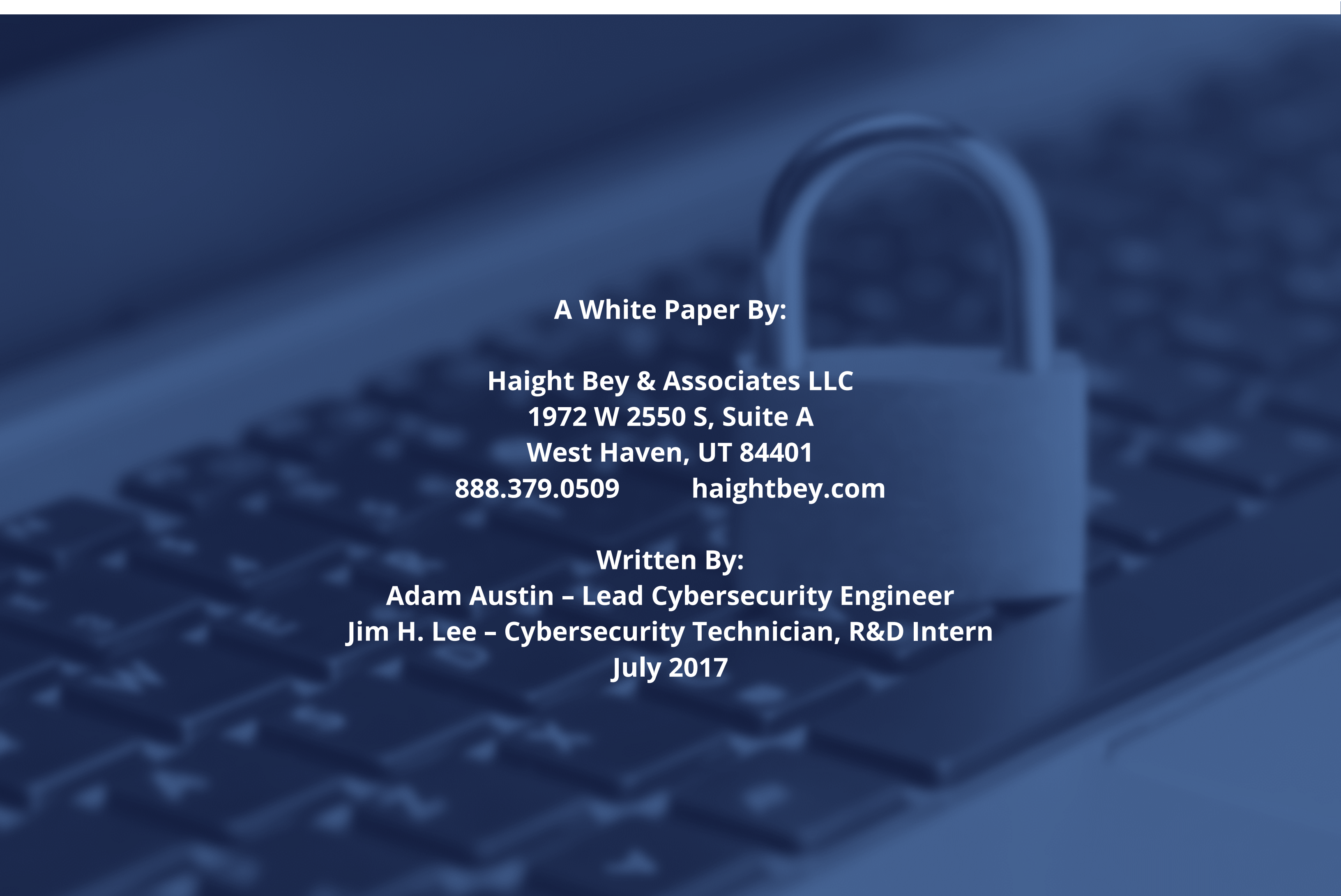
Expanding the Small Business Cybersecurity Hunter's Field of Vision: LogRhythm NetMon Freemium



A White Paper By:

Haight Bey & Associates LLC
1972 W 2550 S, Suite A
West Haven, UT 84401
888.379.0509 haightbey.com

Written By:
Adam Austin – Lead Cybersecurity Engineer
Jim H. Lee – Cybersecurity Technician, R&D Intern
July 2017

Bottom Line Up Front

LogRhythm NetMon Freemium is a valuable tool for someone hunting network traffic anomalies in a small business network, especially since the software is free. Although it has its front/backend limitations and should be one tool of several in a toolkit, it meets many industry standards for network traffic analysis. It can certainly help small business' cybersecurity network security analysts—"Hunters"—quickly baseline network traffic, identify anomalies, and pursue further investigation.

Haight Bey has no affiliation with LogRhythm; we did provide this white paper to LogRhythm for factual review.

Why do what we did?

Cybersecurity and IT professionals who are responsible for securing a network must be fully cognizant of their target environment to distinguish between anticipated and unexpected traffic flow. Often, outlier and anomalous network events require further examination and clarification by a network security analyst—a "Hunter" as we at Haight Bey & Associates LLC (HBA) call such a person. Several critical questions exist to gauge the Hunter's effectiveness in detecting atypical network communications:

- How do Hunters discover and record network traffic anomalies after normal business hours?
- Besides obvious anomalies, such as high volumes of network traffic when a business is closed for the day or week, what other types of network traffic should be considered suspicious?

- When deviations are identified, is it possible to narrow down a time frame of interest and gain full transparency of the communication methods as well as other crucial details?
- What tools/techniques exist to answer these previous questions that are affordable and practical for a small- or medium-sized organization?

At HBA we are always on the lookout for free or low-cost tools to help address these questions, so we were excited when we heard that LogRhythm provides a free version of its Network Monitor product: NetMon Freemium. In July 2017, the research and development team (R&D, "we") at HBA decided to try out this tool. Since HBA offers cybersecurity empowerment services to small to medium sized businesses (SMB), the primary goal was to prototype a LogRhythm NetMon Freemium solution that could bolster the network monitoring systems of SMBs. We intended to configure and integrate this tool into an existing, mature, network using built-in features of the product, including customizable visualizations and dashboards. As a result, we hoped to show that SMBs can supplement their existing cybersecurity tool suite with low-cost or free network monitoring solutions. With assistance from SANS™ Institute resources, we successfully configured the LogRhythm NetMon Freemium interface for Hunters to utilize in the network environment of a small business. The rest of this document outlines the R&D process, including an overall assessment of LogRhythm NetMon Freemium's capabilities in an SMB environment, its technical limitations, suggestions for front/backend improvements, and HBA's future development goals with the product.

What did we want to learn?

To evaluate the functional practicality of LogRhythm NetMon Freemium for an SMB's Hunter, we posed these questions to guide development and implementation:

i) How easy is LogRhythm NetMon Freemium to integrate into a typical SMB network? What costs and human resources are required?

ii) What extent of network traffic is accessible to LogRhythm NetMon Freemium in a typical SMB network? How can this be verified with other network tools?

iii) Given the network traffic accessible to LogRhythm NetMon Freemium, is this information significant to a Hunter?

iv) Is the out-of-the-box user-interface/experience intuitive, enjoyable to use, and easily configurable?

v) Are the product's built-in tables, visualizations, and dashboards useful in establishing a baseline of network traffic against which anomalies stand out?

vi) If a Hunter is dissatisfied with a specific dashboard, can he or she easily create a new table, visualizations, or dashboard in LogRhythm NetMon Freemium's development environment?

What did we do?

Integration

Our first step toward answering these questions was to integrate LogRhythm NetMon Freemium into an SMB's network environment, and to determine the costs of purchasing a computing device that could adequately support the software. In *Packets Don't Lie: LogRhythm NetMon Freemium Review,* author Dave Shackleford expresses that his team finished this process within half-an-hour. We at HBA budgeted an hour for this process, to include some moderate troubleshooting.

We received permission from an SMB client of HBA in Utah to prototype a solution on their internal network. The SMB is a micro-business, with fewer than 20 employees and workstations, and less than 25 Mbps traffic. The network hosts typical office devices, including printers, proprietary internal application and database servers, and a local physical security system.

Integrating LogRhythm NetMon Freemium into this network environment required some collaboration with the SMB's IT service provider to configure a switch to mirror traffic to a SPAN port. Traffic from the following switch ports was mirrored to the SPAN (See Figure 1):

- Network Firewall (network egress/ingress)
- Wireless Access Point (wireless traffic)
- Security Camera System
- Virtual Host VM Network Link (traffic to/from Virtual Server Guests)

