sqrrl

WHITE PAPER

A Framework for Cyber Threat Hunting

TABLE OF CONTENTS

125 Camrbidgepark Drive, Ste 401
Cambridge, MA 02140

p: (617) 902-0784
e: info@sqrrl.com

www.sqrrl.com
@SqrrlData

I. INTRODUCTION - WHAT IS HUNTING?

Many organizations are quickly discovering that cyber threat hunting is the next step in the evolution of the modern Security Operations Center (SOC), but they remain unsure of how to start hunting or how far along they are in developing their hunt capabilities. This white paper formalizes a reference model for how to effectively conduct threat hunting within an organization. We begin with an important question: How can you quantify where your organization stands on the road to effective hunting?

Before we can talk about hunting maturity though, we need to discuss what exactly we mean when we say "hunting". **We define hunting as the process of proactively and iteratively searching through networks to detect and isolate advanced threats that evade existing security solutions.** There are many different techniques hunters might use to find the bad guys, and no single one of them is always "right". The best one often depends on the type of activity you are trying to find.

Hunting consists of manual or machine-assisted techniques, as opposed to relying only on automated systems like SIEMs. Alerting is important, but cannot be the only focus of a detection program. In fact, one of the chief goals of hunting should be to improve automated detection by prototyping new ways to detect malicious activity and then turning those prototypes into effective new automations.

II. THE HUNTING MATURITY MODEL (HMM)

With that definition of hunting in mind, consider what makes up a good hunting infrastructure. There are a number of factors to consider when judging an organization's hunting ability, including:

- The quantity and quality of the data they collect;
- In what ways they can visualize and analyze various types of data;
- What kinds of automated analytic they can apply to data to enhance analyst insights

The quality and quantity of the data that an organization routinely collects from its IT environment is a strong factor in determining their level of **Hunting Maturity**. The higher the volume and the greater the variety of data from around the enterprise that you provide to an analyst, the more results they will find and the more effective they will be as a hunter. The toolsets at your disposal, including the visualizations you can generate and analytics you can use, will shape the style of your hunts and determine what kinds of hunting techniques you will be able to leverage.

