Chapter 6

Prototype in Action

6.1 Interface Overview

This Chapter provides a tour of the CYBEX-P web application. There are many components to the final product, and together they satisfy all of the requirements and usability goals described in prior chapters. First, an overview of the user interface is provided and the specific utility of each part is examined in detail. This Chapter concludes with comprehensive usage scenarios from the perspective of intended users.

Figures 6.1 and 6.2 depict the CYBEX-P homepage. This is a public-facing web page that serves two main purposes. First, it acts as an informational hub for the entire CYBEX-P project. Site content includes feature descriptions, screenshots, and videos for the threat-intelligence graph application. General information about CYBEX-P, such as the project abstract, funding information, and personnel, are also provided. Users are kept up to date with industry trends through cybersecurity-related twitter feeds. CYBEX-P's latest platform and software updates are also presented in this location. The second purpose of this site is to act as a 'portal' to the frontend web application described throughout this work. This homepage is used to sign in or out of the application with two-factor authentication. Then, by selecting the "CTIGraph" button, users are redirected to the main threat-intelligence graph canvas.

The starting point for the main application is depicted in Figure 6.3. The user is greeted with an empty canvas, unless they have any previous graph data auto-saved.

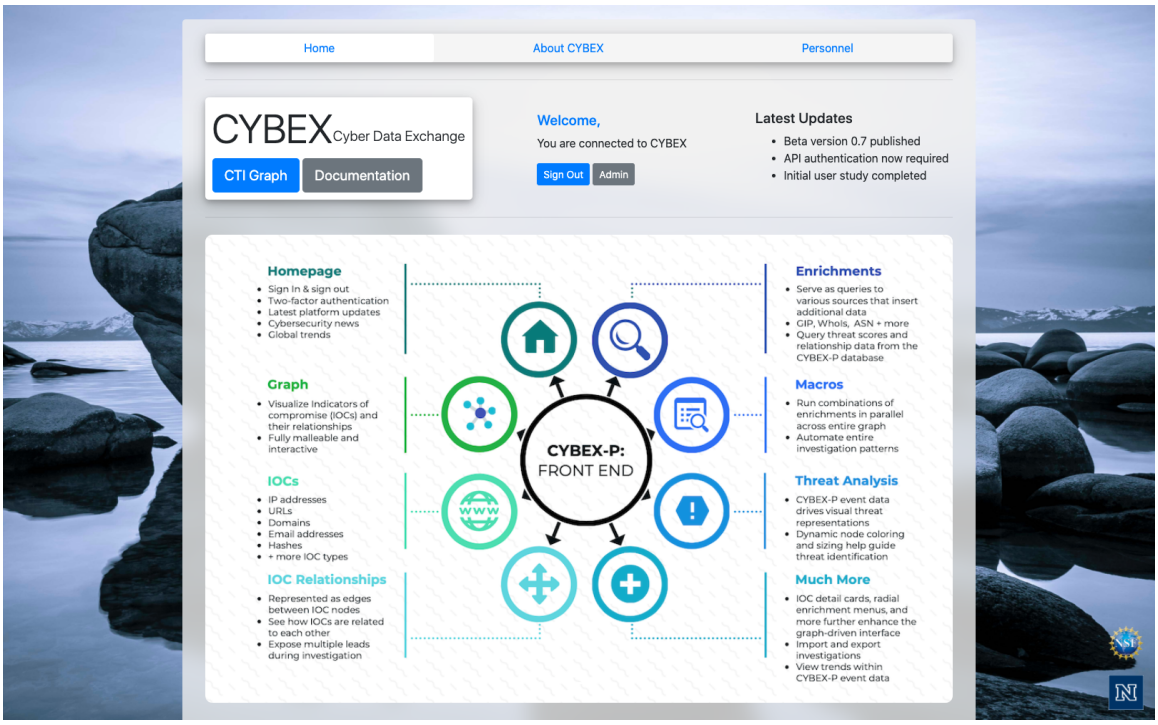


Figure 6.1: Public-facing homepage for the CYBEX-P project.

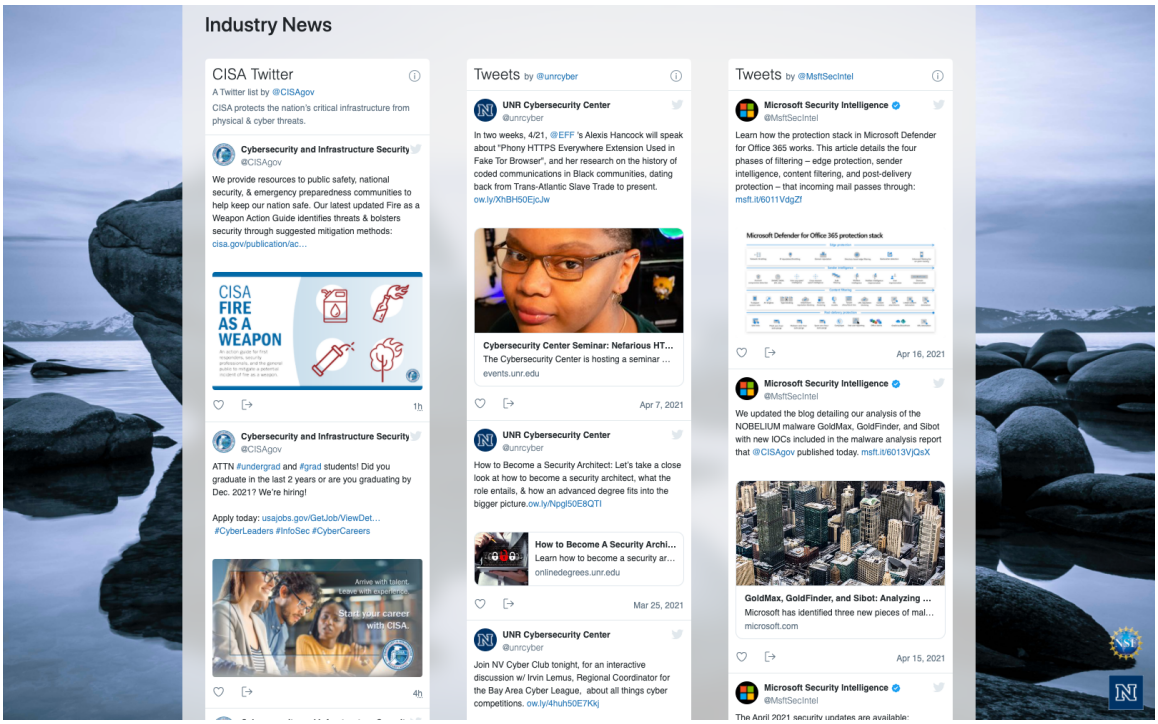


Figure 6.2: Cybersecurity-related news feeds are displayed on the home page.

In such cases, the user's data is persisted from prior sessions and loaded in exactly how they left it. The overall design uses a 'dark' theme, reducing eye strain and emphasizing colors. All peripheral menus and controls are accessible from the perimeter of the canvas. The first class of these controls exist within the top navigation bar. This includes access to functions that are not directly-graph related. Included in these features are links to other CYBEX website pages, user profile information, administrative panel, and a supplemental trends page. These items are depicted later in this section. The second class of controls are those that are deeply related to graph construction and analysis. These are placed on the left, right, and bottom edges of the screen. The three main expandable menus are signified by blue tabs, each with a unique icon. Each tab affords the on-demand display of their associated menus. All three are depicted in their expanded states in Figure 6.4.

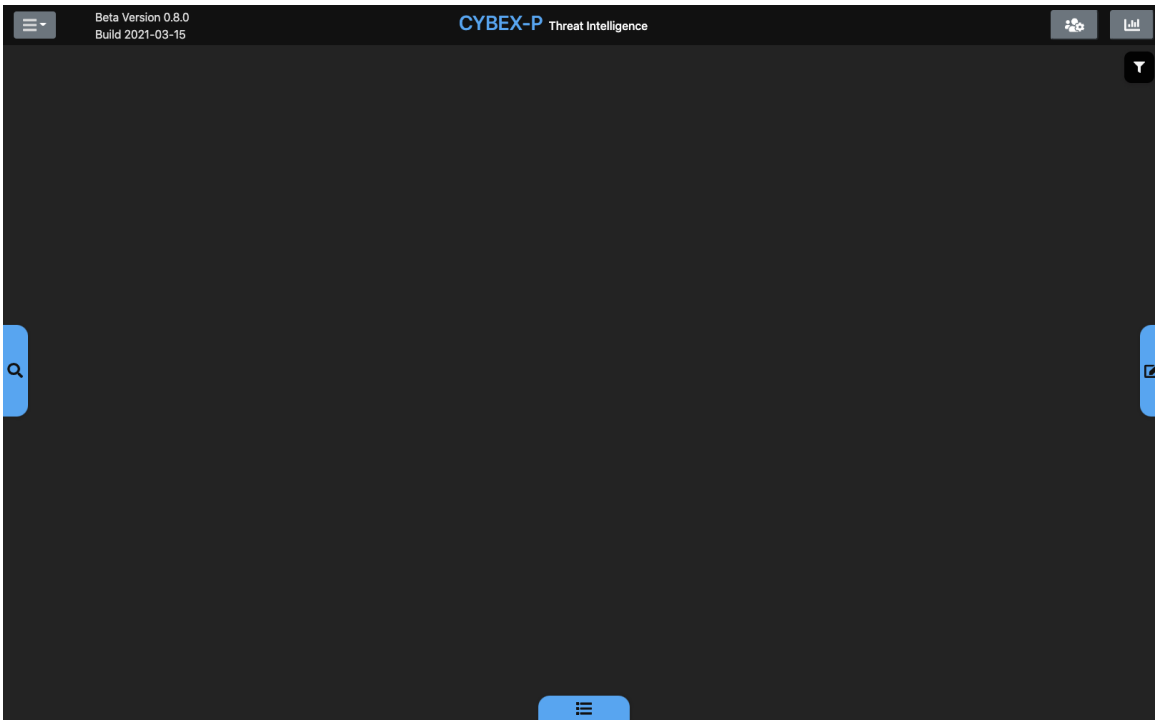
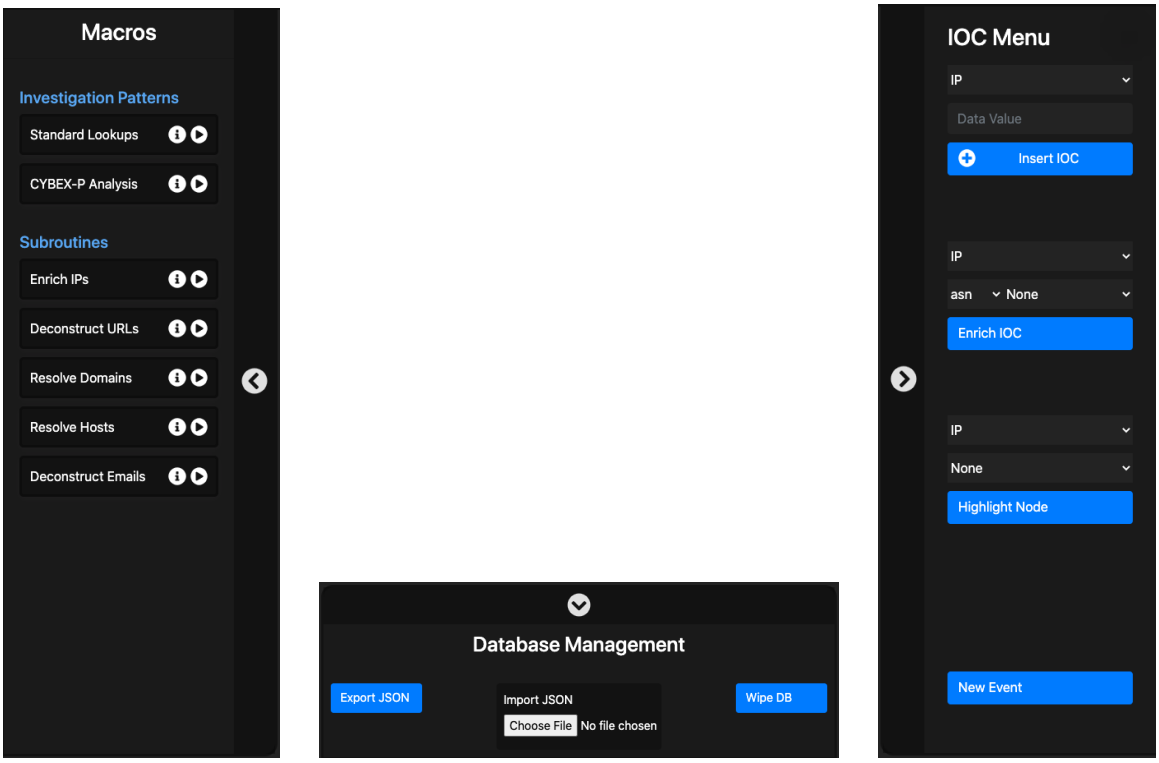


Figure 6.3: CYBEX-P threat intelligence graph interface (empty canvas).