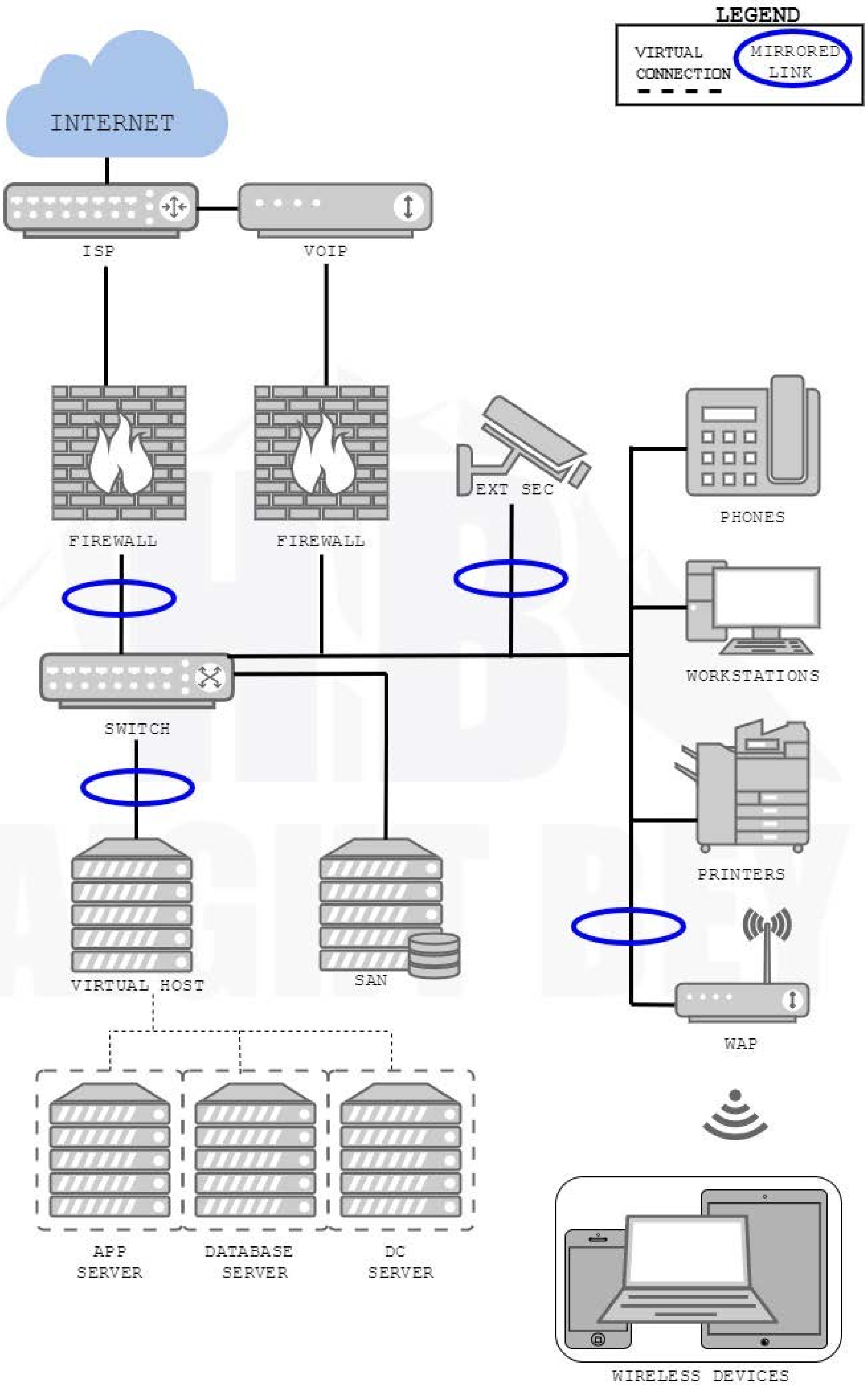


Figure 1: SMB network links mirrored to SPAN port for analysis

Once the SPAN port was configured, we used LogRhythm NetMon Freemium's installation and setup guides, as well as computing hardware suggestions from a LogRhythm blog post, to integrate a standalone LogRhythm NetMon Freemium device into the network. Thus, we were able to view ingress/egress network traffic through a management web-interface in roughly one hour, including the 15 minutes or so it took to install LogRhythm NetMon Freemium onto the chosen hardware device. We used the passively-cooled Vault mini-PC from Protectli, available on Amazon for $359.00 with 8GB RAM and 120GB mSATA SSD. Even after upgrading to a 250GB SSD, the total cost of the hardware solution was under $500; easily affordable for most SMBs. Considering LogRhythm NetMon Freemium is also offered as a virtual machine, the solution could be made even cheaper.

Traffic Visibility

Prior to verifying LogRhythm NetMon Freemium's detectable network perimeter, we used Nmap to confirm IP addresses, device types, and the topology of the target network. We ensured that communications among these devices were also visible by simply querying `SrcIP:X.Y.Z.* OR DestIP:X.Y.Z.*` within LogRhythm NetMon Freemium's search interface and then comparing the results with the output of an Nmap scan. Consequently, we established LogRhythm NetMon Freemium's scope of visibility as presented in Figure 2:

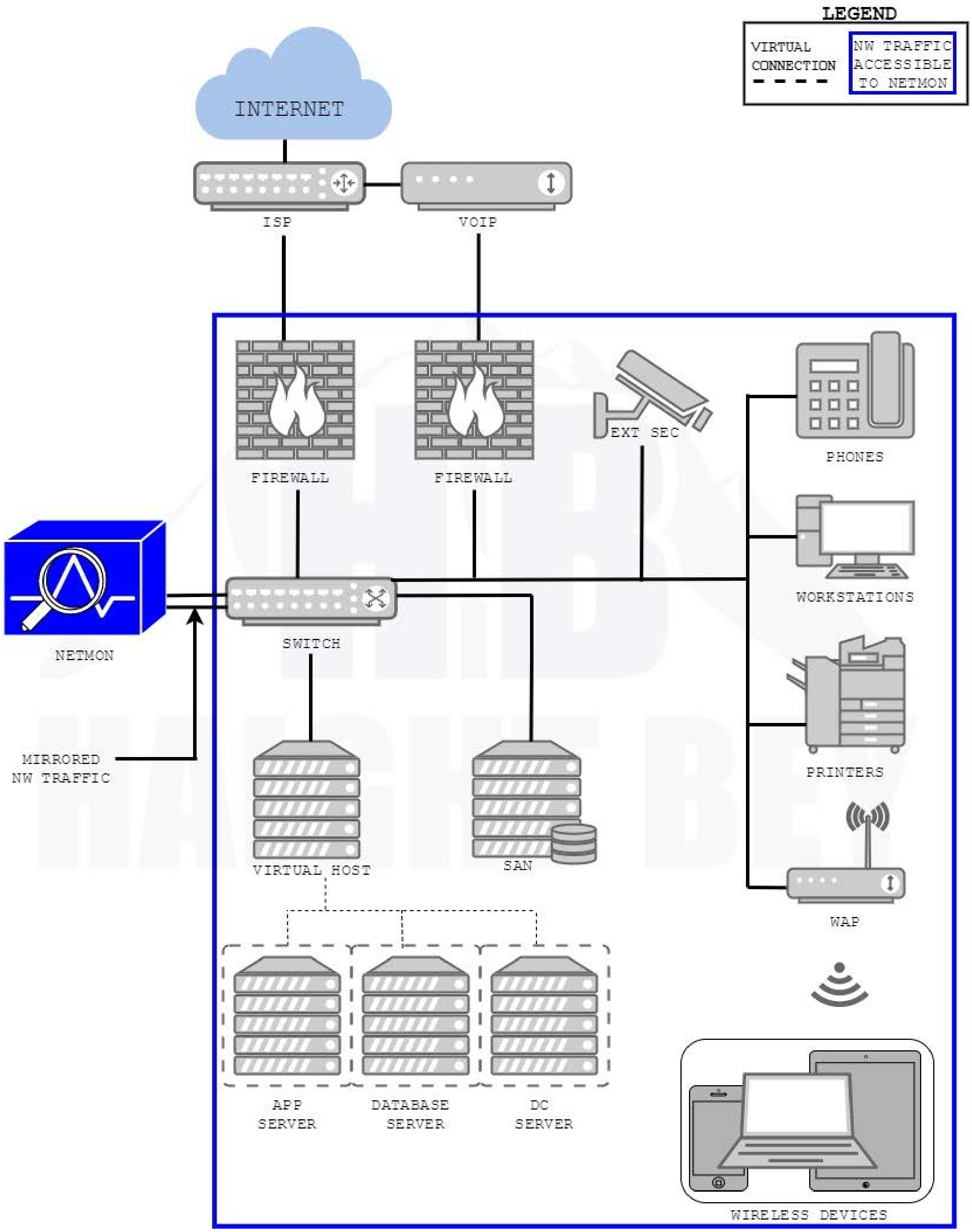


Figure 2: SMB Network traffic visible to LogRhythm NetMon Freemium device

Ultimately, we were able to successfully capture most of the SMB's network traffic flow for analysis in LogRhythm NetMon Freemium, including:

- All ingress/egress traffic through the firewall
- All internal traffic to/from internal servers

Limitations on traffic visibility in this scenario include:

- Intra-virtual machine traffic
- Some internal-only traffic, such as from workstations to printer and vice versa
- Traffic from Virtual Host to SAN on a dedicated VLAN
- VOIP traffic

Despite these limitations, this simple LogRhythm NetMon Freemium setup significantly improved the SMB's visibility into its network traffic, and extended the insight the SMB's Hunter had into the network.

User Interface

LogRhythm NetMon Freemium provides out-of-the-box browser-based interfaces that provide an immense amount of network traffic data that allows the Hunter to adjust the graphical interface to suit his/her needs. The "Analyze" user interface allows, for example:

- Ability to configure displayed data period over minutes, hours, days, weeks, and months
- Granular queries on available detected fields to filter captured traffic
- Metadata summaries of flow sessions which include application path, IPs, MACs, etc.

- Option to turn on packet capture for all analyzed traffic flows
- Ability to download captured file and packet streams in *.pcap format
- Replay functionality with *.pcap formatted streams

Alone, the ability to categorize and download pcaps of traffic of interest makes LogRhythm NetMon Freemium a valuable tool for the Hunter; other free (and not-so-free!) tools do not have this ability out of the box. Reducing a Hunter's analysis time on a forensic task, such as file carving from a TCP stream, for instance, is imperative for an SMB with limited time resources.

Standards-based Analysis

Next, we investigated the usefulness of the visualizations and dashboards within LogRhythm NetMon Freemium for baselining network traffic to facilitate identification of anomalies. As a practical guide and ad hoc standard for baselining network traffic, we utilized the SANS™ Institute's *SANS DFIR Network Forensics and Analysis Poster* (DFIR-Network_v1_4-17, available here for download with SANS™ account) for this stage of the R&D process. Our ambition was to gauge LogRhythm NetMon Freemium's ability to support the analytical methods outlined in the "Network Traffic Anomalies" section of the poster (See Figure 3).

