sqrrl

Target. Hunt. Disrupt.

DATASHEET

SQRRL ENTERPRISE
THE LINKED DATA ADVANTAGE

Linked Data Analysis provides numerous advantages over traditional log analysis methods and tools

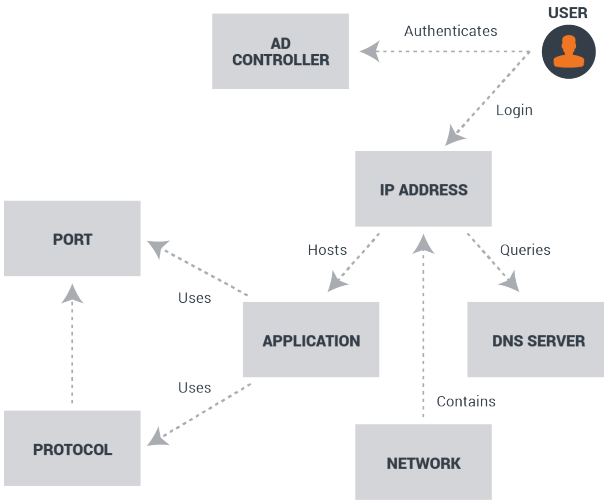
Many enterprise security tools, including SIEMs, Incident Response, and Network Analysis tools are log-based. However, making sense of log files can be tricky, since logs typically exist without context (i.e., it is hard to understand how they relate to the larger cybersecurity environment around them). Luckily, there is a more effective way of organizing your data: Linked Data Analysis.

The Clarity of Context

Linked data describes a format for data representation that highlights the different types of relationships, or links, between entities. In this case, an entity is a logical item of interest, such as a 'user', a 'website', an 'HTTP transaction', and the like. These entities are then linked via different types of relationships – for example, a user can 'know' another user, an employee can 'work for' a manager, etc.

Linked Data Analysis gives cyber "hunters" and incident responders a way to quickly identify the important assets, actors, and events relevant to their organization, accentuating the natural connections between them and providing contextual perspective in an incident response scenario.

Some of the specific advantages of Linked Data Analysis are described below.



An example of a Linked Data model

Easier to Ask Questions of the Data

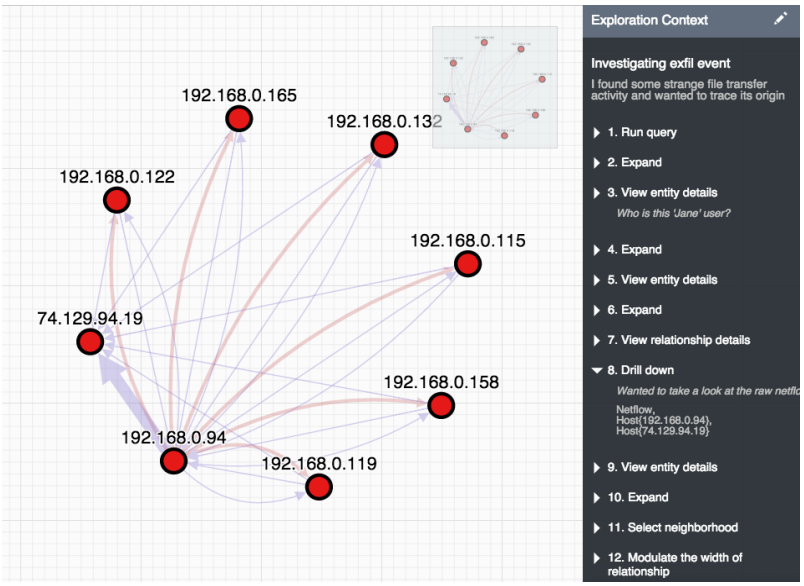
The Linked Data model works particularly well in tandem with threat hunting because it enables you to ask iterative questions more easily. For example, say you are starting with a 'user' and want to ask the question, "Show me all the websites this user has visited in the past day." You can then dynamically expand out relationships from this data, asking questions like "Show me how all the users that have also visited these websites within the same time window" using a simple point-and-click operation. Then, you can further expand and ask to "show me how these users are connected to each other." In this way, linking data can easily facilitate iterative question chaining, which streamlines the process of response and investigation.

Richer, More Intuitive Visualization:

Histograms, bar graphs, and pie charts can only get you so far. Linked Data visualization consists of weighted, directional nodes and edges that can provide compact representations of complex, dense datasets. As opposed to representing just simple trends and comparisons, linked data visualization enables users to easily refer to

relationships and second and third-order connections in the data. This translates to stronger pattern discovery and pattern matching. With a quick glance, analysts can unravel how disparate pieces of data relate and visually "connect-the-dots."

Linked Data visualization naturally aligns to the nature of cyber security data. Network diagrams are typically utilized to outline the structure of an organization's IT systems. Linked Data visualization takes the basic concept of network diagrams and implements it at massive scale and in extreme detail. It also lets an analyst quickly zoom in and out to study both micro- and macro- trends in the data.



Here blue edges represent flow relationships while red edges are logins. The emboldened blue arrow represents larger file transfers between entities.

Faster and More Advanced Analytics

Pattern matching, pattern discovery, and anomaly detection are both faster and more accurate through the use of Linked Data models. These analytics are faster because data points are already connected. Sqrrl's Linked Data solution removes the need for expensive join operations present in relational databases, since data points are pre-joined in the model. This results in much faster cross-graph queries with operations moving through different tables.

Linked data analysis also includes the use of powerful graph algorithms that are not available in traditional log analysis tools. Based on graph mathematical theory, graph algorithms model the strength and direction of relationships within a given system. Graph algorithms can be used not only to detect a correlation, but also to determine its nature and how significant it really is within the overall system.

Massive Scalability

The concept of linked data is not new. However, similar to most log management and analysis solutions, linked data solutions traditionally have been limited by the underlying scalability of the databases that powered them. With the advent of massively scalable non-relational databases, linked data capabilities have taken a leap forward.

Sqrrl's linked data models are deployed on the Apache Accumulo database, which can scale horizontally to thousands of servers and tens of petabytes, while maintaining linear performance. These performance figures enable Sqrrl to provide its customers with interactive search speeds across huge amounts of linked data. Since Accumulo is deployed on low-cost Hadoop hardware, the scaling can be done cost effectively, while not sacrificing durability and resilience.

125 Cambridge Park Dr
Cambridge, MA 02140

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

www.sqrrl.com
@SqrrlData