Figure 6.4(c) shows the IOC Menu, which groups together IOC-specific functions. The first thing it affords is adding new IOCs to the graph canvas. Users select from a full list of IOC types (IPs, URLs, email addresses, and more), type the value of



(a) Macro menu (b) Database management menu (c) IOC Menu

Figure 6.4: Three expandable menus containing graph-related functionality.

that item, then click the insert button. Once nodes are added to the graph, the next group of controls in this menu can be used to execute enrichments on them. The list of available enrichments varies depending on the type of IOC selected. Examples for different IOC types are given in Table 6.1. Descriptions of what each enrichment does are listed in Table 6.2.

Figure 6.4(a) depicts the Macro menu. Macros apply enrichments across every node of the graph at once. There are two categories: investigation patterns and subroutines. The former combines multiple enrichment types together at once, while the latter executes a single enrichment type across the graph. For example, the 'Standard Lookups' investigation pattern includes deconstruction of email addresses (extracting the domain names), but it also performs a number of other things. If the user only wants to deconstruct email addresses, they can run the associated subroutine instead of the full investigation pattern. To get the full list of enrichments

IOC Type	Supported Enrichments
IP Address	asn, gip, hostname, whois, netblock, ports, cybexRelated, cybexCount
URL	deconstructURL, cybexRelated, cybexCount
Domain	whois, resolveHost, nameservers, registrar, mailservers, cybexRelated, cybexCount
Host	whois, resolveHost, nameservers, registrar, mailservers, cybexRelated, cybexCount
Email	deconstructEmail, cybexRelated, cybexCount
Hash	cybexRelated, cybexCount
ASN	cybexRelated, cybexCount

Table 6.1: Supported enrichments for some example IOC types.

Enrichment	Description
asn	Returns the IP's Autonomous System Number (ASN).
gip	Performs a GeoIP lookup to return the country of origin.
hostname	Uses address to return the associated hostname.
whois	Returns owner information from Whois record.
netblock	Returns subnet value.
ports	Performs a Shodan lookup to return a list of open ports.
cybexRelated	Queries the CYBEX-P database for IOCs that are related by common events.
cybexCount	Queries the CYBEX-P database for counts of the IOC's benign and malicious sightings.
deconstructURL	Retrieves the domain name from a particular URL.
deconstructEmail	Retrieves the user and domain name from an email address.
resolveHost	Uses hostname to return associated host address.
nameservers	Returns nameservers.
registrar	Returns registrar from Whois record.
mailservers	Returns mailservers.

Table 6.2: Descriptions of each supported enrichment.

each macro includes, users can click its information button. An example of this on-demand information panel is shown in Figure 6.5. This panel is a semi-transparent and easily-closed popup, reducing the amount of graph area it obstructs.

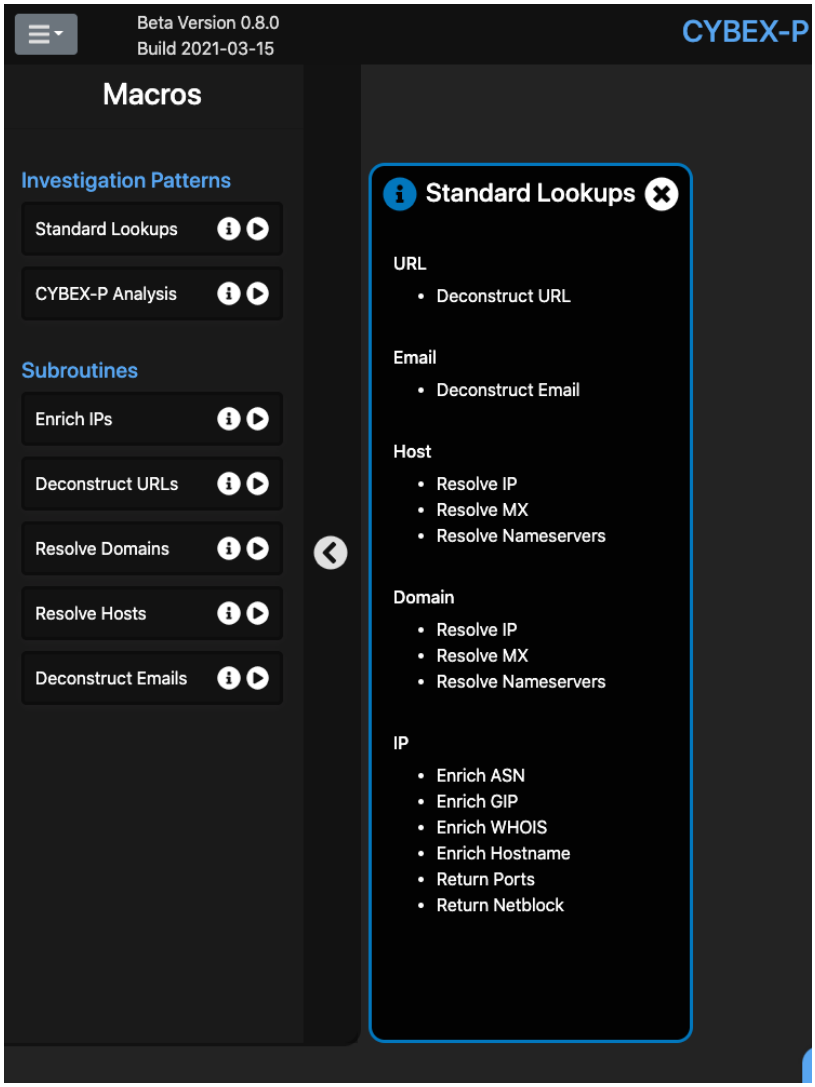


Figure 6.5: Information pop-up for macros is minimally-invasive due to its on-demand nature and semi-transparent background.

The bottom menu, shown in 6.4(b), is for database management. This refers to the graph database that is created for each user (see Chapter 5). Users can export their current graph data as a JSON file, load a new graph from an imported JSON file, or wipe the database for a fresh canvas.

Figure 6.6 shows how nodes appear once they've been added with the IOC menu.

When hovering over each node, a tooltip displays its IOC type and value (see Figure 6.7). When selected, a radial menu renders around the node containing all available enrichments for that node. As demonstrated in Figure 6.8, an informational card is also displayed in the bottom-right portion of the screen. This contains a more complete list of information about the node, as well as some auxiliary options like node commenting. Figure 6.9 shows the result of a GIP enrichment being run from the radial menu. In this case, a single country node is added and connected to the graph. The edge between the two nodes is rendered with an arrow, pointing from the originating node to the new node. This directionality sometimes provides helpful historical context. The arrows allow an investigator to retrace steps, revealing the order enrichments were run. In addition, relationships can sometimes be very directional. For example, if a GIP lookup is run to get country of origin on an IP, the arrow will point from IP to country. The IP is located in the country, not the other way around. In these cases, arrows and descriptive tooltips on edges establish the flow of information.

Figure 6.10 shows a macro in action. Specifically, this is an example of running the 'CYBEX-P Analysis' macro on a single IP address (0.0.0.3). This macro does two things. First, it runs cybexRelated enrichments on all graph nodes. Second, it performs threat analysis of all graph data. When the macro runs, it returns the results of the first step before the second step is done. That way, users can use the new data without waiting for all analysis to complete. Once the second step ends, threat scores are added to each node and the macro is finished. Because macros can have large execution times on large graphs, this sequential approach is key for constant usability.

The resulting graph from a completed CYBEX-P Analysis run is shown in Figure 6.11. Each node's threat score is the percentage of 'malicious' events in which that item was seen (displayed when hovering over analyzed nodes). This is a simple calculation using the 'benign' and 'malicious' sighting counts from the CYBEX-P database. Using this percentage, a categorical threat ranking is assigned to each IOC.

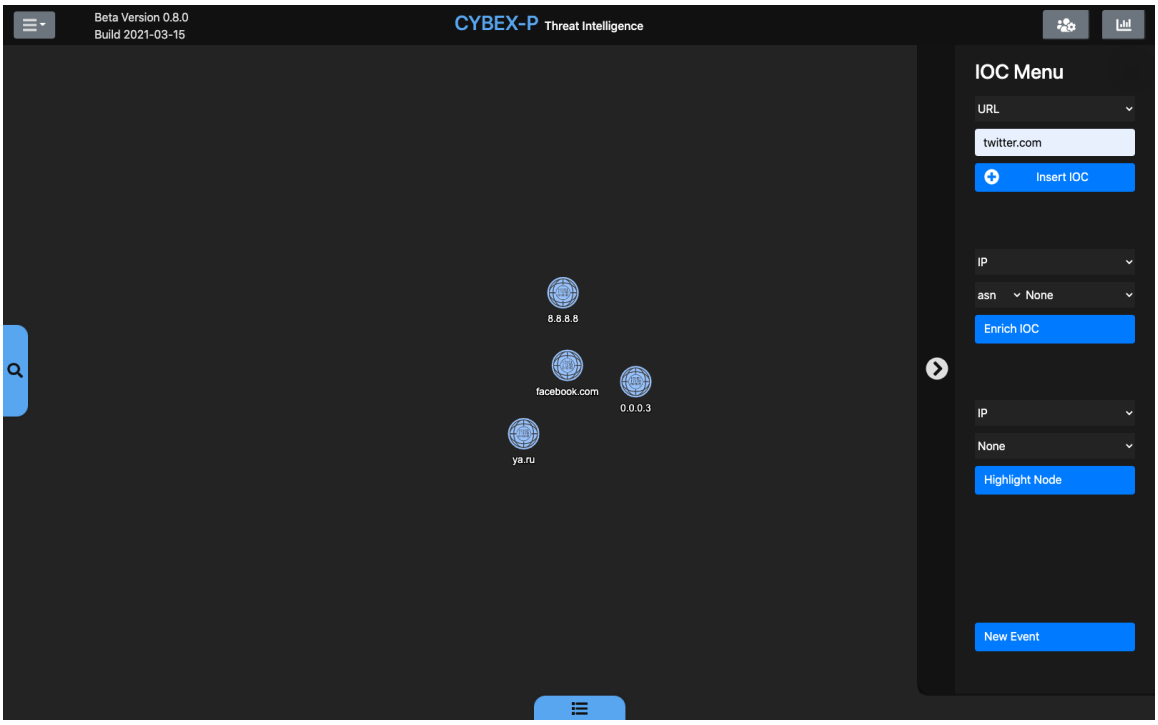


Figure 6.6: IOC menu is used to add some example IP addresses and URLs to the graph.

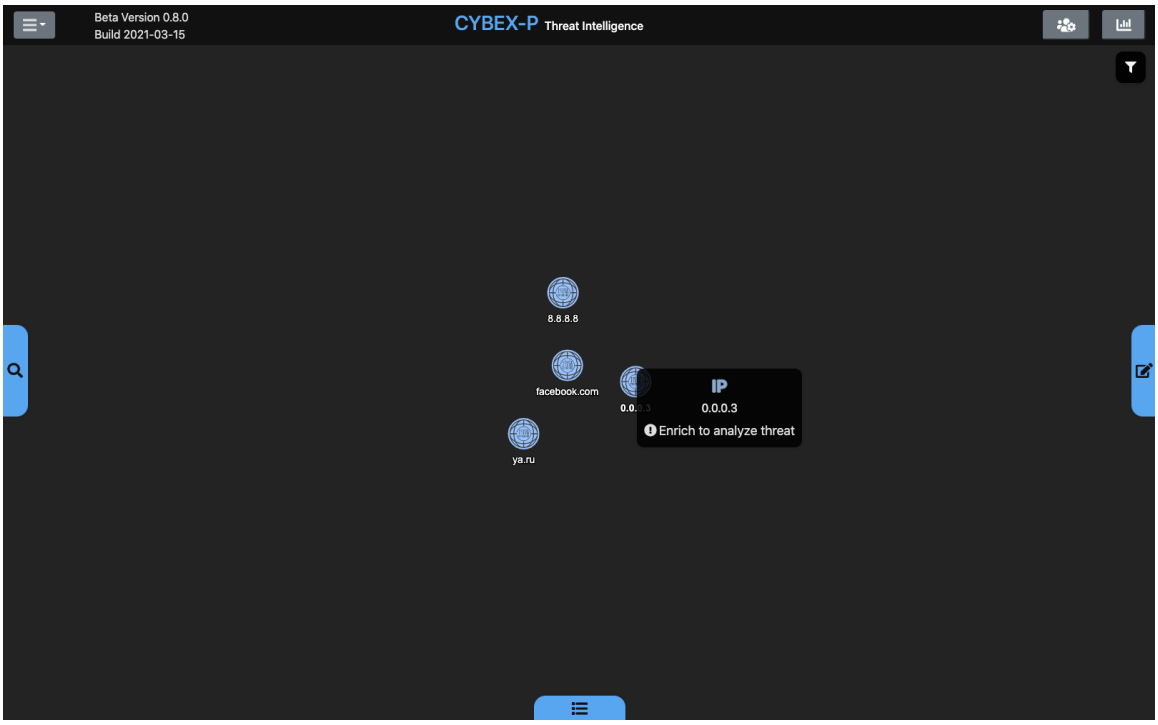


Figure 6.7: Tool-tips containing basic details about nodes are displayed upon mouse hover.