We chose this resource from the SANS™ Institute because:

a) SANS™ Institute is highly respected in the cybersecurity industry, and
b) The Network Traffic Anomalies section of the poster covers a wide spectrum of network traffic to baseline and analyze, and
c) The poster clearly and succinctly outlined the types of network information that could help a Hunter recognize suspicious trends or malicious activities

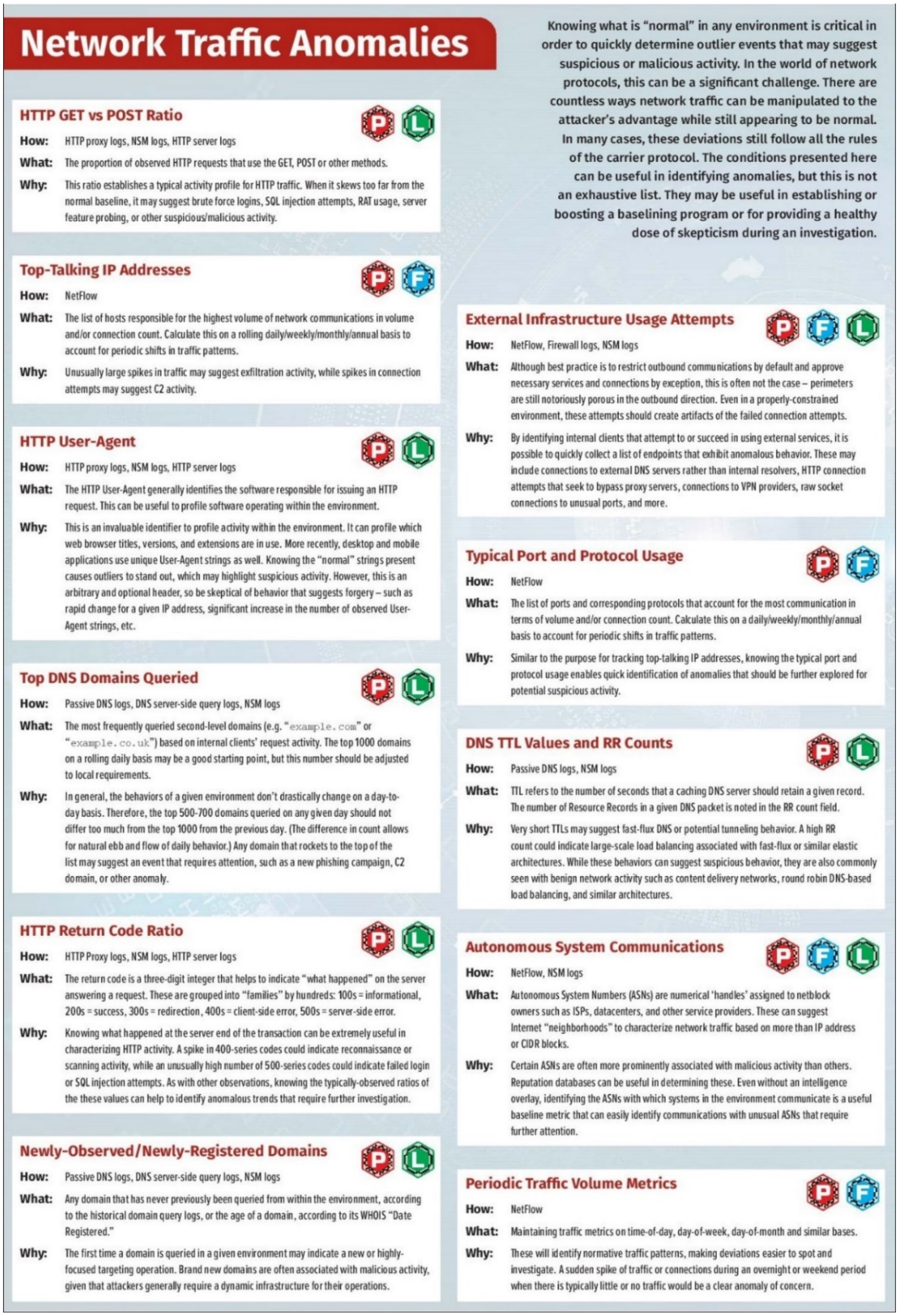
The poster's explanations for each Network Traffic Anomaly (NTA) clarified the relevant data that we needed to display in visualizations and dashboards.

Following Figure 3 are screen captures of the successfully created dashboards with a brief explanation of how a Hunter would use each dashboard to baseline network traffic and subsequently discover anomalies. All visualizations were produced using LogRhythm NetMon Freemium's Visualize development environment.

More about the LogRhythm NetMon Freemium GUI

LogRhythm NetMon Freemium utilizes the Kibana GUI plugin for the open-source Elasticsearch data analysis engine. Kibana uses the term "visualization" for a single graphical display of data, and the term "dashboard" for an aggregation of

Where appropriate, we combined related visualizations of NTA standards into a common dashboard. If you are interested in a general description of how the visualizations and dashboards were created, see Appendix A.



Network Traffic Anomalies

Knowing what is “normal” in any environment is critical in order to quickly determine outlier events that may suggest suspicious or malicious activity. In the world of network protocols, this can be a significant challenge. There are countless ways network traffic can be manipulated to the attacker’s advantage while still appearing to be normal. In many cases, these deviations still follow all the rules of the carrier protocol. The conditions presented here can be useful in identifying anomalies, but this is not an exhaustive list. They may be useful in establishing or boosting a baselining program or for providing a healthy dose of skepticism during an investigation.

HTTP GET vs POST Ratio

P L

How: HTTP proxy logs, NSM logs, HTTP server logs

What: The proportion of observed HTTP requests that use the GET, POST or other methods.

Why: This ratio establishes a typical activity profile for HTTP traffic. When it skews too far from the normal baseline, it may suggest brute force logins, SQL injection attempts, RAT usage, server feature probing, or other suspicious/malicious activity.

Top-Talking IP Addresses

P F

How: NetFlow

What: The list of hosts responsible for the highest volume of network communications in volume and/or connection count. Calculate this on a rolling daily/weekly/monthly/annual basis to account for periodic shifts in traffic patterns.

Why: Unusually large spikes in traffic may suggest exfiltration activity, while spikes in connection attempts may suggest C2 activity.

HTTP User-Agent

P L

How: HTTP proxy logs, NSM logs, HTTP server logs

What: The HTTP User-Agent generally identifies the software responsible for issuing an HTTP request. This can be useful to profile software operating within the environment.

Why: This is an invaluable identifier to profile activity within the environment. It can profile which web browser titles, versions, and extensions are in use. More recently, desktop and mobile applications use unique User-Agent strings as well. Knowing the “normal” strings present causes outliers to stand out, which may highlight suspicious activity. However, this is an arbitrary and optional header, so be skeptical of behavior that suggests forgery – such as rapid change for a given IP address, significant increase in the number of observed User-Agent strings, etc.

Top DNS Domains Queried

P L

How: Passive DNS logs, DNS server-side query logs, NSM logs

What: The most frequently queried second-level domains (e.g. “`example.com`” or “`example.co.uk`”) based on internal clients’ request activity. The top 1000 domains on a rolling daily basis may be a good starting point, but this number should be adjusted to local requirements.

Why: In general, the behaviors of a given environment don’t drastically change on a day-to-day basis. Therefore, the top 500-700 domains queried on any given day should not differ too much from the top 1000 from the previous day. (The difference in count allows for natural ebb and flow of daily behavior.) Any domain that rockets to the top of the list may suggest an event that requires attention, such as a new phishing campaign, C2 domain, or other anomaly.

HTTP Return Code Ratio

P L

How: HTTP Proxy logs, NSM logs, HTTP server logs

What: The return code is a three-digit integer that helps to indicate “what happened” on the server answering a request. These are grouped into “families” by hundreds: 100s = informational, 200s = success, 300s = redirection, 400s = client-side error, 500s = server-side error.

Why: Knowing what happened at the server end of the transaction can be extremely useful in characterizing HTTP activity. A spike in 400-series codes could indicate reconnaissance or scanning activity, while an unusually high number of 500-series codes could indicate failed login or SQL injection attempts. As with other observations, knowing the typically-observed ratios of the these values can help to identify anomalous trends that require further investigation.

Newly-Observed/Newly-Registered Domains

P L

How: Passive DNS logs, DNS server-side query logs, NSM logs

What: Any domain that has never previously been queried from within the environment, according to the historical domain query logs, or the age of a domain, according to its WHOIS “Date Registered.”

Why: The first time a domain is queried in a given environment may indicate a new or highly-focused targeting operation. Brand new domains are often associated with malicious activity, given that attackers generally require a dynamic infrastructure for their operations.

External Infrastructure Usage Attempts

P F L

How: NetFlow, Firewall logs, NSM logs

What: Although best practice is to restrict outbound communications by default and approve necessary services and connections by exception, this is often not the case – perimeters are still notoriously porous in the outbound direction. Even in a properly-constrained environment, these attempts should create artifacts of the failed connection attempts.

Why: By identifying internal clients that attempt to or succeed in using external services, it is possible to quickly collect a list of endpoints that exhibit anomalous behavior. These may include connections to external DNS servers rather than internal resolvers, HTTP connection attempts that seek to bypass proxy servers, connections to VPN providers, raw socket connections to unusual ports, and more.

Typical Port and Protocol Usage

P F

How: NetFlow

What: The list of ports and corresponding protocols that account for the most communication in terms of volume and/or connection count. Calculate this on a daily/weekly/monthly/annual basis to account for periodic shifts in traffic patterns.

Why: Similar to the purpose for tracking top-talking IP addresses, knowing the typical port and protocol usage enables quick identification of anomalies that should be further explored for potential suspicious activity.

DNS TTL Values and RR Counts

P L

How: Passive DNS logs, NSM logs

What: TTL refers to the number of seconds that a caching DNS server should retain a given record. The number of Resource Records in a given DNS packet is noted in the RR count field.

Why: Very short TTLs may suggest fast-flux DNS or potential tunneling behavior. A high RR count could indicate large-scale load balancing associated with fast-flux or similar elastic architectures. While these behaviors can suggest suspicious behavior, they are also commonly seen with benign network activity such as content delivery networks, round robin DNS-based load balancing, and similar architectures.

Autonomous System Communications

P F L

How: NetFlow, NSM logs

What: Autonomous System Numbers (ASNs) are numerical ‘handles’ assigned to netblock owners such as ISPs, datacenters, and other service providers. These can suggest Internet “neighborhoods” to characterize network traffic based on more than IP address or CIDR blocks.

Why: Certain ASNs are often more prominently associated with malicious activity than others. Reputation databases can be useful in determining these. Even without an intelligence overlay, identifying the ASNs with which systems in the environment communicate is a useful baseline metric that can easily identify communications with unusual ASNs that require further attention.

