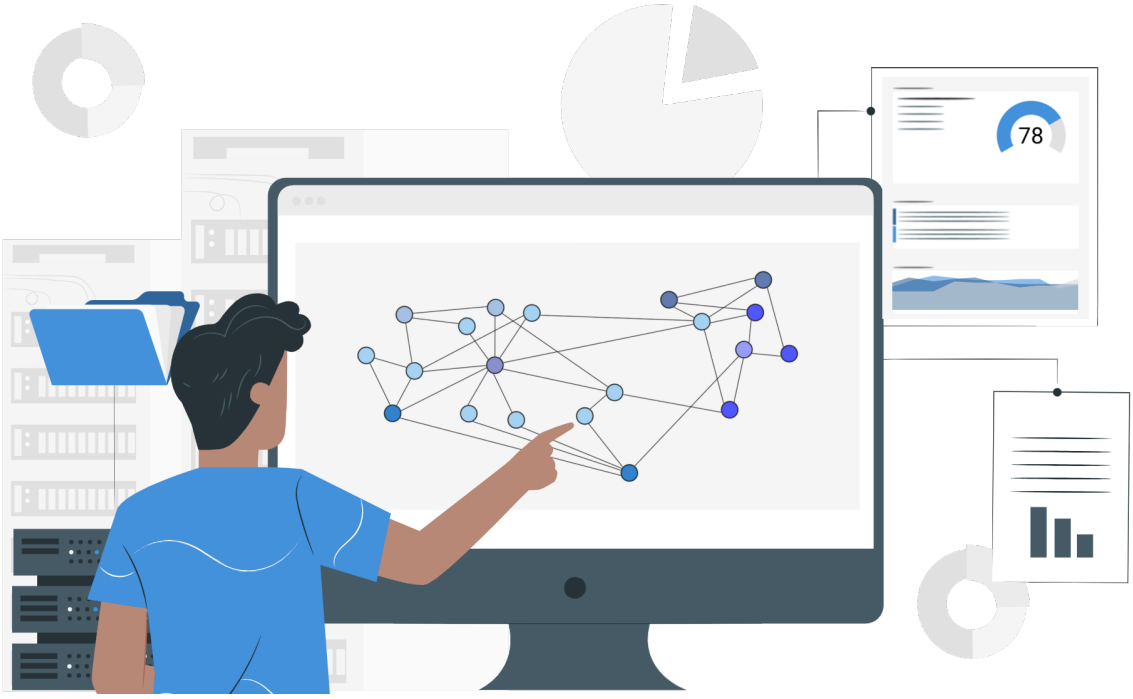
CHALMERS
UNIVERSITY OF TECHNOLOGY



Connecting the dots

Designing an interactive network graph for threat intelligence investigations and research

Master's thesis in Industrial Design Engineering

LUDVIG ANDERSSON & ERIK MARBERG

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

Master's thesis 2021

Connecting the dots

Designing an interactive network graph for threat intelligence investigations and research

LUDVIG ANDERSSON & ERIK MARBERG

Department of Industrial and Materials Science
Division of Design and Human Factors
Chalmers University of Technology
Gothenburg, Sweden 2021

Connecting the dots
Designing an interactive network graph for threat intelligence investigations and research

Department of Industrial and Materials Science
Chalmers University of Technology
SE-412 96 Gothenburg Sweden
Telephone + 46 (0)31-772 1000

Cover: Illustration by Freepik Storyset, https://storyset.com/business edited by Ludvig Andersson & Erik Marberg

Print: Reproservice Chalmers
Gothenburg, Sweden 2021

ABSTRACT

Cyber security and threat intelligence is a constant struggle of trying to stay ahead of threats in order to mitigate risk and disrupt adversaries. Each year billions are spent to ensure this security yet enormous amounts of money still makes it to the hands of Threat Actors. Understanding the surrounding threat landscape and assessing it fast to ensure actionable intelligence is of essence for all businesses and governments.

This thesis aims to design a conceptual solution for interactive, visual representation and exploration of the data and information building up this threat landscape. This to fulfill the purpose of enhancing analyst's and security operation's ability to quickly form actionable intelligence in order to defend their assets.

Interviews were conducted with Threat Intelligence Analysts throughout the project. Initially to form an understanding of the domain and empathizing with their process and needs. This followed by two consecutive Create- and Evaluate phases where the analyst's could express their thoughts about the proposed concepts and solutions.

The result was not only a design concept but also a defined General Use Case from which a set of user requirements were defined. These requirements were used as guidelines for the design concept which was visualized as a set of wireframes. The project resulted in a concept for an interactive network graph allowing users to explore and control large amounts of data in a comprehensible interface. The concept suggests both designs and interactions which will aid Threat Intelligence Analysts when conducting investigations.

Keywords:
network graph, threat intelligence, interaction design, concept development, cyber security

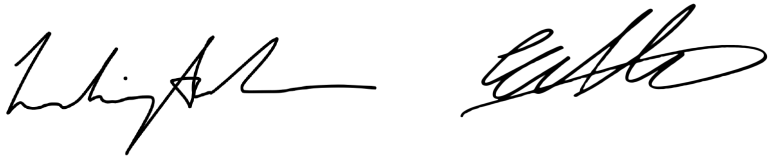
ACKNOWLEDGEMENTS

This Master's thesis project was conducted in the spring of 2021 by two graduate students enrolled in the master's program Industrial Design Engineering and the master's program Interaction Design and Technologies, respectively. The project was conducted in collaboration with the threat intelligence company Recorded Future. We would like to begin this report by expressing our gratitude to all people who have contributed throughout the project.

Firstly, we would like to thank our mentors Hanna Johansson and Peter Erhard from the Product Design team at Recorded Future. We value and appreciate all your input and guidance in helping us to stay on track throughout this project. A big thank you to the rest of the Product Design team as well, for welcoming us and sharing your thoughts, it has been a blast to spend the spring together with you.

We would also like to thank all others from Recorded Future who have helped us with inspiration and bringing this project together. A special thanks to the four threat intelligence analysts from the Insikt group that took their time to participate in our interviews, the project would not have been possible without your contribution.

Moreover, we would like to express our gratitude towards our supervisor from Chalmers University of Technology, Lars-Ola Bligård, not only for your support and concise feedback but also for your patience.



Ludvig Andersson & Erik Marberg, Gothenburg, June 2021

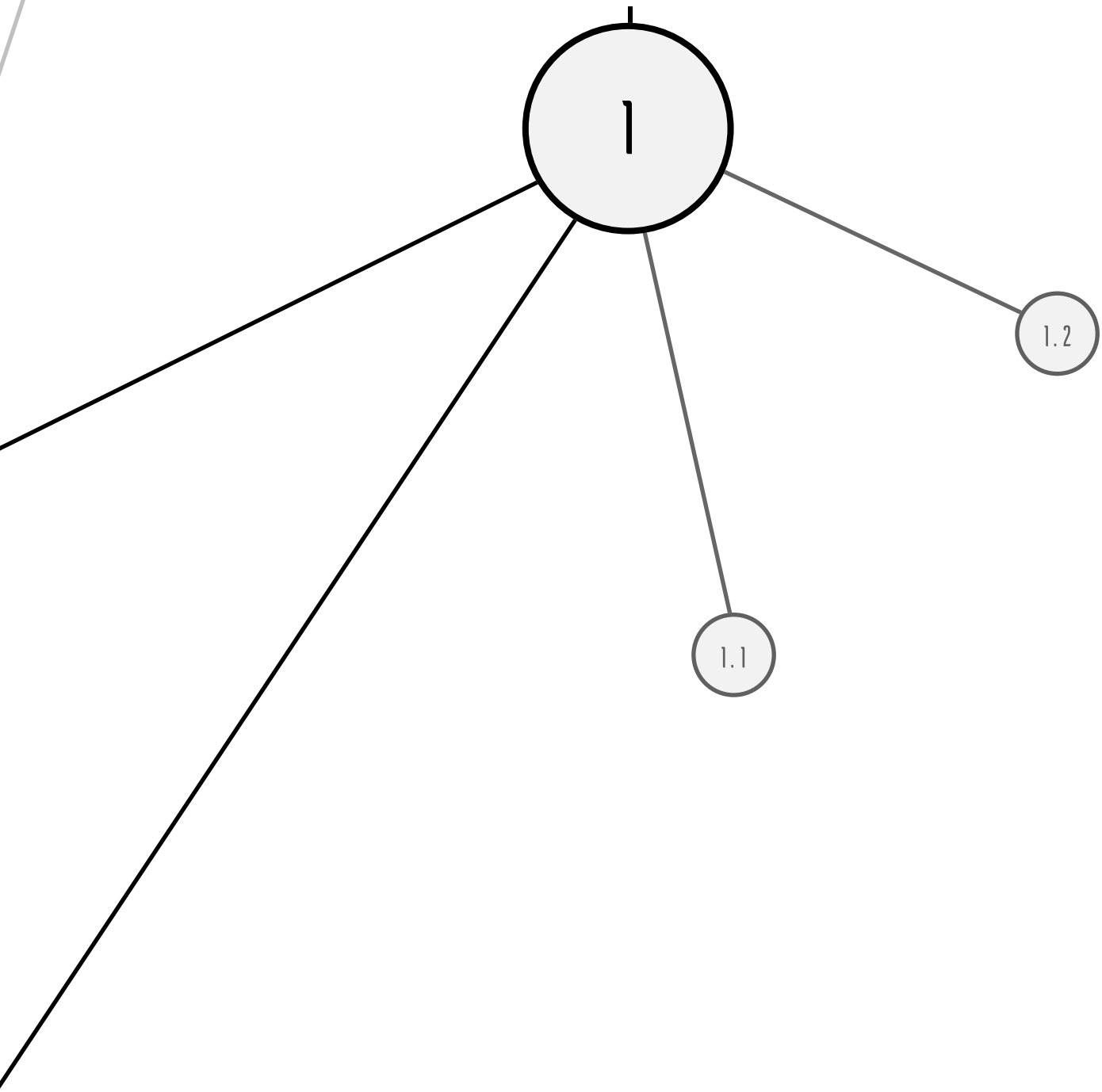
Table of content

Terminology

Advanced Persistent Threat (APT)	A Threat Actor conducting large-scale operations over extended periods of time. They are often, but not necessarily, nation-state sponsored groups and typically they are driven by political or economic motives.
Advanced Query Builder (AQB)	A feature within the Recorded Future portal allowing users to ask for specific information using several customizable inputs and filters.
Connection	Any association between entities found through references within the Recorded Future platform.
Domain	A text-string designed to define and identify a reference to an IP address hosting content online.
Edge	In the context of a network graph, the edge is the object connecting the nodes within the graph, i.e. the line between the dots.
Entity	Any separately identifiable item having its own independent existence.
Indicator of compromise (IOC)	Any indicator, often in the form of an entity or event, that leads security operations to believe it might be posing a risk to one's assets.
Intelligence	Intelligence is formed by interpreting data and information forming an understanding of the bigger picture of complex issues to inform decision making.
Intelligence Card	A component within the Recorded Future platform, each entity has their own Intelligence Card. The Intelligence Card contains information such as the risk score, references, links, trends, connections and much more.
IP address	Internet Protocol (IP) addresses are the identifying schema for IP traffic and are used to assign specific computers or servers to receive traffic. The IP address is of a numerical value.
Links	Within the Recorded Future platform, high confidence connections defined either through internal analyst research reports or some specific automatically generated sources are defined as Links.
Malware	Originating from combining the words malicious software, malware is any software intentionally causing harm to computers, networks or servers.

Network graph	Also known as a node-link diagram is a type of diagram which can indicate complex relationships between a large number of nodes.
Node	In the context of a network graph, the nodes are the objects being connected by the edges of the graph, i.e. the dots connected by lines.
Pivoting	Within threat intelligence research pivoting is the act of searching and moving between different associated entities to find further information.
Ransomware	A type of malware which as the name suggests restricts the victim from accessing their asset to blackmail them for the ransom.
Risk Score	A feature within the Recorded Future platform assigned to certain types of entities. All information about an entity is compared to a set of risk rules, triggering certain rules resolves to a certain Risk Score.
Security Operations Center (SOC)	A part of an organization's security office focusing on detecting, mitigating and disrupting intrusions or other types of incidents.
Tactics, Techniques, and Procedures (TTP)	A composite object of a Threat Actor's common work procedure, summarizing the way they cause harm.
Threat Actor	An individual or group posing a threat towards an organization.
Threat Intelligence Analyst (TIA)	A specific role within security operations focusing on trying to define more information about incidents. This by collecting information, analyzing it and turning it into actionable intelligence for their organization.
Vulnerability	A weakness within a computer system which can be exploited by a Threat Actor.
Wireframe	A mock-up or guide of a graphical interface used to convey its design. Can be of a wide fidelity range, from just boxes and labels representing the layout to fully detailed static images.

INTRODUCTION



1. Introduction

Understanding that the important question is not *if* you will be hacked but *when* is common knowledge within the world of cyber security. This is the reason that governments and companies all over the world spend large sums of money on security infrastructure and employing cyber security teams. The need for cyber security professionals is so large that in 2020 there was a deficit of over 3 million cyber security professionals worldwide (ISC2, 2020). The need to spend these large sums of money can be understood with insights to the risks an organization faces when it comes to cyber-crime. Ransomware is a currently very common method where criminals infect a victim's computers, networks, or servers with malicious code that will lock the victim out of their own system. The Threat Actor then demands a ransom payment to give access back to the victim. Reports claim that one Threat Actor achieved thefts of up to 90 million dollars in a single campaign (Cimpanu, 2021) and Purplesec estimates the annual cost of ransomware attacks to be up to 20 billion dollars in 2020 (Purplesec, 2021). Ransomware attacks even reached the mainstream news when several American hospitals were attacked in 2020 (Salama et al., 2020). While financially motivated attacks, like the previously described ransomware incidents, are common. There are also countless other ways to be attacked and Threat Actors can be motivated by basically anything. To minimize the risks when an organization can't employ enough security staff, the tools available to the security teams are vital.

Recorded Future is offering a Software-as-a-Service solution for aiding security teams worldwide to assess their threat environment and stay ahead of threats. Through their platform, users can access actionable intelligence (Recorded Future, 2021). The data and information building up this intelligence are all contained in the Recorded Future Intelligence Graph. By utilizing a combination of machine learning and human analysts', relationships between different entities and events mentioned online are defined and it is all arranged in the Recorded Future Intelligence Graph. Accessing and interpreting this information, however, can be time-consuming. As time is a crucial pressure point in threat operations, speeding up this process is imperative.

1.1 Purpose & Aim

The purpose of this project is to aid analysts in their investigations by visualizing the relationships and associations between different entities. This increases the analysts' ability to understand and evaluate their current threat landscape. Through this intelligence, they can make more informed and actionable decisions and by that stay ahead of threats and adversaries.

The aim of this project is to create a conceptual design solution for the interactive visualization of the Recorded Future Intelligence Graph. The proposed concept shall identify requirements and key functionalities needed in a solution to aid Threat Intelligence Analysts (TIA) in their

investigations. The main elements of the concept should be presented at a design level consisting of solution sketches or wireframes.
The concept should be based on the current Recorded Future data model and data sources available yet can suggest possible changes or additions if they are deemed key factors for improving the user experience. The design solution should allow users to efficiently interact with the data in order to make quicker, more confident, and informed decisions based on more actionable intelligence.

1.2 Disposition

This report contains eight main chapters excluding the discussion and conclusions. Each chapter will be briefly introduced in this disposition.

1.2.1 Introduction to Cyber security

This chapter will contain information about the domain of cyber security such as important concepts and frameworks. This is meant to introduce the reader to the domain of cyber security and create an understanding for the specific users that the final concept was designed for. While there is far more to cyber security and threat intelligence than can be fit into a chapter like this, the basic understanding required for the project and the contents of this report will be obtained in this and the following chapter.

1.2.2 Project background

This part of the report presents the background to the master thesis project. It includes an introduction to the project itself, an overview of the collaborating company, Recorded Future, and an introduction to their current platform and its most prominent components which formed essential preconditions to the project.

1.2.3 Design theory

Within design theory, there were two main concepts that had a major impact on the project. These were theories regarding how to visualize information and how to strive for usability, both will be described in this section.

1.2.4 Methods

In this chapter, methods used will be briefly defined and described. These methods were used during the project to enable efficient collection of data and reduce the risk of bias. They were also used to facilitate ideation and decision-making during the process. The methods are presented in order of appearance.

1.2.5 Design process

This chapter will describe the process of the project, visualized in figure 1, as well as describing which methods were applied and how, for each phase.

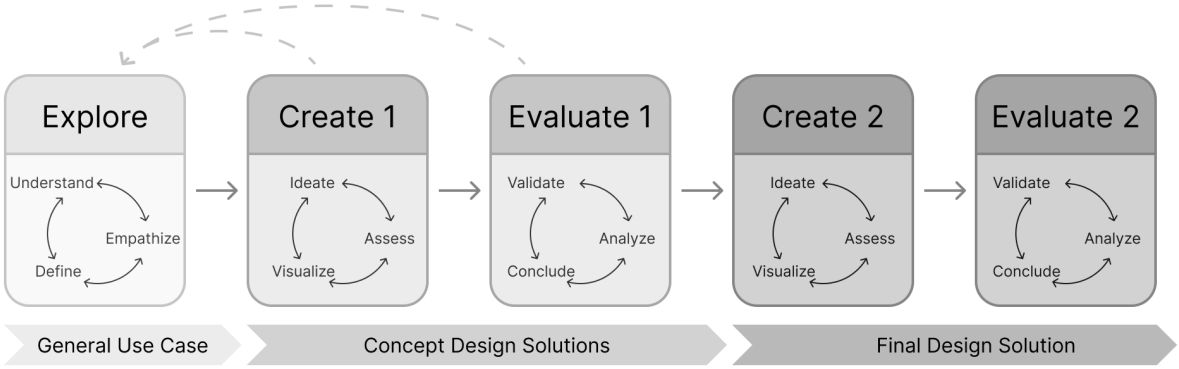


Figure 1: Visualization of the project process

1.2.6 Result: General Use Case

The first result of the project, which then served as an outline for all following phases and methods, were the General Use Case. An initial version of this use case was evaluated during the user interviews conducted during the Explore phase. The General Use Case was then iterated and adapted based on the feedback from the interviews forming a new more clarified version. This version was subject to evaluation during the first Evaluate phase where it was validated and confirmed as a well-defined representation by all interviewed users. The defined General Use Case of a Threat Intelligence Analyst consists of five phases.

1.2.7 Result: Final Design Solutions

Throughout the process and its phases conclusions were drawn, decisions were made, and design solutions were shaped and evaluated, those results will be presented here. This section will describe an overall overview of the design solution and then scoping down to describe specific parts of the solution in detail. The project resulted in a concept of an interactive network graph for threat intelligence investigations and research as seen in figure 2. In this section, the overall functionality of the concept will be presented as well as defining specific key features and how they aid analysts in their work. The features will be described based on where in the interface they are found apart from some functionalities which are seen as composites of the overall concept. The interface has been divided into five main components, each with its own specific functionalities, designs, and ways of interaction.