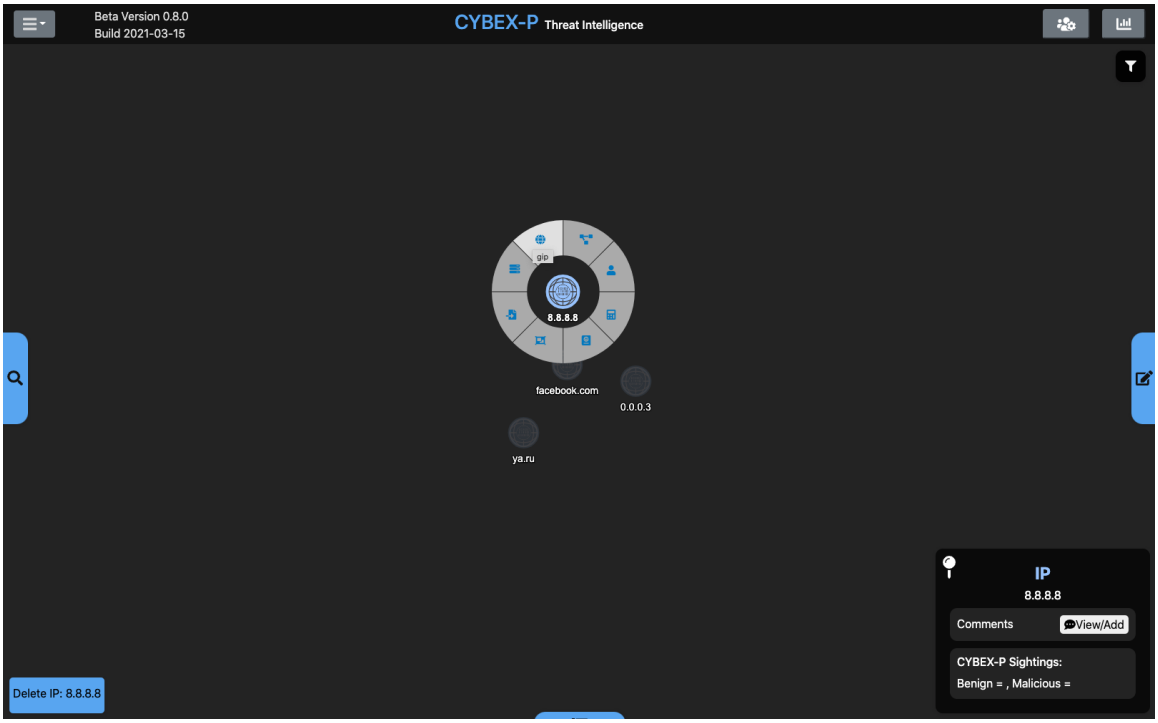


Figure 6.8: The full details and available actions for a node are displayed when clicked on.

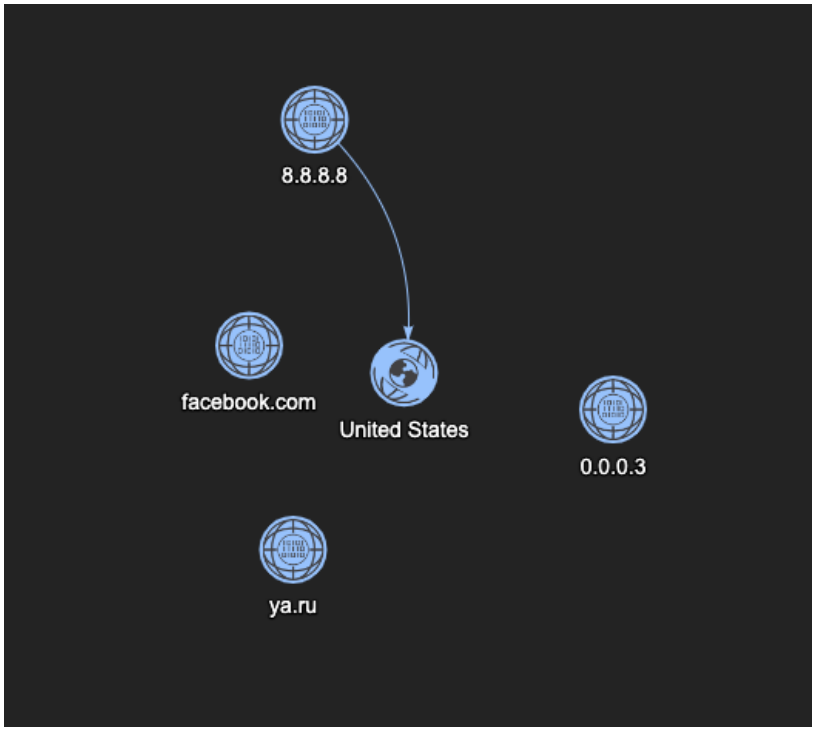


Figure 6.9: The resulting graph after running a single enrichment on a single node.

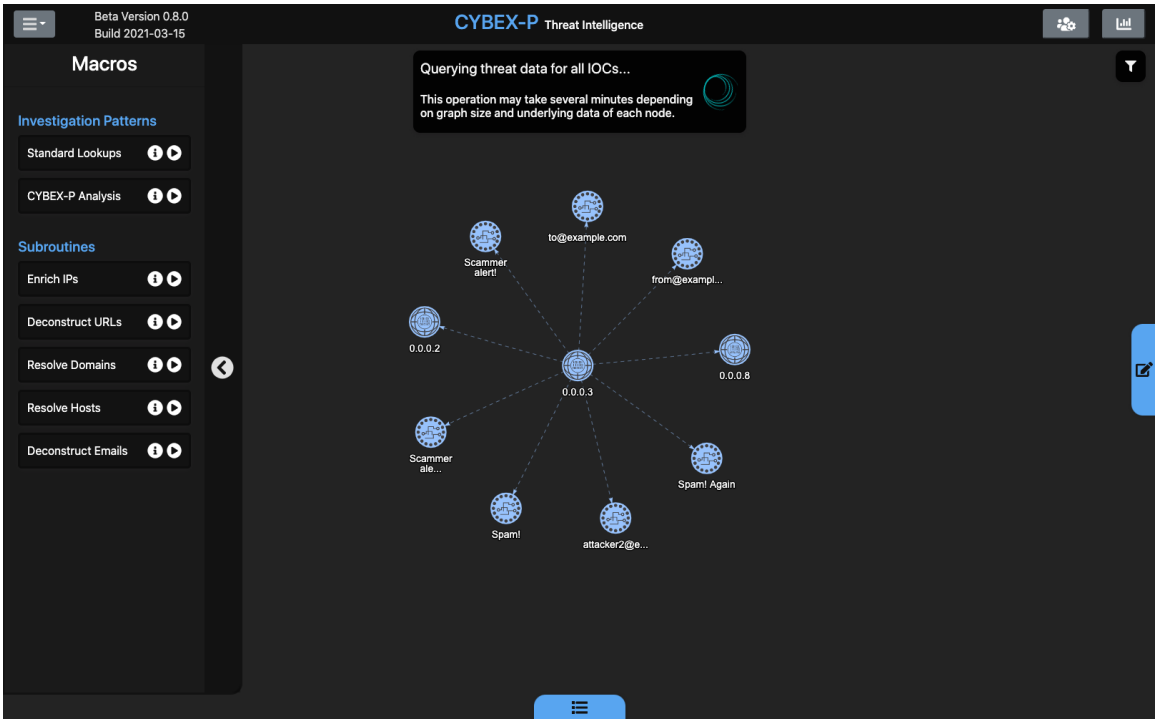


Figure 6.10: The resulting graph after the first operation (adding related IOCs) of the 'CYBEX Analysis' macro has completed.

These are 'benign', 'moderately malicious', and 'very malicious'. These are assigned node colors of green, yellow, and red, respectively. Blue nodes are the default color, and they remain blue if no threat information can be gathered. Threat isn't the only thing that is visualized in this case, however. When the count values are returned, they also are used to create a 'total sightings' value (malicious + benign sightings). Total sightings are visualized using the size of each node. A larger node equates to a larger number of overall sightings in the CYBEX-P database. This feature attempts to add useful context, which is one of the design goals discussed in Chapter 5. However, as discovered in Chapter 7, this feature may need to be adjusted to prevent analyst confusion.

CYBEX-P analysis also provides event context. When obeserving two nodes with a CYBEX relationship, their common edge can be hovered over. The resulting tool-tip goes beyond a simple textual description. Instead, it displays a chain of color-coded boxes. From one end of the chain to the other, this represents exactly how the

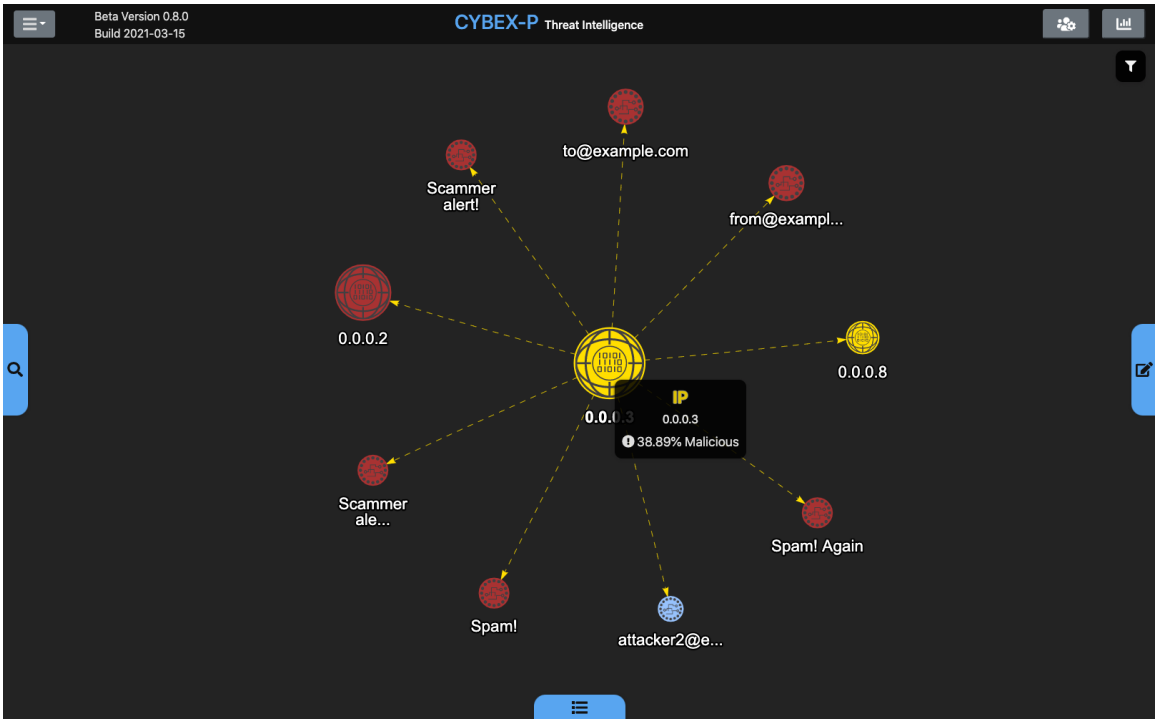


Figure 6.11: The resulting graph after the 'CYBEX Analysis' macro has performed threat analysis. A 'threat score' is given when hovering over the node.

two nodes are connected. The example shown in Figure 6.12 shows the chain between two IP addresses. Attributes are purple, objects are blue-green, and events are gold. Using these classifications requires some knowledge of how CYBEX-P tracks items in its database. The given chain shows that a destination IP (0.0.0.3) was connected to a network traffic event. Note that 'dst' (destination) is the object, with 'ipv4' being its associated attribute. Likewise, it is easy to see that a source IP was part of this same event. Serving as the common ground for these two IPs, the network event exposes this potentially unknown relationship. There are many types of events, objects, and attributes that can form these types of relationships. Dashed lines are used to distinguish CYBEX event relationships from all other enrichment types.