Periodic Traffic Volume Metrics

P F

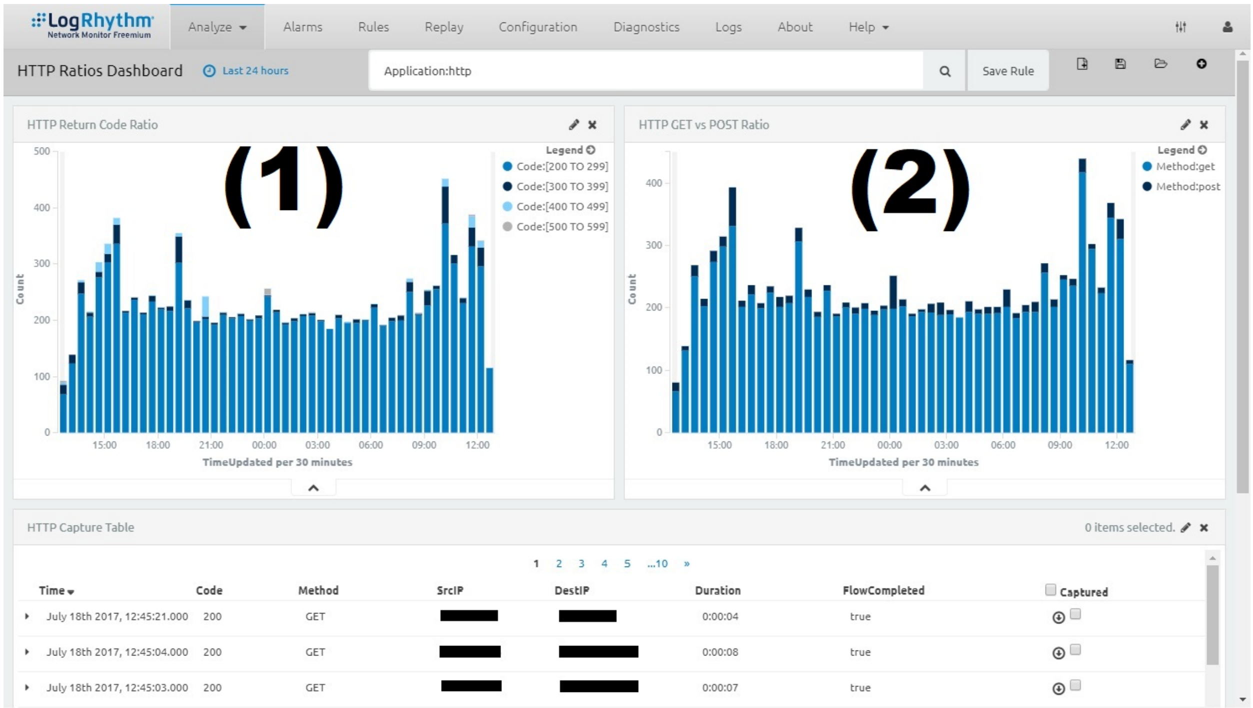
How: NetFlow

What: Maintaining traffic metrics on time-of-day, day-of-week, day-of-month and similar bases.

Why: These will identify normative traffic patterns, making deviations easier to spot and investigate. A sudden spike of traffic or connections during an overnight or weekend period when there is typically little or no traffic would be a clear anomaly of concern.

Figure 3: SANS DFIR poster Network Traffic Anomalies section (reprinted with permission from SANS™ Institute)

Dashboard for “HTTP GET vs POST Ratio” and “HTTP Return Code Ratio”



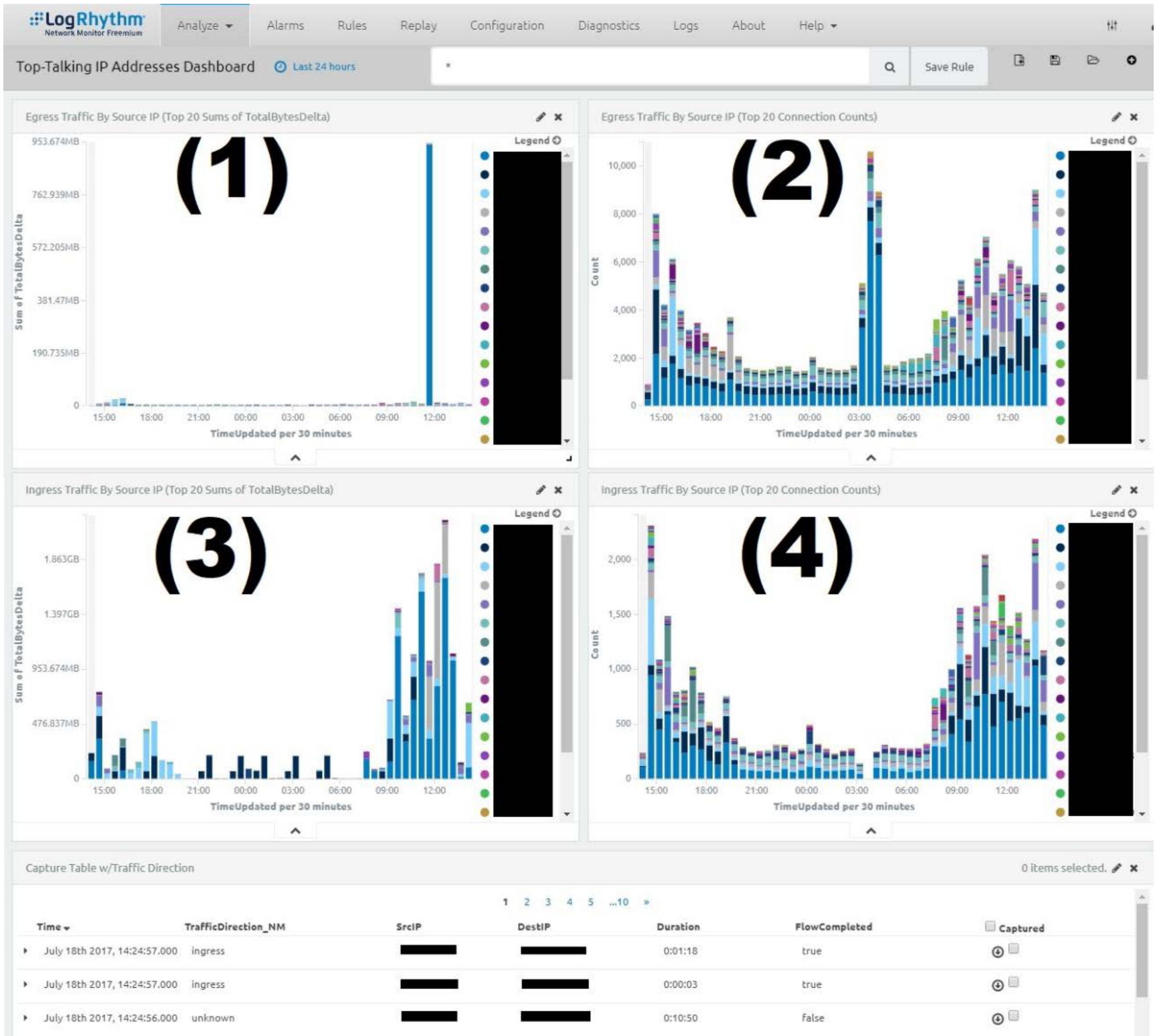
Time	Code	Method	SrcIP	DestIP	Duration	FlowCompleted	□ Captured
July 18th 2017, 12:45:21.000	200	GET	[redacted]	[redacted]	0:00:04	true	□
July 18th 2017, 12:45:04.000	200	GET	[redacted]	[redacted]	0:00:08	true	□
July 18th 2017, 12:45:03.000	200	GET	[redacted]	[redacted]	0:00:07	true	□

Figure 4: HTTP GET/POST and Return Code Dashboard

How would a Hunter use this dashboard?[1]	
Over time, establish a baseline of:	**Investigate further when:**
Typical proportion(s) between HTTP GET and POST request methods	Observed ratio deviates from normal baseline
Typical proportion(s) among #00-series return codes	Observed frequency distribution displays one or multiple spikes in #00-series return codes

[1] The contents of these tables i.e. “How would a Hunter use this dashboard?” consist of text directly from Figure 3.

Dashboard for “Top-Talking IP Addresses”

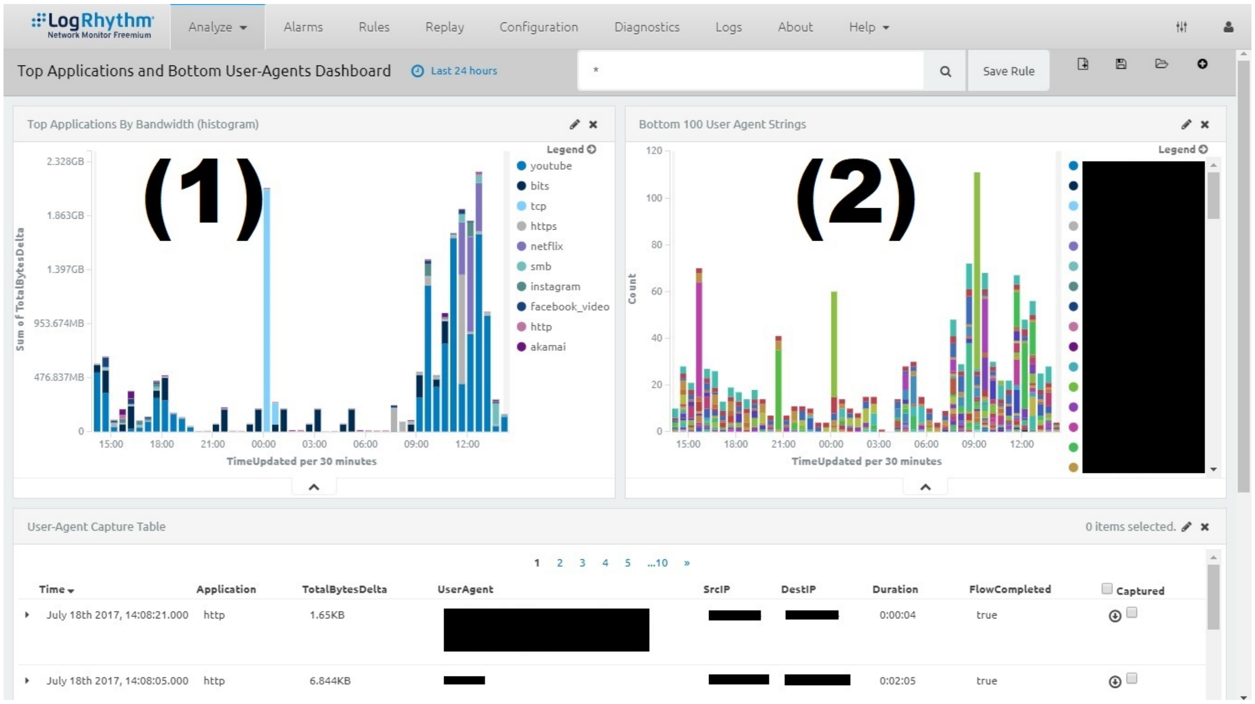


Time	TrafficDirection_NM	SrcIP	DestIP	Duration	FlowCompleted	□ Captured
July 18th 2017, 14:24:57.000	ingress	[redacted]	[redacted]	0:01:18	true	□
July 18th 2017, 14:24:57.000	ingress	[redacted]	[redacted]	0:00:03	true	□
July 18th 2017, 14:24:56.000	unknown	[redacted]	[redacted]	0:10:50	false	□