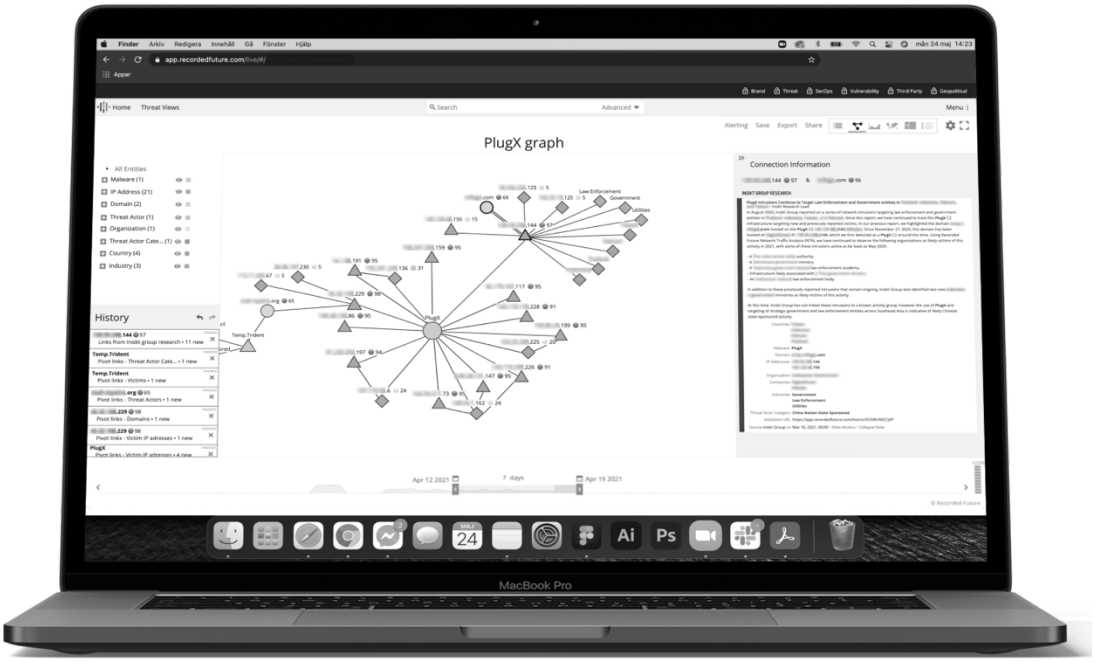
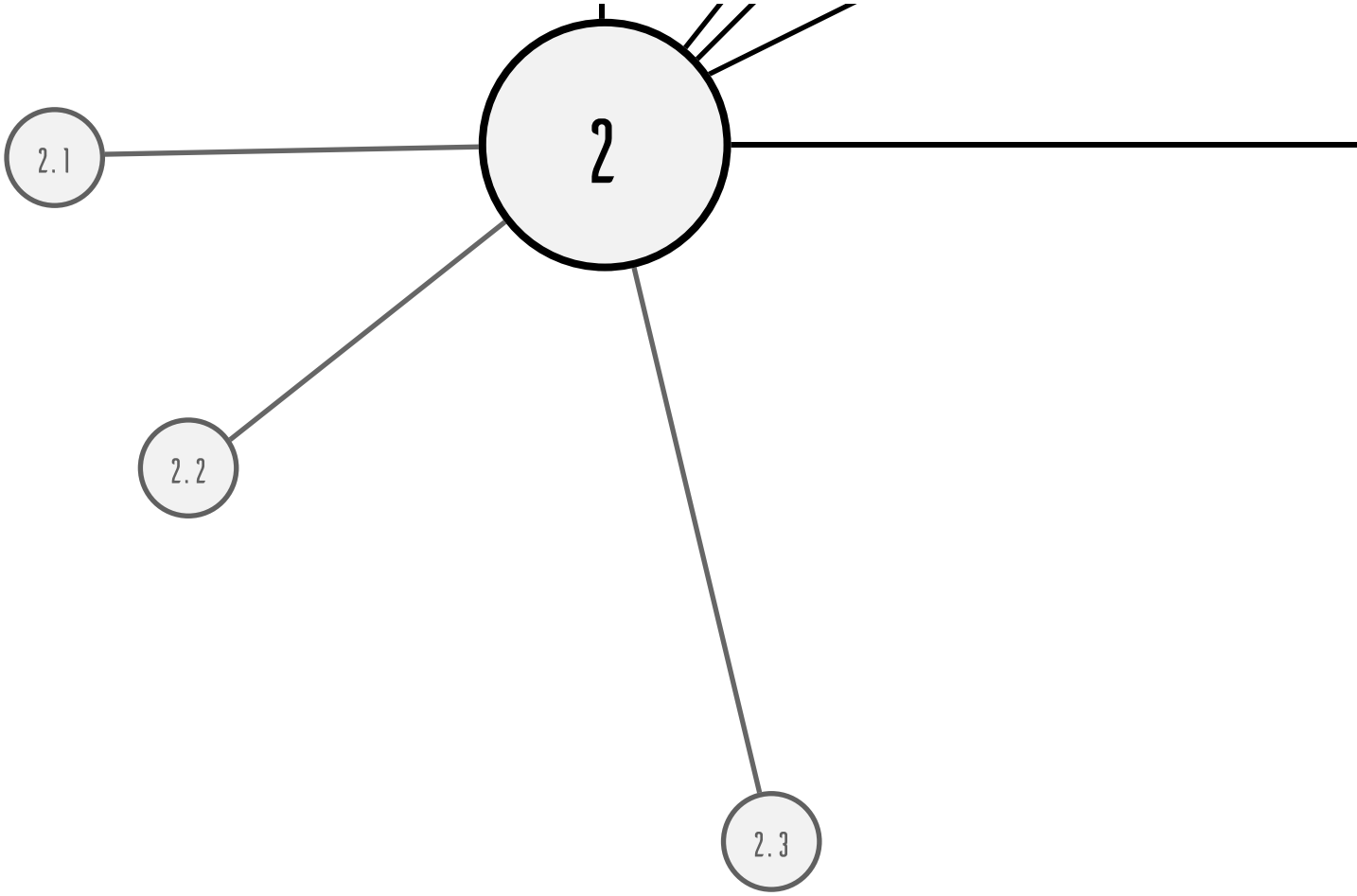


Figure 2: Mockup of the final concept

1.2.8 Recommendations for future iterations

During the project, multiple design suggestions were constructed and evaluated while it is still clear that some features had to be prioritized lower than others. This section provides the design recommendations for future iterations of the Network graph.

INTRODUCTION TO CYBER SECURITY



2. Introduction to Cyber security

This chapter will contain information about the domain of cyber security such as important concepts and frameworks. This is meant to introduce the reader to the domain of cyber security and create an understanding for the specific users that the final concept was designed for. While there is far more to cyber security and threat intelligence than can be fit into a chapter like this, the basic understanding required for the project and the contents of this report will be obtained in this and the following chapter.

2.1 Threat intelligence

> *"Threat intelligence is evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing or emerging menace or hazard to assets that can be used to inform decisions regarding the subject's response to that menace or hazard."*
> *(Gartner, 2013)*

Threat intelligence is a complicated thing, including many parts and it is interpreted to mean different things by different people. The comic book, Threat Intelligence and Me by Robert M. Lee (2017) explains it in a clear manner and this section will follow his structure, starting off by explaining threats.

2.1.1 Risk or Threat?

A threat is a person or group with the intent, opportunity, and capability to cause harm to someone or to organizations (Rynes & Bjornard, 2011). The capabilities of a threat could be the specific malware they use or any skills they have and lastly, the opportunity could be a weakness in a tool used or sloppy security requirements. Things like weak passwords or requiring too difficult passwords, leading to post-it notes with passwords written on them, is an opportunity for a threat. The last thing someone needs to be a threat is the intent to cause harm. If someone only fulfills two out of the three requirements mentioned they will not pose a threat to an organization. A Threat Actor without the opportunity will not be able to cause harm no matter their intent and capabilities.

Closely related to threats are *risks* and for communication, risk can be a more useful tool than threats. The reason risk can be more useful is because it is quantifiable and can be communicated as the risk of monetary loss which everyone understands (Gundert, 2020). Similar to how threats were defined previously the risk can be seen as a sum of the asset one is trying to protect, the threat to that asset, and any vulnerability that can be exploited by that threat to attack the asset (Threat Analysis Group, N/A). In the same way as with threats all three parts need to be present for there to be any risk. With an understanding of potential threats and vulnerabilities related to an asset, one can say that there is a low, moderate, or high risk of the asset being harmed or losing it entirely which will have a monetary cost.

To begin identifying threats and assessing risks, an organization needs to know itself. It needs to know what assets it has to lose and potential opportunities and vulnerabilities concerning those assets. With gained knowledge of the organization and its systems, people, security efforts, and anything internal that can become an opportunity or vulnerability, the organization can then begin to look outward to identify potential threats.

2.1.2 From Data to Information to Intelligence

Every organization has access to excessive amounts of data from its internal systems and processes. This data can be anything from network traffic logs to usage data and it describes one individual, inarguable fact (Recorded Future, 2017). A security team cannot make decisions based on individual pieces of data since one data point could simply be *IP address A* communicated with *IP address B* at time *X*. By combining and processing multiple points of data one can form information that answers a simple question. Using the same example, there might be data indicating that *IP address A* belongs to the company while *IP-address B* belongs to a supplier. When asking "Is it bad that these two are communicating?" with the information available and combination of data, the answer could be that it is not bad. The next step would be to analyze a lot more information together with this and in context to create intelligence to make an assessment. Intelligence is more difficult to obtain than data or information and has a lower quantity of outputs from it while the value of said output is larger. To create valuable intelligence, large amounts of data need to be processed to create information. The process is both difficult and time-consuming for humans to do. This is where computers excel in the threat intelligence world as they are capable of processing the huge volumes of data required. Computers on the other hand are unable to create intelligence (for now) and this is where the human analysts come in with its ability to assess the information.

2.1.3 Intelligence informs action

Previously in the example, it was concluded that it was not bad for the two IP addresses to communicate based on the available information. With more information such as understanding the frequent targeting of companies in the supplier's industry by certain Threat Actors. Combining this with the type of communication between the IPs, matching the habits of a certain malware said Threat Actors are known to use one might come to a different conclusion. This conclusion might not be that the communication actually was bad but that actions need to be taken to stay ahead of possible future threats. This to ensure both the severity as of now, and assessing how this might change in the future. With good intelligence an organization can confidently decide actions, prioritize security efforts, and changes that need to be made within the organization to reduce risk (Recorded Future, 2017). However, before an organization is able to take action on intelligence they need to have a certain maturity. The organization needs to understand its threat landscape and be able to process its own data and create its own intelligence. Another criterion is that the organization empowers its security staff to take action and accept the value of threat intelligence (Lee, 2017).

2.2 The Threat Intelligence Analyst

The Threat Intelligence Analyst is the specific Recorded Future web portal user considered in this project, and if the distinction needs to be made it is the Cyber Threat Intelligence Analyst. The image given to cyber security professionals in popular culture is that they sit in a room with big screens, massive data flows and maps with arbitrary lines, trying to counter-hack the hackers. The first part with alerts and massive amounts of flowing data fits the description of working in a Security Operations Center (SOC) (De Groot, 2020). For an organization, the SOC is there to detect and stop incidents. They work at a high pace, triage alerts, and make quick decisions but this is usually not where the TIA is found. Somewhat depending on the organization as the SOC looks different in every organization based on its needs. When something gets escalated from the SOC, for example suspicious traffic or that a novel malware has been found and there is a need for further investigation, then the TIA steps in.

2.2.1 The work of a Threat Intelligence Analyst

The first of the two main tasks of a Threat Intelligence Analyst is to collect information, analyze it and turn it into actionable intelligence for their organization to use. This could be any kind of intelligence, from analyzing a vulnerability in a system that recently was exploited in the wild to trend research concerning Threat Actors targeting their industry. The second task is to communicate this intelligence to the right recipients. How they communicate the result of their research is highly dependent on who the intended receiver is and what kind of decision they should make as a result of the intelligence. When the recipient is intended to implement the actions suggested they might need a highly detailed and technical form of communication. On the other hand, when communicating with management the focus could be shifted to risk or adapted to a level that they are expected to understand.

An investigation is usually conducted using multiple tools and strategies, from sandboxing malware to further understand it, to investigating suspicious IP addresses that communicate with internal systems (Check Point, N/D). The experienced TIA has a large bank of knowledge of where to look for the type of information that is necessary for their current investigation.

2.2.2 Pivoting in the context of a TIA

When conducting investigations a common analytic technique is pivoting. Pivoting is the act of exploiting a data element, often in the form of an indicator of compromise (IOC), to locate other elements and related connections. It could, for example, be the act of investigating an IP address to discover domains resolving to it. Pivoting often consists of an iterative workflow of discovering new information followed by discarding irrelevant or untrusted information. Through these iterations, the analyst searches for information that might move their investigation forward and that eventually can sum up to conclusions and actions to mitigate the threats (Slowik, 2020)(Caltagirone et al., 2013).

2.3 Analytical Frameworks

There are several analytical frameworks commonly used within threat intelligence. These frameworks provide a structured way of thinking about attacks and adversaries as well as a common grammar and syntax to be used when communicating within and between teams and organizations. The thesis will present three of these frameworks as they were considered to have aided in the process, especially in understanding the context of the users, their thought processes, and workflows.

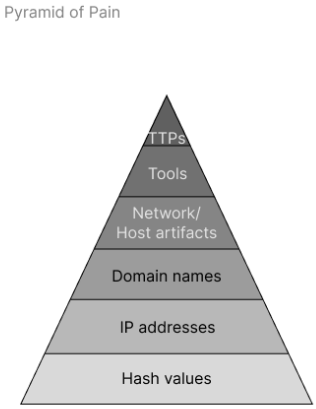


Figure 3: Recreation of the pyramid of pain (Bianco, 2013)

2.3.1 Pyramid of pain

The Pyramid of Pain (figure 3) is a concept developed by David Bianco which illustrates the relationship between different types of indicators (Bianco, 2013). The concept showcases indicators at different levels of the pyramid where indicators at a higher level will cause the adversaries more pain if denied from them. The contents of the lower tiers are simpler for the targets to identify and block but they are also easier for the adversary to change. On the higher tiers, the contents are more difficult to identify but are more powerful for the victim as it makes attacks easier to identify, mitigate, and disrupt. The tactics, techniques and procedures (TTPs) and tools an Advanced Persistent Threat (APT) uses are composites of objects such as habits and skills which are more difficult and time-consuming for the adversary to change.

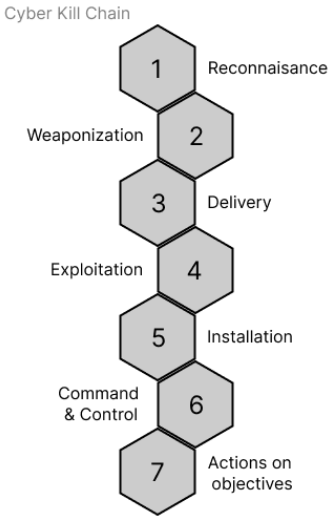


Figure 4: Recreation of the Cyber Kill Chain (Hutchins et al., N/D)

2.3.2 Cyber Kill Chain

The Cyber Kill Chain (figure 4) by Lockheed Martin (Hutchins et al., N/D), breaks down the structure of an attack into seven stages. The framework theory relies on the fact that if you as a defender can disrupt the chain at any stage the attack cannot move on to the further stages (MacGregor, 2015). The simplicity of the framework makes it easy to digest and apply but it is also limited as it does not take into account that many modern attacks might skip stages or perform multiple stages at the same time (The Recorded Future Team, 2020).

2.3.3 Diamond Model

The Diamond Model (figure 5), published by the Centre for Cyber Threat Intelligence and Threat Research (Caltagirone et al., 2013), focuses on adversaries and the development of their TTPs over a longer time period rather than focusing on the progress of one single attack event. The Diamond Model consists of four features; Adversary, Infrastructure, Capability, and Victim. The model consists of two axes, the socio-political axis with Adversary and Victim on its ends, and the technical axis containing Infrastructure and Capability. As described by (Caltagirone et al., 2013) "In its simplest form, the model describes that an *adversary* deploys a *capability* over some *infrastructure* against a *victim.*" Keeping track of these core features and how they evolve over time helps analysts to understand the adversaries and threats and develop defense strategies that can disrupt these adversaries (MacGregor, 2015) (The Recorded Future Team, 2020).

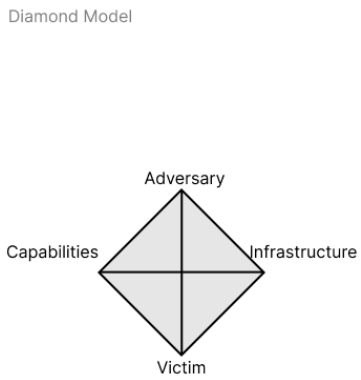
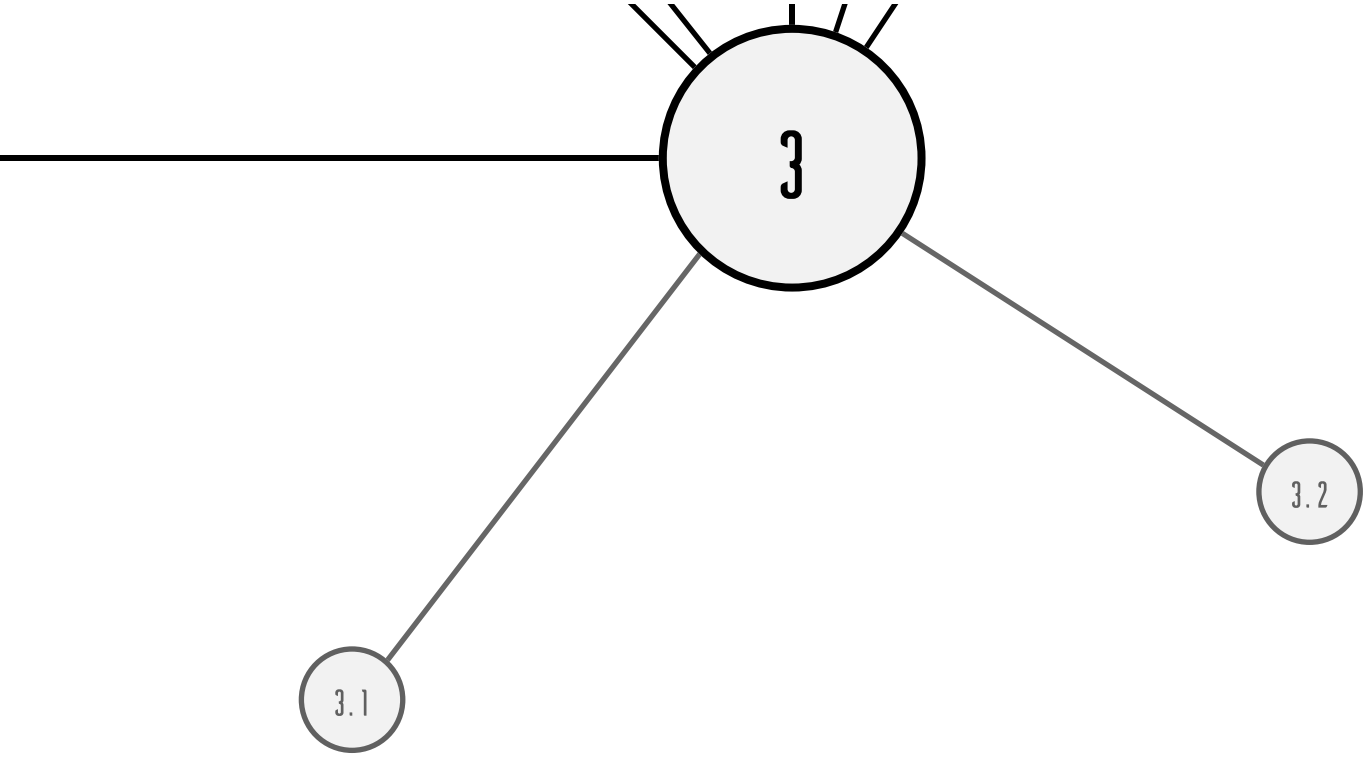


Figure 5: The Diamond model (Caltagirone et al., 2013)

This introduction was meant to create a basic understanding of Cyber Security, Threat Intelligence, and the work of a Threat Intelligence Analyst. The following chapter will contain a background to the project and introduce the collaborating company, Recorded Future, and its portal which formed the outline of the project.