Users can customize CYBEX-P enrichments using the expandable filter in the top-right portion of the screen. Displayed in Figure 6.13, this component affords adding restrictions to the data that is returned. Specifically, start and end dates can be selected, according to a desired timezone. When returning related IOCs or IOC

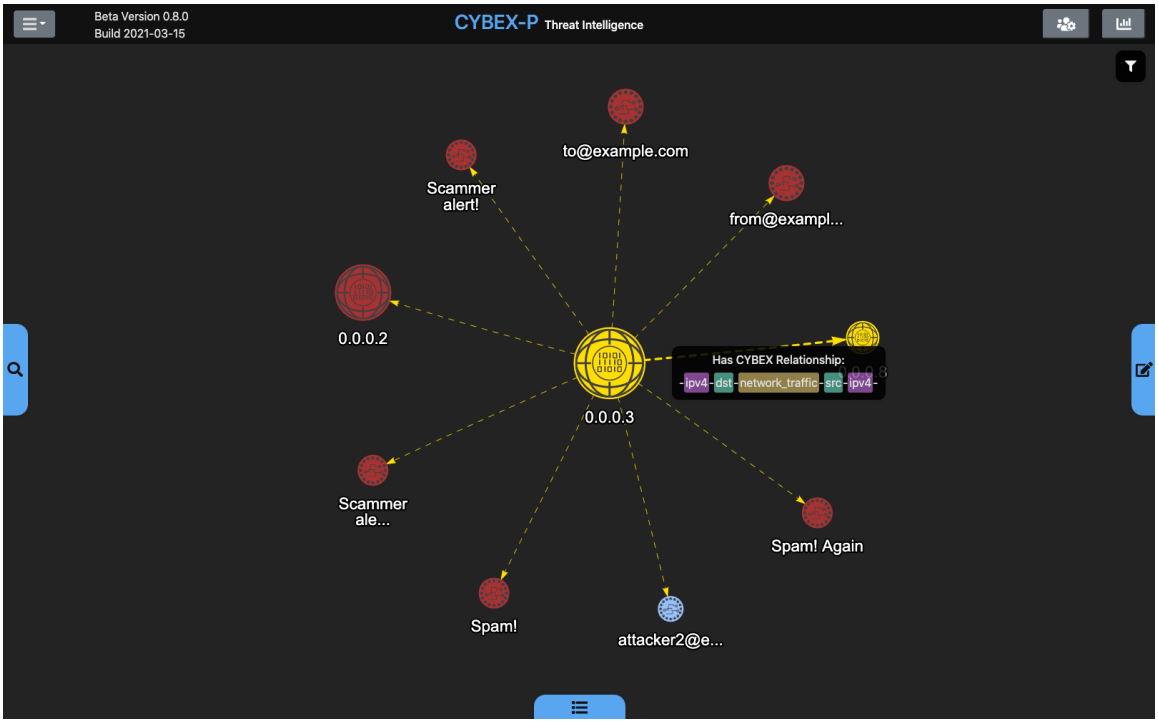


Figure 6.12: When hovering over an edge that represents a CYBEX relationship, the event data that connects the two nodes is displayed.

sightings data, only results gathered within these ranges be used. This can be helpful if an analyst wants to limit an investigation to a known timeframe, such as when some suspicious activity occured. For example, consider an IOC that is only seen in benign contexts for many years. However, it then briefly appears in several malicious events. If too large of a time range is used (such as a year), the many benign counts may outweigh the few malicious ones. This may make an analyst wrongly think the IOC is safe. Instead, if they limit the range to a particular week of interest, they will get threat classifications that are more relevant to their investigation.

Aside from the core graph features, adjacent functionality is provided in the top navbar. First, an expandable menu is depicted in Figure 6.14. This provides links that redirect to the CYBEX homepage and various documentation pages. Another option allows users to download cowrie honeypot log files that are hosted on the same server as the application. These files are purely for experimental purposes and not intended as a main feature of the application. Lastly, users can open a window for

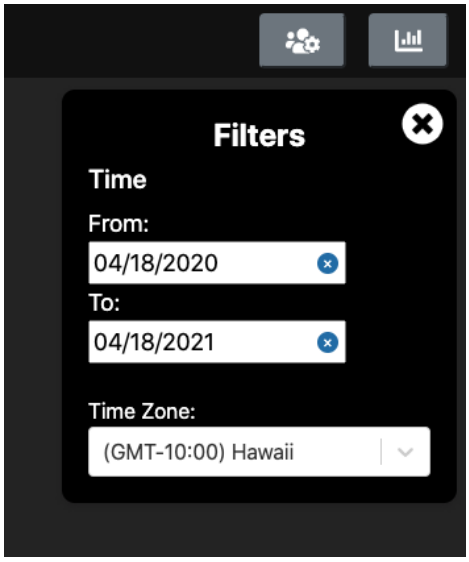


Figure 6.13: Expandable window allows users to filter CYBEX enrichments and macros to particular time ranges.

their user profile. This is depicted in 6.15 and shows the username, email address, and a unique identifier of the person who is signed in. Any organizations the user belongs to is also shown here.

There are two buttons on the rightmost part of the navbar. The first signifies administrative functions, and opens the user management window. Shown in Figure 6.16, this panel allows admins to add and remove users to their organizations. The final buton on the navbar opens the trends panel, displayed in Figure 6.17. This is a space for plots that show high-level trends on the overall CYBEX-P database. The plots shown here are just examples, but can represent things like events over time or the most prevalent event types.

The last major feature is user event data submission. Shown in Figure 6.18, users can upload supported file types. The example shown here is a JSON file from captured Cowrie honeypot logs. The user must also supply a timezone to associate the submitted event(s) with. Strict validation is built into this form. First, the application will not allow users to submit forms that have incomplete fields. If so, required fields are highlighted to users so they can fix their mistake. Next, the actual contents of the file are validated against a list of supported schemas. For example,

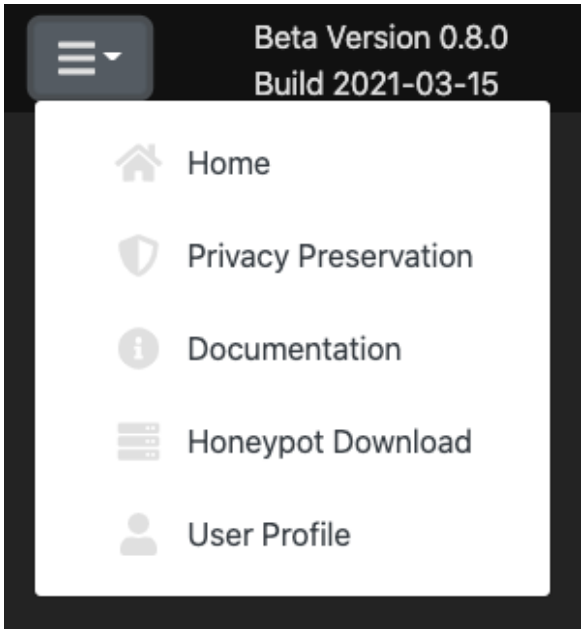


Figure 6.14: Expandable menu in top-left corner reveals some navigation options and secondary features.

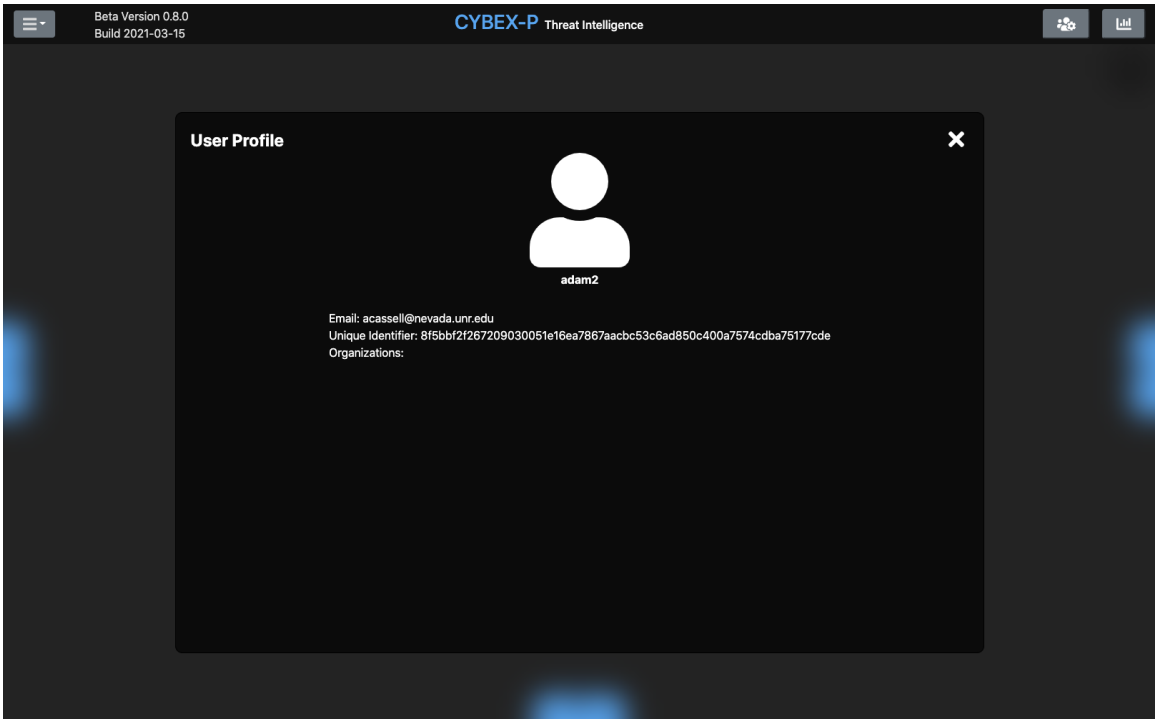


Figure 6.15: User profile panel displays basic account information about the currently signed-in user.

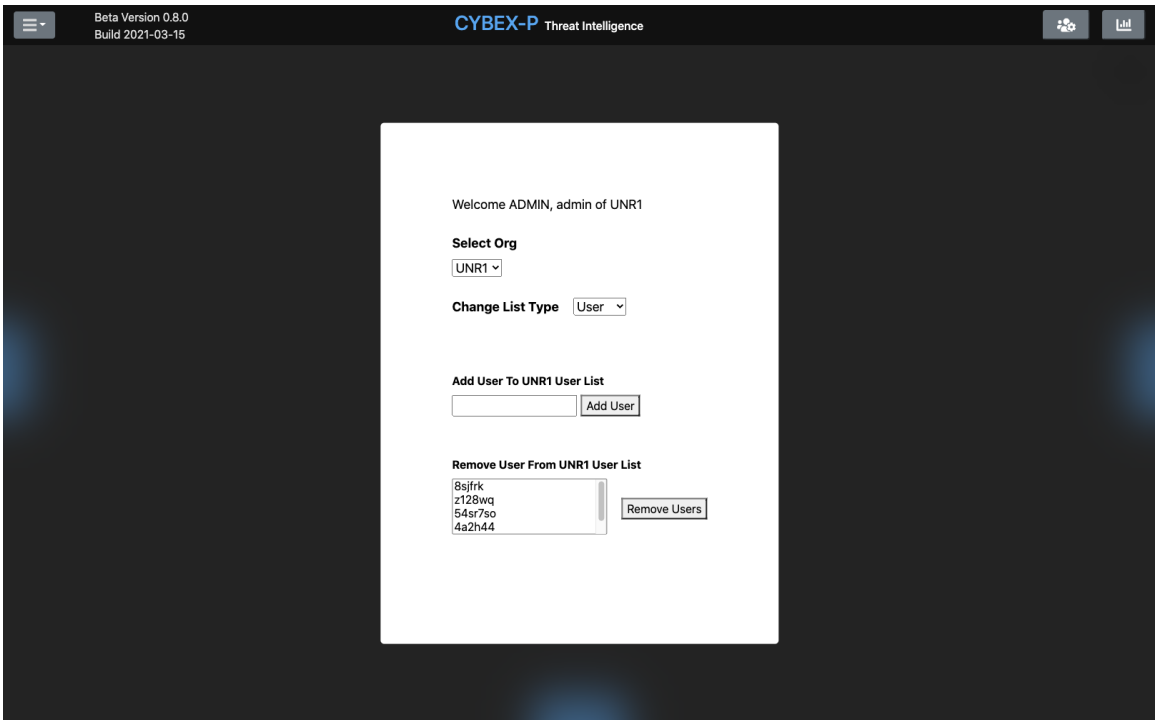


Figure 6.16: User management panel allows admins to add/remove users to their organizations' lists.

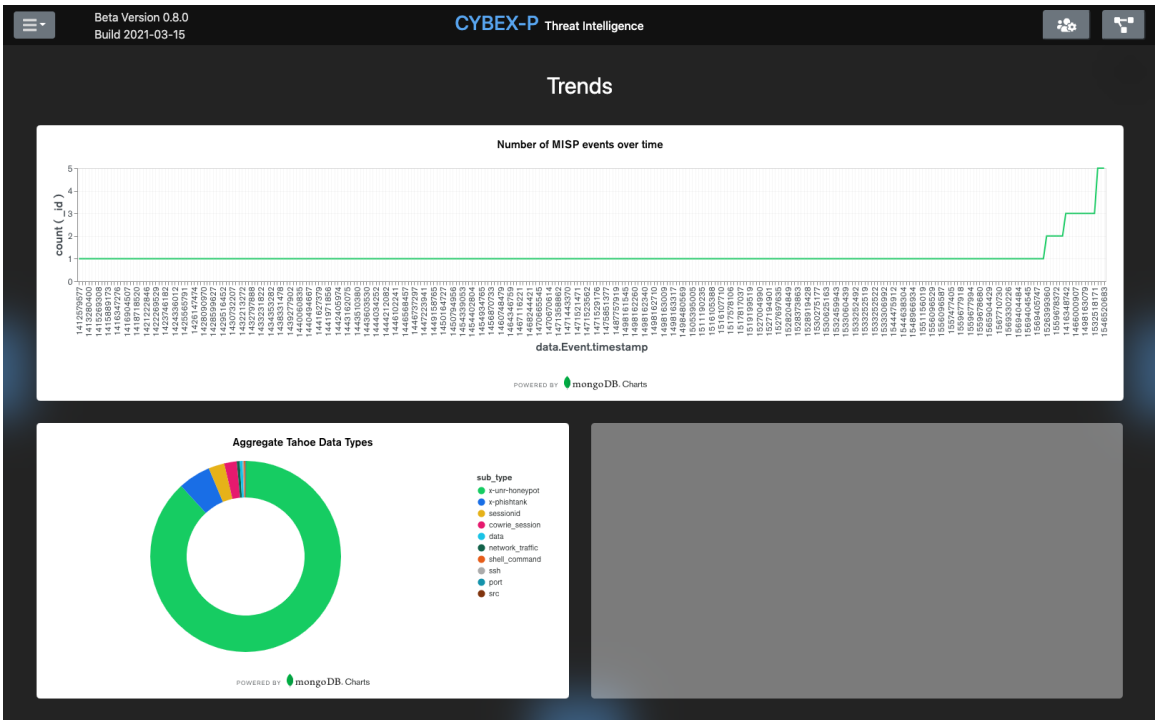


Figure 6.17: Trends panel displays high-level statistics and time-series information about CYBEX-P's data sources.