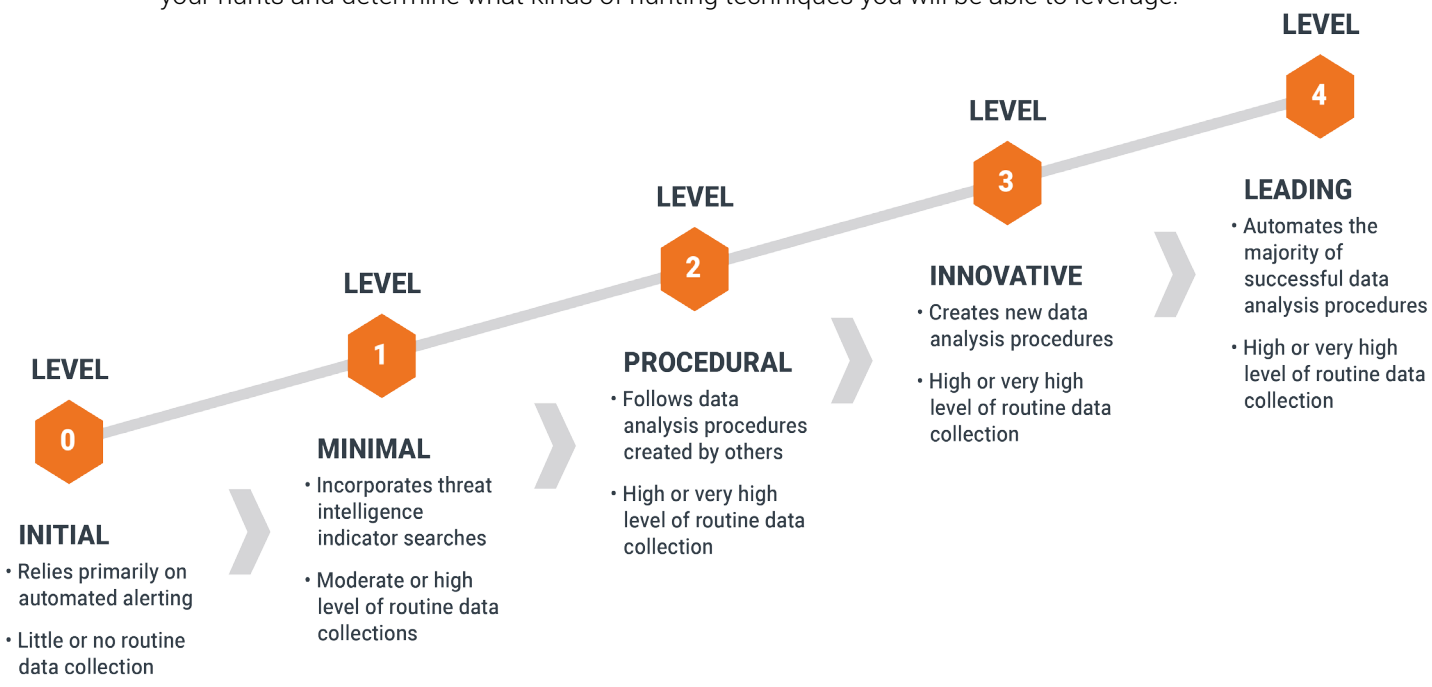

Figure 1. The Hunting Maturity Model (HMM)

A. Steps of the HMM

The Hunting Maturity Model, first developed by Sqrrl's own security technologist and chief hunter, David J. Bianco, describes five levels of organizational hunting capability, ranging from HM0 (the least capable) to HM4 (the most).

HM0 - INITIAL

At HM0, an organization relies primarily on automated alerting tools such as IDS, SIEM or antivirus to detect malicious activity across the enterprise. They may incorporate feeds of signature updates or threat intelligence indicators, and they may even create their own signatures or indicators, but these are fed directly into the monitoring systems. The human effort at HM0 is directed primarily toward alert resolution.

HM0 organizations also do not collect much information from their IT systems so their ability to proactively find threats is severely limited. Organizations at HM0 are not considered to be capable of hunting.

HM1 - MINIMAL

An organization at HM1 still relies primarily on automated alerting to drive their incident response process, but they are actually doing at least some routine collection of IT data. These organizations often aspire to intel-driven detection (that is, they base their detection decisions in large part upon their available threat intelligence). They often track the latest threat reports from a combination of open and closed sources.

HM1 organizations routinely collect at least a few types of data from around their enterprise into a central location such as a SIEM or log management product. Some may actually collect a lot of information. Thus, when new threats come to their attention, analysts are able to extract the key indicators from these reports and search historical data to find out if they have been seen in at least the recent past.

Because of this search capability, HM1 is the first level in which any type of hunting occurs, even though it is minimal.

HM2 - PROCEDURAL

If you search the Internet for hunting procedures, you will find several great ones[1]. These procedures most often combine an expected type of input data with a specific analysis technique to discover a single type of malicious activity (e.g., detecting malware by gathering data about which programs are set to automatically start on hosts). Organizations at HM2 are able to learn and apply procedures developed by others on a somewhat regular basis, and may make minor changes, but are not yet capable of creating wholly new procedures themselves.

Because most of the commonly available procedures rely in some way on least-frequency analysis (as of this writing, anyway), HM2 organizations usually collect a large (sometimes very large) amount of data from across their enterprise.

HM2 is the most common level of capability among organizations that have active hunting programs.

HM3 - INNOVATIVE

HM3 organizations have at least a few hunters who understand a variety of different types of data analysis techniques and are able to apply them to identify malicious activity. Instead of relying on procedures developed by others (as is the case with HM2), these organizations are usually the ones who are creating and publishing the procedures. Analytic skills may be as simple as basic statistics or involve more advanced topics such as linked data analysis, data visualization or machine learning. The key at this stage is for analysts to apply these techniques to create repeatable procedures, which are documented and performed on a frequent basis.

Data collection at HM3 at least as comprehensive as it is at HM2, if not more advanced.

[1] Three good examples are here, here, and here.

HM3 organizations can be quite effective at finding and combating threat actor activity. However, as the number of hunting processes they develop increases over time, they may face scalability problems trying to perform them all on a reasonable schedule unless they increase the number of available analysts to match.