PROJECT BACKGROUND



3. Project background

This part of the report presents the background to the master thesis project. It includes an introduction to the project itself, an overview of the collaborating company, Recorded Future, and an introduction to their current platform and its most prominent components which formed essential preconditions to the project.

3.1 The project - Visualizing the Recorded Future Intelligence Graph

This project revolves around the Recorded Future Intelligence Graph, as of now the Recorded Future Intelligence Graph is the core of the Recorded Future platform containing all data and information which then is interpreted and presented as intelligence to the user through the existing user interface (UI). In the existing UI of the Recorded Future portal, the intelligence is presented in various views but the graph itself is inaccessible to the users. This project will explore the possibilities of allowing the user to interact in a more direct way with the Recorded Future Intelligence Graph. It will explore ways of presenting the Recorded Future Intelligence Graph and how this can be made in an intuitive and accessible manner allowing the user to explore and interact with the data. Visualizing and allowing direct interaction with the Recorded Future Intelligence Graph will raise Intelligence by unraveling answers to questions beyond the users' questions and will provide them with a visual representation of the associations between different points of interest. Through visual representation, the exploration of complex data will lead to more meaningful and actionable insights (Shneiderman, 2014). This will provide value to a wide variety of work areas both within and outside of security departments, spanning from management to roles such as those within a Security Operations Center, especially to Threat Intelligence Analysts.

3.2 The collaborating company - Recorded Future

Recorded Future describes itself as “The Threat Intelligence Company”. They were founded in 2009 but filed their first patent in 2007. Recorded Future was founded by Christopher Ahlberg and Staffan Truvé, both Ph.Ds. in computer science from Chalmers University of Technology in Gothenburg, Sweden. The company currently has offices in seven locations with its headquarters located near Boston, Massachusetts in the United States.

Recorded Future provides a threat intelligence platform helping security teams worldwide to reduce risks and stay ahead of threats. In order to do this, Recorded Future utilizes current and historic data to define relationships between different entities and events mentioned online. Using machine-learning for harvesting data in real-time, not only from open sources but also from closed forums and the so-called deep and dark web. Entities such as IP addresses, domains, vulnerabilities, and Threat Actors as well as the relationships between them form a threat-oriented digital twin of the world, represented by the Recorded Future Security

Intelligence Graph. The users can get access to information about past, current, and upcoming events with relations to their organization that is mentioned online. The combination of algorithms and human analysts leads to actionable threat intelligence which helps users make quicker and more confident decisions to reduce risks and mitigate threats (Porkorny et al., 2019).

3.2.1 The Recorded Future Platform and Intelligence Graph

The Recorded Future Security Intelligence Platform is based on the continuously evolving Recorded Future Intelligence Graph and the intelligence can be accessed through multiple interaction points such as a web portal, a browser extension, a mobile application, and different types of APIs and integrations. The intelligence is arranged in different modules, each with its own area of interest within security intelligence, for example; Brand Intelligence, Threat Intelligence, and Vulnerability Intelligence (Recorded Future, 2021). This project focuses on the web portal and will not be tied to a specific module.

Within the platform there are certain features that have been of high interest, forming a solid baseline for this project. The Recorded Future Intelligence Graph is a combination of two sub-graphs; an ontology graph and an event graph. Different entities can therefore be associated with one another through events, but they can also be associated through ontological connections. For example; the fact that Paris is located in France forms an ontological connection through their geographical dependency rather than needing to be mentioned in a certain event. In general, the ontological graph has a higher degree of confidence since it covers information that changes slower and less frequently. Apart from geographical dependencies the ontology graph also covers relationships between entities such as ownership, technical dependencies, and industry categories.

The second part of the Recorded Future Intelligence Graph is the events graph, in this subgraph, each event is established by one or multiple references mentioning the same activity, discovered through any of the sources Recorded Future scours. This part of the graph changes continuously since it represents all the references of activities happening online. Within each of these events, there are one or multiple entities who are part of or mentioned, and by that associated with said event. The Recorded Future Intelligence Graph contains over a billion entities connected through over 70 billion references, forming over 100 billion relationships. All of the data building up this event-part of the Recorded Future Intelligence Graph is harvested in real-time by machine learning and classified through natural language processing (NLP). The collected data is processed through NLP to combine and interpret the data into information. This allows the information to be automatically classified and categorized in the Recorded Future Intelligence Graph based on its origin, content, and context.

3.2.2 Crucial Components in the Recorded Future Platform

In the platform, one major component is the Intelligence Cards. All entities in the Recorded Future Intelligence Graph have their own Intelligence Card as presented in figure 6 granting access to all information Recorded Future knows about that certain entity. It contains

information such as; references mentioning the entity, events it has been part of, and associations to other entities through said references or events. All this information about an entity is compared to a set of rules which results in a Risk Score for that certain entity. Triggering one or more rules resolves to a certain Risk Score based on the severity of those rules. The Risk Score is a live object and is updated continuously in real-time, based on the collected information, where different rules have different time-spans for triggering. The Risk Score, as the name implies, is an indicator of the risk tied to that entity and whether it is categorized as suspicious, malicious, or similar. The risk score is merely an indicator and whether the entity in question currently poses a threat towards the user's assets is not communicated through this value. To find events and the entities connected to them the portal contains the Advanced Query Builder (AQB). It allows highly customizable queries with options to limit the time frame, what sources to include and many more.

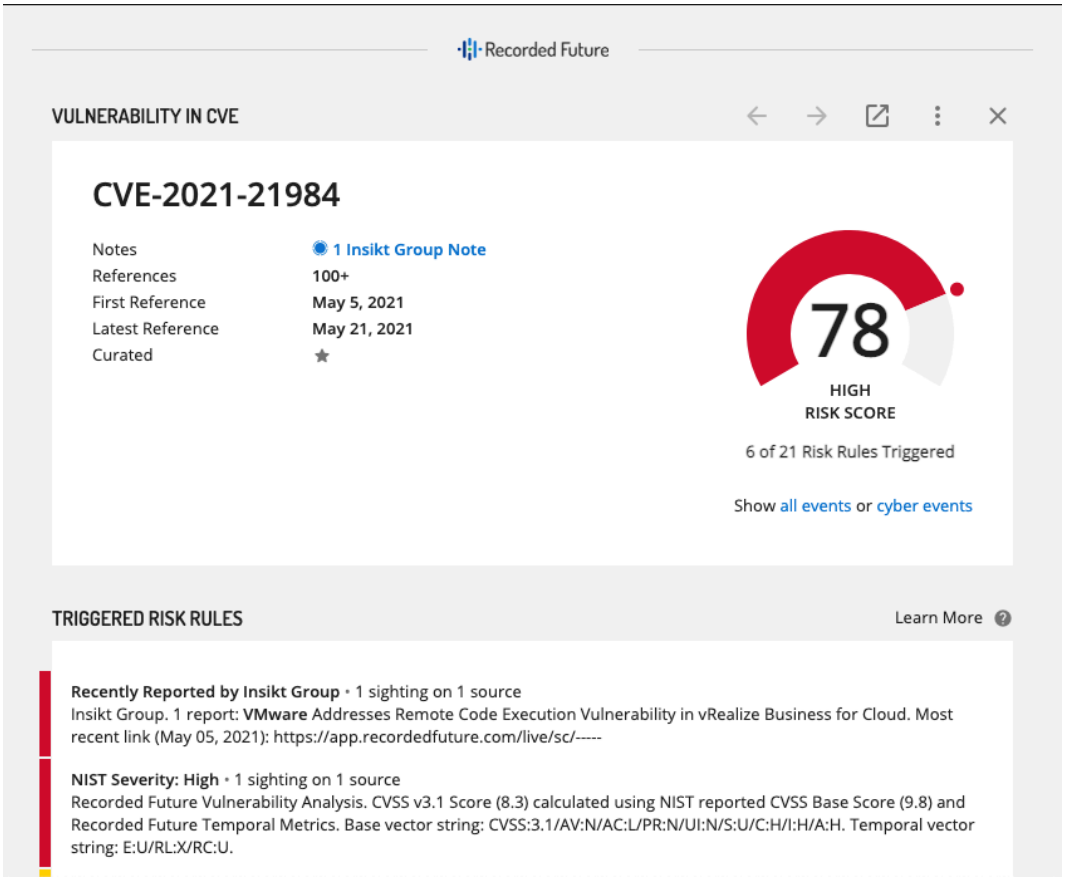
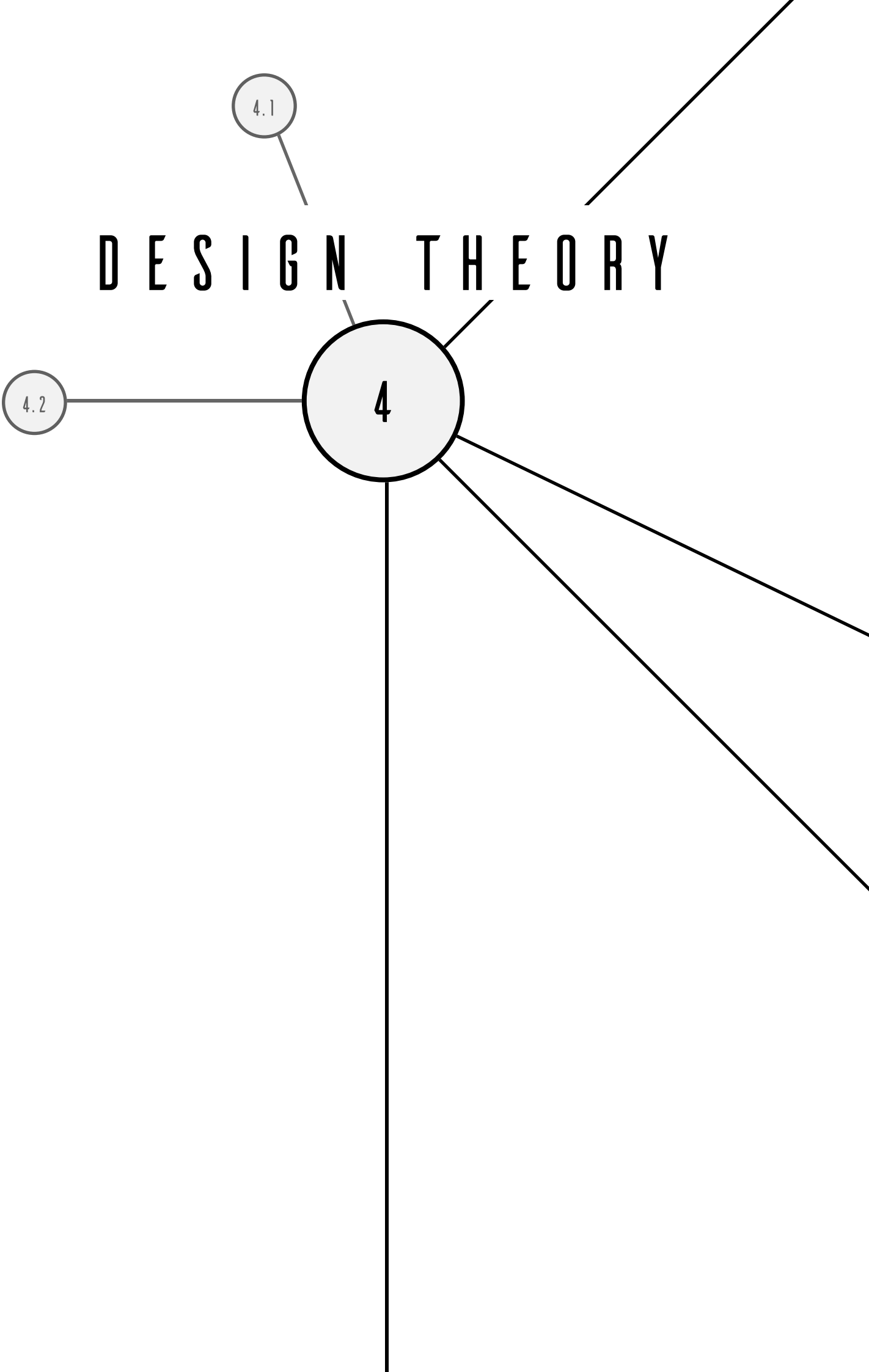


Figure 6: An Intelligence Card showing the vulnerability CVE-2021-21984

Collecting data from a large number of sources with a wide range spanning from the open and dark web, to technical sources as well as internally written analyst reports, leads to huge amounts of information with a large variety in confidence. For example, an analyst report from Insikt, Recorded Future's internal analyst team, is more trusted and has a higher confidence level than information found in, for example, a tweet or other open-source web data. Recorded Future has distinguished the difference between these sorts of associations by introducing what they refer to as *Links*. Links are defined as associations with a high confidence level versus connections that have a lower confidence level in the platform. In reality, the confidence level is not binary, i.e. confident or unconfident, but rather a span.

These formed the preconditions of the project and set the scope for the following phases. By combining this with the design theory and methodology presented in the upcoming sections, the project foundation was set up.


4.1
DESIGN THEORY
4.2
4

4. Design theory

Within design theory, there were two main concepts that had a major impact on the project. These were theories regarding how to visualize information and how to strive for usability, both will be described in this section.

4.1 Visualizing Information

As this project will focus on visualizing large quantities of data and information and designing interactions to manipulate said information, there are some design theories that are applicable. This section will focus on the human perception of information visualizations and how design can aid that process.

4.1.1 Visual Salience

Firstly, let us describe and define the difference between symbols and glyphs, two elements widely used in graphical data visualizations. A symbol is a graphical element used to represent a certain object in a specific context. While a symbol only represents the object itself, a glyph is used to represent both an object as well as certain attributes tied to that entity. Common ways of representing these numerical attributes are for example; size, vertical position, or contrast (Ware, 2013) (Dunne & Shneiderman, 2013).

When visualizing large amounts of data and information it is important for the user to be able to distinguish between different types of information. This can be accomplished by encoding elements (Nowell et al., N/A). Encoding can be achieved in numerous ways but some of these are processed differently compared to each other. This as they are said to belong to different channels, these are; color, shape, and motion. That these are processed separately means that encoding using one channel will not constrain encoding using any of the other channels which can be useful when designing information containing multiple attributes (Ware, 2013).

By utilizing redundant coding, the design of symbols or glyphs that are maximally distinctive can be achieved. Redundant encoding is achieved by making the symbol distinct in multiple channels, allowing the user to search based on any of the properties attached to each symbol (Nowell et al., N/A)(Ware, 2013). In contrast to redundant encoding there is conjunction coding, this is when symbols are encoded using multiple feature dimensions but they are not specific for each type of symbol. Instead, conjunction encoding requires the user to perform a sequence of searches, one for each featured property to find the correct symbol. This type of encoding will therefore complicate searches and make symbols less distinctive compared to using single property encoding or redundant encoding (Ware, 2013). See example in figure 7.

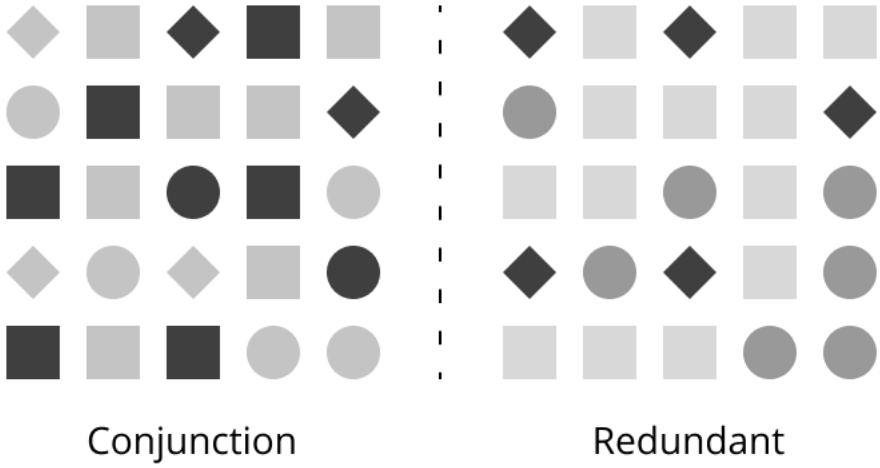


Figure 7: Image showing the difference between Conjunction and Redundance when encoding

One concept regarding the distinctiveness in these sorts of encoding is preattentive processing which describes how the human mind intuitively can focus its attention and distinguish between different visual objects. The concept of preattentive processing is important since it can radically decrease the processing time. This is especially the case when there are a lot of distractors, in other words, objects that are not the intended focus for the user, which otherwise increases the response time. There are multiple ways of achieving preattentive distinct features, for example by altering; size, color, orientation, contrast, or motion (Treisman, 1985) (Ware, 2013).

4.1.2 Interacting with visualizations

By performing actions and interacting with interfaces users can affect their current situation and which information currently is presented to them through the interface. Kirsh and Maglio (1994) distinguish between two types of actions; epistemic and pragmatic actions. Epistemic actions are described as activities leading the user towards discovering new information, to improve their cognition by making their mental computation simpler. A pragmatic action, on the other hand, intends to lead the user closer towards their physical goal. This distinction results in the conclusion that actions can be both epistemic and pragmatic since a specific activity can alter both the physical and informational state of an object. The distinction of whether an action is to be seen as epistemic or pragmatic lies in the underlying goal of the activity and defining whether the goal is to change the physical or the informational state.

By designing interactive visualizations and allowing epistemic actions, interfaces can grant user's opportunities to explore the data that seems most important or interesting to them. An extensively cited mantra proposed by Ben Shneiderman (1998) for these information-seeking activities states; "Overview first, zoom and filter, then details on demand". Allowing these types of interactions stated in the mantra in an iterative manner rather than in that specific order will allow users a continuous workflow of unraveling information and exploring different parts of the interface as they please. To design an effective interface for analytics of large quantities

of information, allowing the user to focus their attention will be just as important as it is for a camera to be able to focus on specific objects when taking a picture (Shneiderman & Plaisant, 2015). Ideally, all visualized data points in the interface should be interactive and allow the user to explore or manipulate them further (Ware, 2013).

4.2 Usability Heuristics

Well-renowned author Jakob Nielsen published his "10 Usability Heuristics" (1994) defining general guidelines for interaction design. These are well-known and commonly used and cited within the field of usability and user experience. Some of these heuristics have been considered and used in the design process of this project and the ones with the most impact will be described in more detail below.

4.2.1 Visibility of System Status (Heuristic #1)

This heuristic describes the importance of feedback and system communication to keep users informed of the current state, allowing users to retain their feeling of control. Each interaction between user and interface must be clearly communicated and feedback must be presented within a reasonable time to keep the user in the loop. Not providing this transparency of the system status will degrade the trust for the system (Harley, 2018). By providing a design with clear communication of the system state the interface can provide good opportunities for the user to overcome "*the gulf of evaluation*", allowing successful interaction with the system (Whitenton, 2018).

4.2.2 User Control and Freedom (Heuristic #3)

The third of Nielsen's heuristics revolves around the supporting functionality of allowing the user to cancel or undo actions they might have taken by mistake or taken unwillingly (Rosala, 2020). Users should always have the ability to opt-out of actions they've begun or fulfilled to quickly recover from mistakes or slips. Common controls for these types of actions found in UIs are *back*, *cancel*, *close* and *undo* which all, in different ways, lets the user abandon an action and return to the previous system state.

4.2.3 Consistency and Standards (Heuristic #4)

Following both internal and external standards is the focus of the fourth heuristic. By following conventions the user is relieved from wondering what a certain word or action means. This aids users as it makes their preconceived mental models accurate and useful even when experiencing a completely new interface. Consistency can be divided into internal and external consistency and both should be accounted for to form an intuitive interface that is both easy to learn and to use. Internal consistency regards having consistency within a product, this means having similar colors, layouts, and interactive elements throughout the product. External consistency is achieved by following common standards for the type of platform or industry as a whole. By following these patterns and conventions, both internal and external, user's expectations are met and higher learnability is achieved (Krause, 2021) (Cooper et al., 2014).