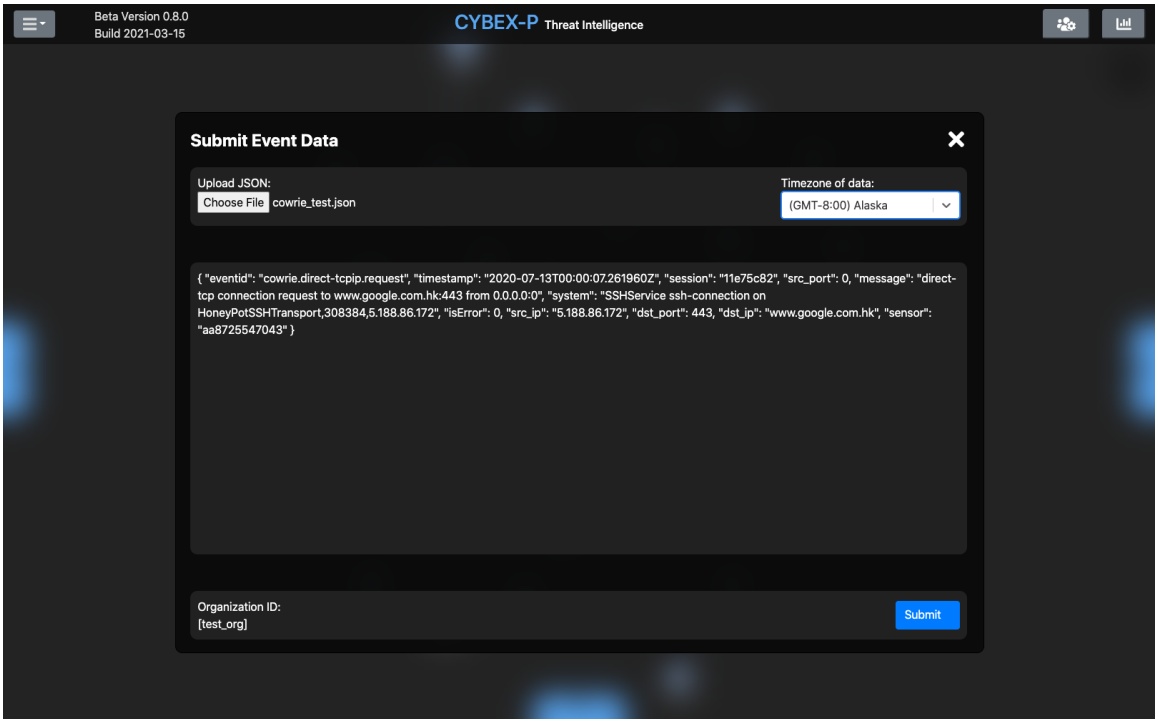


Figure 6.18: Users can upload their own event files as contributions to the shared CYBEX-P database.

Cowrie log files follow a known pattern of key/value pairs. If the data types and key types don't match exactly, the data is rejected. Users are alerted if something is wrong with their data, and can then address it. In addition, there are safeguards in the backend CYBEX-P system to prevent bad data from being introduced. These are outside the scope of this work, but they complete a robust series of validation steps that all data is subjected to.

With all the major features now introduced, it is clear that there is a wide variety of functions that the application can fan facilitate. These satisfy the requirements and intended use cases defined in Chapter 5. Figure 6.19 is an activity diagram representing behavior of the application. It shows how some key user actions drive the main features of the application. Note that this diagram was drawn during early development stages, so a few details have since changed. The following sections walk through some specific usage scenarios step-by-step, demonstrating the intended utility of the application.

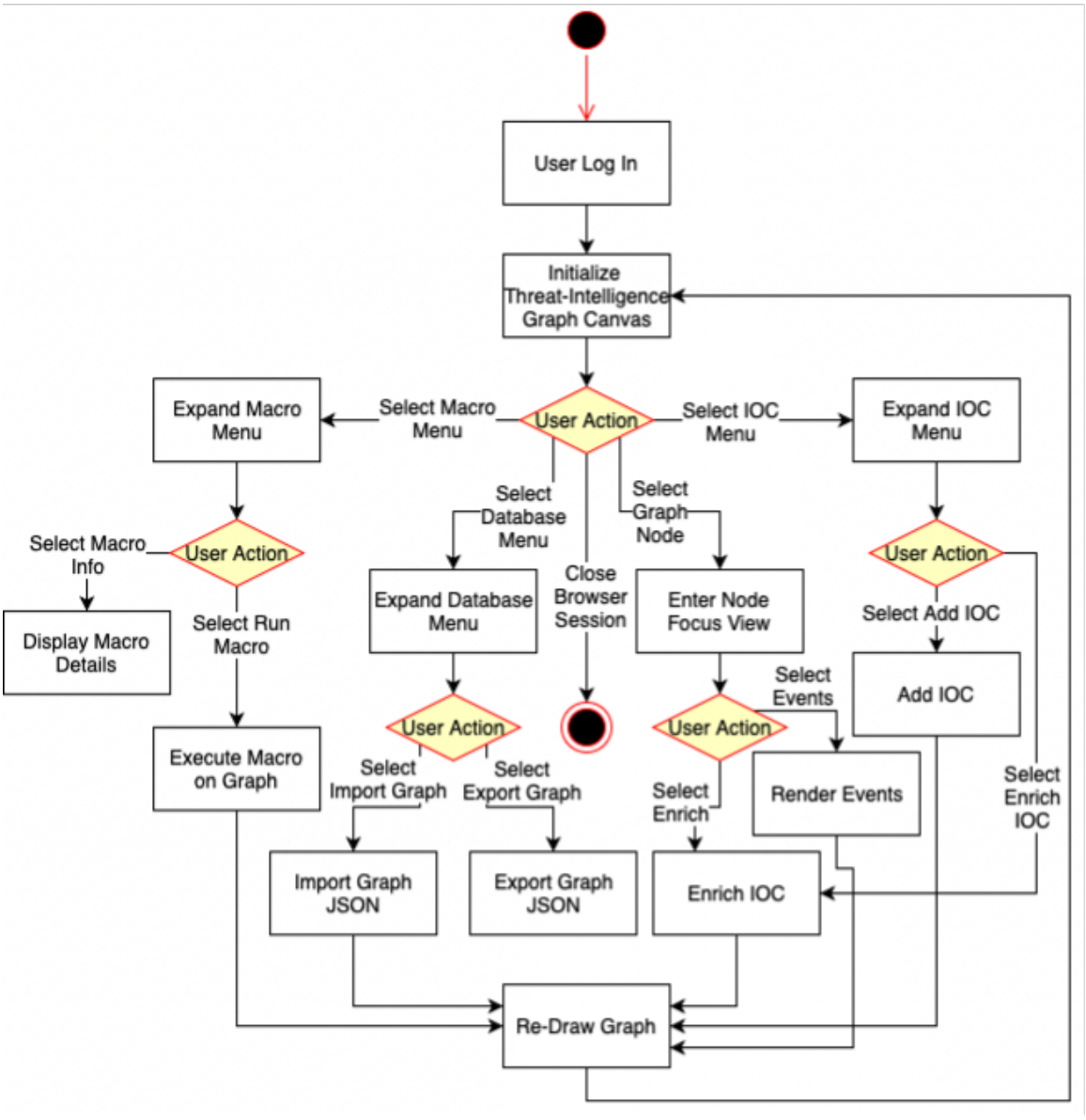


Figure 6.19: Activity diagram illustrating the expected behavior of the product.

6.2 Scenario A: Large Graph Construction

The first scenario resembles some exploratory graph construction, rather than specific analysis. This approach may be taken by analysts who simply want to get a big-picture view of a particular network graph. In this scenario, they are mostly concerned with visualizing connections between IOCs of interest and any adjacent items. This is a case where the investigator doesn't have any particular leads; rather they would simply like to picture their entire network topology.

This scenario begins with the same set of nodes pictured in Figure 6.9. These nodes are all of the known IOCs in the network the analyst monitors. From this starting point, the analyst runs the 'Standard Lookups' macro to return as much lookup data as possible for all nodes. This macro includes all of the non-CYBEX enrichments listed in Table 6.2. Figure 6.20 shows what the screen looks like while this macro executes. Figure 6.21 shows the result. The investigator decides to run this macro recursively, calling it a couple additional times. This expands the graph space with multiple layers of additional nodes, growing the scope exponentially each time. The investigator wanted to get a big-picture view of their network, so this approach helps them quickly achieve that. Figures 6.22 and 6.23 show the results of the second and third macro runs, respectively.

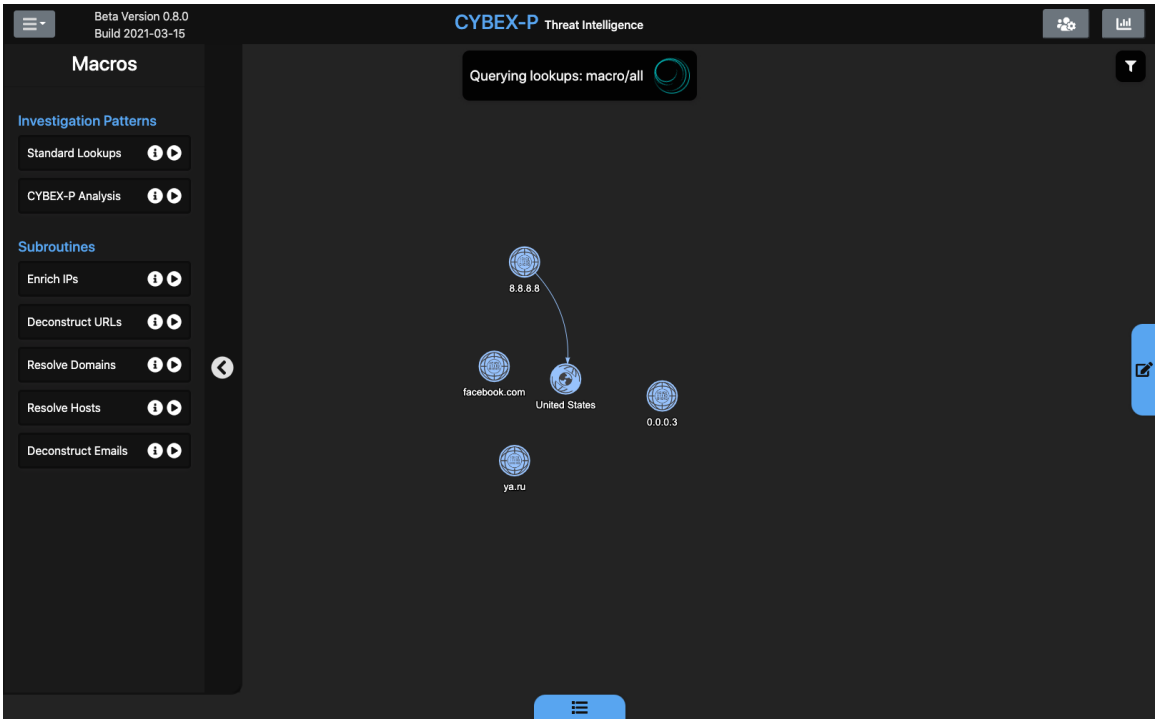


Figure 6.20: The 'loading' state for the 'standard lookups' macro. Users can continue using the graph while the process completes.

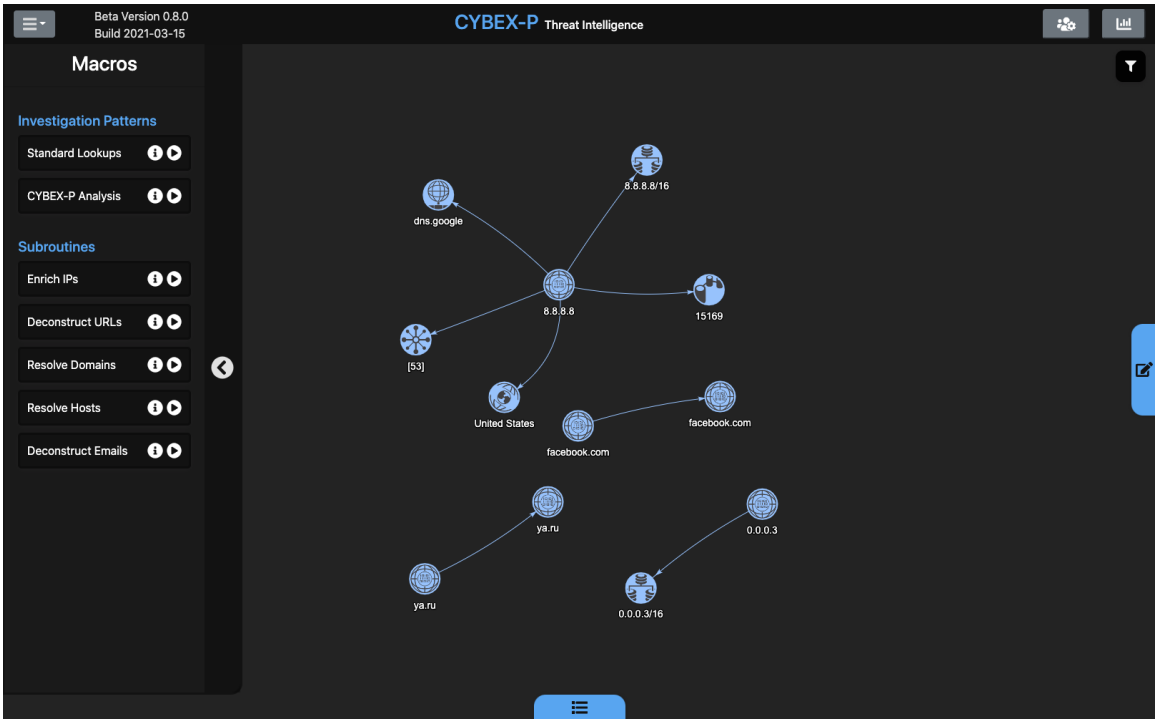


Figure 6.21: The result of a single run of the 'standard lookups' macro.