HM4 - LEADING

An HM4 organization is essentially the same as one at HM3, with one important difference: automation. At HM4, any successful hunting process will be operationalized and turned into automated detection. This frees the analysts from the burden of running the same processes over and over, and allows them instead to concentrate on improving existing processes or creating new ones.

HM4 organizations are extremely effective at resisting adversary actions. The high level of automation allows them to focus their efforts on creating a stream of new hunting processes, which results in constant improvement to the detection program as a whole.

B. Automation and the HMM

It may seem confusing at first that the descriptions for both HM0 and HM4 have a lot to say about automation. Indeed, an HM4 organization always has automation in the front of their minds as they create new hunting techniques. The difference, though, is that HM0 organizations rely entirely on their automated detection, whether it's provided by a vendor or created in house. They may spend time improving their detection by creating new signatures or looking for new threat intel feeds to consume, but they are not fundamentally changing the way they find adversaries in their network. Even if they employ the most sophisticated security analytics tools available, if they are not constantly improving their automated approach, they are not hunting.

HM4 organizations, on the other hand, are actively trying new methods to find the threat actors in their systems. They try new ideas all the time, knowing that some won't pan out, but that others will. They are inventive, curious and agile, qualities you can't get from a purely automated detection product. Although you can't simply buy an automated system that will get you to HM4, a good hunting platform can certainly give your team and analysts an enormous boost in sophistication.

C. Using the HMM

CISOs that hear that their organization needs to "get a hunt team" may legitimately be convinced that an active detection strategy is the right move, and yet still be confused about how to describe what a hunt team's capability should actually be. A maturity model will ideally help anyone thinking of getting into hunting get a good idea of what an appropriate initial capability would be.

More importantly for those organizations who already hunt, the HMM can be used both to measure their current maturity and provide a roadmap for improvement. Hunt teams can match their current capabilities to those described in the model, then look ahead one step to see ideas for how they can develop their skills and/or data collection abilities in order to achieve the next level of maturity. In order to get anywhere, you must first know where you are and where you want to be.

III. THE HUNTING LOOP

As we saw above, hunting maturity is based on a number of criteria that determine how effectively an organization can get through the hunting process. But what is involved in the actual process of hunting? If automating the hunt is the ultimate goal of a hunt team, how do you determine what steps must be automated on a tactical level?

To avoid one-off, potentially ineffective "hunting trips," it is important for your team to implement a formal cyber hunting process. Sqrrl has developed a Threat Hunting Loop (depicted below) consisting of four stages that define an effective hunting approach. The goal of a hunt team should be to get through the loop as quickly and effectively as possible. The more efficiently you can iterate, the more you can automate new processes and move on to finding new threats.