Figure 5: Top-Talking IP addresses Dashboard

How would a Hunter use this dashboard?	
Over time, establish a baseline of:	**Investigate further when:**
Hosts associated with the highest network communications in terms of bytes transferred and connection counts	Observed distribution displays large spikes in traffic from unusual hosts during and after regular business hours

Dashboard for “HTTP User-Agent”*



[redacted]

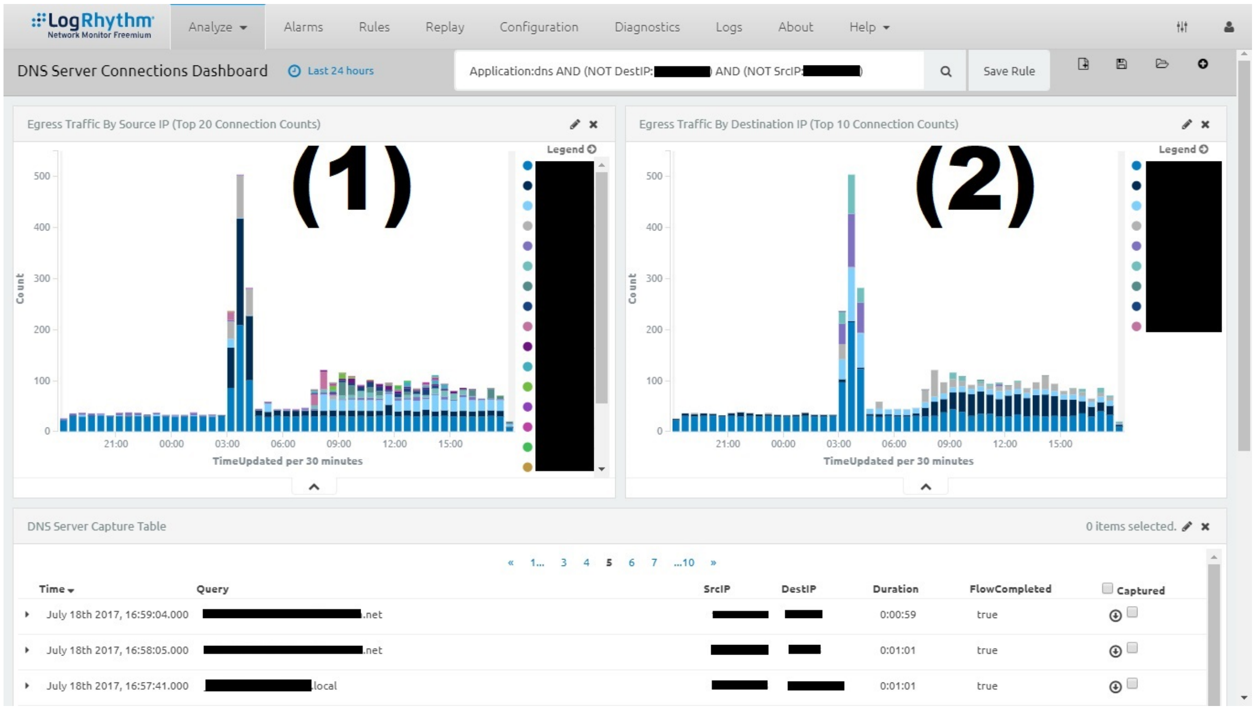
Time	Application	TotalBytesDelta	UserAgent	SrcIP	DestIP	Duration	FlowCompleted	□ Captured
July 18th 2017, 14:08:21.000	http	1.65KB	[redacted]	[redacted]	[redacted]	0:00:04	true	□
July 18th 2017, 14:08:05.000	http	6.844KB	[redacted]	[redacted]	[redacted]	0:02:05	true	□

Figure 6: HTTP User-Agent Dashboard

*Dashboard configured to capture user-agent strings used by all application protocols

How would a Hunter use this dashboard?	
Over time, establish a baseline of:	**Investigate further when:**
Typical proportion(s) among applications and protocols in terms of bytes transferred	Observed frequency distributions display large spikes in bandwidth consumption from clear text protocols or unusual applications
Most common and verified user-agent strings	Outlier user-agent(s) that display abnormal distinctions including acronyms, misspellings, strings mixed with numbers/other unusual symbols, unidentified entities (“Agent”), etc.

Dashboard for “External Infrastructure Usage Attempts”**



DNS Server Connections Dashboard

Last 24 hours

Application:dns AND (NOT DestIP:[redacted]) AND (NOT SrcIP:[redacted])

[redacted]

[redacted]

DNS Server Capture Table

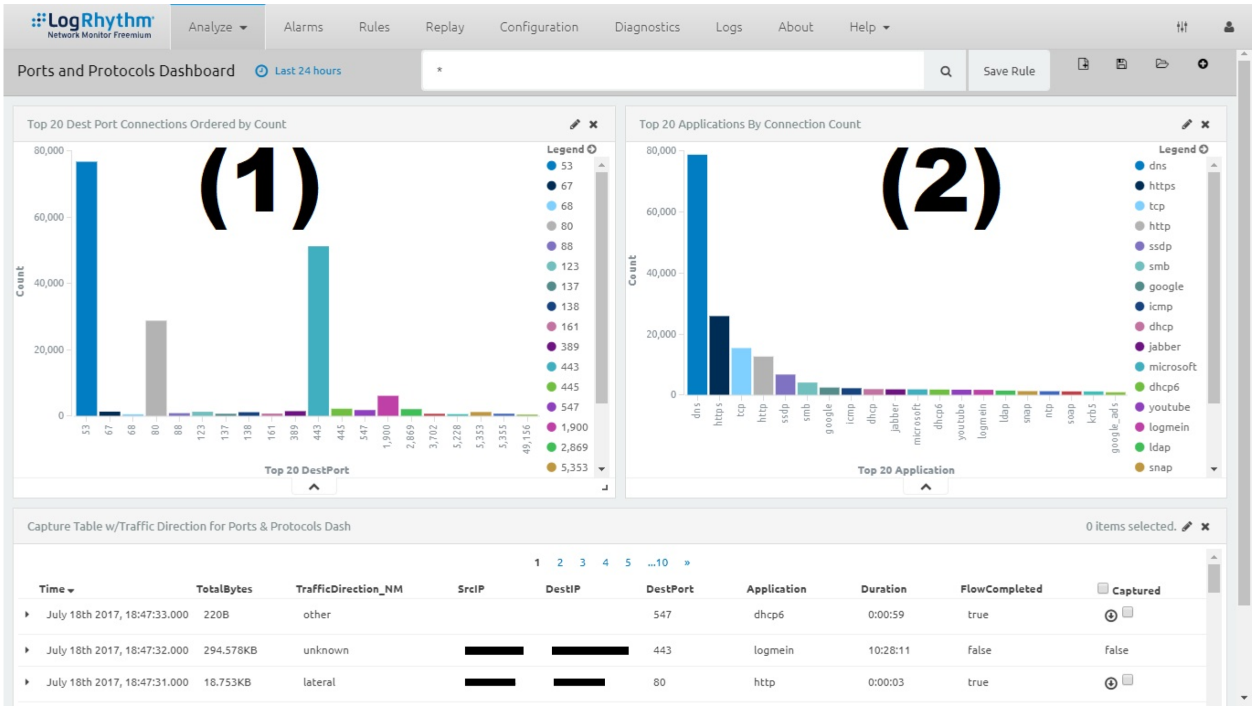
Time	Query	SrcIP	DestIP	Duration	FlowCompleted	□ Captured
July 18th 2017, 16:59:04.000	[redacted].net	[redacted]	[redacted]	0:00:59	true	□
July 18th 2017, 16:58:05.000	[redacted].net	[redacted]	[redacted]	0:01:01	true	□
July 18th 2017, 16:57:41.000	[redacted].local	[redacted]	[redacted]	0:01:01	true	□

Figure 7: External Infrastructure Usage Attempts Dashboard

**Dashboard configured to focus only on discovering DNS traffic to external resolvers; the other NTA standards proved challenging to implement in the Visualize development environment in the time frame for this R&D project. See ‘What else do we want to do’ section below (page 30) for more information.

How would a Hunter use this dashboard?	
Over time, establish a baseline of:	**Investigate further when:**
Typical usage of external resolvers	An internal client attempts to gain DNS resolution from an external source instead of the internal DNS server

Dashboard for “Typical Port and Protocol Usage”