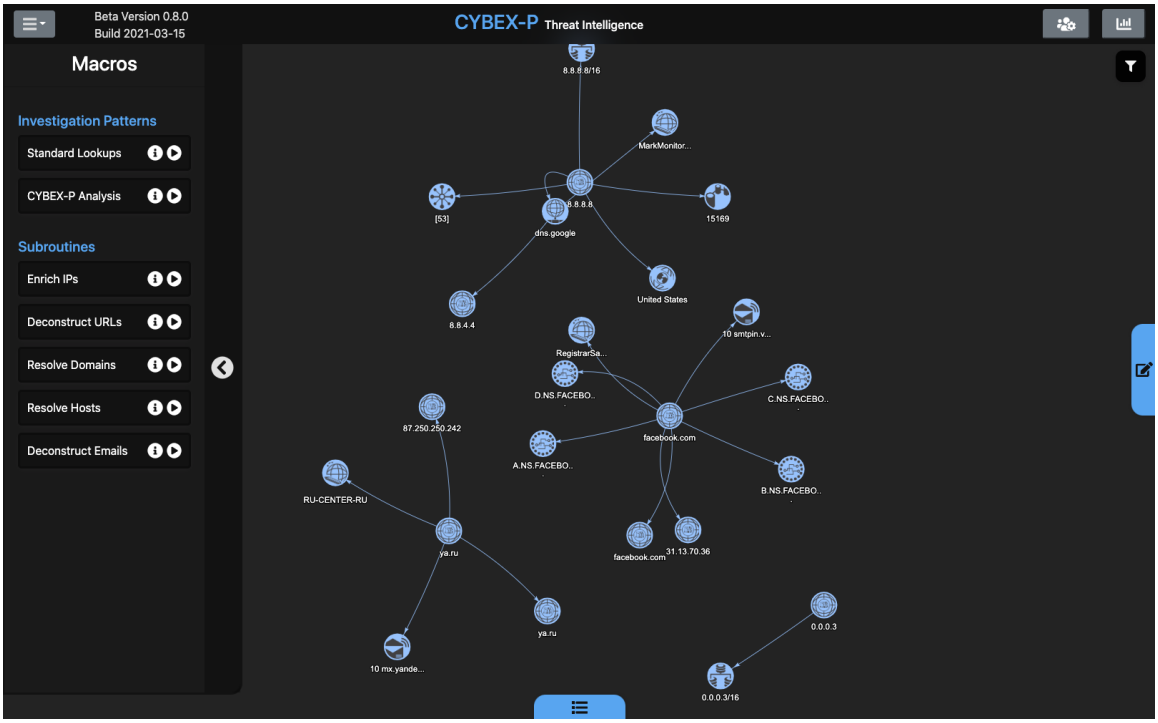


Figure 6.22: The result of two runs of the 'standard lookups' macro.

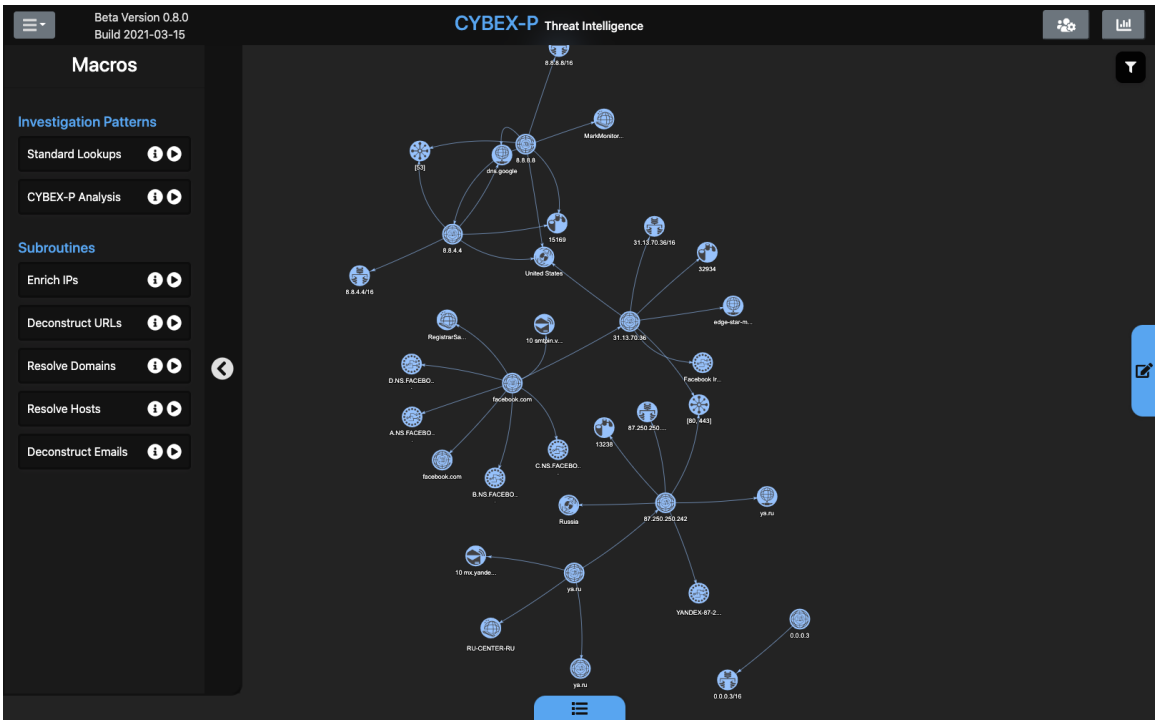


Figure 6.23: The result of three runs of the 'standard lookups' macro.

At this point, the investigator wants to zoom in to part of the graph. They also want to manipulate the graph so that the nodes can still largely be viewed concurrently in this zoomed state. Figure 6.24 shows the graph after the user clicked and dragged one of the bottom node clusters to the upper-right of the canvas. The graph physics also pulls the connected nodes in the same direction as the one the user dragged. When the user lets go, the graph stabilizes and the new positions of all nodes are stored. Figure 6.25 displays a zoomed-in view of this repositioned graph. The investigator is curious about the dual-cluster shape formed in the bottom-left portion of the graph. Zooming in and hovering over the common edge gives a better view of the relationship between facebook.com and the IP address 31.13.70.36. The anlayst wants to eventually figure out if any other IPs are associated with the same host as 31.13.70.36. However, they will need to gather some more data to add to the graph before returning to this question another time. To make it easy to locate later, the node in question is highlighted using the IOC menu controls. Node highlighting of this host name is shown in Figure 6.26.

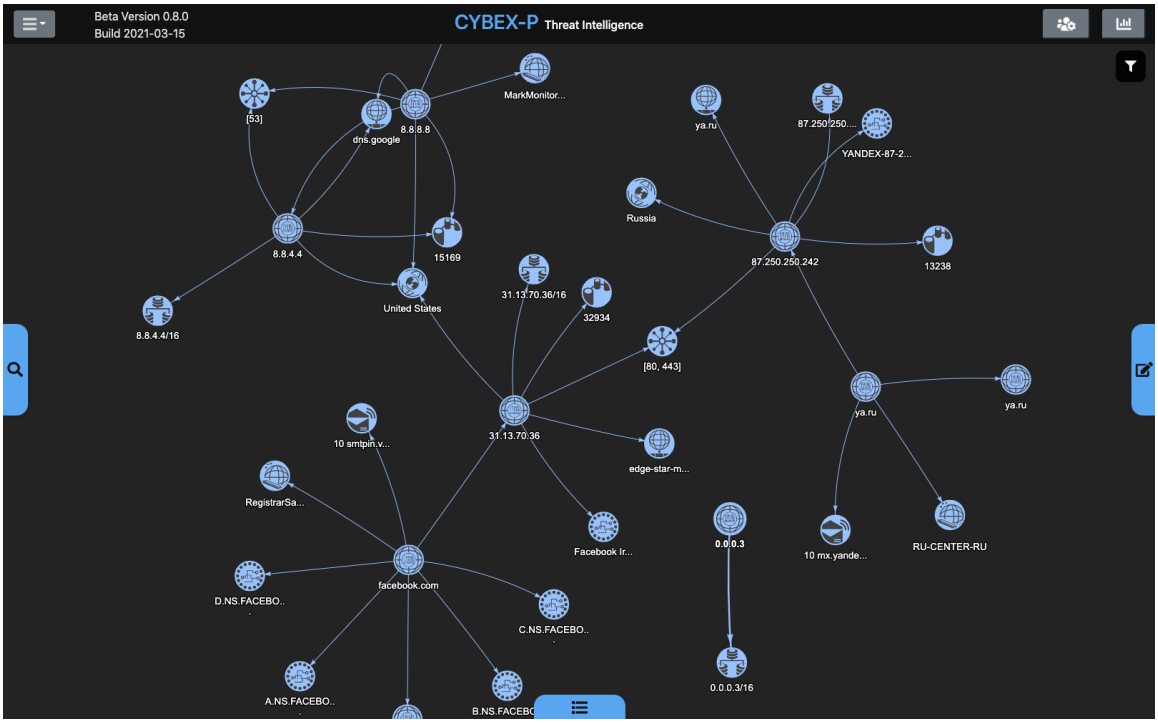


Figure 6.24: Graph nodes can be repositioned however users like, and users can pan/zoom the canvas.

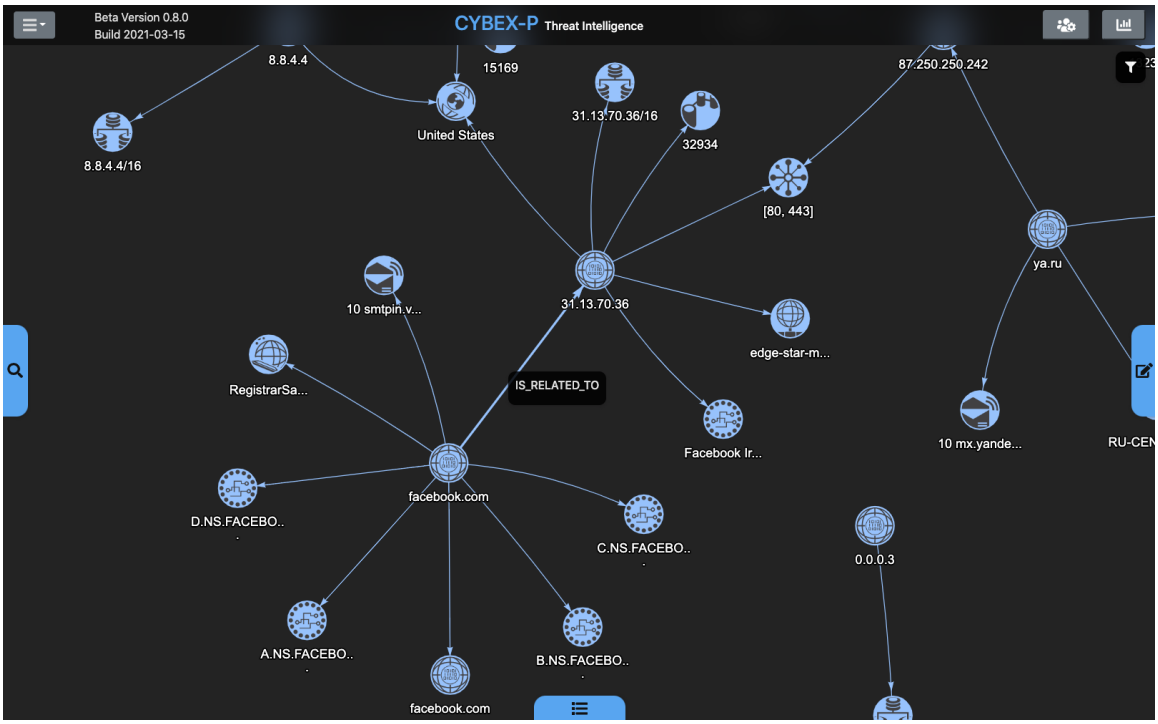


Figure 6.25: Users can pan/zoom the canvas. Hovering over edges reveal information about the relationship between nodes.