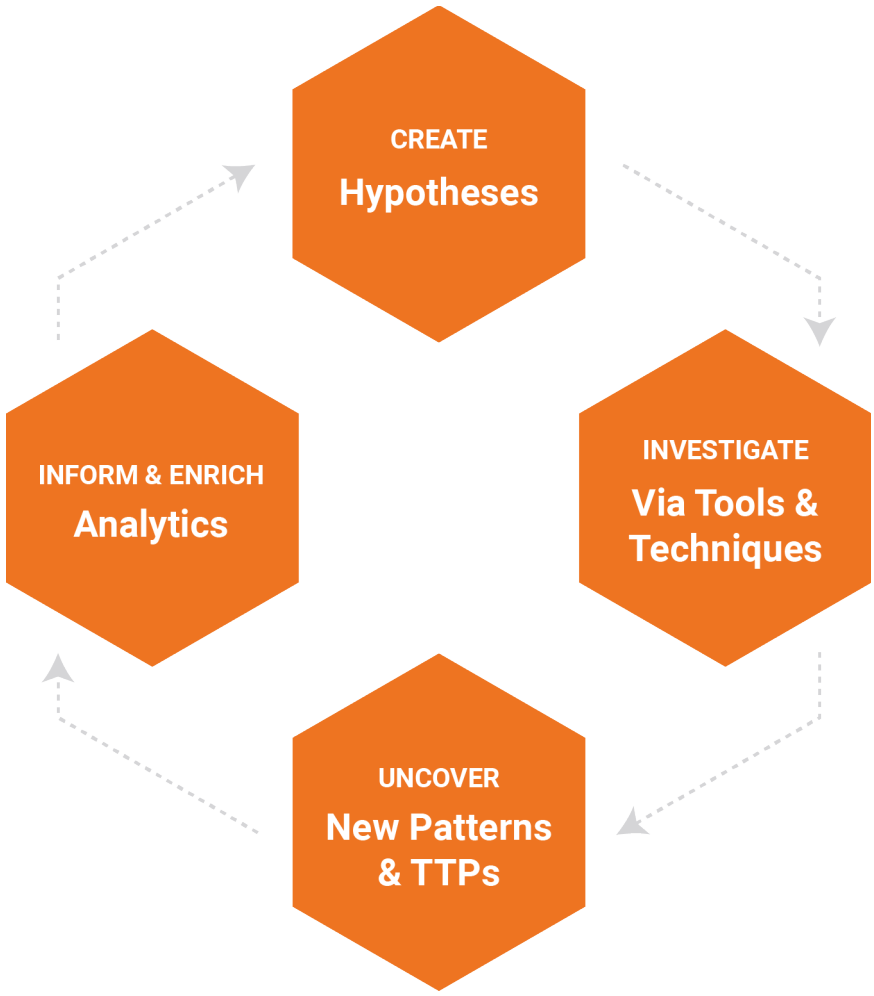

Figure 2. The Hunting Loop

A. Hypothesis Generation

A hunt starts with creating a hypothesis, or an educated guess, about some type of activity that might be going on in your IT environment. An example of a hypothesis could be that users who have recently traveled abroad are at elevated risk of being targeted by state-sponsored threat actors, so you might begin your hunt by planning to look for signs of compromise on their laptops or assuming that their accounts are being misused around your network. Each of these subhypotheses would be tested individually. Analysts can develop hypotheses manually based on this type of intelligence.

At a more advanced level, hypotheses might also be generated automatically by risk algorithms that flag specific users or entities as suspicious based on a variety of factors. For example, a risk assessment algorithm could take the outputs of various Kill Chain behavioral analytics (e.g., beaconing behavior, lateral movement behavior, exfiltration behavior) and combine them into a single risk score for a user or entity that will provide a good starting point for a hunt.

B. Tool and Technique Enabled Investigation

Second, hypotheses are investigated via various tools and techniques, including Linked Data Analysis and visualizations. Effective tools will leverage both raw and linked data analysis techniques such as visualizations, statistical analysis, or machine learning to fuse disparate cybersecurity datasets. Linked Data Analysis is particularly effective at laying out the data necessary to address the hypotheses in an understandable way, and so is a critical component for a hunting platform. Linked data can even add weights and directionality to visualizations, making it easier to search large data sets and use more powerful analytics.

Many other complementary techniques exist, including row-oriented techniques like stack counting and datapoint clustering. Analysts can use these various techniques to easily discover new malicious patterns in their data and reconstruct complex attack paths to reveal an attacker's Tactics, Techniques, and Procedures (TTPs).

C. Pattern and TTP Discovery

Next, tools and techniques uncover new malicious patterns of behavior and adversary TTPs. This is a critical part of the hunting cycle. An example of this process could be that a previous investigation revealed that a user account has been behaving anomalously, with the account sending an unusually high amount of outbound traffic. After conducting a Linked Data investigation, it is discovered that the user's account was initially compromised via an exploit targeting a third party service provider of the organization. This TTP (initial compromise via a third party system via particular type of malware) should be recorded, shared (both internally and externally), and tracked within the context of a larger attack campaign. Linked data relationships will also contextually reveal what other accounts were associated with the compromised third party service.

D. Automated Analytics

Lastly, successful hunts form the basis for informing and enriching automated analytics. Don't waste your team's time doing the same hunts over and over. Once you find a technique that works to find threats, automate it via an analytic so that your team can continue to focus on the next new hunt. There are many ways this can be done, including developing a saved search to run regularly, creating new analytics using tools like Sqrrl, Apache Spark, R, or Python, or by even providing feedback to a supervised machine learning algorithm confirming that an identified pattern is malicious.

