5e:00:53:66), password: 00:00:5e:00:53:66
08:00:27:4c:91:df	MAC: USDepart_00:53:61 (00:00:5e:00:53:61), password: 00:00:5e:00:53:61
08:00:27:4c:91:df	MAC: USDepart_00:53:66 (00:00:5e:00:53:66), password: 00:00:5e:00:53:66

Unstructured File Query

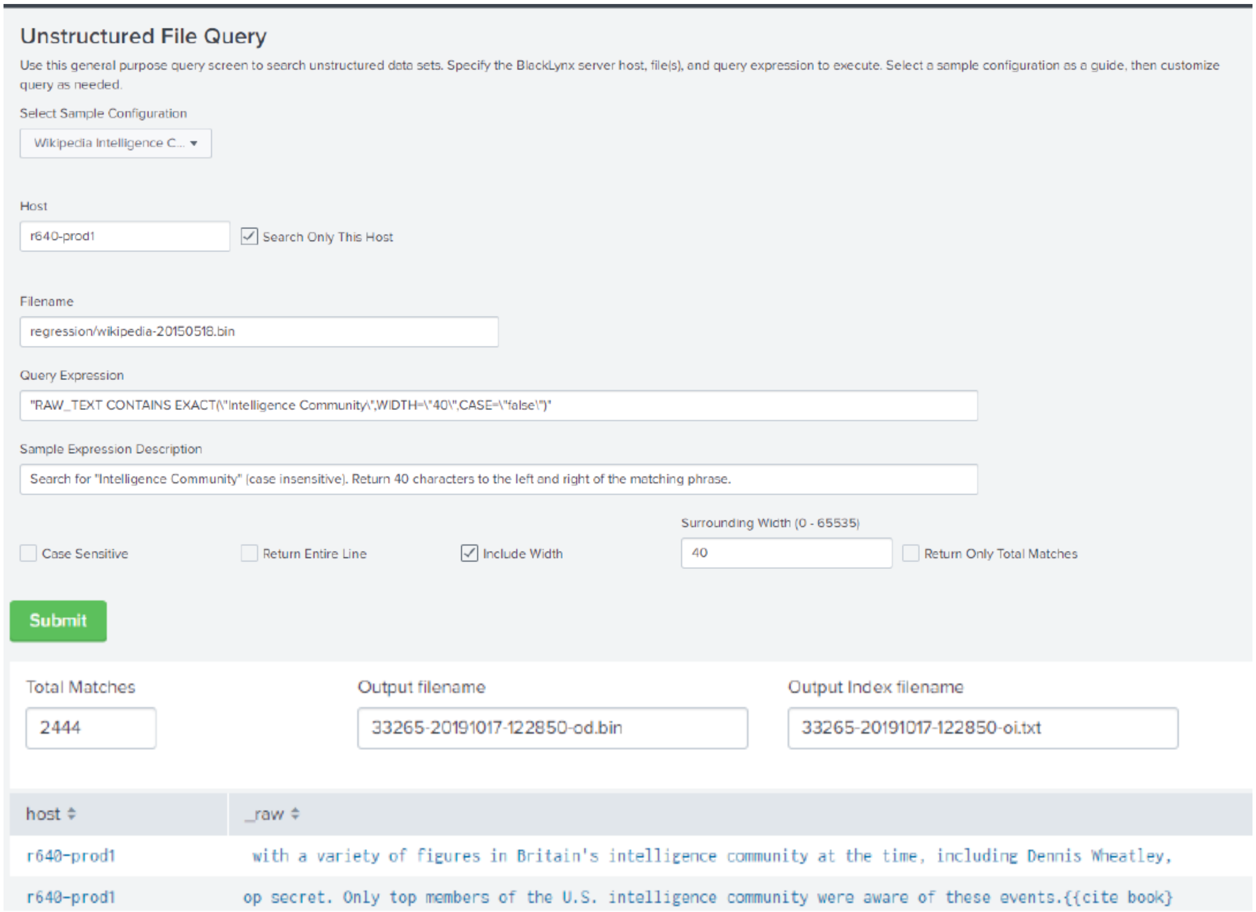
The "Unstructured File Query" can be used across a wide range of use cases for incident detection, analysis, and containment using unstructured data.

The Dashboard in the figure below provides a general purpose query screen which can be used to search unstructured data sets, like system logs, user logs, and application logs. Specify the Extreme Search server host, file(s), and query expression to execute. Select a sample configuration as a guide, then customize the query as needed.

> Query Inputs:
- Host –Extreme Search server hostname or IP Address
- Filename –Unstructured file(s) to be searched
- Query Expression –search criteria, examples of exact, fuzzy hamming distance and regular expression.

> Operation and Output:
- Search results of unstructured file(s) using a complex set of fuzzy search, exact, and regular expressions, without indexing or transformation.
- Return of the results to Splunk in a JSON format

The unstructured file query tool can be applied to a wide variety of attack types including malware outbreak, virus outbreak, data theft, unauthorized access, or improper usage.



Unstructured File Query

Use this general purpose query screen to search unstructured data sets. Specify the BlackLynx server host, file(s), and query expression to execute. Select a sample configuration as a guide, then customize query as needed.