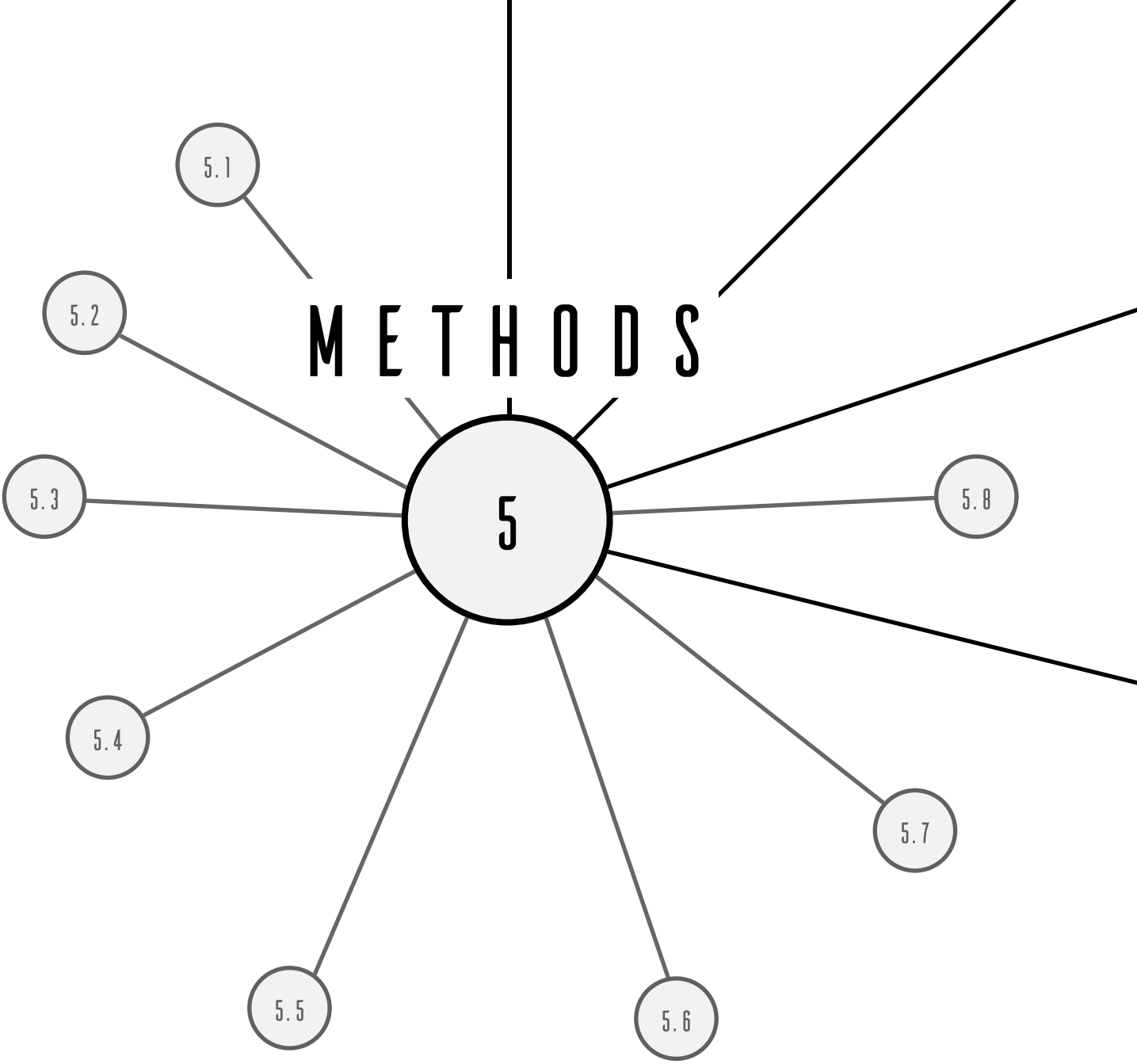
4.2.4 Error Prevention (Heuristic #5)

The fifth heuristic, Error prevention, advises that interfaces should not only allow users to recover from errors, as in the third heuristic, but it should also preferably try to prevent them from happening in the first place. Nielsen argues that the term “user error” actually focuses attention poorly since it implies that it is the user who is to blame for the problem. Instead, the blame should be directed towards the designer and the interface since it is allowing these errors to be committed. There are different types of errors that can occur and preventing them also differs but by providing suggestions, constraints, and warnings, the risks of committing errors can be reduced. Nielsen suggests that preventing errors with high costs should be prioritized since these are harder to recover from (Laubheimer, 2015)(Cooper et al., 2014).

4.2.5 Recognition rather than Recall (Heuristic #6)

This heuristic implies that users should not be restricted to their short-term memory and having to remember things from one part of the interface to another to interact effectively (Budiu, 2014). Providing additional context allows for smoother interaction and reduces cognitive load since they allow the user to rely on recognition rather than having to recall information from their memory from scratch. By providing appropriate associations and cues the interface can guide the mind of the users to effectively recognize the correct chunk of information.

These design theories combined with the earlier presented project background, were applied together through the methods presented next. Through this a solid foundation for decision making and designing of the concept solutions was defined, and suitable methods could be chosen and applied.


METHODS
5
5.1
5.2
5.3
5.4
5.5
5.6
5.7
5.8

5. Methods

In this chapter, methods used will be briefly defined and described. These methods were used during the project to enable efficient collection of data and reduce the risk of bias. They were also used to facilitate ideation and decision-making during the process. The methods are presented in order of appearance.

5.1 Interviews

Interviews in general can be described as a conversation between an interviewer and one or more interviewees focusing on the interviewees' thoughts or experiences of a certain matter. This methodology is commonly used to gain qualitative insights within the user-centered design in order to empathize and form a better understanding of the potential user and their workflows or thought processes. Interviews can be conducted in person or over distance using various mediums such as phone calls, digital video-conferencing tools, or even through written conversations. When conducting interviews with people located in different parts of the world these video-conferencing tools are very useful as they are a close representation of meeting in person utilizing both video and audio to communicate. These video-conferencing tools also provide opportunities for both interviewer and interviewee to share their screen content allowing them to show visual elements to clarify statements regarding specific visuals. Interviews can be conducted in multiple formats of which two will be described below (Wood, 1997).

5.1.1 Semi-structured

Performing a so-called semi-structured interview relies on using an interview template to guide the interview. The template contains a set of questions that will be investigated and discussed during the interview. In order for the interview to be semi-structured, it is important that the template contains open-ended questions, as well as, both allowing and encouraging follow-up questions. This encourages discussion to form an even deeper understanding of the interviewee's thoughts and feelings (Sharp et al., 2019).

5.1.2 Unstructured

An unstructured interview is less structured in the way that it does not follow a predetermined template or set of questions. Instead, it might revolve around a specific topic and let both interviewer and interviewees lead the conversation in any direction that seems appropriate or interesting. This type of unstructured interview can be useful when the interviewer is uncertain of what to ask and just has a curiosity to learn more about anything within a specific topic (Sharp et al., 2019).

5.2 Solution Sketches

A solution sketch is used to clearly visualize an idea with more details, preferably it should show multiple frames or states to indicate its functionalities. The idea to be visualized through

solution sketching usually originates from an earlier ideation method or session but solution sketching also allows for new ideas or the combination of previous ideas. Allowing iterations of previous sketches can be helpful as ideas of new, even more, useful solutions might emerge (Google LLC, N/A).

5.3 Brainstorming

Brainstorming focuses on widening the idea space and raising the number of possible solutions (Sharp et al., 2019). Brainstorming can be conducted both as an individual or as a group activity and is most commonly used in the early to mid-stages of a project. Chauncy Wilson (2013) describes three principles of brainstorming; aim for quantity over quality, abandon criticism, and stimulate out-of-the-box thinking. The goal of a brainstorming session should be to produce a large number of ideas or solutions. Ideas generated can then be used to spark further ideas by synthesizing or altering them to solve other problems. This manipulation of ideas can be utilized either within the same session or in the following iteration. To allow all participants to think freely and not hinder themselves, no type of criticism should be expressed, neither in words or in body language. After a session has been finished, ideas can be reviewed by a critique session, voting, or something similar to establish a ranking or decision regarding how to proceed (Wilson, 2013).

5.4 Braindrawing

Much similar to the above-presented brainstorming is braindrawing, another method for idea generation and problem-solving. Braindrawing utilizes the power of rough sketches to illustrate and convey ideas and solutions. Braindrawing can be especially useful when focusing on ideating for graphical interfaces or layouts. The procedure of a braindrawing session is very similar to that of brainstorming, firstly stating the problem or topic to iterate upon. This is followed by an iterative process of quick, rough sketching and passing said sketches to be continued by another participant. After some rounds of sketching the ideate session can be concluded by reviewing and discussing the solutions to decide how to move forward (Wilson, 2013).

5.5 MoSCoW

To prioritize features, the MoSCoW rules can be applied. The acronym of MoSCoW represents “Must haves”, “Should haves”, “Could haves”, and “Want to have but will not have this time round” (DSDM Consortium, 2003).

- The “Must have”-features represents the minimal viable product, without them the system will be unfunctional or useless.
- The “Should have”-features are not required to build a functional product but it represents features which are deemed important to add value to users.

- “Could have”-features could also be valuable to users and nice to have if they do not require too much time or effort to implement. These features will be the first to be left out if the project’s time frame will not suffice.
- “Want to have but will not have this time round” also referred to as “Won’t have but Would like to have”-features are the ideas that have been expressed but can be pushed to a future development phase.

Arranging and prioritizing features according to these rules may seem simple but it is important to have a sufficient amount of features within each category. The lower prioritized features can be useful to add some contingency to the project both regarding the scope and its timeframe, as they can be altered, added, or removed in order to comply with the project plan (DSDM Consortium, 2003) (Waters, 2009).

5.6 Now-Wow-How

Another framework for prioritizing features or decision-making is the Now-Wow-How framework. The method is focusing on two requirements, the degree of innovation and the difficulty of implementing. These two are placed on two perpendicular axes, the horizontal for representing implementation difficulty and the vertical representing innovation degree. The axes form a 2 by 2 matrix, the bottom left will represent the “Now”-ideas, as they are both easy to implement and considered less innovative, suggesting they can be used as-is. The bottom right quadrant represents the “Wow!”-ideas, they are innovative yet easy to implement. Lastly, the top right represents the “How?”-ideas, these ideas are considered innovative but tough or uncertain how to possibly implement them. The top left corner ideas represent the ideas that are hard to implement but with a low amount of innovation and are therefore considered less useful for the project. All ideas or features are placed within the matrix to be able to distinguish their characteristics and to prioritize between them before moving on with the project (Friis Dam & Yu Siang, 2021).

5.7 Morphological Matrix

This is a method focusing on defining key functions of a task, exploring possible solutions to these, and finally combining the various solutions to form full concepts to be evaluated. First, a set of key functionalities are defined. Then a number of possible solutions achieving the needs of that functionality are generated and added to the matrix. This step of creating possible solutions is then repeated for all key functions building up a matrix. From this matrix, different solutions can then be combined freely into concepts that should fulfill all needs and functionalities defined. The concepts can then be evaluated and as the solutions can be combined in various ways a lot of possibilities can be explored (Fargnoli et al., 2006) (Treffinger et al., 2006).

5.8 Wireframing

When designing graphical user interfaces wireframes are used to quickly visualize the functionality and flow of interactions. Wireframing can be drawn both using pen and paper as well as using digital tools or software. The wireframes should clearly communicate the layout of the interface and its interactions and navigation. As with sketching they can be used at a variety of fidelities based on the use case. All in all a wireframe on its own is a static image that at a higher fidelity represents a realistic look into an interface. Combining several wireframes in a sequence with arrows or another type of narrative can indicate the flow between different parts of the interface (Interaction Design Foundation, N/A).

While this is a theoretical description of the methods most were adjusted to be applied as seen most appropriate during the process of the project. How the methods were applied and how they affected the project will be described in the Design Process chapter.

DESIGN PROCESS

6

6.1

6.2

6.3

6.4

6.5

6. Design Process

This chapter will describe the process of the project, visualized in figure 8, as well as describing which methods were applied and how, for each phase.

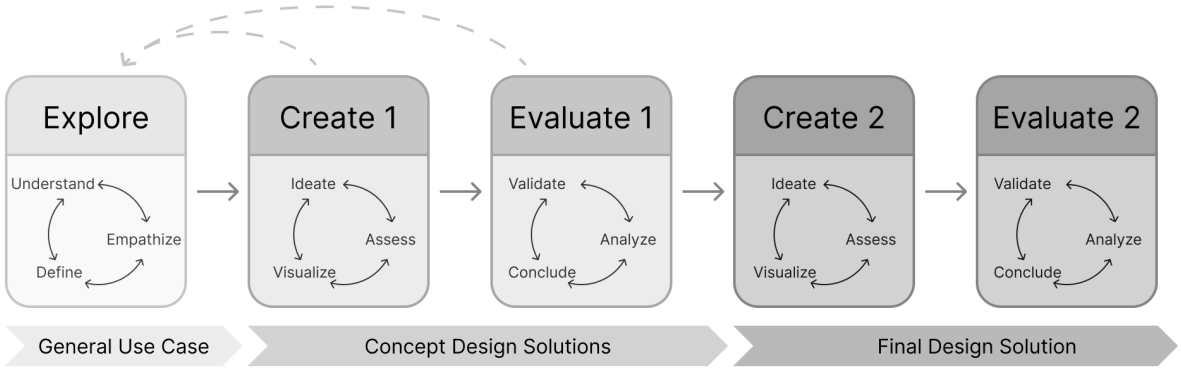


Figure 8: A visualization of the project process

The overall process for the project was planned based on two existing process models; the Double Diamond model and the Design Thinking model. The double diamond model consists of two major phases, each divided into two subphases. The first phase is named Research and consists of the subphases Discover and Define. Discover revolves around understanding the problem by investigating user needs and requirements. Then, in the Define subphase, these insights are gathered and boiled down to a concluded problem definition which forms the foundation of the second phase, the Design phase. First in the Design phase comes the Develop subphase which explores and ideates around the problem definition to come up with possible solutions. The second subphase of Design, last in the whole iterative process is Deliver which includes testing and evaluating solutions with users (British Design Council, 2005).

The Design Thinking model consists of three phases; Understand, Explore, and Materialize, each consisting of two subphases (Gibbons, 2016). The first phase, Understand, involves the subphases Empathize and Define. These involve investigating user needs and defining them and by that, they have a clear similarity to the Research phase of the Double Diamond model. The second phase, Explore, consists of ideate and Prototype. This is where the two models start to differ, as the Design Thinking model has divided this into two subphases while in the Diamond Model it is represented by one, the Development subphase. In the same manner, the last phase of Design Thinking called Materialize consists of two subphases, Test and Implement, which has a clear similarity to the Deliver subphase from the Double Diamond. There are similarities between these models and they both served as inspiration when shaping the process for this thesis project.

This project's design process consisted of three main phases; Explore, Create and Evaluate. The phases are described as separate, chronological parts but in some sense, they partly overlap and there are also iterations between them. Each phase had its specific goals and the deliverables of one phase would affect and be used in the succeeding phases. In the same manner, as in the two models used as inspiration, the thesis process starts with exploring.

Initiating the project focusing on forming an understanding of the domain, and empathizing with its users. This first phase accounts for about half of the overall project time. This as the domain is completely new to both project members and it is considered necessary to form a good and solid foundation and understanding before proceeding to the following phases. From the exploring phase and the gained knowledge, needs and requirements were defined. Continuing with a creative phase by ideating, assessing, and visualizing possible ideas and features. The visualizations would then be used as mediating objects in the evaluations with users conducted in the last phase. The evaluation focuses on assessing suggested solutions, analyzing the outcomes, and lastly drawing conclusions regarding how to proceed. As mentioned the project process consists of three main phases, each with internal iterations within them, while the project as a whole also iterates on the last two phases, conducting five phases in total. Meaning the conclusions from the first evaluation phase will be used as a foundation in a second Create phase followed by a second evaluative phase before forming and concluding the final result. By following this process, gained knowledge and understanding can be incorporated throughout the process, and through several iterations within and between phases the ideas are allowed to grow and adjust according to the results from the evaluations.

6.1 Explore

The initial phase of the project, the Explore phase, and all its internal subphases revolved around three themes; understanding the domain, understanding the users and their current workflows, and lastly, understanding the current tools used in general and specifically the Recorded Future platform. As all these subthemes, more or less, are dependent on each other, this exploration was an iterative process just as the upcoming phases. It is important to note that the majority of the time during the project was spent in this phase. This resulted in the decision to focus on Threat Intelligence Analysts, a General Use Case and a set of requirements for the design solution.

6.1.1 Understand

The thesis project began by trying to understand the domain of cyber security and threat intelligence in general. This learning period consisted mainly of consuming articles, literature, and blogs of different kinds, focusing on understanding what threat intelligence is, which different types of applications of threat intelligence there are, as well as, how they are used to disrupt or mitigate threats. Since this focused mainly on forming a basic understanding and a solid foundation for upcoming activities, most time was spent reading different security vendor blogs as these often explained the subject in a way suitable for beginners in the domain. This way, a domain-specific vocabulary started to form and step-by-step different perspectives of threat intelligence and its most common subjects were explored.

In parallel with this overview of the domain, the Recorded Future portal was explored, using an internal learning module with introductory videos to learn the basic functionalities of the portal. A more unstructured individual exploration of the portal was also conducted with the aim of trying to get a better understanding of the possibilities and restrictions of the portal and

the underlying data model. To gain an even better understanding of the Recorded Future Intelligence Graph and the underlying data model, a one-hour-long interview was conducted with the CTO and a Senior Architect. A semi-structured interview template, leaning towards unstructured rather than structured, was prepared. This with the aim of trying to allow the interviewees to lead the conversation, as they were the ones with expert knowledge of the domain. The interview was managed and structured enough to stay focused on the predetermined areas of focus, the possibilities, and restrictions of the data model. While it still let the interviewees elaborate on the subjects within those areas as they pleased and thought were most interesting from their perspective.

The company provided two previously recorded interviews or presentations which had been held with employees from other teams with the aim of explaining the different workflows and responsibilities of different roles. One of the recordings presented an introduction to what the Insikt team, Recorded Future's internal research team, does, and the other an introduction to what a SOC team does. As these were just recordings of meetings conducted in the past by the product design team, they provided answers to some concerns but did not allow further questions or exploration. The recordings together with the reading gave an initial understanding of the users but there were still some areas that remained unclear.

To complement this user-centered perspective on the portal, a one-hour-long online meeting was held with a Threat Intelligence Analyst from the Insikt group. During this meeting, the analyst held a presentation that provided an overview of the workflow used by that specific team within Insikt. A compressed version of one of their previous investigations served as a baseline and showed step-by-step tasks performed both within and outside of the Recorded Future portal, combining different tools and resources. Rather than just watching recordings or reading blogs or literature, this actual meeting with a user gave the possibility to ask follow-up questions. Both questions that arose during the meeting but also concerns or questions formed during reading or from the recorded video interviews were brought up and answered. This resulted in a greater understanding regarding the specific tasks and thought processes behind different actions that could be clarified.

All these insights from reading, watching recordings, and holding meetings could then be compared and contrasted to shape and define an initial version of a generally applicable use case, seen in figure 9.