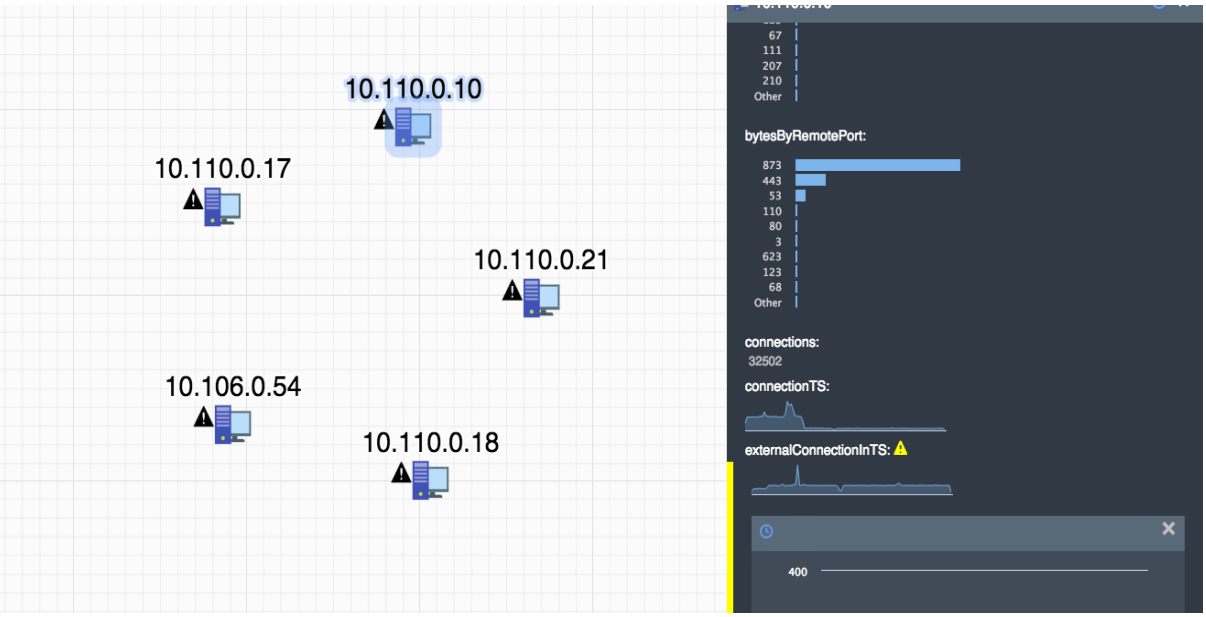

Figure 3. An example Sqrrl analytic for detecting outliers in a set of host machines

The hunting loop is a simple but effective process that can radically enhance an organization's network defense. It is most effective when it is used together with traditional security systems, complementing detection efforts that most organizations already have in place. The ultimate objective of a hunting team should always be to get through the loop as efficiently as possible.

IV. THE HUNT MATRIX

Having laid out the hunting maturity model and the hunting loop, we can now look at how these two concepts fit together. In this section of the paper, we combine the capabilities of each level of the maturity model mapped to different steps of the hunting loop in a practical matrix.

A. Bringing it all Together

We already know that hunting is comprised of four steps and that hunting is most effective when these four steps are carried out iteratively, constantly building on each other. Organizations at different levels of the hunting maturity model will execute steps of the hunting loop in various ways. The matrix combines the four steps of the Hunting Loop and the five steps of the maturity model.

