



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.

The screenshot shows the Splunk Enterprise interface for the Blacklynx search application. The dashboard is titled 'Blacklisted DNS' and includes a search bar with the following inputs: Host: Herndon-VA-0313, Filename: PCAP/wired-lan-session.pcap, and Query Expression: 'udp.srcport == 53'. Below the search bar, there is a table of results with the following columns: DNS Name and blacklisted. The table contains 15 rows of domain names, with the first three rows having a 'True' value in the 'blacklisted' column.

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	
at162.d.akamai.net	
accounts.google.com	
ad.adlegend.com	
ad.wsod.com	
ads.us-east-1.prod.a.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



time	http.request.uri	filename	blacklisted
Sep 11, 2018 14:04:16.851824000 EDT	/releases/ryftx/old/ryftx-tgz-installer-x86_64-U1664-r736.tgz	ryftx-tgz-installer-x86_64-U1664-r736.tgz	True
Sep 11, 2018 14:03:50.576225600 EDT	/icons/blank.gif	blank.gif	
Sep 11, 2018 14:03:50.578299000 EDT	/icons/back.gif	back.gif	
Sep 11, 2018 14:03:50.578335600 EDT	/icons/folder.gif	folder.gif	
Sep 11, 2018 14:03:52.923626000 EDT	/icons/text.gif	text.gif	
Sep 11, 2018 14:03:52.924190000 EDT	/icons/compressed.gif	compressed.gif	
Sep 11, 2018 14:03:58.893232000 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-U1664-r839.tgz	ryftx-tgz-installer-x86_64-U1664-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.

The screenshot shows the Splunk Enterprise interface for a search titled "Cybersquatting Example For DNS Bro Logs Using Edit Distance". The search criteria are: Host: Herndon-VA-0313, Filename: DNS*.log, and 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>



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.JPEG7SSN=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.JPEG7SSN=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.JPEG7SSN=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.JPEG7SSN=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. Your 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

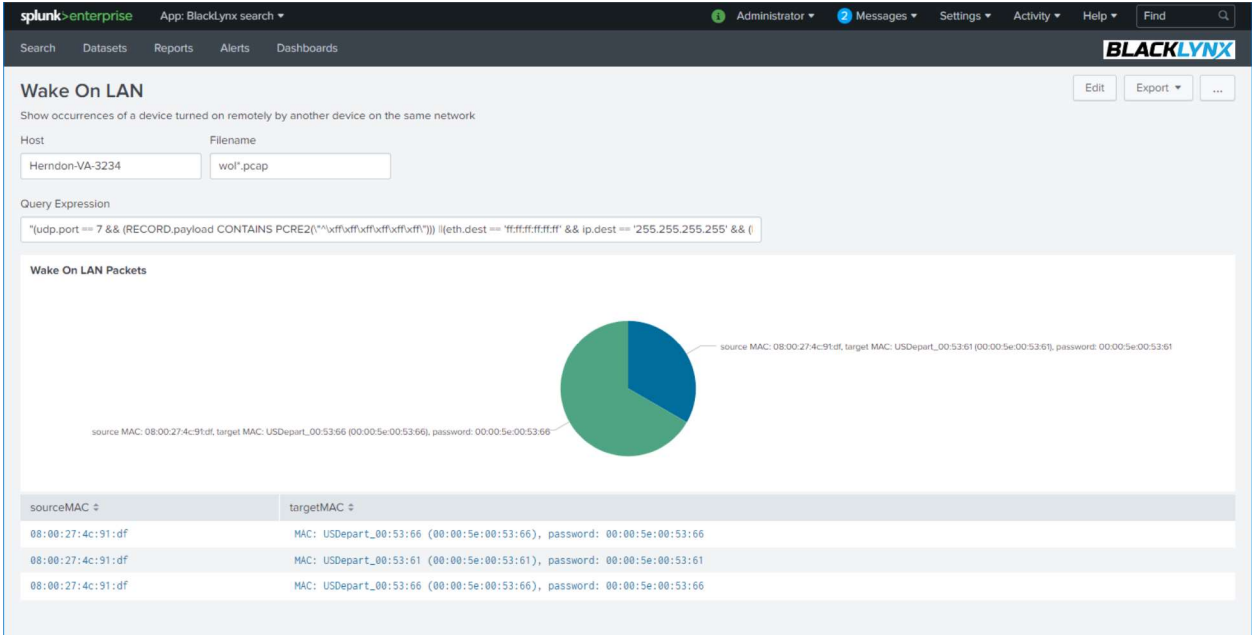
Wake on LAN is an ethernet networking standard that allows a computer to be turned on or awakened by a network message. This type of unauthorized access is a particular security concern for utilities, manufacturing facilities, internet of things, and edge devices. The CyberLynx dashboard searches a PCAP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



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 Return Only Total Matches

Surrounding Width (0 - 65535)
40

Submit

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,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.



splunk>enterprise App: BlackLynx search Administrator Messages Settings Activity Help Find

Search Datasets Reports Alerts Dashboards **BLACKLYNX**

Non-Indexed Search with Regexp

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

Host: R01-0001 Filename: splunk_407* Query Expression: *RAW_TEXT CONTAINS PCRE2(!^client-(?!127)\d+\.\d+\.\d+)(\d+\.\d+\.\d+)|k

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%2F20180536%2F1527716762.json&catalog=%2Ftwitter%2FmonthlyCatalogs%2F201805.json&local=true" status=200 host= R01-0001
>	6/30/18 8:15:03.000 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%2F20180536%2F1527716762.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%2F20180536%2F1527716762.json&catalog=%2Ftwitter%2FmonthlyCatalogs%2F201805.json&local=true" status=200 host= R01-0001