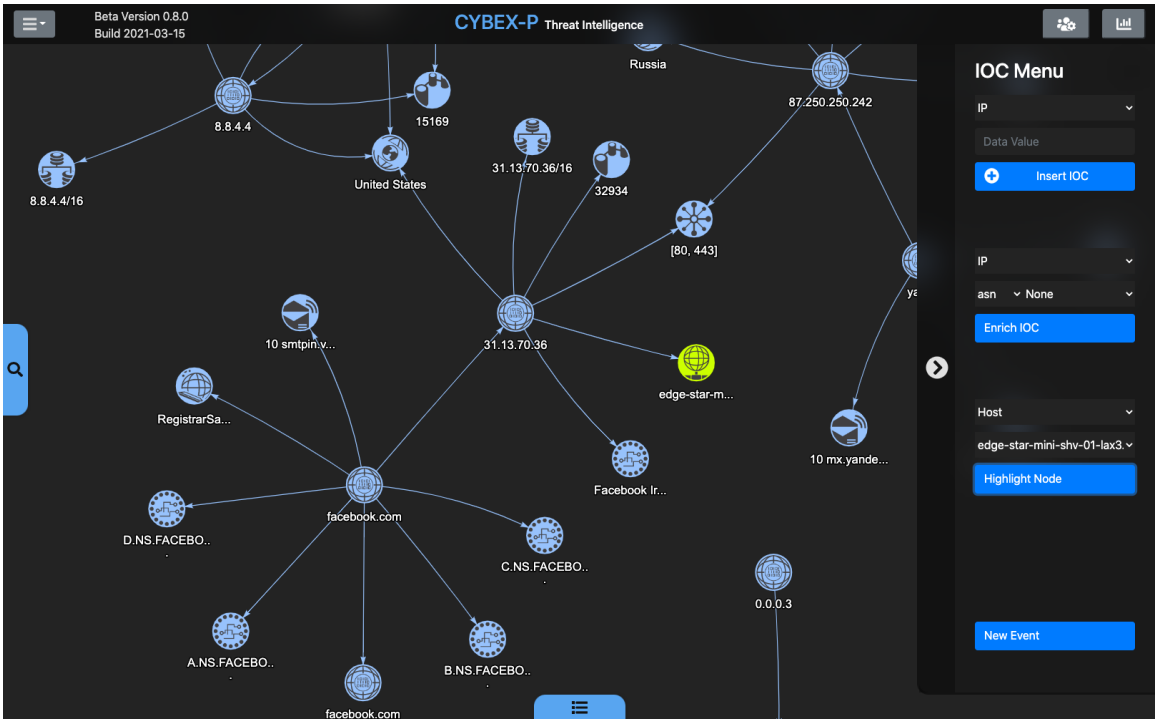


Figure 6.26: Users can mark a certain node to remain highlighted and easily traceable during an investigation.

Next, the analyst wishes to add even more data to the graph. They have already run the standard lookups a few times, but they haven't yet tapped into the CYBEX-P database. They decide to run the cybexRelated enrichment on a few key IOCs. As pictured in Figure 6.27, additional context has now been provided for facebook.com. The previously related data is connected with solid lines, and the new data related by CYBEX is connected with dashed lines. This clearly shows the level of extra context CYBEX-P can provide, relating additional items that were found in common event data. To understand exactly how facebook.com was related to one of these nodes, the user hovers over the edge.

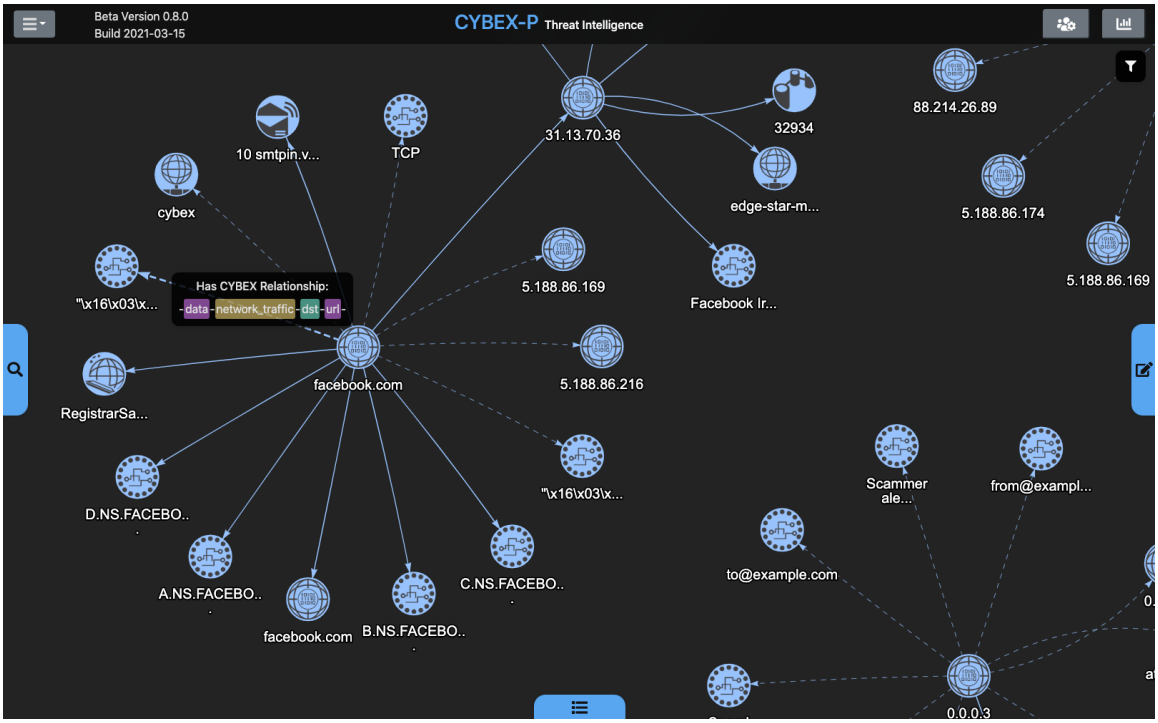


Figure 6.27: The graph is further expanding by using cybexRelated enrichments. The results of these are connected by dashed edges. Context from event data is provided in edge tool-tips to communicate how two objects were connected.

The analyst decides that the graph is now properly visualizing the desired scope of their network. However, there are a few things they would like to trim before sharing this graph with others. For example, the analyst notices that some IP addresses in the graph are not helpful, and instead are a distraction. This could be for many reasons, such as an IP having no meaningful significance to their organization's assets. Perhaps they would like to simplify the graph with only the most pertinent connections so that an executive can review it easily. One IP the user would like to remove is 0.0.0.2, and Figure 6.28 shows them selecting it. They use the delete button in the bottom-left portion of the screen to remove the node and all its connected edges. Figure 6.29 displays the result of this operation.

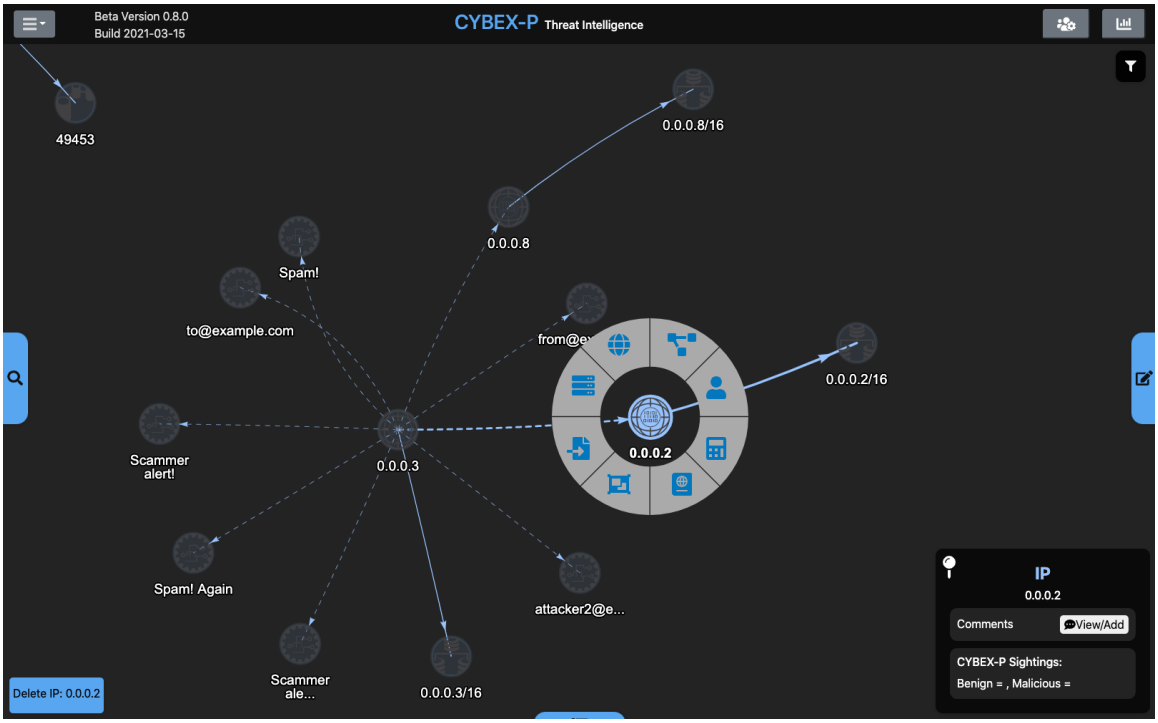


Figure 6.28: To delete a node, the user can select it and then click the bottom-left 'delete' button.

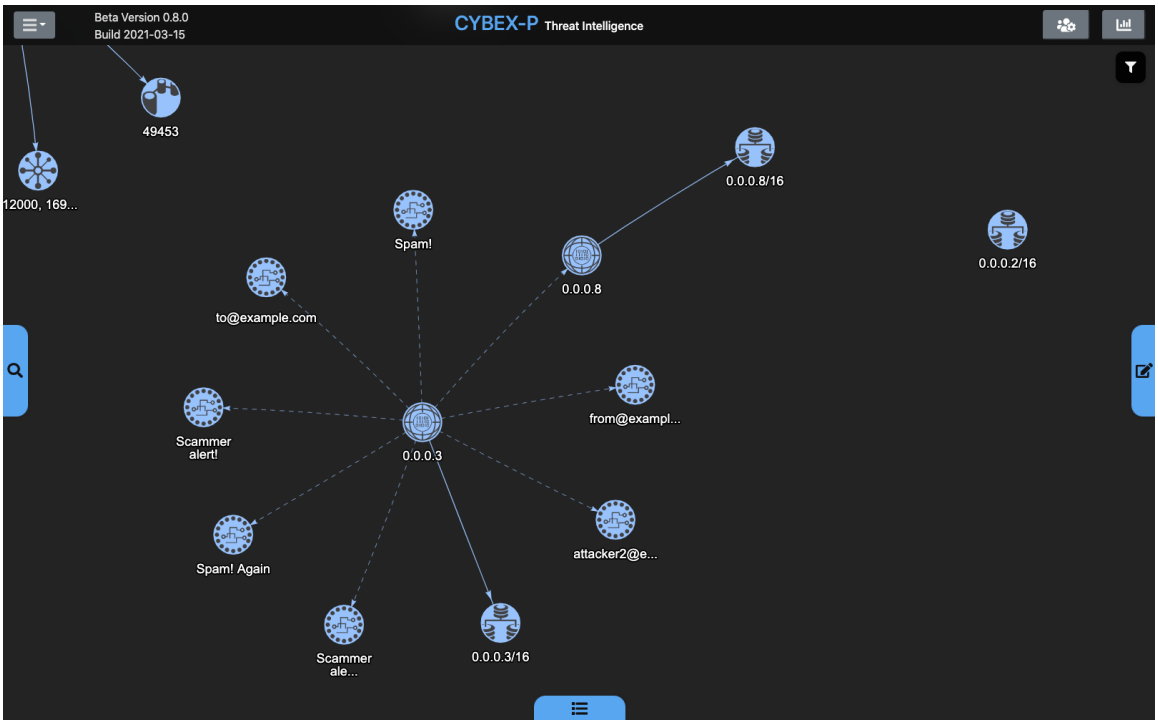


Figure 6.29: The results of deleting a node. Note that all connected edges are removed and related nodes are re-rendered appropriately.

This scenario was entirely exploratory, and focused on building a big-picture visual representation of a particular network. This graph can be referenced when explaining the network's topology, or can even serve as a starting point for threat analysis. Every participating organization of CYBEX-P may want to construct neutral graph templates representing their network assets. Then, when needed, investigators can load these graphs and quickly start running threat analysis. The scenario in the next section focuses specifically on a threat analysis use-case.

6.3 Scenario B: Focused Threat Analysis

For the second scenario, the investigator is specifically concerned about just two IOCs. They have already received information that these IOCs are possibly suspicious, so they want to start from these two nodes rather than some larger network topology. The main goal in this case is to simply display as much threat context about these items as possible. Figure 6.30 shows the two suspicious nodes that the investigator starts with. They immediately run the CYBEX-P Analysis macro, and the result of this operation is shown in Figure 6.31. It is already apparent that the two original IOCs are part of data clusters with very different threat signatures. The IP address 0.0.0.3 itself is colored yellow and classified as moderately malicious. Several other items are found to be related to this IP through CYBEX event data. Almost all of them are colored red, indicating they pose very strong threats (having greater than 50% malicious scores). Meanwhile, the other suspect IOC, a URL called ya.ru, seems to be benign. Most of its related IOCs are classified and colored green.