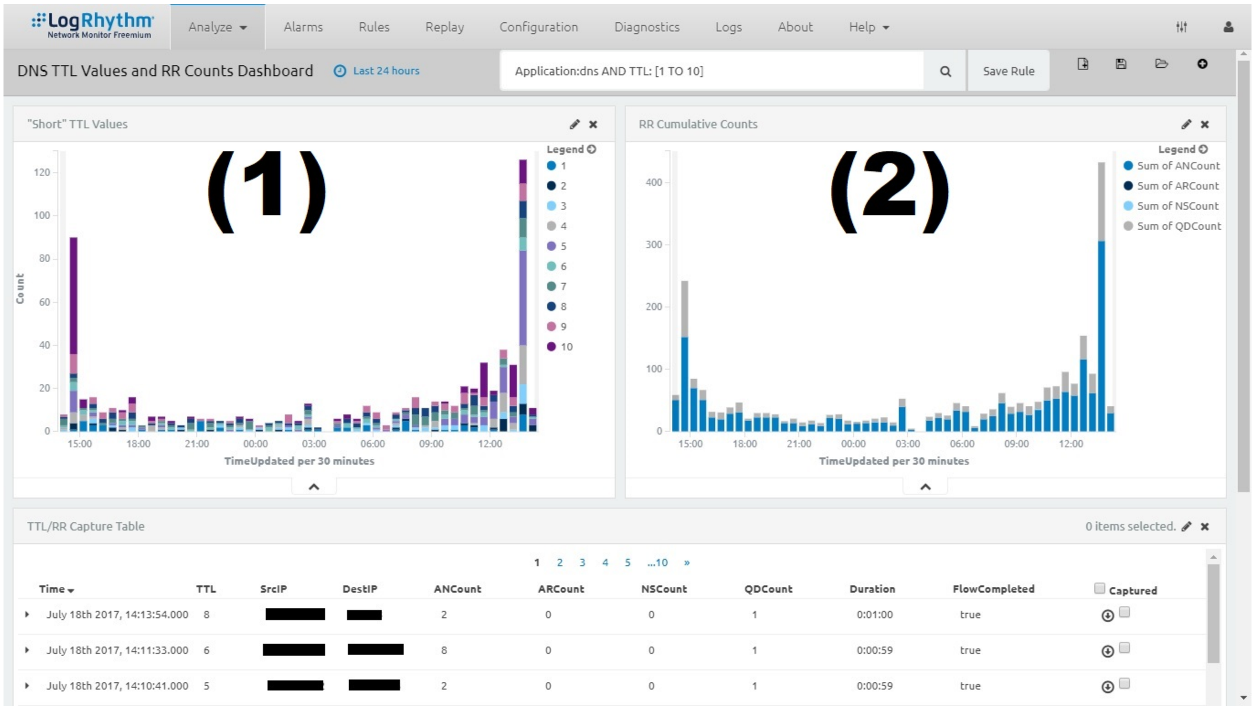


Time	TotalBytes	TrafficDirection_NM	SrcIP	DestIP	DestPort	Application	Duration	FlowCompleted	□ Captured
July 18th 2017, 18:47:33.000	220B	other			547	dhcp6	0:00:59	true	⊕ □
July 18th 2017, 18:47:32.000	294.578KB	unknown	[redacted]	[redacted]	443	logmein	10:28:11	false	false
July 18th 2017, 18:47:31.000	18.753KB	lateral	[redacted]	[redacted]	80	http	0:00:03	true	⊕ □

Figure 8: Typical Port and Protocol Usage Dashboard

How would a Hunter use this dashboard?	
Over time, establish a baseline of:	**Investigate further when:**
Most common ports and corresponding protocols that are associated with highest number of connection counts	Ports and protocols appear in the dashboard that were previously ranked as uncommon Suspicious ports/protocols, e.g. 23/telnet, appear in the dashboard

Dashboard for “DNS TTL Values and RR Counts”:



DNS TTL Values and RR Counts Dashboard   Last 24 hours

Application:dns AND TTL: [1 TO 10]

TTL/RR Capture Table   0 items selected.

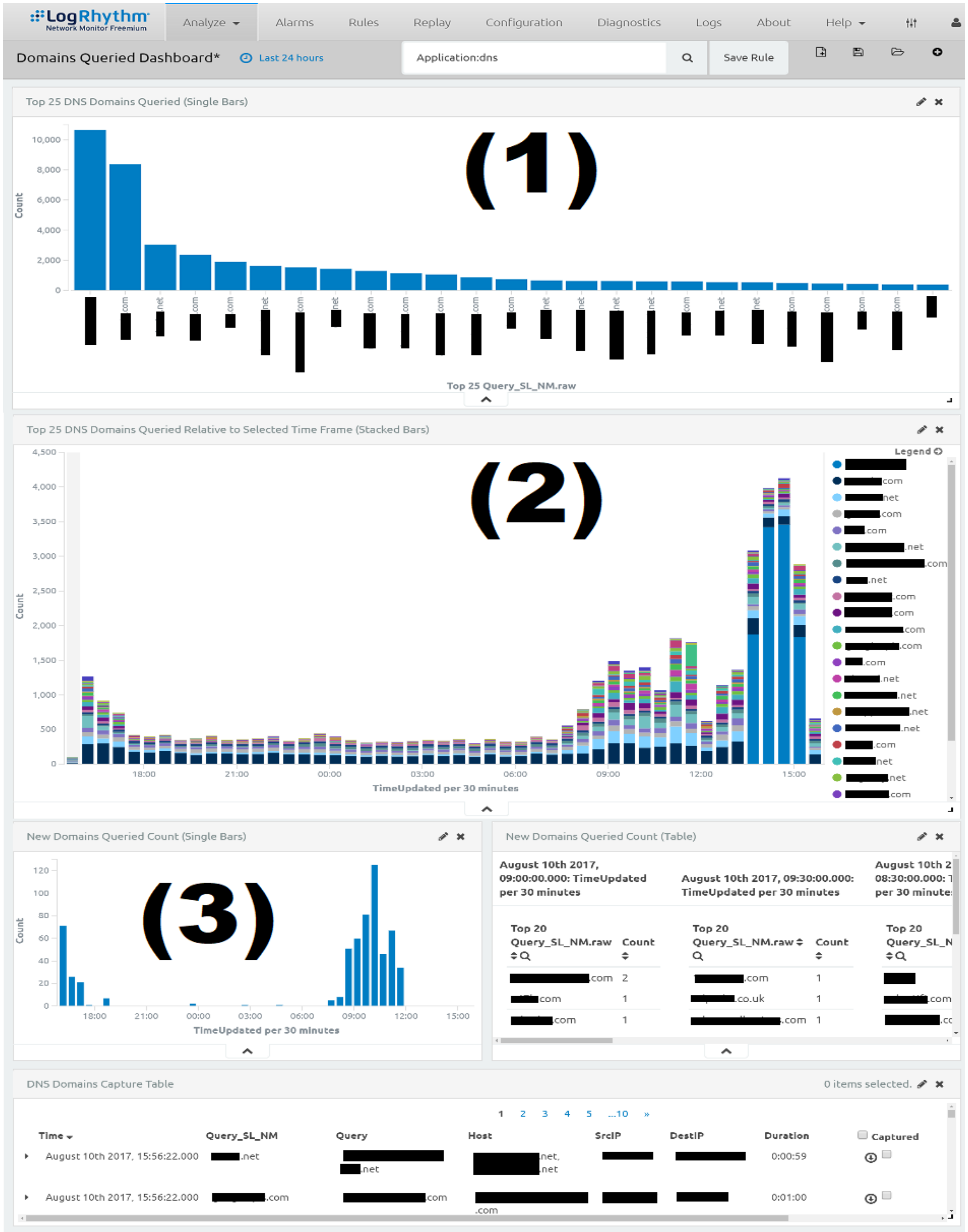
1 2 3 4 5 ...10 »

Time	TTL	SrcIP	DestIP	ANCount	ARCount	NSCount	QDCount	Duration	FlowCompleted	□ Captured
July 18th 2017, 14:13:54.000	8	[redacted]	[redacted]	2	0	0	1	0:01:00	true	□
July 18th 2017, 14:11:33.000	6	[redacted]	[redacted]	8	0	0	1	0:00:59	true	□
July 18th 2017, 14:10:41.000	5	[redacted]	[redacted]	2	0	0	1	0:00:59	true	□

Figure 9: DNS TTL Values and RR Counts Dashboard

How would a Hunter use this dashboard?	
Over time, establish a baseline of:	**Investigate further when:**
Typical proportion(s) of “short” time to live values (TTLs) and different resource records (RRs)	Observed frequency distribution displays one or multiple spikes in short TTLs or RRs

Dashboard for “Top DNS Domains Queried” and “Newly-Observed Domains” ***



August 10th 2017, 09:00:00.000: TimeUpdated per 30 minutes		August 10th 2017, 09:30:00.000: TimeUpdated per 30 minutes		August 10th 2 08:30:00.000: T per 30 minute:
Top 20 Query_SL_NM.raw	Count	Top 20 Query_SL_NM.raw	Count	Top 20 Query_SL_N
[redacted].com	2	[redacted].com	1	[redacted]
[redacted].com	1	[redacted].co.uk	1	[redacted].com
[redacted].com	1	[redacted].com	1	[redacted].cc

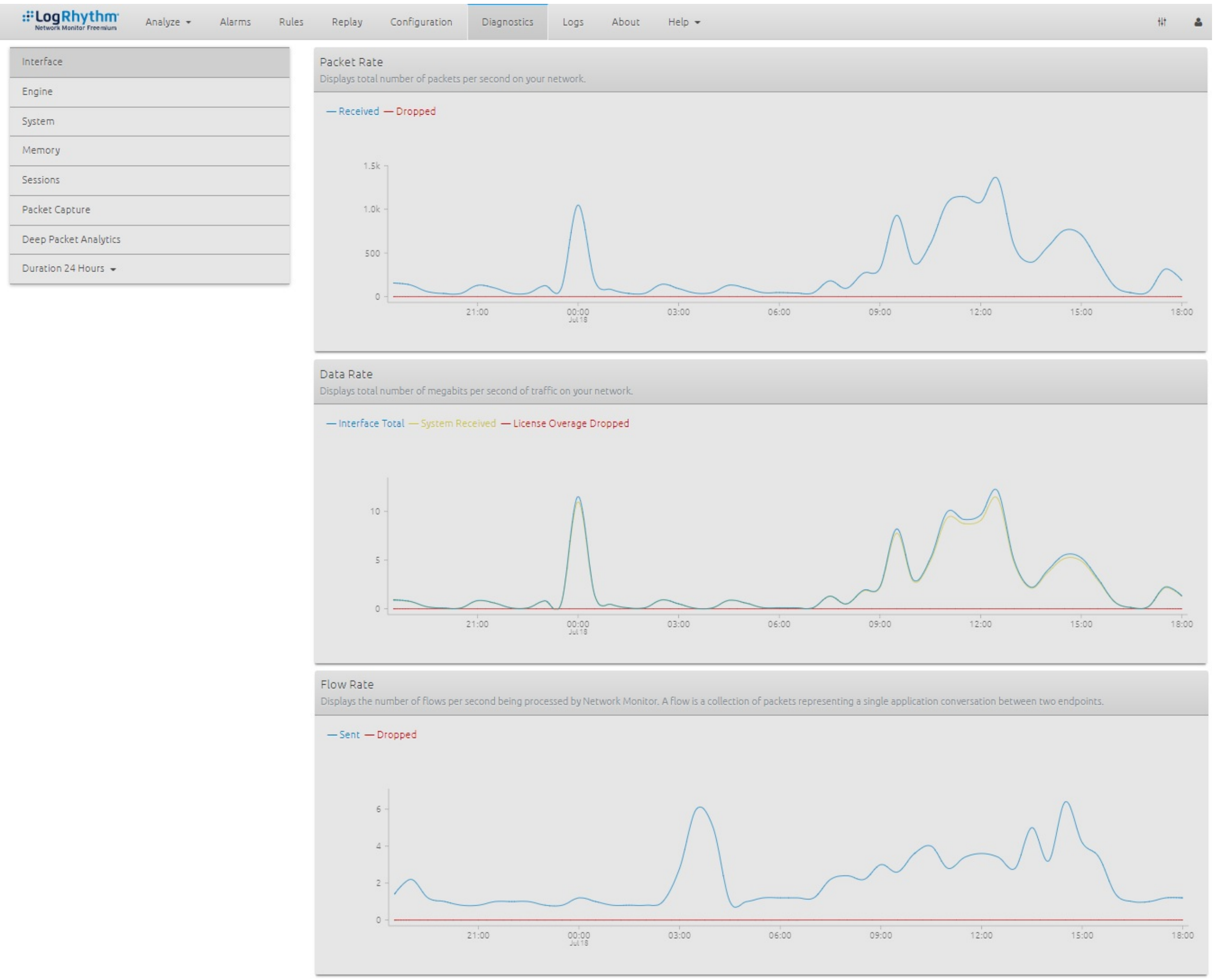
Time	Query_SL_NM	Query	Host	SrcIP	DestIP	Duration	□ Captured
August 10th 2017, 15:56:22.000	[redacted].net	[redacted] [redacted].net	[redacted].net, .net	[redacted]	[redacted]	0:00:59	□
August 10th 2017, 15:56:22.000	[redacted].com	[redacted].com	[redacted] .com	[redacted]	[redacted]	0:01:00	□