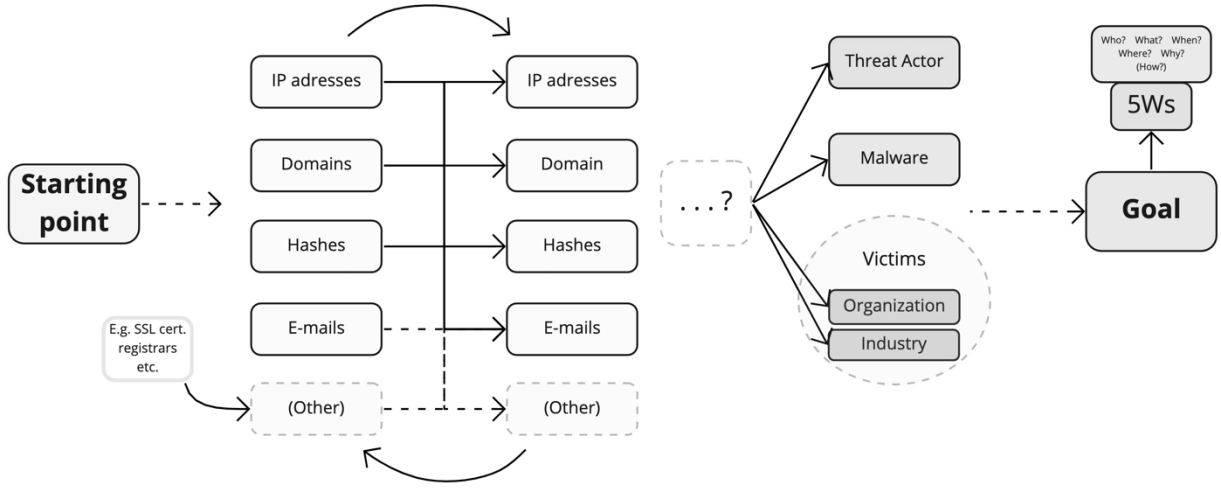


Figure 9: The first iteration of the General Use Case as shown to users

6.1.2 Empathize

In order to gain further knowledge about users and their workflows, a set of four interviews were conducted with four different Threat Intelligence Analysts from Insikt. One of these interviewees was the same as from the meeting which was held previously during the Understand subphase. All interviews followed a semi-structured interview template, they lasted about an hour and were conducted online. The interviews were recorded to reduce the risk of losing information and to allow transcriptions afterward. Each interview had one project member leading the interview based on the prepared template and the other project member focused on listening, taking notes, and helping with asking follow-up questions. The project members took turns acting out the two roles equally throughout all interviews conducted during the entire project.

These four interviews focused on the analyst's workflows when researching or investigating certain events, IOCs, or trends. The aim was to define a more grounded definition of a General Use Case and to form a broader understanding of the analysts' motivations and processes when researching. As all of these interviews were conducted with personnel from an internal team they were all considered experts in using the Recorded Future portal. Their experience both in using the portal and as analysts in general varied, but they all had at least one year of experience with working with Insikt. During the interviews, they were encouraged to describe, and if possible also show, other tools they use regularly and how these aid them, explaining specific functionalities they use and why they choose these over the Recorded Future portal. The users were also asked to evaluate the General Use Case formulated during the understand subphase, based on the literature, recorded interviews, and the Analyst's presentation. This to either confirm, or suggest possible adjustments and improvements to it to provide a better representation and generalization of their overall workflows.

Throughout this exploration, ideas that emerged were written down or sketched to not lose them in the process. One of these ideas (figure 10) functioned as a mediating object presented at the end of these four interviews. This initial solution sketch, of very low fidelity, was

presented to show the current intended direction of the project and evaluate whether the users saw any value or interest in such a solution. The visualization was a very quick, rough paper sketch just giving some hints of a possible solution but still in a clear enough way to convey the ideas of the overall functionality to the users.

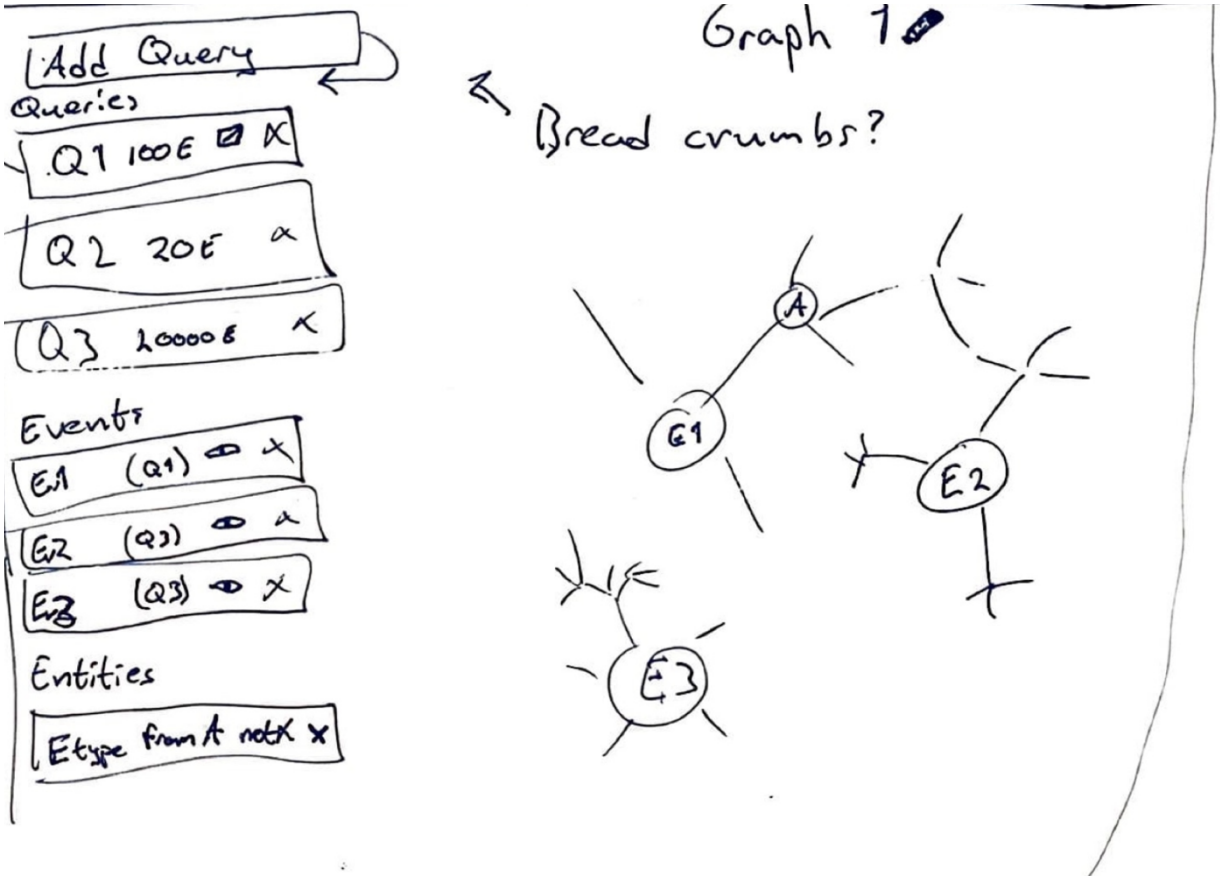


Figure 10: The initial sketch shown to users

6.1.3 Define

From these initial subphases within the Understand phase, there were some definitions suggested that would become the foundation for the upcoming Create and Evaluate phases. To narrow down the scope of the project it was decided to focus on one specific user group rather than trying to aim for a generally applicable solution across different threat intelligence applications. The targeted user group was narrowed down to focus specifically on Threat Intelligence Analysts based on multiple factors. The main factor in this decision was to form a clear problem definition with a defined corresponding use case and applicable hypothetical solution. This type of solution was considered to be most useful in investigating and doing research that has a longer timeframe than a use case focusing on responding to alerts. The decision of the target group was also supported by the factor of having access to users to interview, in which Threat Intelligence Analysts from the Insikt group showed interest in being part of the project.

Based on the feedback from the user interviews the General Use Case was iterated and adjusted. In the initial version, there were some gaps in the chain since it was considered unclear what type of actions lead the use case onwards. These gaps could be filled or removed by the lessons

learned from the interview and a new, adjusted version of the General Use Case was prepared. This would be used as an outline when ideating and designing the solutions and would be subject to another validation in the interviews during the first evaluation phase.

These decisions regarding target user and use case combined led to the definition of a set of user needs and requirements for such a solution. It was defined that the solution should help analysts define all or parts of the *5Ws* of a certain risk by allowing pivoting indicators and visualizing relationships between entities and events found through said pivoting. It should provide a continuous workflow to allow clearer tracking of indicators and associations over time.
The solution should indicate differences in importance between indicators and connections when investigating to speed up the workflow. The solution should also guide user actions without inhibiting their individual preferences and workflows.

List of requirements to fulfill the needs of analysts in reaching the aims of an investigation:

I. Allow pivoting between indicators
II. Visualize indicators
III. Visualize associations
IV. Provide a continuous workflow
V. Indicate differences in importance
VI. Guide user actions without inhibiting them
VII. Seamless experience between graph and other portal features
VIII. Ability to discard and filter information
IX. Provide options to share investigations

These requirements are presented without internal prioritization since all are considered equally important for the final solution.

6.2 Create 1

At this stage of the project, the user needs and requirements were processed in order to form functionalities that would solve these needs. The produced features were discussed and evaluated from a standpoint regarding what key features are needed for the concept to function. With that standpoint in mind, three separate concepts were defined, created, and visualized based on different variants of features that were deemed to be key features for achieving and fulfilling the most crucial user needs. These concepts were then to be tested during the Evaluate 1 phase.

6.2.1 Ideate

In this initial subphase, the identified user needs and requirements were transformed through brainstorming into features fulfilling these needs in the design solution. Each requirement functioned as individual starting points for ideation, and through a combination of

brainstorming and braindrawing one or several functionalities solving each need was established (figure 11). Each of the features was then defined more clearly to clarify and avoid misconceptions regarding their functionalities. Some were combined as they were considered to be very similar or since the same feature could solve multiple requirements. This ideation session took about three hours where both project members started by writing or sketching their individual ideas regarding each need. After all the individual ideas had been put on paper, the tables were switched to be able to iterate on each other's ideas. This sparked new ideas both within and between needs or features by using them as inspiration or by combining different ideas.

Figure 11: Braindrawing sketches

6.2.2 Assess

After the formulation of ideas and features, the MoSCoW-model (figure 12) was performed by adding each specific feature from the ideation to a sticky note and then having a board with four quadrants, one for each category in the MoSCoW-model. The sticky notes could then be placed in the appropriate category and after a few iterations allowing re-categorization, the result stabilized and formed a consensus. The MoSCoW-model focuses solely on the need for certain features for the system to function. As it does not take usability or difficulty of

implementation into account and by that, it was decided to evaluate the features using another framework with a different focus as well.

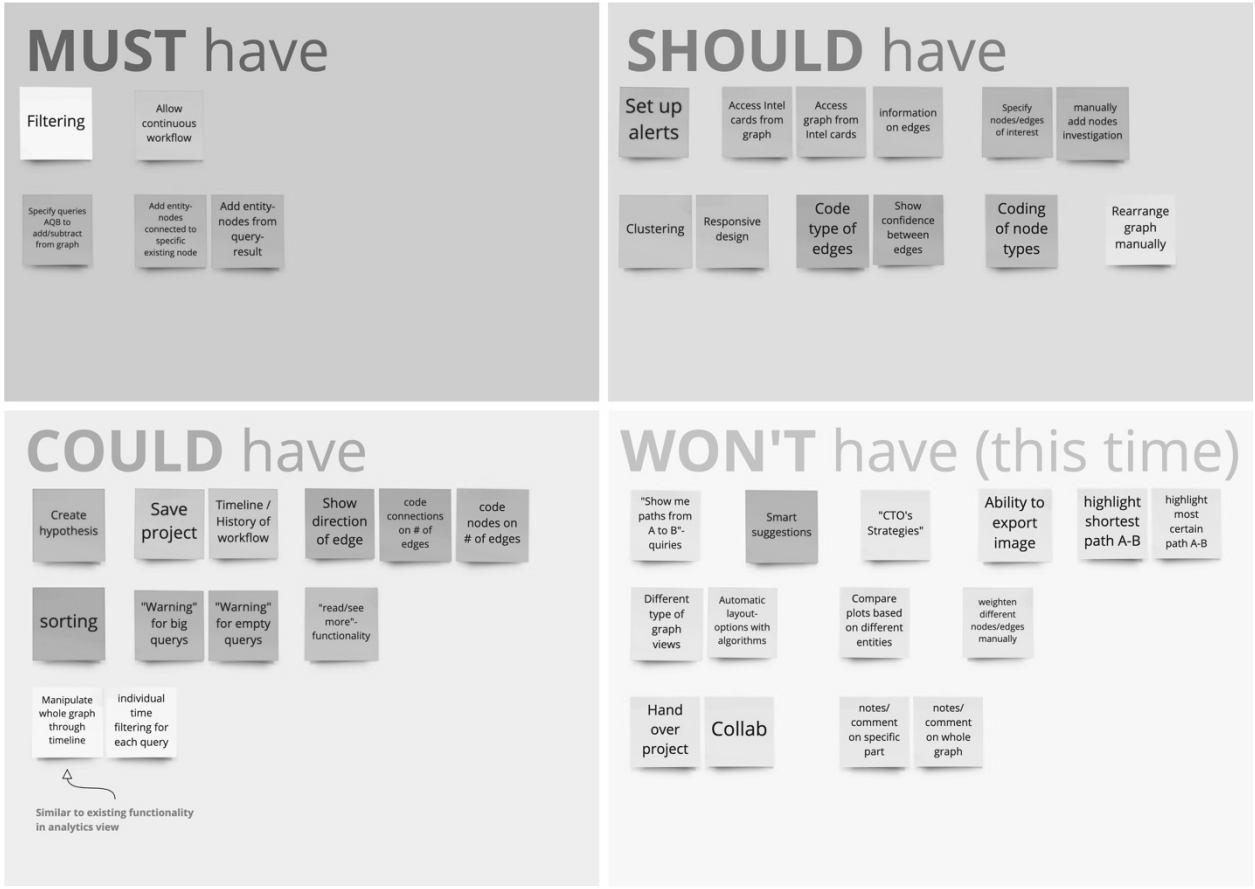


Figure 12: The MoSCoW model from Create 1

The second evaluation of features utilized a sort of decision matrix called the Now-Wow-How (figure 13) framework. This framework is used to define and compare the level of innovation and difficulty of implementation between different ideas. Similar to the previous method the features were listed in another set of individual sticky notes and then categorized on a large board with two axes, with one representing innovation and the other representing implementation. After discussion and some iterations, this result formed a consensus regarding the placement and categorization of each feature.

Figure 13: The Now-How-Wow framework from create one

By combining the results of the MoSCoW-model and the Now-Wow-How Framework a prioritization of the functionalities was defined. This prioritization resulted in four specific functionalities that were considered crucial for the design solution and therefore would be the focus of the last steps in this initial create-phase. The four functionalities, chosen to be in focus where;

A. Visualizing events containing several associated entities.
B. Handling of the AQB within the Network graph.
C. Manual pivoting between entities.
D. Access to underlying entity information.

Functionality A incorporated ideas aiming to solve requirements II and III while functionality B focused on requirement VII. Requirements I, IV, and VI were the focus of functionality C, and lastly functionality D aimed to fulfill requirements V and VII. In accordance with this requirement, VIII and IX were not solved by any functionality and were instead decided to be tackled in later iterations.

6.2.3 Visualize

By brainstorming ideas for each of the four most crucial functionalities and sketching possible solutions, a range of possibilities was generated. By defining the possible solutions of each

functionality A-D and combining them using a Morphological Matrix, seen in figure 14, three concepts were formed with different combinations of the four functionalities. Some additional design choices needed to be made regarding some other, lower prioritized, features of the concepts. Since these weren't the focus of the upcoming evaluation phase, a single solution was simply chosen and applied to all three concepts to maintain consistency and keep the focus on the higher prioritized functionalities. When applicable, the choices were based on either design theory, insights from the user research, or a combination. These lower prioritized design choices were also defined in a separate column of the Morphological Matrix to have a clear and consistent plan to follow when wireframing the composite concepts.

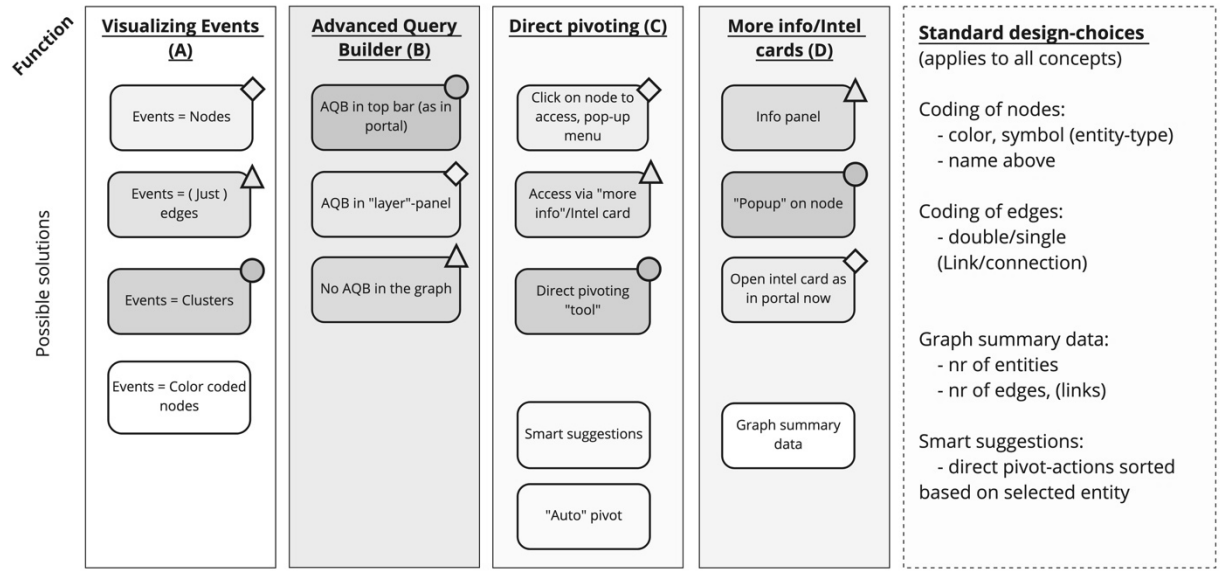


Figure 14: The Morphological matrix from Create 1

The three concepts referred to as concept blue, green, and orange (circle, diamond and triangle) were wireframed using Miro, an online whiteboard platform allowing collaborative work. To use colors when referring to the concepts was decided since calling them numbers or letters was thought to risk affecting how users would perceive them. It was considered that the evaluations might get affected since that sort of naming could imply a predetermined priority or rating, which was not the case, and therefore the use of colors was chosen to avoid this sort of misconception or biasing.

The wireframe designs were constructed with low fidelity since they sought to focus on the functionality and interaction of the features rather than the look and feel. The level of fidelity was sought to be the same for all three concepts to allow fair comparisons and avoid having users prefer one concept based on its fidelity level rather than the actual functionalities. Since the design choice regarding the encoding of nodes was defined as a lower priority feature, all nodes could be prepared beforehand and then simply copied and pasted to be re-used in each of the concepts. The concepts were prepared individually by the project members, before peer-reviewing them to unveil and correct any misconceptions that occurred from the transition between the morphological matrix to the wireframes (Appendix 1).

6.2.4 Hackathon

In parallel to this phase, the collaborating company arranged a 48-hour hackathon for its employees which aligned timely with the end of this project phase. One idea to be explored during the hackathon was to try to visualize the Recorded Future Intelligence Graph, in other words, something more or less similar to this thesis project. This was considered a great opportunity to receive more viewpoints and feedback and perhaps even getting somewhat of a functional prototype produced. In the end, the hackathon project in itself had little impact on the final results of the thesis project but some of the resulting wireframes from the hackathon came to be used in the interviews in the following phase (figure 15).