Select Sample Configuration: Wikipedia Intelligence C... ▾

Host: r640-prod1 ☑ Search Only This Host

Filename: regression/wikipedia-20150518.bin

Query Expression: "RAW_TEXT CONTAINS EXACT(\"Intelligence Community\",WIDTH=\"40\",CASE=\"false\")"

Sample Expression Description: Search for "Intelligence Community" (case insensitive). Return 40 characters to the left and right of the matching phrase.

☐ Case Sensitive ☐ Return Entire Line ☑ Include Width Surrounding Width (0 - 65535): 40 ☐ Return Only Total Matches

Total Matches: 2444

Output filename: 33265-20191017-122850-od.bin

Output Index filename: 33265-20191017-122850-oi.txt

host ⇕	_raw ⇕
r640-prod1	with a variety of figures in Britain's intelligence community at the time, including Dennis Wheatley,
r640-prod1	op secret. Only top members of the U.S. intelligence community were aware of these events.{{cite book}

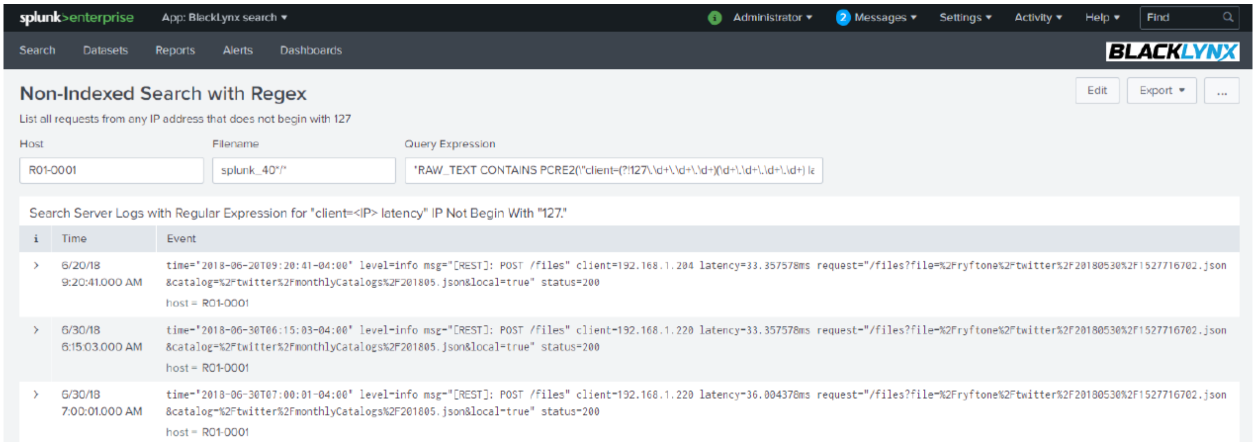
Non-indexed Search with Regex

The “Non-indexed Search with Regular Expression” performs a negative PCRE2 search on a semi-structured log file looking for any client address that does not begin with ‘127’ or the localhost. It returns any lines for the log that match the query and extracts the time from the line. For this search type CyberLynx leverages NPUSearch,$^{\text{TM}}$ a neuromorphic computing algorithm developed by Lewis Rhodes Labs.

> Query Inputs:
- Host –LRL server hostname or IP Address
- Filename –PCAP file(s) to be searched
- Query Expression –search criteria, in this case client=<IP address> where the IP address does not begin 127.*.*.*

> Operation and Output:
- Search of text log files using a regular expression
- Return of the results to Splunk in a JSON format

This type of search can be used to detect a wide range of cyber incidents.



Non-Indexed Search with Regex

List all requests from any IP address that does not begin with 127

Host	Filename	Query Expression
R01-0001	splunk_40*/*	"RAW_TEXT CONTAINS PCRE2(\"client=(?!127\.\d+\.\d+\.\d+)(\d+\.\d+\.\d+\.\d+) la

Search Server Logs with Regular Expression for "client=<IP> latency" IP Not Begin With "127."

i	Time	Event

>	6/20/18
	9:20:41.000 AM
		`time='2018-06-20T09:20:41-04:00' level=info msg="[REST]: POST /files" client=192.168.1.204 latency=33.357578ms request="/files?file=%2Fryftone%2Ftwitter%2F20180530%2F1527716702.json`
		`&catalog=%2Ftwitter%2FmonthlyCatalogs%2F201805.json&local=true" status=200`
		host = R01-0001

>	6/30/18
	6:15:03.000 AM
		`time='2018-06-30T06:15:03-04:00' level=info msg="[REST]: POST /files" client=192.168.1.220 latency=33.357578ms request="/files?file=%2Fryftone%2Ftwitter%2F20180530%2F1527716702.json`
		`&catalog=%2Ftwitter%2FmonthlyCatalogs%2F201805.json&local=true" status=200`
		host = R01-0001

>	6/30/18
	7:00:01.000 AM
		`time='2018-06-30T07:00:01-04:00' level=info msg="[REST]: POST /files" client=192.168.1.220 latency=36.004378ms request="/files?file=%2Fryftone%2Ftwitter%2F20180530%2F1527716702.json`
		`&catalog=%2Ftwitter%2FmonthlyCatalogs%2F201805.json&local=true" status=200`
		host = R01-0001