Visualizing Cyber Threats with KeyLines
Contents

Visualizing Cyber Threats with KeyLines 4
What is network visualization? 4
What is KeyLines? 5
Visualizing Cyber Threats with KeyLines 6
  1. Software vulnerabilities 6
  2. Anomalous login behaviors 9
  3. Data breaches 11
  4. Malware propagation 13
Cyber security is a data-led battle. Every day, organizations collate terabytes of disparate information into centralized operations centers. What’s often missing from these systems, however, is a sophisticated way to analyze and understand connections.

When analyzing cyber security data, the connections – between devices, events, locations, IPs, signatures, and so on – hold the key to uncovering anomalies, threats and vulnerabilities. The best way to understand these connections is to visualize them.

**What is network visualization?**

Network visualization is the process of visually representing complex connected data as interactive node-link diagrams. KeyLines, by Cambridge Intelligence, is a toolkit for building powerful web applications for network visualization.

Network visualization allows users to perform better, more efficient data analysis – which in turn enables them to take better decisions more quickly.

In this white paper, we are going to look at some examples of how KeyLines has been used to extract insight from complex connected cyber data, and explore the four key advantages of this approach:

1. Rapidly transform data into actionable, communicable intelligence
2. Harness the brain’s pattern-recognition ability
3. See data in full connected context
4. Intuitively explore intricate data detail.
What is the KeyLines toolkit?

KeyLines, by Cambridge Intelligence, is a unique technology for building powerful web applications for network visualization.

Using the KeyLines toolkit, developers can rapidly incorporate visualization components to deploy alongside existing cyber security dashboards. These applications turn raw connected data into powerful interactive charts, empowering users to ‘join the dots’ to discover patterns and anomalies.

The applications built using KeyLines run in virtually any web browser and any device. A flexible architecture means they can be deployed into existing IT environments, as part of a dashboard or as standalone tools. Data can be pulled from multiple sources and interactive charts can be shared for reporting purposes.

The advanced functionality available includes:

- Easy filtering
- Multiple automated layouts
- Social network analysis (SNA) measures
- Node and link aggregation and grouping
- A time bar for visualizing temporal data
- A mapping integration to visualize connected data with a geospatial dimension

To register for a free trial, visit https://cambridge-intelligence.com/try-keylines.
Visualizing Cyber Threats with KeyLines

1. Analyzing software vulnerabilities

Network visualization is an interactive and intuitive process. This example shows how it makes actionable data insight quick to extract and easy to communicate.

The densely connected nature of network data means it can be complex to unravel. A few simple techniques, like filtering and layouts, can make its insight easier to understand and communicate.

This example dataset shows the web of relations between malware (exploit kits) and the vulnerabilities they exploit. The vulnerabilities are color coded by their technology.

There are a lot of connections here – too many to unravel at once. Let’s start by focusing just one technology that does not have the best reputation: Internet Explorer.
Now we have highlighted all malware in this dataset known to target vulnerabilities in Internet Explorer. Hovering over a node displays a tooltip with a link to resources describing the malware in question:

Let’s approach the chart from another angle. By selecting a Malware node, we can highlight the technologies they are known to exploit:
Nuclear is an example of a more sophisticated threat, exploiting vulnerabilities in Adobe Flash, Internet Explorer, Oracle Java, Microsoft Windows and Google Chrome.

In this example, we can see the most basic network visualization can provide a powerful way for organizations to understand their exposure to risks from unpatched or vulnerable technologies. The charts can be easily shared too – as static PNG image files, or interactive network charts – making it easier and faster to communicate rich insight to a wide audience.
2. Detecting anomalous login behaviors

This example shows how network visualization allows us to harness the human ability to recognize patterns and anomalies that could otherwise go undetected.

Unusual login patterns are a significant indicator of compromise. For example, a UK-based organization with staff located only in Europe would not expect any significant volume of logins from outside that region. Likewise, most users should only access their account from a relatively narrow range of IP addresses. With techniques more commonly used to detect fraud in transactional data, it’s possible to find suspicious login behavior by visualizing data as a network.

This screenshot shows an overview of thousands of logins to an online currency exchange system:

The central node of each structure indicates an online account; each connected node is an IP address that has been used to access that account. We can see that most accounts have been accessed by 1-4 different IP addresses. Specific ‘star’ structures stand out. These should be investigated as they indicate an account has been accessed from many different locations.
Here we have zoomed in on two ‘star’ structures. At this level, we can see more detail:

- Green nodes are the user
- Yellow nodes are devices used
- Purple nodes are the account.

Looking closer still, we can see that the user node uses a glyph to indicate the country of registration for the account. The node connected by a thick yellow link is the account’s ‘original’ IP address.

In this example, we should look at this account and ask why this user has logged into the system from more than 20 locations.

This technique allows an analyst to rapidly ‘scan’ huge volumes of data and pick out specific anomalies for further investigation.
In this example, we have taken a dataset from the Verizon Data Breach report, as curated by the Veris Project. It shows the relationships between groups of attackers (green people nodes), their attack vectors (color coded links) and their victims (building nodes).

The addition of a time bar allows us to overview the trends in the dataset over time, comparing vectors and finding patterns. For example, email is a lesser-used attack vector – with one spike in Jan 2014. Whereas basic technology – defined here as breaches using means such as phone or LAN access – is more widely used, but decreasing. In particular, it is the most common approach in breaches involving end-users or employees (the large blue cluster to the right of chart center).
‘Advanced tech’ – defined here as web application, remote access, backdoor, C2, command shell, VPN – is the most widely used and is particularly favored by the Activist Group included in the dataset (top left red cluster).

In this example, we have also included a search function to find specific companies, and node sizing by degree (in this case the number of outbound links) to help reveal particularly unlucky nodes:

For example, Microsoft has been more unlucky than most:

This is a simple example that demonstrates the advantage of a visual, graph-based approach to data analysis.

In a single chart, we can combine multiple facets of a complex dataset: Attackers, attacks, victims, vectors and times. This allows us to easily compare and contrast patterns, and see specific incidents in their wider contextual environment. The result is a faster route to data insight, and more advanced analysis.
4. Understanding patterns in malware propagation over time

As well as allowing the rapid overview of large datasets, network visualization provides an intuitive way to hone in on specific details, and understand complex relationships.

This demo illustrates a complex dataset, but through the network chart, time bar and a variety of controls we can quickly discover broad trends and precise details.

On the right hand side of the demo, you can see 8 examples of malware – each color-coded. In the KeyLines chart itself there are a number of node-link structures showing relationships between servers, ISPs and hosts:

The IP address nodes represent servers acting as ‘Command and Control’ machines. These are associated with an ISP, represented by their location’s country flag, and one or more hosts – the contact point(s) for their botnets. The color of the links represents the malware family.

The dataset covers a 5-year period. When the chart opens the default view is the display of every infected machine for the whole period. The ‘Online only’ button provides a filter that highlights only machines known to be active at the end of the period - e.g. the end of December 2015.
If we want to dig in to one specific malware, we can just select it in our filter panel. This reveals all infections for that malware and a trend-line in our time bar:

This macro view of the dataset allows us to see long term patterns and trends. For example, here we can see a dramatic peak in Feodo Version C infections in H1 of 2015, compared to a steady rise and fall in ZeuS infections, peaking in H2 2014.

We can also zoom into specific anomalies. This French ISP, for example, has been particularly unlucky:
We can clearly see an extensive history of infections, plus a number of current infections spanning four malware families.

In this example, we can see how a visual approach to graph analysis has allowed us to quickly move from overview to detail, giving us insight into broad trends and specific incidents.