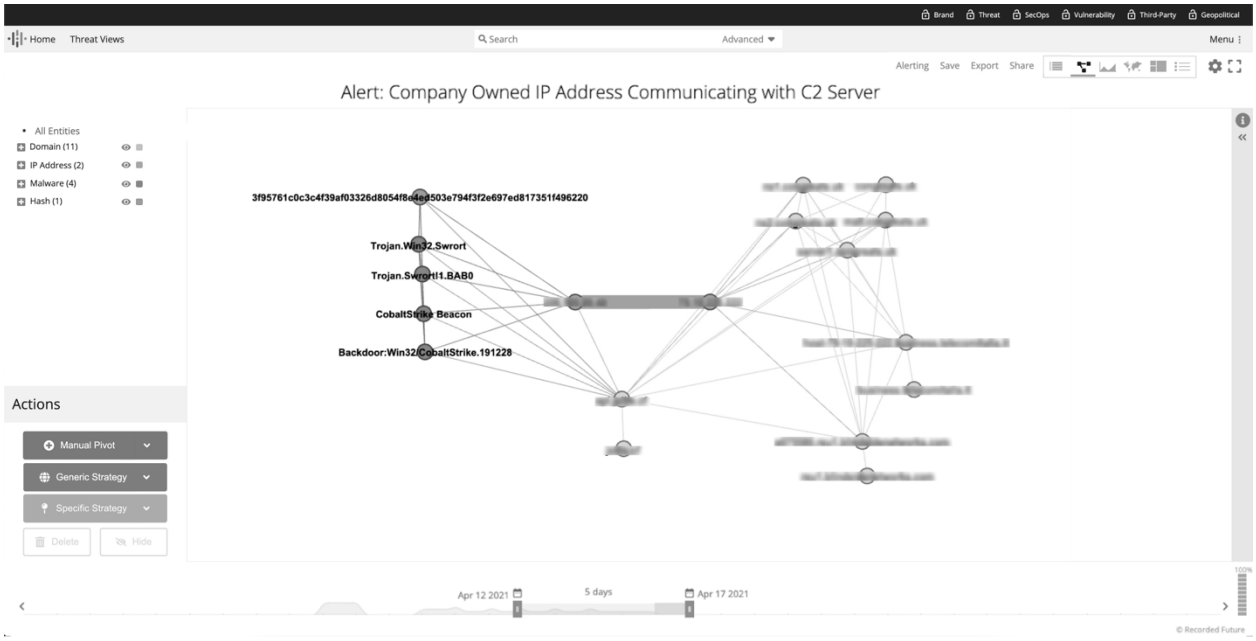


Figure 15: Image shown to users during Evaluate 1. Parts of the result from the hackathon.

6.3 Evaluate 1

This phase utilized the concepts created in the previous phase by evaluating them with users through semi-structured interviews. From this, new requirements were brought forward and old requirements were improved upon. The iterated General Use Case defined at the end of the Explore phase was also re-evaluated and confirmed in the interviews with the TIAs in this phase.

6.3.1 Validate

To begin this validation part of the Evaluate phase, four interviews were conducted with the same TIAs from Insikt as in the Explore phase and performed in the same way. The interviews were once again conducted online, this time with a presentation containing visualizations of the concepts and took approximately one hour each. At the beginning of each interview, the focus was on evaluating the General Use Case. After evaluating the use case, the focus turned to the three concepts. To avoid any bias caused by the order in which the concepts were shown each user got to see the concepts in a different order. Each concept was presented with a walkthrough following the same simplified use case. The use case was based on a real

investigation from Insikt, in order to focus on the concepts rather than the investigation in figure 16. The reason for it to be based on a real investigation was to keep it realistic and to remove risks of the interviewees being distracted by the path of the investigation. It followed an investigation of a phishing attack, starting with five IOCs, and doing five consecutive pivots from the same IOCs and in the same order, in each of the concepts. The interviewees were asked the same questions concerning each concept and their respective variants of each of the four functionalities. At the end of every interview, there was also a discussion based on images showing the concept produced in the hackathon project. The images from the hackathon placed the graph into a more realistic context and the goal was to get another set of reactions. Unfortunately, the wireframes of the hackathon concept sometimes caused a bit of confusion as they were not a higher fidelity version of one of the previously presented concepts but rather could be seen as an entirely new, fourth, concept.

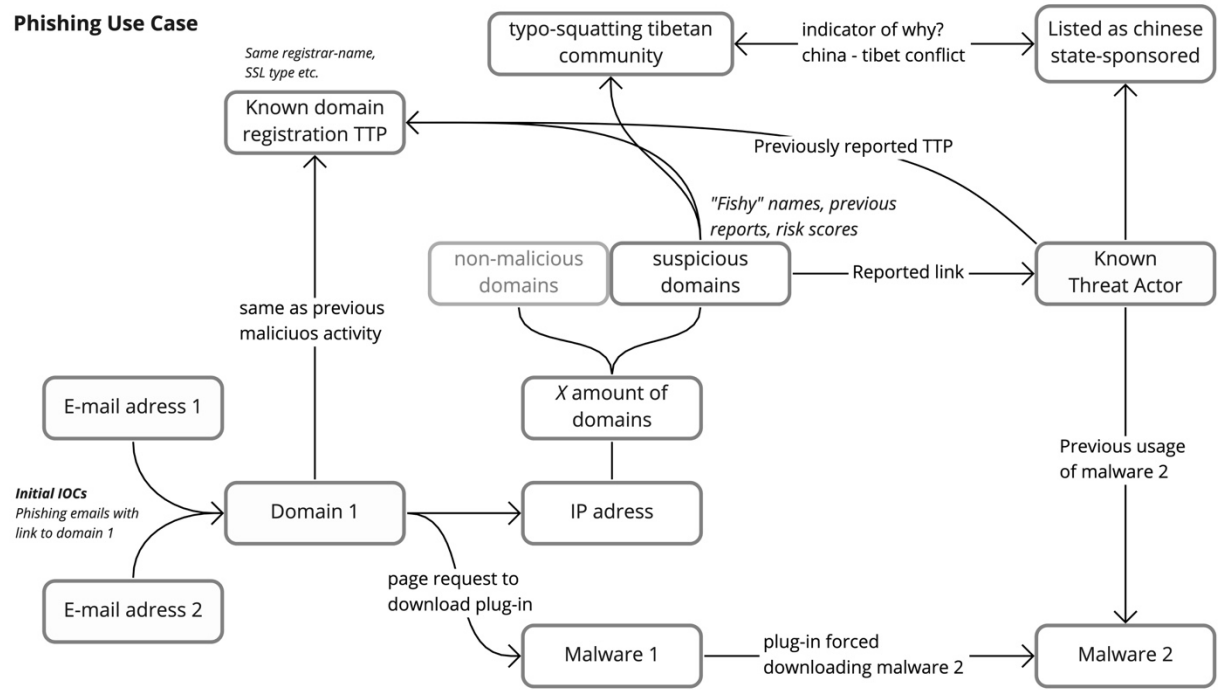


Figure 16: The use case used when showing the different concepts

In parallel with the TIA interviews, an opportunity arose to join a call with one of Recorded Future's clients from a global organization to conduct a short interview. The client was not a TIA but rather held a management position in the whole organization's security operations. This interview gave an opportunity to, not only, interview someone not employed by Recorded Future but also someone with an understanding of the TIA role while still perceiving it from a different perspective. The time available during the call was only about 20 minutes which meant that what could be covered would be less than the previous interviews with the TIAs from Insikt. The focus was the General Use Case and what the customer believed could be seen as keys for this tool to become useful to their organization.

Towards the end of this subphase another interview with the CTO and Senior Architect was held, once again, to improve upon the concepts by benefiting from their comprehensive

knowledge of the Recorded Future Intelligence Graph. Like the previous interview with these two, it was allowed to go pretty much wherever it wanted to and often it was more of a discussion than an interview. The purpose was to explore possibilities and opportunities rather than to find user requirements.

6.3.2 Analyze

As every interview with the TIAs was recorded they could be transcribed afterward. When the transcription was finished the interviews were analyzed by reviewing the transcriptions together with the notes from the spectator taken during the interviews. There were specifically four crucial functionalities, functionality A-D, that were in focus in these evaluations with the Insikt users. The users' sentiments towards functionalities A-D were evaluated and compared to discover any aligning preferences. Whenever the users expressed likes, dislikes, or ideas of new features those comments were brought forward in the analysis. Each one of these comments was transferred to a separate document summarizing all such insights from the four interviews. This way it could easily be aligned if several users expressed the same thoughts or if there were disagreements regarding a specific feature. This analysis document was brought to the following subphase to decide which way to move forward in the project.

6.3.3 Conclude

A result of the analysis was that two out of the four crucial functionalities, functionalities B and D, had two feature variants each of which were equally liked or preferred by the users. The two versions of functionality B were relatively simple to combine which solved that issue. However, the two versions of D were not possible to combine. A decision was made to use the version that was seen as more adaptable as the tool evolves. For both functionality A and C, the users had clear preferences of one variant for each functionality. These preferred options were determined to be brought to the next creative iteration which would follow this phase (figure 17).

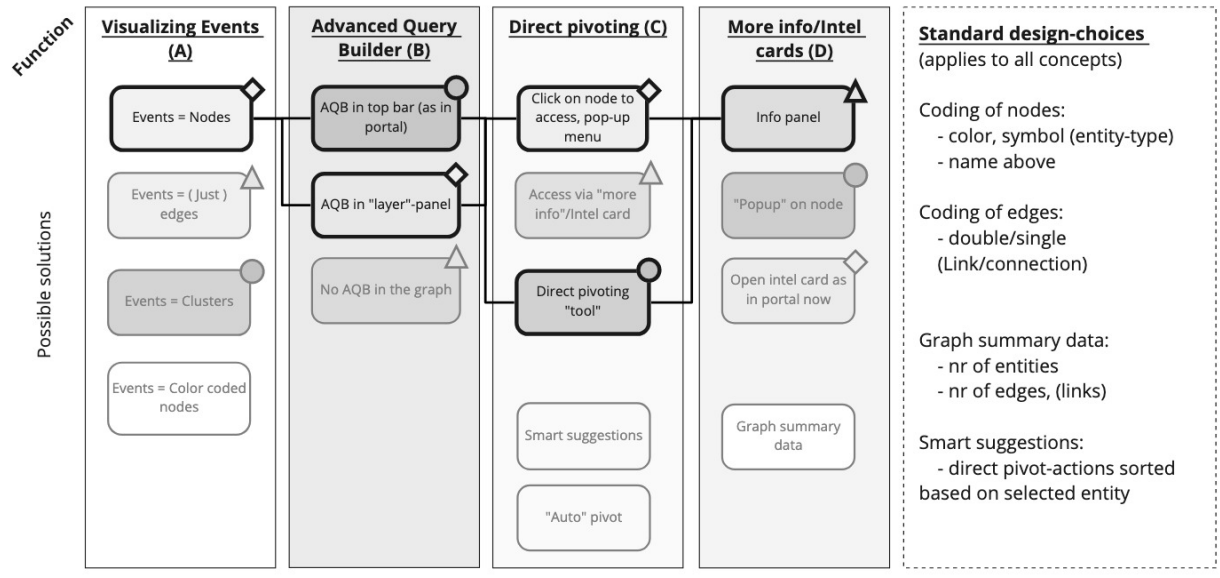


Figure 17: Morphological matrix with the preferred versions of features highlighted

The additional comments and ideas brought up in the interviews were both bringing up previously discovered user needs which were not targeted in these concepts but also completely new ideas. In total there were too many to include and test efficiently within the next iteration, requiring a decision regarding which ideas to continue developing and exploring and which ones would simply be saved for possible iterations after this project. To make a decision, a comparison of how many users had expressed a need for such a feature and assumptions on whether the feature would have a large impact on the users. It was also considered whether it was feasible to test and if it aligned with the scope of the project. With those guidelines, it was decided to add one newly discovered requirement that would be a subject in the following second Create phase.

Updated list of requirements:

I. Allow pivoting between indicators
II. Visualize indicators
III. Visualize associations
IV. Provide a continuous workflow
V. Indicate differences in importance
VI. Guide user actions without inhibiting them
VII. Seamless experience between graph and other portal features
VIII. Ability to discard and filter information
IX. Provide options to share investigations

New:

X. Transparency to underlying information

6.4 Create 2

By merging the best solution variant for each key functionality from the initial three concepts a final concept was to be designed in this phase. There were also some new functionalities added based on the feedback from the previous evaluations as well as some who had been lower prioritized in the first Create iteration. The new features and ideas were explored and ideated upon followed by two peer-review sessions, one with a technical approach and the other with a design approach. As in the first Create phase, wireframes were designed showcasing the features and solutions to be presented and evaluated in the upcoming Evaluate 2 phase.

6.4.1 Ideate

From the conclusions drawn from the evaluations regarding the previous designs, a new concept was to be shaped. This concept included the best variants of each key functionality from the previous evaluation as well as adding some new features to allow evaluating other functionalities. Some of the feedback revolved around needs that had been discussed already before the initial creation phase. These although had not been prioritized to make it into those concepts but there were some ideas and sketches regarding these which could be revised. Other

needs were new discoveries and since that needed to be iterated from scratch. Firstly these new user needs were iterated and ideated upon to invent features that could fulfill these needs. A brainstorming session was held, firstly focusing on trying to define these new functionalities and then sketching different variants. During this brainstorming, the sketches of previously defined features were incorporated to be iterated once more. This resulted in a set of new features which could be described and presented together with the previously chosen key features to show another dimension of the functionalities this overall concept could allow.

6.4.2 Assess

After the new functionalities had been defined an iteration of the feature prioritization was conducted. The new features were added to the list before comparing them and seeing where they would be placed in accordance with the previously mentioned features. The old features were also subject to re-assessment to see if any insights gathered from the user interviews in the last phase had impacted their prioritization.

A peer-review session regarding the feasibility of the functionalities was held together with the head architect. It was done to evaluate whether these ideas and suggestions formulated to fulfill certain user requirements actually could be implemented with a feasible amount of work. According to the meeting, the difficulty of implementing some of the suggested features would require more changes to the current model than others but all was considered to be reasonable. This led to the conclusion that all of the preferred features from the internal prioritization could be implemented in the wireframing.

During the process of wireframing, a mostly finished concept was peer-reviewed by the product design team at Recorded Future. Through this session features, workflow, and the overall look and feel of the concept were discussed. An expected outcome of the review was that the concept would need to be altered before being shown to users because of the feedback. For this reason, the session was held before every part of the concept was completely finished to keep a buffer of time. The feedback led to some re-prioritization of what should be included in the concept, which mainly meant raising the prioritization level of some features leading to more things being added. It also became clear that the expectations on the solution were not completely aligned between the design team and the TIA users interviewed. This misalignment is most likely due to the design team considering a wider range of users while the concept is designed with a specific set of users in mind.

From the functionalities defined in Create 1, the definition of functionality A and D were updated and five new functionalities were added to be incorporated into the design in this second Create phase. Forming the following table (table 1) of functionalities and corresponding requirements they aim to fulfill.

Functionality	Requirement(s)
A. Visualizing entities and associations.	II, III
B. Handling of the AQB within the Network graph.	VII
C. Manual pivoting between entities	I, IV, VI
D. Access to underlying entity and association information.	V, VI, VII, X
E. Visualized attributions for entities and associations	V, VI, VII
F. Ability to undo or discard actions	IV, VIII
G. Tracking of work path	IV, VIII, IX
H. Time filtering	VII, VIII
I. Saving graphs	IV, IX

Table 1: Functionalities and corresponding requirements

6.4.3 Visualize

As described above, the visualizations and construction of wireframes and the assessment process were conducted in an iterative manner. Some initial wireframes of a mostly finished concept were constructed to be able to be discussed with the product design team. From that design-oriented feedback, the final concept was visualized through a set of adapted wireframes. These were of higher fidelity than in the first Create phase and were placed in a more realistic setting by visualizing it "within" the portal, using current portal elements as a sort of frame. Some design choices regarding the look and feel were made based on the ability to keep consistency towards the rest of the portal but the functionalities were based on the previous decisions taken in the project. Visualizing the wireframes as if they were a part of the current interface in the portal was seen as somewhat of a risk since the users might get distracted or influenced by their current opinions regarding the portal.

As in the previous creation phase, the wireframes in figure 18 were produced with a specified, predetermined path of investigation in mind. The path was determined so that the features that would be the subject of the following evaluations could be presented in a clear way. Some features were not incorporated into the use case but were still visualized to be able to evaluate them as well although they would not be incorporated in the walkthrough. This decision was made since it did not seem feasible to include all specific features within such a short use case that would be presented during the interviews. The highest prioritized features were included while the slightly lower prioritized features were visualized as separate wireframes.

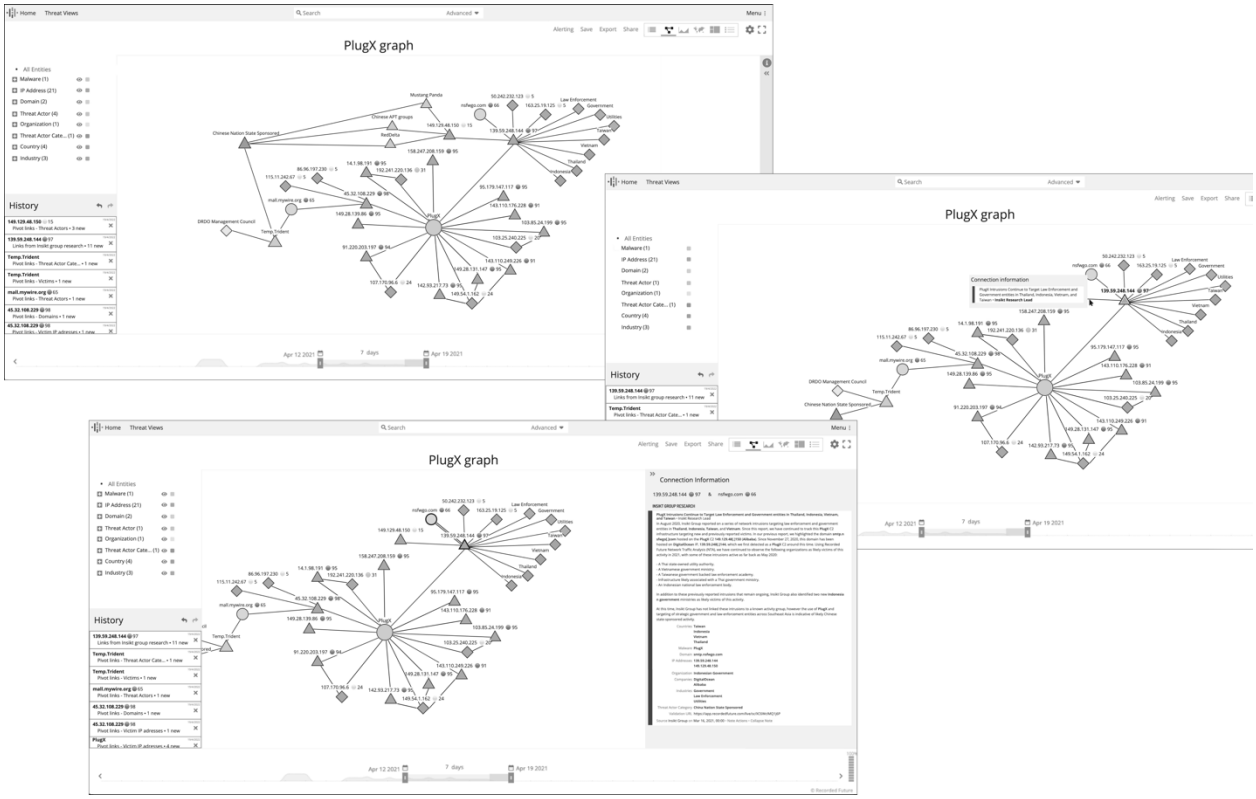


Figure 18: Three wireframes produced during Create 2

6.5 Evaluate 2

A second evaluation phase was conducted as the last phase of the whole project. This included a new set of online interviews with the same users as in the previous phases. The results from the interviews were analyzed and resulted in further conclusions regarding the presented concept. The feedback from the interviews will be discussed in the final concept chapter.