Figure 10

***The “Newly-Registered Domains” NTA was not included in this dashboard because we are currently developing a DPA for this. See ‘What else do we want to do’ section below (page 30) for more information.

How would a Hunter use this dashboard?	
Over time, establish a baseline of:	**Investigate further when:**
Most common domains queried by internal clients on daily	Highly queried domains that were previously ranked as uncommon appear Domains from foreign countries (example.co.xx) are queried
Typical time periods when total number of newly queried domains by internal clients fluctuate	Spikes of newly queried domains occur during unusual time periods

Out-of-the-box viewer for “Periodic Traffic Volume Metrics”****



Displays total number of packets per second on your network.

Displays total number of megabits per second of traffic on your network.

Displays the number of flows per second being processed by Network Monitor. A flow is a collection of packets representing a single application conversation between two endpoints.

Figure 11

****This chart is available in the Diagnostics page of the LogRhythm NetMon Freemium interface (as opposed to the “Analyze” interface) and does not require any additional configuration

How would a Hunter use this dashboard?	
Over time, establish a baseline of:	**Investigate further when:**
Typical values for Packet and Data rate on the network	Rate charts show abnormally large volumes of traffic on the network

We successfully developed 7 dashboards to display baseline network traffic against which 8.5 out of the 11 Network Traffic Anomalies presented in the SANS™ poster will stand out; 9.5 out of 11 if you consider that the out-of-the-box Diagnostics viewer presents "Periodic Traffic Volume Metrics", another recommended trend of interest. Since we created a visualization for "Newly-Observed Domains" but none for "Newly-Registered Domains", we consider this NTA half complete.

Creating a visualization for the "Top DNS Domains Queried" NTA proved more difficult than one would expect, considering there is a `Query` field available to inject in a dashboard that filters on `Application:dns`. However, the way in which the `Query` field is parsed adds too much "noise" in the list of, say, top 100 domains queried, as all levels of the domain are displayed as individual queries. For example, a DNS query of `a.b.com` will display "`a`", "`b`", "`com`", and "`b.com`" as separate queries. In attempt to resolve this, we implemented a Deep Packet Analysis Rule (DPA) with the built-in Lua language to parse the `Query` metadata field to a second-level domain representation, which we designated as `Query_SL`, but the undesired parsing from visualizations persisted. We presented this issue in a post on the LogRhythm Community NetMon Discussions page, and were given a thorough and straightforward solution by a LogRhythm software engineer. He explained that the ElasticSearch interface is initially configured to automatically parse strings when they are selected as the analyzed field in a visualization; to disable this, the ElasticSearch mappings for visualizations must be modified so that a custom field is not subject to any analysis process or parsed into tokens.

We were unable to create a visualization for the "Newly-Registered Domains" NTA because this required access to a WHOIS database providing dates of when domains were registered, although we do have some ideas on how this data can be gained through implementation of a DPA. Again, see the "What else do we want to do" section below (page 30). The search logic required to detect a newly queried domain, with respect to a history of queried domains, could not be implemented in LogRhythm NetMon Freemium's "Visualize" development environment. To compensate for this, we created a text file within a local directory and then manually appended a short list of second-level domains previously queried. After determining permission requirements for this file with LogRhythm software engineers (again via the beneficial LogRhythm Community NetMon Discussions forum), we implemented a DPA which would read this text file, check if the currently queried second-level domain existed, and append this domain if it was not found. This allowed us to successfully create a `Previously_Queried` field which serves to indicate whether a second-level domain had been queried within our internal environment.

Additionally, we could not produce a useful dashboard for the "Autonomous System Communications" NTA.   Within LogRhythm NetMon Freemium's Help page, we confirmed that a metadata field for Autonomous System Numbers (ASNs) existed. However, after implementing a DPA for this field, the returned values did not exhibit any variability, i.e. the only value returned was 0 which is a reserved ASN. We suspect we need to feed traffic between the firewall and router to achieve more relevant results.

How would a Hunter use this tool?

We expect the SMB Hunter to review each of the created dashboards and the Diagnostics viewer daily, to establish acceptable trends as well as investigate any anomalous events. Of course, many of these anomalies will be false positives, but, just like a bowhunter in the forest, the network traffic Hunter must understand the reality that “not all that moves is prey”. Over time, the Hunter will improve at quickly culling false positives. We predict the Hunter would typically need to dedicate no more than 1/2 hour daily to LogRhythm NetMon Freemium dashboard review; anomaly investigation will take longer. The Hunter would add LogRhythm NetMon Freemium to his or her daily review of a suite network monitoring tools.

During the daily review session, we recommend setting the displayed time period to 24 hours. This yields a better scope of the patterns in ingress/egress traffic during and after business hours. Weekends and holiday periods would require respectively longer display periods. If the SMB Hunter were to find a suspicious event, they could quickly narrow down the time frame of interest, then utilize the drill down tools in the interfaces. For instance, if the Hunter spotted a suspicious User-Agent string, he or she would click on the section of the chart (Visualization (2) in Figure 6) displaying the User-Agent, and subsequently the capture table below would drill down to the relevant network traffic. The Hunter could then analyze the captured flow, and even download the pcap file for further analysis.

Over time, the Hunter will use the dashboards to establish a visual baseline and spot anomalies worth investigating as deviations from these baselines. Although we didn’t utilize the Alarm feature of LogRhythm NetMon Freemium, we predict a typical Hunter, as he or she becomes more experienced with the tool and confident in the baselines, will likely explore this feature. LogRhythm NetMon Freemium’s DPA Rules engine can be used to analyze traffic and trigger Alarms that appear in the “Alarms” user interface. Working with Alarms is a goal for our own future development with the software—see the “What else do we want to do?” section on page 30 for more information.

Our evaluation on the Development Process, Results, & Relevancy to SANS' standards

In the matrix below we evaluate how LogRhythm NetMon Freemium can help a SMB Hunter analyze network traffic against each SANS™ NTA standard.   We used three criteria for the evaluation:

1) **Ease of Implementation**—How difficult was the implementation process for the visualizations?

- Easy: Visualization exists out-of-the-box, or only required simple re-configuration
- Moderate: Required some novel development and re-configuration
- Challenging: May require significant development within the tool

2) **Ease of Anomaly Detection**—How quickly can the Hunter discover anomalies after establishing a baseline?

- At-a-glance: Deviations are immediately apparent within the visualizations, at-a-glance
- Deeper look: Deviations require cross-referencing with other organized information or more in-depth examination of visualizations and associated capture table to recognize oddities; for instance, a scroll through the chart legend may be required to spot deviations

3) **Additional Tools Required?**—Are additional software or network tools required to complete analysis?

- Yes: At least one other tool (Nmap, Wireshark, SIEM, Firewall logs, etc.) is required
- No: The out-of-the-box interfaces and drill down tools are sufficient

Evaluation Results Matrix	Evaluation Criteria		
	Ease of Implementation	Ease of Anomaly Detection	Additional Tools Required?
HTTP GET vs. POST Ratio	Moderate	At-a-glance	No
Top-Talking IP Addresses	Moderate	Deeper look	No
HTTP User- Agent	Easy	Deeper look	No
Top DNS Domains Queried	Challenging	Deeper look	No
HTTP Return Code Ratio	Moderate	At-a-glance	No
Newly-Observed/Registered Domains	Challenging	Deeper look	Yes (WHOIS)
External Infrastructure Usage	Moderate	Deeper look	Yes (WHOIS)
Typical Port and Protocol Usage	Moderate	Deeper look	No
DNS TTL Values and RR Counts	Moderate	At-a-glance	No
Autonomous System Communications	Challenging	N/A	Yes (WHOIS)
Periodic Traffic Volume Metrics	Easy	At-a-glance	No

What would we change?

In this section we list some limiting aspects of LogRhythm NetMon Freemium that we felt would be a challenge for a typical SMB to overcome, and we also list some ideas for general improvements or nice-to-have features for the product:

Limitations

- Regular Expression Queries: Filtering and searching is easy for simple queries like IP address, but anything more complicated is not intuitive; some queries may even be impossible (e.g. querying for a colon ":" in a URL). The query tool uses ElasticSearch as integrated into the Kibana GUI, which supports the Lucene regular expression engine and is not Perl-compatible. Those of us used to powerful regexp searches in scripting engines are quickly hamstrung with Lucene. For instance, trying to create a query to show only HTTP User-Agent strings with all lowercase letters is not intuitive. This is a result of how ElasticSearch tokenizes the parsed fields. And for those of us not familiar with the syntax, it can take some time getting used to.
- Domain Query Records: To analyze new domain queries against historical records, such as suggested by the SANS™ "Newly-Observed/Newly-Registered Domains" NTA, the Hunter must, using a tool, do the following:
    1. record a running tally of historical domain information,
    2. compare each newly received domain against the historical record, and
    3. update the record with the new domain

  Doing so is not possible with the fields readily available for analysis in LogRhythm NetMon Freemium. This requires a DPA to be implemented immediately after initial installation, so that a list of previously queried second-level domains can be appended right when the web user-interface is launched. Furthermore, another DPA must be implemented to gain information regarding the date of when a second-level domain was registered which requires connection to the WHOIS database.