HUNTING MATURITY LEVEL

HUNTING LOOP STEPS

	HM0 Initial	HM1 Minimal	HM2 Procedural	HM3 Innovative	HM4 Leading
DATA COLLECTION	Little or no data collection	Moderate collection of some types of data from a few key points in the IT environment	High collection of certain types of data throughout the IT environment	High collection of certain types of data throughout the IT environment	High collection of many types of data throughout the IT environment
HYPOTHESIS CREATION	Respond to existing automated alerts from SIEM, IDS, Firewall, etc.	Review threat intelligence to develop new hypotheses	Review threat intelligence and "friendly intelligence" to develop new hypotheses	Review threat intelligence, "friendly intelligence", and manual cyber risk scoring (i.e. "crown jewel analysis") to develop new hypotheses	Review threat intelligence, "friendly intelligence", and automated cyber risk scoring to develop new hypotheses
TOOLS & TECHNIQUES FOR HYPOTHESIS TESTING	Alert consoles, SIEM searches; No proactive investigation	Utilize SIEM or log analysis tools to conduct basic search via full-text or SQL-like queries	Utilize simple tools and histograms to search and analyze data based on existing hunting procedures	Leverage visualizations and graph searches. Develop new hunting procedures	Advanced visualizations and graph searches. Publish, and automate new hunting procedures
PATTERN & TTP DETECTION	None; Only SIEM/ IDS alerts	Identifying IOCs at bottom of PoP like domains, URLs, and hashes	Identification of IOCs at bottom and middle of PoP and mapping trends of those IOCs over time	Able to detect adversary TTPs and other IoCs at the top of the PoP	Automatic complex TTP discovery and campaign tracking; Active sharing of IoCs with information sharing organization
ANALYTICS AUTOMATION	None	Integrates threat intel feeds into automated alerting for basic matching	Build a library of effective hunting procedures and performs them on a regular schedule	Build a library of effective hunting procedures and performs them frequently; basic data science (standard deviation, outlier detection)	Automate effective hunting procedures to continuously improve alerting capabilities; advanced data science (machine learning)

Figure 4. The Hunt Matrix - Combining the HMM and Hunting Loop

The matrix includes data collection as an important part of the hunting process. After all, you can't hunt if you can't see anything. Data collection from HM0 to HM4 matures in a linear way, from collecting little to no data to collecting many different types of data from throughout your IT environment.

B. Scaling through Hunting Maturity

Scaling up hunting maturity through the hunting loop depends on certain key focus points for each step.

» Maturing **hypothesis creation** is dependent on crafting increasingly more dynamic questions, and moving from manual hypothesis creation to automatic generation via risk scoring analytics.

» Maturing the **tools and techniques** used to follow up on hypotheses is dependent on moving from using simple searches and histograms to using tools with advanced visualizations and graph search capabilities.

» Maturing your **pattern and TTP detection** is dependent on expanding the kinds of Indicators of compromise (IoCs) you can collect from the Pyramid of Pain. This involves moving from collection of simple IoCs like malicious IPs to more complex tasks like tracking adversary TTPs.

» Finally, maturing **analytics and automation** is dependent on moving from simple analytics such as stack counting, to advanced and automated analytics powered by machine learning, and moving from repeated searches to feeding gathered information back into your automated detection systems.

V. CONCLUSION - THE SQRRL ADVANTAGE

Sqrrl Enterprise is a uniquely sophisticated, best-in-class hunting tool that leverages linked data analysis, powerful visualizations, and advanced analytics to more effectively execute each step of the hunting loop. Sqrrl's Big Data technology roots enable easy, flexible, and scalable storage of the data you need to hunt. By giving analysts intuitive ways to explore their data and collaborate with their colleagues, Sqrrl helps to narrow the window between when events occur inside an organization's network and being able to take action on them.