6.5.1 Validate

The final evaluations consisted of interviewing the same four TIAs a third and final time. This time, however, the interviews were conducted in pairs of two analysts in each interview. Both interviews were conducted online, recorded, and lasted for one hour. The decision of conducting the interviews in pairs was made to encourage discussion during the interviews and to further explore what the interviewees did or did not agree about. The interviews were semi-structured and began by presenting a scenario where a certain malware was explored through showing a sequence of wireframes following a predetermined path of investigation. The use case scenario this time did not follow the path of a specific, previously presented, investigation. Instead, it was made up by the project members using multiple investigations as inspiration following specific strategies presented in both previous interviews and in literature. The use case was based solely on connections specified as *Links* in the platform and utilized real data from both automatically generated sources and manually added Insikt reports or notes. The decision to follow a new imaginary investigation was based on the fact that it would allow more freedom to follow the available Links in the platform while still being interpreted as real as any already existing investigation that could have been copied. The scenario allowed the

new features to be displayed and the interview template had questions regarding each of these added features.

After the scenario had been presented there were additional questions regarding some general difficulties or concerns of using the tool. A few additional wireframes of other features not incorporated within the scenario were also presented to allow evaluation of those features as well. These interviews focused on one single concept compared to having three concepts in the previous evaluation phase. The concept presented in these interviews had more features incorporated, as well as, including two rather than one interviewee led to it being just a bit limited in time. All questions from the template were managed within the timeframe but all additional discussion or follow-up questions at the end could not be incorporated.

6.5.2 Analyze

As with the previous interviews conducted in the earlier phases, these two were also transcribed separately by one project member for each interview. The transcriptions were then reviewed together and specific comments about insights, questions, and ideas were combined from both interviews in a new separate document. This way any similarities or disagreements regarding different features could be found as well as different insights mentioned from just one of the interviews. Since the interview template contained questions regarding each of the new features added in the second creation phase the feedback could be collected in a clear and consistent manner from both interviews. This allowed clear comparisons with respect to most of the feedback while it still contained some specific insights discovered through only one of the interviews as they then had resulted from follow-up questions. This summary of expressions and thoughts formed the baseline for the internal discussions held to decide how to proceed.

6.5.3 Conclude

From the feedback gathered from the interviews some of the features could be validated while others were concluded to need some further iterations and alterations to fulfill the needs of the users. As some of the functionalities and features had been evaluated in the first evaluation phase and not chosen to be iterated from there, they were not in focus to allocate more time to the new or updated functionalities. All in all, every functionality but functionality B received feedback through the interviews.

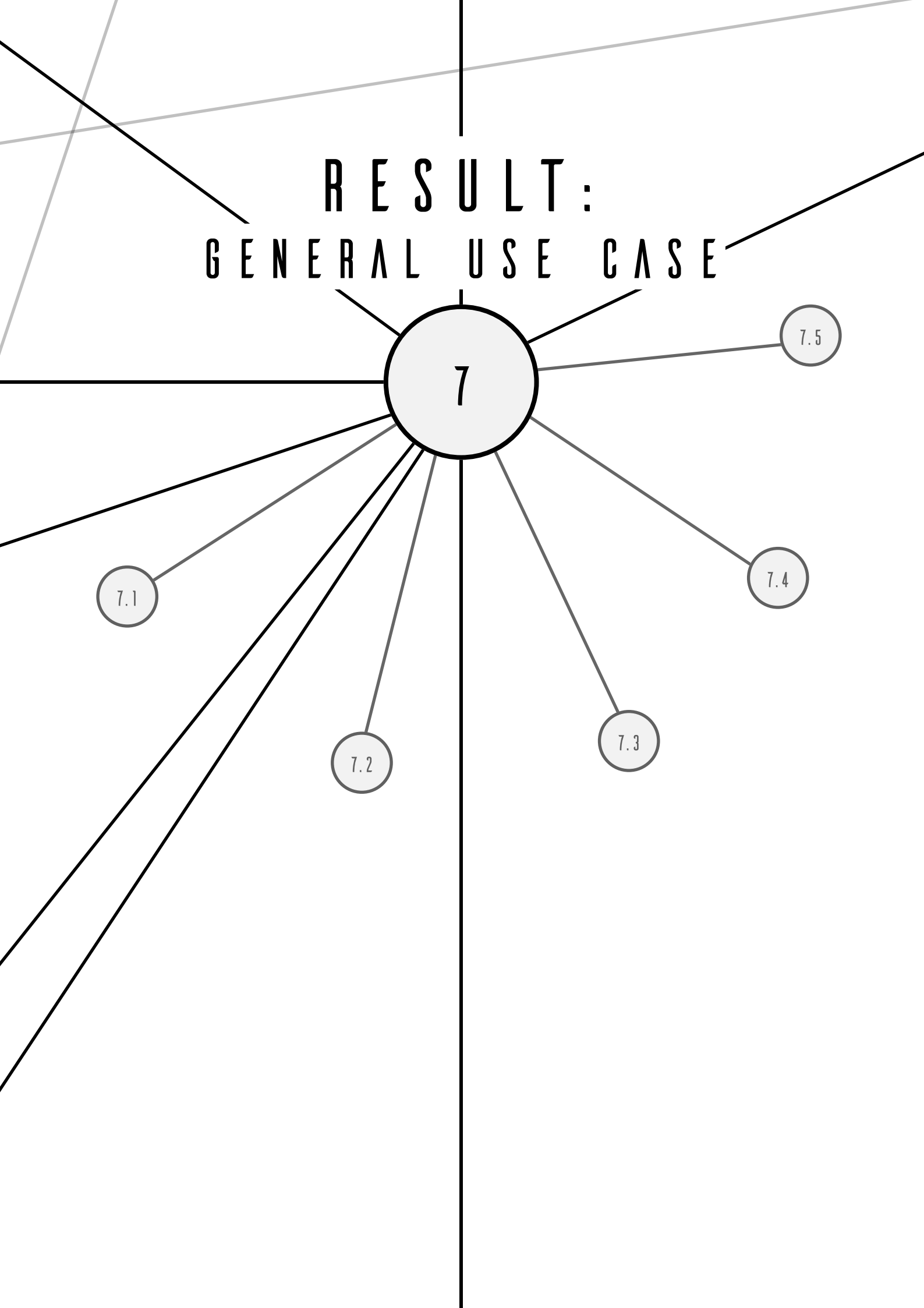
Functionalities A, C, D, E, received mainly positive feedback regarding their functionalities and how well they seemed to fulfill their respective requirements. Although fulfilling the needs the features were thought to need some further iteration and usability testing regarding their visual representations. Some received feedback regarding minor improvements while others could make use of evaluating some more possible design options since they were considered to may cause some confusion, especially as the graph grows bigger.

Functionalities F, G, H, and I were all considered to be fulfilling their targeted requirements and were validated by the users, the only constructive criticism regarding those was towards functionality H, the time filtering. The users highlighted and requested that it needs to be

conveyed clearly what the filter is currently set to and how it would act if they save their investigation and revisit it another day, which was considered somewhat unclear in the presented wireframes.

The specific designs of all of these features and how they fulfill the overall project aim and specific functionality requirements will be described below in the chapter, Result: Final Design Solutions, but first the defined General Use Case will be described in detail.

RESULT:
GENERAL USE CASE



7. Result: General Use Case

The first result of the project, which then served as an outline for all following phases and methods, were the General Use Case. An initial version of this use case was evaluated during the user interviews conducted during the Explore phase. The General Use Case was then iterated and adapted based on the feedback from the interviews forming a new more clarified version. This version was subject to evaluation during the first Evaluate phase where it was validated and confirmed as a well-defined representation by all interviewed users. The defined General Use Case of a Threat Intelligence Analyst consists of five phases as seen in figure 19.

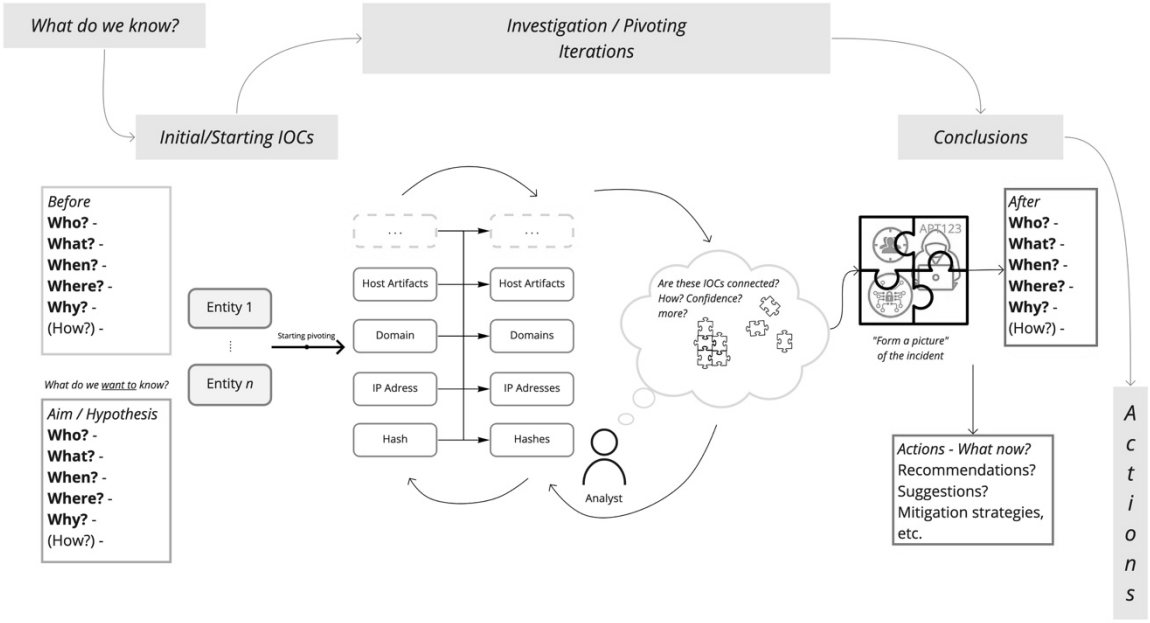


Figure 19: Visualization of the Threat Intelligence Analyst's General Use Case

7.1 Step 1: What is known? What is the aim?

The first phase in any investigation focuses on trying to define what is currently known about the incident, event, trend that is the area of focus of the investigation. A common way of working when defining what is known is to start with the 5Ws; *Who?, What?, When?, Where?, And Why?* (figure 20)

For the TIAs, the question of *Who* in an investigation could imply multiple sides or roles of the investigation. Utilizing the Diamond model perspective, the answers to this question are found on the points of the socio-political axis, thus the adversary and the victim. In investigations, there are usually both adversaries and victims, and there can also be several of each of these. The targets in this context can for example be specific organizations, certain industries, or a person. The adversaries are often some kind of Threat Actor, they could be unassociated aliases or known hacker groups, often referred to as APTs. In some investigations, both the adversary and the victim are known and in some investigations, one or both of these indicators are unknown. Threat intelligence analysts need to be very certain before attributing specific organizations or even persons as responsible for attacks. Faulty attribution is very dangerous as it can lead to actions that an organization believes will protect them but in reality will not.

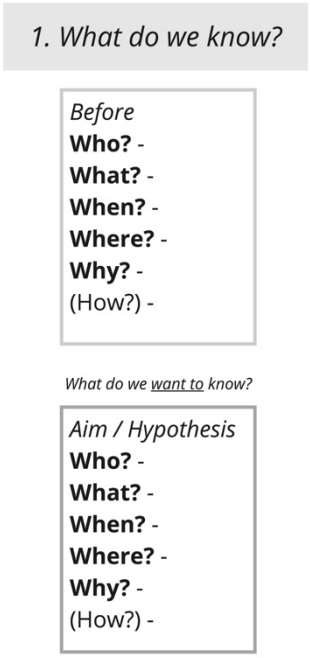


Figure 20: Step 1 - What do we know?

What has happened, is happening right now, or even might happen in the future, is the next question the analyst seeks the answers to. Again, looking at this from the Diamond model perspective, this question instead includes the aspects of the technical axis. Thus, trying to define the kind of attack, what capabilities in form of methods or attack vectors, which kind of infrastructure is used, and so on. The IOCs found and defined here will be crucial for the upcoming steps of the investigation.

The *When* is probably more clear regarding what it refers to in this case. This focuses on defining the timeframe of the investigation, trying to answer questions like; when was the incident identified? For how long have these infrastructures been active? The defined time frame of the investigation will have a major impact on the number of indicators the investigation will result in. It is important for the analyst to continuously revise this timeframe in order to reduce the risk of missing crucial information.

Defining *Where* these incidents are happening might be hard to imagine or define for someone outside of the domain, as these incidents are most often occurring online. But all online infrastructure is hosted somewhere in the real world and both adversaries and victims are also found in the real world. Many of the APT groups are linked to specific areas or states as well

as organizations or businesses that have their specific locations. Where these indicators are located can sometimes indicate answers to the last W, the *Why?*.

Why can sometimes be the hardest and sometimes the easiest to define. As described in the introduction to this thesis, Ransomware is a rising and very common type of attack. As the name implies, the why in this case is usually economical gain. Thus, if the attack vector used is known as Ransomware, money could often be the answer. In other cases, the why can be explained as a result of other ongoing conflicts in the world. Conflicts and wars no longer just take place on the battleground but just as much behind computer screens. By this, defining where adversaries and victims are located can indicate these sorts of reasons behind an attack.

After exploring and investigating answers to all of these Ws and establishing a starting ground of the investigation the analyst needs to define what they are searching for. Setting an aim based on these five questions will set the direction of the investigation and can also indicate to the analyst what they are striving for.

7.2 Step 2: Initial set of IOCs

The second step consists of defining an initial set of IOCs from which the investigation will start. These IOCs are found in the answers to the questions of the first step (figure 21). This set of IOCs could be any number of entities, spanning from one single entity up to hundreds depending on the type of investigation and the timeframe it focuses on. The IOCs can be various different entities, they can be of technical nature such as IP addresses, e-mail servers, domains, or hashes or of a social-political nature like Threat Actors or victims. The most common and still comprehensible number of initial indicators according to the interviews seemed to be about five to ten indicators. When investigating large amounts of IOCs many analysts tended to divide them into smaller chunks to get a clearer overview and reduce the risk of missing or forgetting to examine all IOCs.

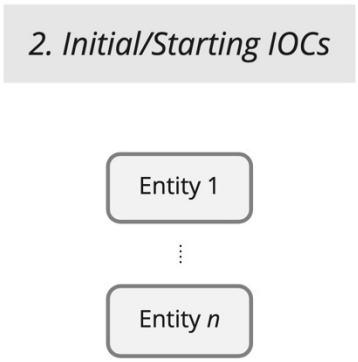


Figure 21: Step 2 - Initial set of IOCs

7.3 Step 3: Pivoting

From this initial set of IOCs the third and, most likely, the most time-consuming phase will take place. This third phase involves pivoting these IOCs as visualized in figure 22. Pivoting will be performed iteratively from every slightly suspicious entity, uncovering related entities of the same or other types. For each conducted pivot the analyst will gain more information and insight into the investigation by finding more associated entities. These newly discovered entities then have to be evaluated whether they are seen as relevant or not, to be able to discard entities that will not bring the investigation forward. This type of iteration between discovering and discarding entities will be part of all pivots. The analyst will continue pivoting until

reaching its initial or revised aim, with time available to the analyst as a major limiting factor. Pivoting is performed from one indicator at a time or several at once depending on the analyst's preferences, tools available, and the amount and type of indicators. The pivoting will in most cases rely mostly on the technical indicators as these often contain clearer associations to other technical indicators. Finding these associations are usually easiest among the types of entities that are found at the bottom of the Pyramid of Pain. The further up you move in the Pyramid of Pain the harder it is to determine the associations between them. On the other hand, the technical indicators found through the pivoting can be added to the Diamond Model, building up a clearer picture of the event and its adversary. To be able to distinguish or pin-point certain TTPs to a specific Threat Actor relies on forming composite objects of multiple indicators found in the lower tiers of the Pyramid of Pain. Meaning each piece of information found through pivoting can in the end help build a complete picture of the event.

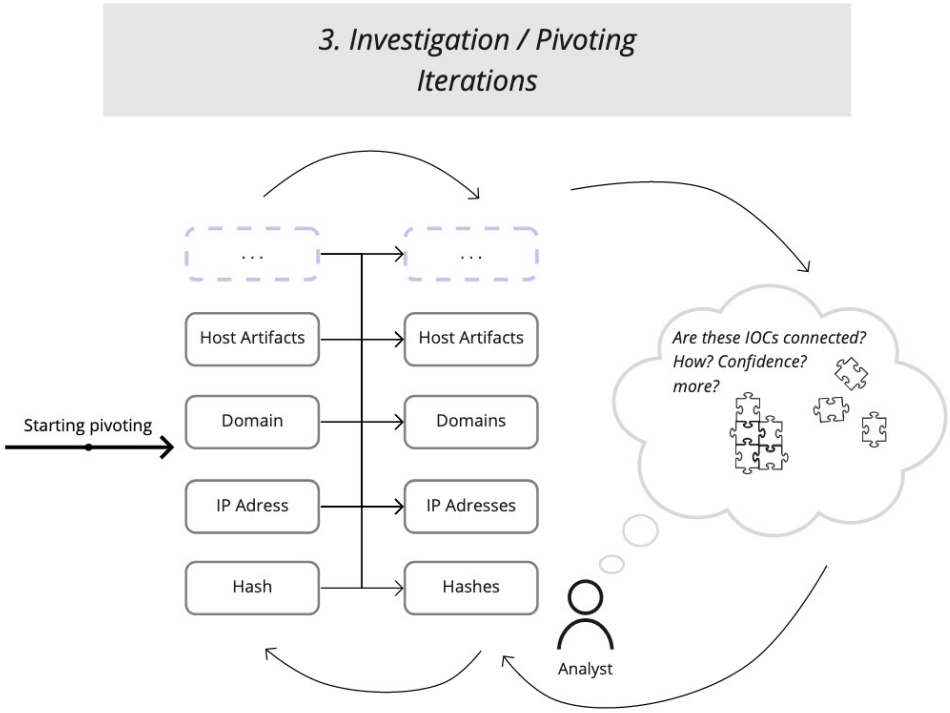


Figure 22: Step 3 - Pivoting

7.4 Step 4: Conclusions

Partly overlapping or iterating between the third phase, the fourth will occur as well, which revolves around the conclusions drawn from said pivoting. As the investigation proceeds through each pivot, new pieces of information will be discovered and added to the complete picture of the investigation. Eventually, this fourth phase aims to result in some conclusions around the investigation, answering one or more of the initial aims set at the first phase, see figure 23. During the third and fourth phases, it is also common for the initial aim to be revised in accordance with information that is discovered during pivoting. The ability to draw these conclusions will rely on the confidence level of the different associations found through pivoting and the analyst's capabilities to summarize these insights into conclusions.

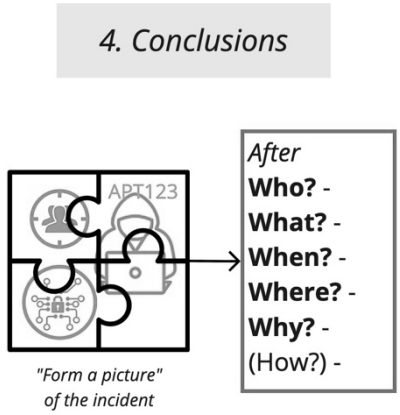


Figure 23: Step 4 - Conclusions

7.5 Step 5: Actions

Finally, in the fifth and last step, actions to disrupt or mitigate threats are suggested based on the conclusions, see figure 24. Such actions could for example be blocking certain technical indicators such as IP addresses or domains from the internal networks. As reducing risks and mitigating threats is one of the main purposes of security operations, these suggested actions are a very important step in moving forward and staying ahead of the adversaries. As described by the Cyber Kill Chain framework, if a victim can stop a potential threat at any point in the chain none of the following steps can continue and the threat is mitigated, at least for some time.