- Distinguishing Raw Socket Connections: Displaying raw socket connections is a standard provided by SANS™ in the "External Infrastructure Usage Attempts" NTA. Doing so in LogRhythm NetMon Freemium is not simple using readily accessible visualization fields. Displaying such connections could be facilitated if LogRhythm implemented a boolean field indicating whether a received packet utilized a raw socket or not; however, it is unclear whether or not LogRhythm NetMon Freemium can even perform this type of analysis.

Potential Improvements or "Nice-to-haves"

- Drilling Down on Long-Tail Data: LogRhythm NetMon Freemium is very powerful for data aggregation, but when a lot of discrete data is shown in a visualization it is difficult to achieve the necessary preciseness with the mouse to hover over or select small discrete data points. This can make long-tail analysis frustrating. Hovering over legend entries to highlight chart data is helpful to locate the data in the chart, but to drill down on a time or to find captures in an associated capture table, one must click on the associated chart data, not the legend entry. In some cases, placing the mouse in the precise location on the chart proves impossible, and one must manually filter the chart to achieve greater resolution. An automated resolution-increase function for data-laden charts would be useful.
- Create a Dashboard Directly from a Visualization: The ability to create a new dashboard directly from a visualization screen does not exist. Right now, one has to save a visualization and then enter the dashboard interface, create a

new dashboard, and then add the visualization to the dashboard. A similar mechanic needs to be engaged whenever a visualization is updated or reconfigured. Skipping the exit visualization→enter dashboard step would save some time, and automatically updated visualizations across dashboards would save frustration.

- Lucene regexp Help: Expand the coverage of Lucene regexp in the LogRhythm NetMon Freemium built-in help. Some additional help is available at the LogRhythm Community forum (requires account), and certainly from Google searches, but it's nice to be able to access specific help topics right in the tool.
- Visualization mark-up or annotations: Provide a note-taking, screen mark-up, or other annotation tool within a visualization or dashboard for the Hunter to highlight anomalies, or otherwise take notes. This will be especially valuable when the Hunter is trying to correlate anomalies between different dashboards.

What else do we want to do?

HBA is excited to continue R&D with LogRhythm NetMon Freemium in SMB network environments. In the immediate future, we plan on concentrating on the following areas:

- Using JSON to create more powerful queries
- R&D on the Deep Packet Analysis (DPA) Rules syntax (using the Lua language), to determine:
    - How to expand the visualization and reporting to more completely fulfill the External Infrastructure Usage Attempts NTA, specifically ASN metadata, and raw socket connections.
- R&D on the DPA Rules syntax to simplify the SMB Hunter's task of triggering Alarms on anomaly criteria specific to that SMB network environment.
- Create visualizations and dashboards focused on data loss prevention (DLP), or spotting large or suspiciously-timed transfers of data or files. DLP isn't explicitly included in the SANS™ NTA standards, but identifying such activity is certainly important to any organization.
- In the SMB scenario described in this paper, LogRhythm NetMon Freemium can't view all the internal network traffic, such as traffic from workstations to printers, or vice versa, intra-virtual machine traffic, and VOIP traffic, which is a significant subset of traffic. To give the Hunter full visibility into all network traffic, we need to configure additional SPAN port(s) or a network tap to feed missed traffic to the other bonded interface on the microcomputer on which we installed LogRhythm NetMon Freemium.
- What if the SMB in this scenario migrated all of their infrastructure to the cloud? How could we use LogRhythm NetMon Freemium in this case?

What do we think?

We are satisfied with the LogRhythm NetMon Freemium product as a low-cost addition to the cybersecurity toolkit for an SMB Hunter. Free software can be a boon (with obvious exceptions: laden with spyware, not supported, etc.), and the cost of the necessary hardware was reasonable for a typical SMB. The process to install and integrate the product into a SMB's network environment was painless.

Although full analysis of all internal network traffic was hampered by the chosen installation location in the SMB network, the LogRhythm NetMon Freemium tool provided transparency into most network traffic of interest. With such out-of-the-box capabilities, a typical Hunter or network administrator could use the tool to establish a baseline of typical behaviors of his/her network and that of its internal clients.

The user interface and experience in LogRhythm NetMon Freemium is intuitive, and relatively easily configurable for most of the SANS™ Institute Network Traffic Anomalies standards we judged the tool against. A couple of the anomaly standards were met by LogRhythm NetMon Freemium right out-of-the-box, and in short-order we were able to create novel dashboards that could help the SMB Hunter identify other anomalies. The ability to drill down into a packet, flow, or pcap within seconds of identifying and anomaly is of much value to a Hunter, and automatic packet capture of (all, if desired) traffic flows distinguishes LogRhythm NetMon Freemium from other tools. For several of the NTA standards, LogRhythm NetMon Freemium would be the only tool the Hunter would need to identify an anomaly and determine the root cause.

However, it would be a mistake for a Hunter to rely solely on LogRhythm NetMon Freemium for network traffic monitoring, as the tool has limitations in meeting several of the SANS™ standards, such as identifying external infrastructure attributes and usage attempts. Additionally, there is no out-of-the-box capability to record historical lists of domain queries for future comparison.

The user interface needs improvements in some areas as well. For example, the main query interface has limitations, and users that are familiar with regular expression searches in scripting engines may become frustrated as they familiarize themselves with the Lucene regex syntax. Although we utilized and explored whether the product's DPA Rules engine could be used to overcome the limitations noted above, we approached the integration and standards test from the perspective of a SMB Hunter, who may not have the time or skill set to perform such development.

Overall, LogRhythm NetMon Freemium is a valuable tool for a SMB Hunter, especially since the software is free.   It can certainly help the Hunter baseline network traffic, identify anomalies, and pursue further investigation. Once the tool has been installed and configured for the SMB environment, we predict a SMB Hunter would spend ½-hour or so a day using the dashboards to establish baselines of normal network traffic. Once the Hunter is comfortable with the baselines, he or she should be able to quickly spot anomalies during that same ½-hour period, and then use the tool to plan and execute further investigations. LogRhythm NetMon Freemium is a great example of why a SMB need not necessarily pursue high-cost software solutions to gain visibility into network traffic.

For more information on LogRhythm NetMon Freemium and to download it, see: https://logrhythm.com/network-monitor-freemium/

We found the very active NetMon Community extremely helpful with deployment: https://community.logrhythm.com/

Appendix

HTTP Dashboard Get/Post and Return Code Dashboard (Figure 4) General Configuration Description	
i.	Created a blank dashboard to display only HTTP traffic by setting default Lucene query to `Application:http`
ii.	Created visualization (1) to show ratios of HTTP response status codes based on specified ranges from SANS™ poster.
iii.	Created visualization (2) to show ratios of HTTP GET/POST request methods.
iv.	Added visualizations (1) and (2) to dashboard.
v.	Edited the out-of-the-box capture table to provide more in-depth information about the visualizations; added to dashboard.
Top-Talking IPs Dashboard (Figure 5) General Configuration Description	
i.	Edited out-of-the-box visualization (1) to display top 20 sums of bytes transferred from a local client to the server since the last update.
ii.	Edited out-of-the-box visualization (2) to display top 20 connection counts from a local client to a server.
iii.	Created visualization (3) to display top 20 sums of bytes transferred from the server to a local client since the last update.
iv.	Created visualization (4) to display top 20 connection counts from a server to local client.
v.	Edited the out-of-the-box capture table to provide more in-depth information about the visualizations.
vi.	Added the visualizations (1) – (4) and modified capture table into a blank dashboard.
HTTP User-Agent Dashboard (Figure 6) General Configuration Description	
i.	Added the out-of-the-box visualization (1) to a blank dashboard.
ii.	Edited the out-of-the-box visualization (2) to display bottom 100 user agent entities.
iii.	Added visualization (2) to dashboard.
iv.	Edited the out-of-the-box capture table to provide more in-depth information about the visualizations; added to dashboard.
External Infrastructure Usage Attempt Dashboard (Figure 7) General Configuration Description	
i.	Created a blank dashboard to display only DNS traffic which excluded connections to internal resolvers; accomplished by setting default Lucene query to `Application:dns AND (NOT DestIP:A.B.C.D) AND (NOT SrcIP:A.B.C.D)`
ii.	Added previously created visualizations (1) and (2) to dashboard.
iii.	Edited the out-of-the-box capture table to provide more in-depth information about the visualizations; added to dashboard.
Typical Port and Protocol Usage Dashboard (Figure 8) General Configuration Description	
i.	Created visualization (1) to show top 20 destination port connections.
ii.	Created visualization (2) to show top 20 application connections.

iii.	Edited the out-of-the-box capture table to provide more in-depth information about the visualizations.
iv.	Added visualizations (1) and (2) as well modified capture table to a blank dashboard.
DNS TTL Value and RR Count Dashboard (Figure 9) General Configuration Description	
i.	Created a blank dashboard to display only DNS traffic which excluded packets with times to live (TTLs) greater than 10; accomplished by setting default Lucene query to `Application:dns AND TTL:[1 TO 10]`
ii.	Created visualization (1) to show the frequency of packets with TTLs in the range [1 – 10].
iii.	Created visualization (2) to show the cumulative resource records (RRs).
iv.	Edited the out-of-the-box capture table to provide more in-depth information about the visualizations.
v.	Added visualizations (1) and (2) as well as modified capture table to dashboard.
Top DNS Domains Queried and Newly-Observed Domains (Figure 10) General Configuration Description	
i.	Created a blank dashboard to display only DNS traffic; accomplished by setting default Lucene query to `Application:dns.`
ii.	Created visualization (1) to show overall top 25 second-level domains queried by internal clients; added this visualization to dashboard
iii.	Created visualization (2) to show time relative top 25 second-level domains queried by internal clients; added this visualization to dashboard
iv.	Created visualization (3) to show trends of newly queried second-level domains based on selected time period; added this visualization to dashboard
v.	Edited the out-of-the-box capture table to provide more in-depth information about the visualizations; added this modified capture table to dashboard