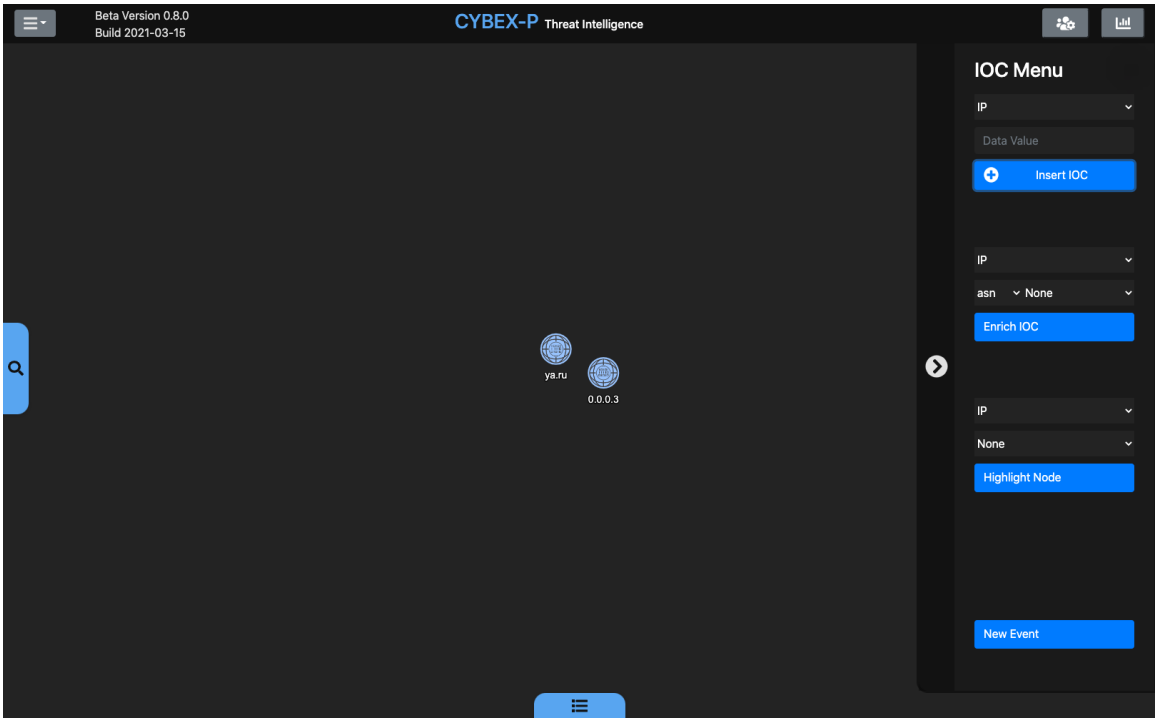


Figure 6.30: The user adds two items of interest to to the graph. The intention is to perform focused threat analysis on these IOCs.

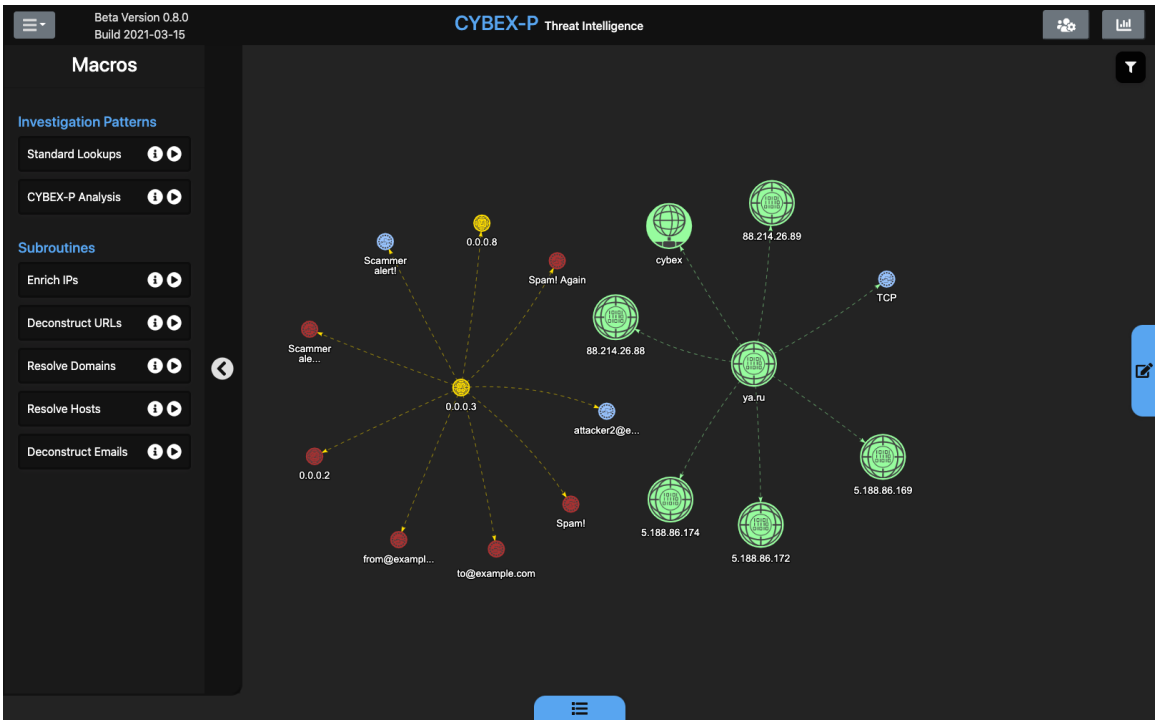


Figure 6.31: The user runs the 'CYBEX Analysis' macro. Related IOCs and attributes are added, alongside automatic threat classification.

Before jumping to conclusions, the analyst wants to perform a few subsequent operations to rule out a threat near ya.ru. They decide to expand the graph scope by running one iteration of the Standard Lookups macro. The result of this is shown in Figure 6.32. This does two things. First, it helps determine if ya.ru has any basic network relationships with the malicious cluster. In this case, no such connection is drawn. Second, there is a new layer of peripheral data from which to run further threat analysis. The investigator decides to run the CYBEX Analysis macro again on this newly expanded graph. Figure 6.33 shows this final graph stage. Some additional data is added via CYBEX event relationships, but no significant new threat classifications are made. The investigator has not been able to show any connection between the benign cluster and the malicious one.

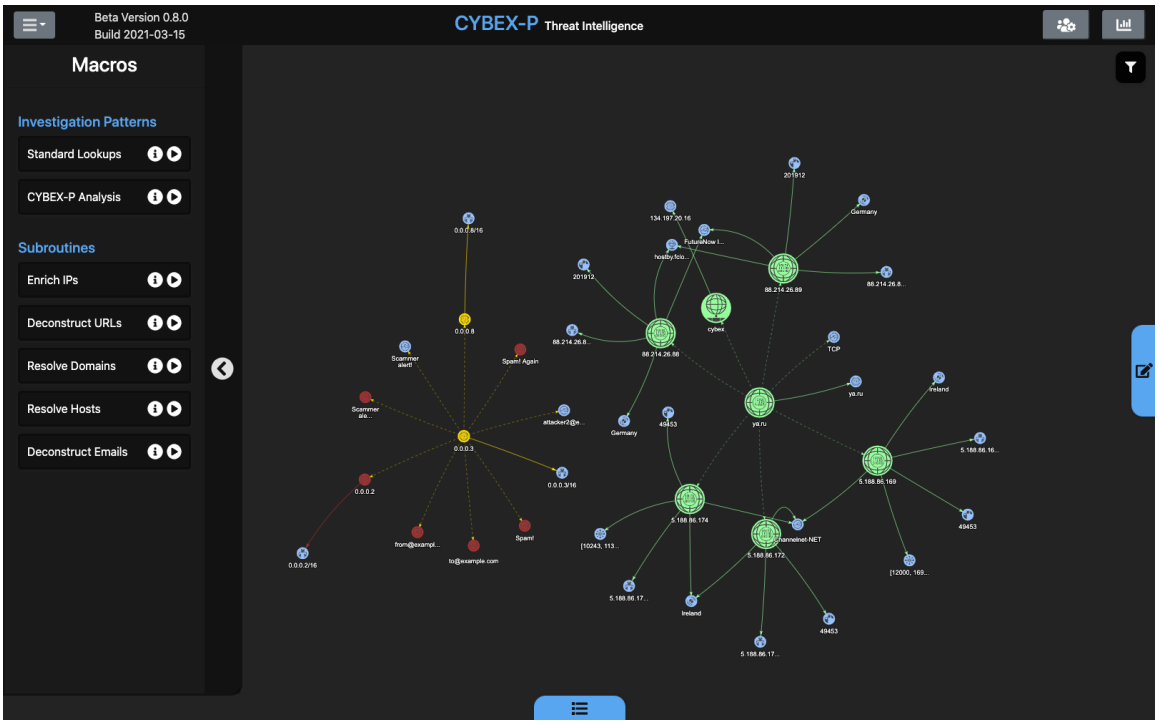


Figure 6.32: The user expands the graph scope by next using the 'Standard Lookups' macro. New data is added that can also be subjected to threat analysis.

The results of this analysis could, at the investigator's discretion, be sufficient to conclude that ya.ru is not a threat. Meanwhile, they can confirm that 0.0.0.3 a threat that should be addressed. The investigator may save this graph and share it

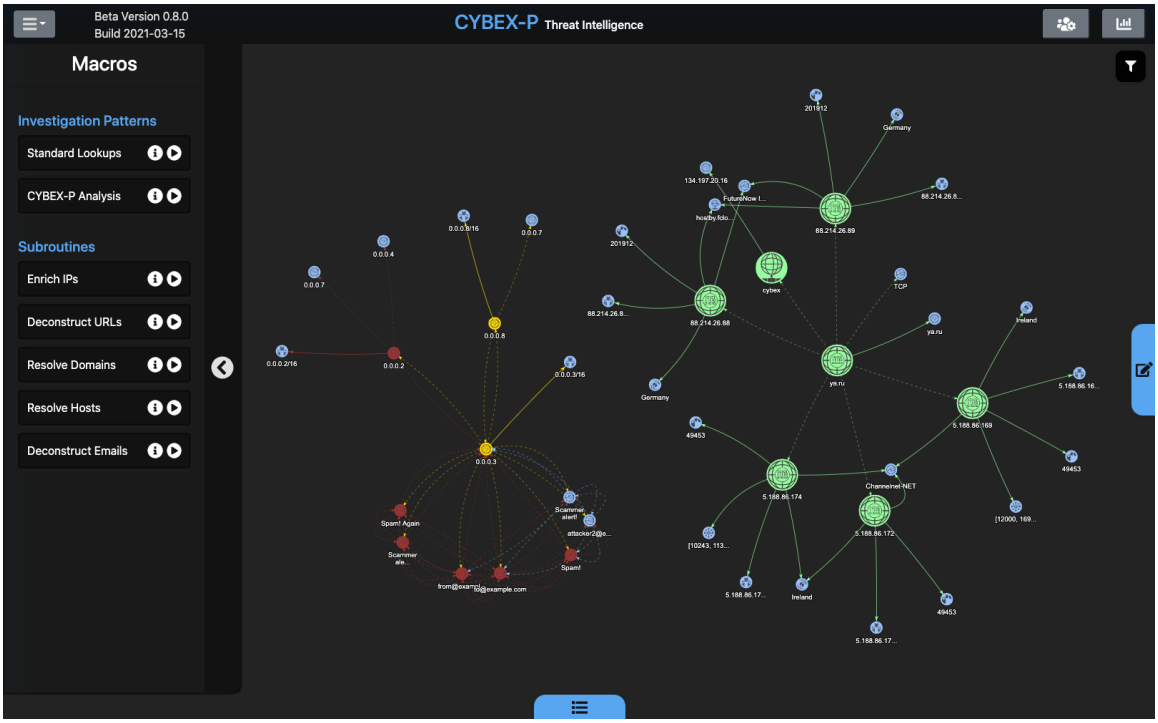


Figure 6.33: The user runs the 'CYBEX Analysis' macro once more. The user can now see the full scope of available threat knowledge and a comprehensive number of important relationships. Each relationship contains on-demand event context that the investigator can selectively inspect.

with their colleagues or superiors so that a threat mitigation plan can be developed. Note that there are two key elements of subjectivity that users of this system must always be aware of. First, the thoroughness of each investigation may vary. In the example in this section, the user ran three total macros (CYBEX-P Analysis twice and Standard Lookups once). In high-stakes situations, these macros may need to be run several more times, in case some malicious relationship exists far at the edges of the initial graph scope. The second consideration is that not all IOCs will be known to the CYBEX-P database. In these cases, neutral classifications are given and the default blue color is retained. Investigators can never assume that neutral items are benign. Instead, they must simply conclude that there is "no evidence of maliciousness". These neutral items should be monitored over time, in case some relevant event data eventually gets submitted to the system.