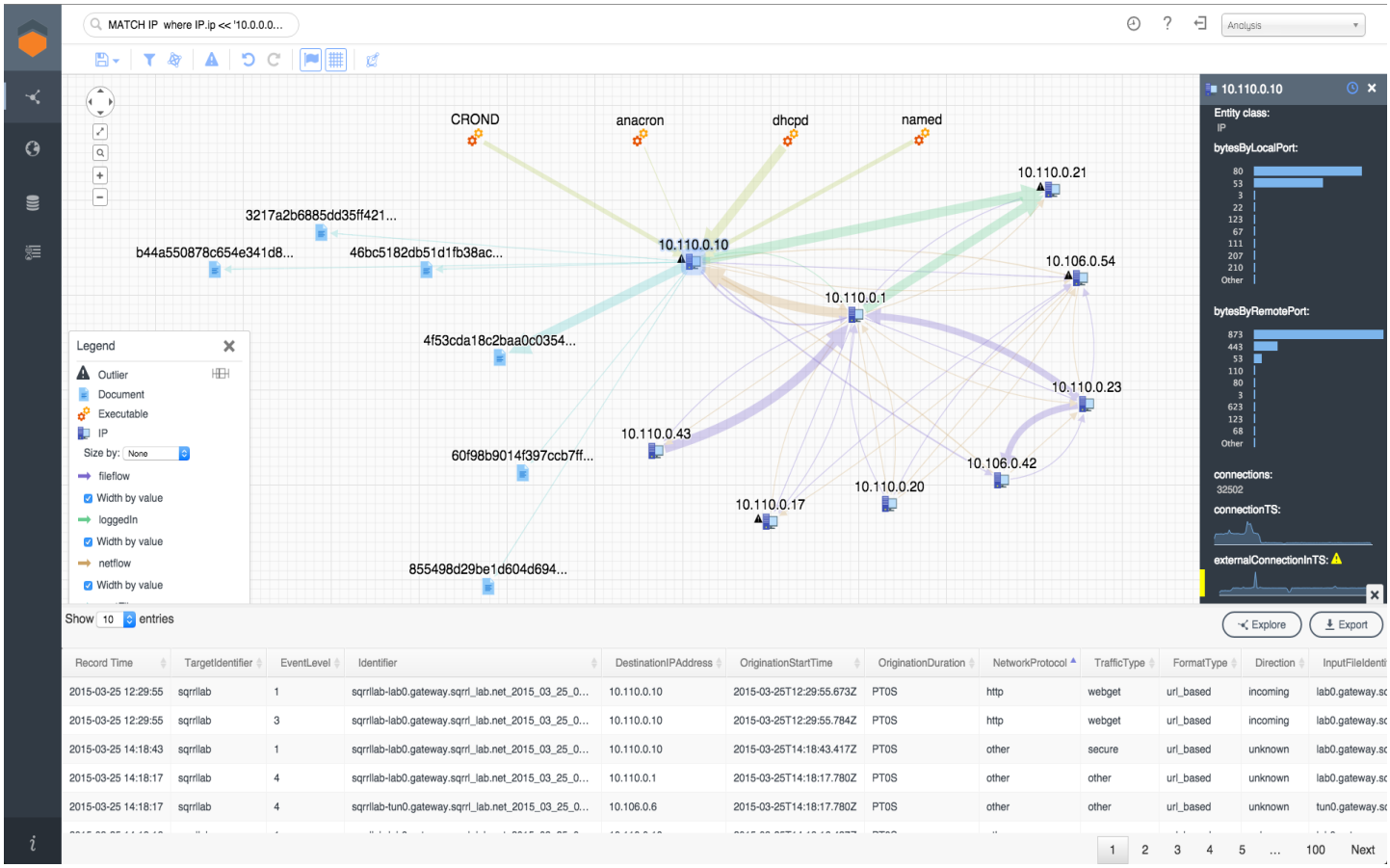

Figure 5. The Sqrrl Enterprise UI, featuring graph visualization, detailed entity information and drill downs on underlying raw data

Sqrrl Enterprise can not only speed up an analyst's ability to carry out the hunt at each step of the loop, but by ramping up data storage, data analysis and analytics deployment, it can also transform the maturity of an organization by several factors.

Sqrrl's benefits include:

- Contextual, intuitive graph visualizations of even the most complex networks
- Aggregating and fusing petabytes of disparate data sets
- Realtime search, query, and analysis of entity behaviors
- Advanced question chaining through context graphs
- Fast drill downs into connected datasets
- Automated detection of anomalies and adversary tactics

With Sqrrl, you can leverage powerful analytics on datasets that are petabytes in size. This lets you make sense of even the most complex networks, and enables you to hunt for threats wherever they may hide. Sqrrl unites all the aspects of the hunting framework into a single, unparalleled tool.

Discover Sqrrl Enterprise at sqrrl.com.

Learn More

—Cyber Hunting eBook: **What Security Executives Need to Know**

—Sqrrl's Hunting Site Page

—Webinar on Hunting **by David J. Bianco**

—Linked Data White Paper

—Sqrrl Enterprise Test Drive VM

—Cyber Hunting Use Case

125 Camrbidgepark Drive, Ste 401
Cambridge, MA 02140

p: (617) 902-0784
e: info@sqrrl.com

www.sqrrl.com
@SqrrlData