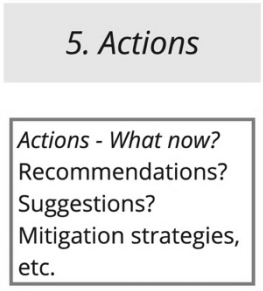
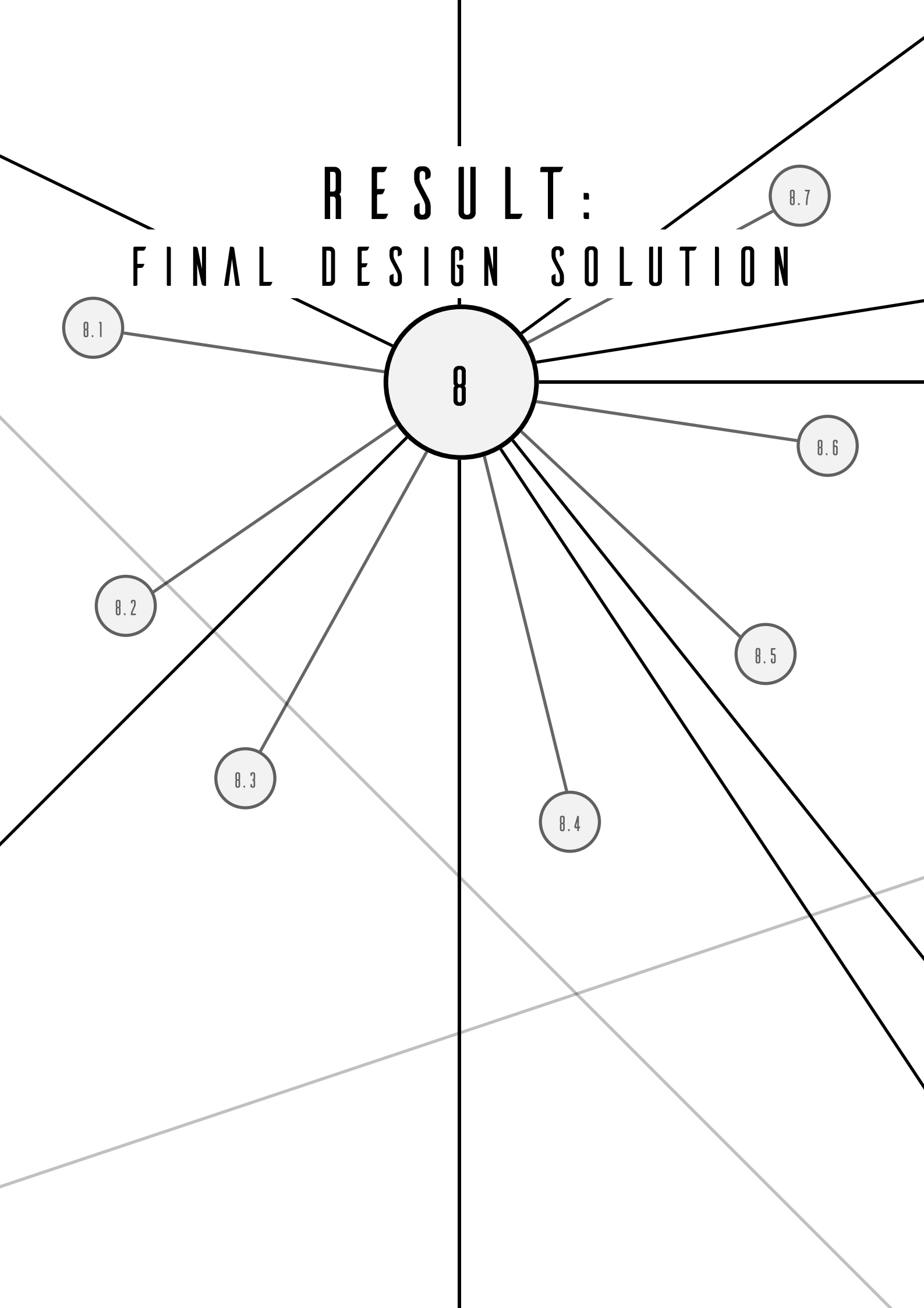


Figure 24: Step 5 - Actions

This General Use Case was used as a ground point when visualizing the interactions and intended workflow of the Network graph which was presented during user evaluations. It also formed a solid basis when working with designing the Network graph and was used to keep the users' workflow in mind throughout the ideation and design methods forming the final design concept presented next.

RESULT:
FINAL DESIGN SOLUTION



8. Result: Final Design Solution

Throughout the process and its phases conclusions were drawn, decisions were made, and design solutions were shaped and evaluated, those results will be presented here. This section will describe an overall overview of the design solution and then scoping down to describe specific parts of the solution in detail. The project resulted in a concept of an interactive network graph for threat intelligence investigations and research as seen in figure 25. In this section, the overall functionality of the concept will be presented as well as defining specific key features and how they aid analysts in their work. The features will be described based on where in the interface they are found apart from some functionalities which are seen as composites of the overall concept. The interface has been divided into five main components, each with its own specific functionalities, designs, and ways of interaction.

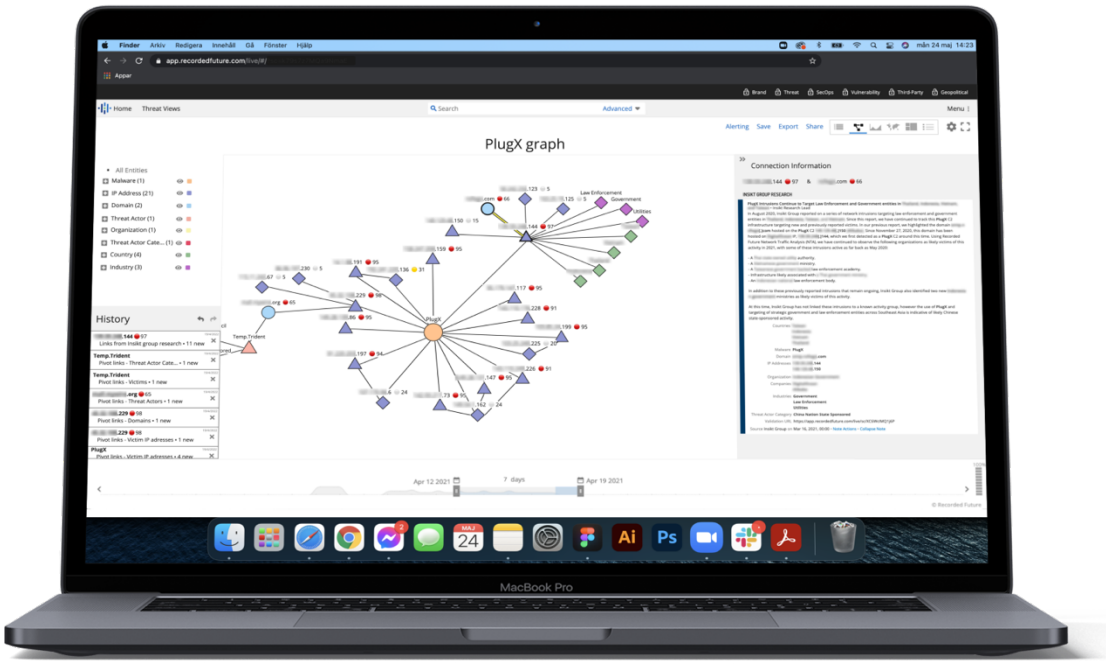


Figure 25: Mockup of the final concept

8.1 Overall design solution and features

During the Understand phase, every user mentioned switching back and forth between different tools when describing their workflow. When pivoting from indicator to indicator different tools provide different values depending on what kind of indicator they are specialized on. The Network graph, as seen in figure 26, will provide the most value if it allows users to access information of any kind available in the portal while it still is presented to the user in a way that the most common actions are the easiest to access. The interviews made it clear that the users preferred a tool that might require more effort to learn as long as it allows them to do more types of actions in accordance with their own preferences. In most instances where there was a simpler, less powerful option and a more complex but also more powerful option they preferred the more complex and powerful one. The comparison they made was that to be efficient in the portal today they had to learn how to use the Advanced Query Builder to their advantage. The AQB is not a simple feature to use efficiently and it requires the user to understand both the tool itself, the sources of the Recorded Future Intelligence Graph, and how to interpret the results. They preferred there to be a learning curve if that meant that their investigations would yield better results. A reason for this might be their highly technical skills and the fact that working with threat intelligence investigations is already considered to be both complex and time-consuming. Being limited by the tools or options available might cause them to switch to another tool that allows them to perform those types of tasks meaning it is important to let the users take control and perform actions as they please.

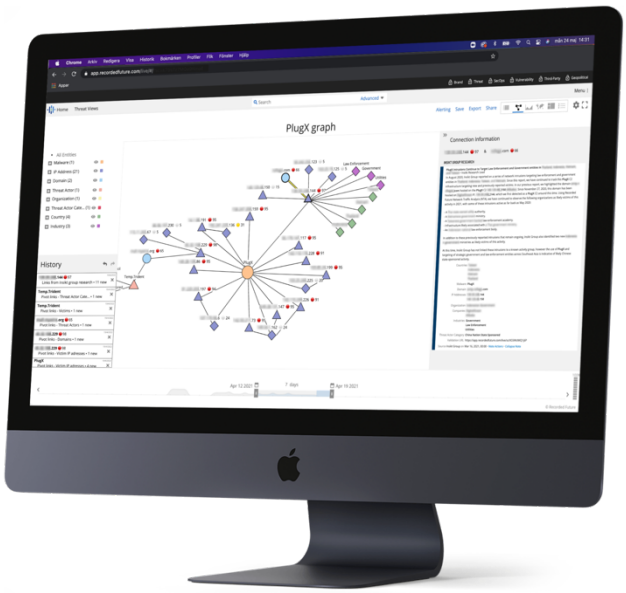


Figure 26: Mockup of the Network graph on a screen

An assumption can be made that an investigation can be performed in one or multiple sessions in the graph. Each session can take just a few minutes but can also last several hours when TIAs investigate the indicators found through pivoting the graph. Based on this it is clear that there is a need for a continuous workflow within the graph. When a TIA is using the Network graph they need to be able to quickly and continuously assess the information they have and take the next step. If the experience when working with the graph is that it is disjointed and backtracking or changing the direction of the investigation or accessing information is difficult the tool will be frustrating to use. It should also be easily accessible from other places on the portal.

There are multiple reasons for a user to start an investigation and one could be as simple as reading about a new malware and wanting or needing to look further into it. Accordingly, allowing the users to start a new graph from an Intelligence Card or any other place where they

interact with entities or events in the platform is necessary. The phrase “start a new graph” implies that old graphs exist which has to be true and allowing users to save graphs and come back to them is absolutely necessary. When a TIA is working on an investigation it could be completed in one sitting but most commonly it could be spread over at least a couple of days. For there to be multiple investigations active at the same time is also common. To accommodate multiple investigations, naming graphs will be necessary. It is not uncommon for there to be investigations that continue indefinitely, for example when tracking trends. The ability to visually see how the threat landscape changes as new data are introduced into the Recorded Future Intelligence Graph will be powerful for users.

Another facet of saving graphs is the ability to share what has been saved, this can be useful for peers evaluating an investigation to see if things were missed or when the result of an investigation is being delivered. Depending on who the recipient is, and most likely their technical knowledge, how the graph is shared will differ. If the recipient is highly technical like a senior analyst they will request the ability to access the data and be able to do their own pivots. On the other hand, if the recipient is not as technical or does not have the time to thoroughly walk through the graph, an exported image with annotations on what is important might be the best way to allow users to share their results. Exporting images will also be useful for incorporating them into reports written by the analyst which is a common deliverable of investigations.

8.1.1 Binary, discrete or continuous levels of confidence?

The following designs are based on an assumption that there are merely two levels of confidence for associations between entities. They are either *Links*, which are high confidence associations where Recorded Future claims that the connection exists, or *connections*, where the confidence is considered to be lower. This assumption was made to enable the testing of other features but it was always known that it did not reflect reality as a whole. To build the Network graph there needs to be a decision made on how to attribute confidence to all connections. This however is not chiefly a design decision but rather one that is based on how the data is gathered, categorized, and how the confidence of a reference can be evaluated. From the interviews and meetings with both users and engineers, this is a difficult task and will need work on all levels of the platform to form a conclusive way to distinguish and categorize confidence before being able to decide how to design such an attribution. One such effort made is the platform-convention of Links, which are distinguished from other connections with less confidence. In reality, the confidence level, as described earlier, can be of a wide range especially within the associations that this paper refers to as connections. By choosing to limit the confidence range to just two levels a conceptual design which has some indication of confidence could be presented.

In the future, there could, or even should be more levels of confidence but moving to a way of assigning confidence to associations in a continuous span will have its difficulties. By that, having distinct levels or categories allows for clearer definitions of the design task and clearer distinctions within the graph which is important to aid the users when they decide what

connections to investigate further. If they know that a Link is based on research made by an analyst they might trust it more than one that is simply a connection drawn from entities co-occurring in a news article. Designing around the confidence put in connections will always be a key in the graph and it will enable users working with the graph to draw quicker and more confident conclusions based on the indications of different levels of confidence.

8.1.2 Time is critical

Possibly the most powerful way of limiting the amount of data to show is by deciding what timeframe to access data within. As this is going to be necessary on any solution to be built, the time filter that already exists in the platform was adapted to the Network graph. The way the filter currently works is that the user specifies a time frame when they build their queries that filters the results. The time frame can be set to either specific dates or a more fluid time frame like “the last seven days”, tomorrow, or -15 days to +15 days. The current time filtering in the portal was copied and re-applied to the graph as it was considered easy to use and intuitive for the users. Though when using the more fluid timeframe in the graph it becomes more difficult as there will be noticeable change of the graph content between every session or even during a session if a connection becomes older than the set time frame. This fluidness is less intuitive to use in a graph as compared to the query in the portal where the results are presented as a static list. However, the users felt that it would be useful to use the fluid time frame in certain investigations and as long as they were the ones that defined the time frame. They also expressed interest in the ability to use a combination and set a time frame like from “YY-MM-DD” to “now”. This filter is applied to every result in the graph and this can cause problems when conducting an investigation. Some results are very time-dependent like when an IP address was flagged as malicious. Resolving IP addresses can change a lot and connections to an IP address when it was not known to be malicious are rarely useful to pivot to in an investigation. So, IPs need a very specific timeframe, Threat Actors on the other hand might not. A Threat Actor's TTPs change much less frequently and while the specific time frame makes the results interesting for an IP address it will most likely risk leaving out important connections for a Threat Actor. A more adaptive way of filtering would most likely be powerful for the users but be difficult to implement in a way that would be easy to use. Assuming that a graph could include hundreds of entities, the number of filters would rise quickly and be difficult to keep track of for even the most skilled and experienced user.

8.2 Network graph

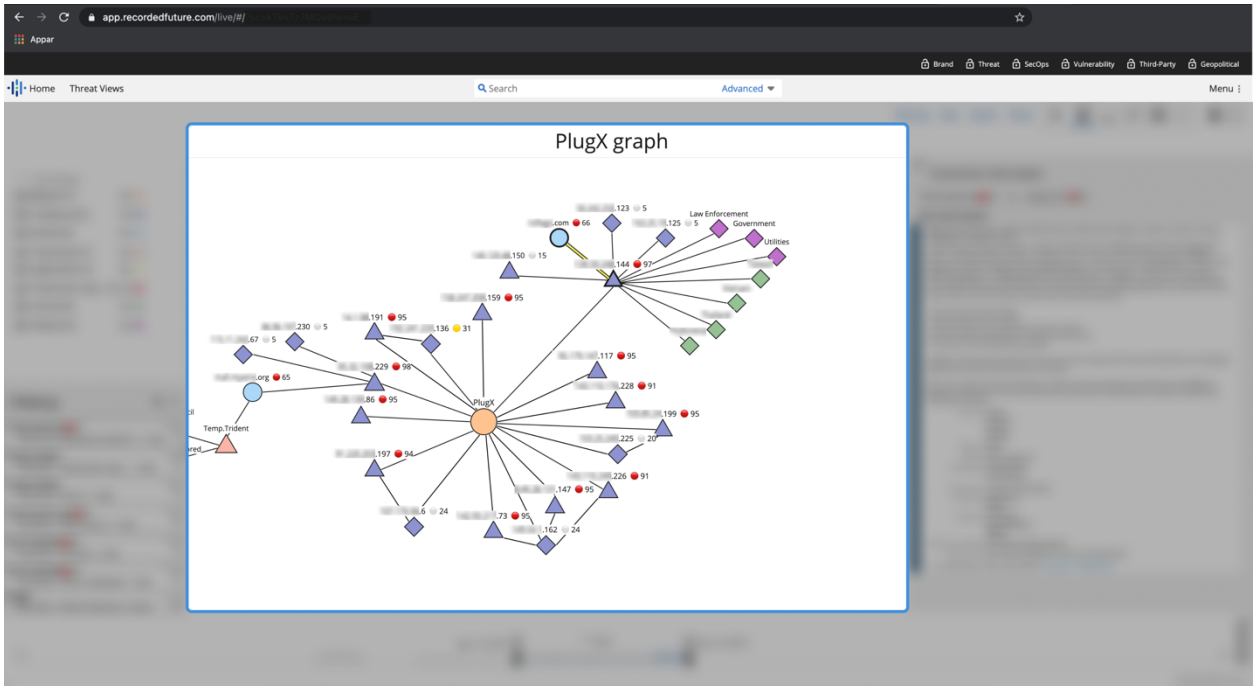


Figure 27: The network graph in focus

The main feature of the concept is the Network graph, taking up most screen real estate (figure 27), where relationships between entities are visualized. In the graph, any entity type is visualized as a node and any connection between entities is visualized as an edge connecting them. All nodes and edges in the graph are interactive and allow the users to perform a set of actions on them.

Even though there are multiple ways or reasons to start an investigation, the most common one is that the analyst has a list of IOCs from an incident, as explained in the General Use Case. As such the most common start of the graph will be that an analyst imports every entity within these initial IOCs from a list that they have created in the portal or by uploading, for example, a CSV file. At this point, the graph is often easily overviewed since the number of nodes is simply, as the analysts expressed it, a handful. The graph, however, will grow rather quickly and can become cluttered with hundreds of edges and nodes, becoming both more difficult to use and understand. When a decision is made to include information within the graph, what value it brings needs to be carefully evaluated against how much clutter it adds to the graph. An example of such a decision is the inclusion of risk scores on each entity, within the graph as in figure 28. There was consensus from the users that it would be useful to guide focus and to speed up the workflow for the analysts. Some deemed it to be so important that even if it added too much clutter it should be included in the graph in some form. Others were afraid that it would add too much clutter and make it too difficult to use.

As mentioned above the Network graph will in many or most cases grow to contain a large number of nodes and even more edges. With that in mind, there is a clear need for the users to be able to pan around in the graph and to zoom in and out to overview the whole graph or to focus on specific parts. This feature complies with the, within the field of information visualization, well-known Shneiderman (2005) mantra; “Overview first, zoom and filter, then details on demand”. Especially when combined together with other graph controls and features for accessing information that will be described in detail later in the actions part. In accordance with other types of tools used for constructing network graphs, one can assume that there will be an algorithm that arranges all nodes into the graph but there will still be a need for users to be able to manually rearrange the graph. The users should be able to select one or multiple nodes and, through drag-and-drop, move them as they wish within the graph. This should be allowed both to follow usability heuristics three and four but also to let users to direct focus to the parts they believe is most important. An example of when the user-, or viewer-focus is important would be when an image of the graph is exported for another person's benefit. The ability to isolate the most critical branch of an investigation will make the communication more efficient and allow for quicker interpretations and actions.

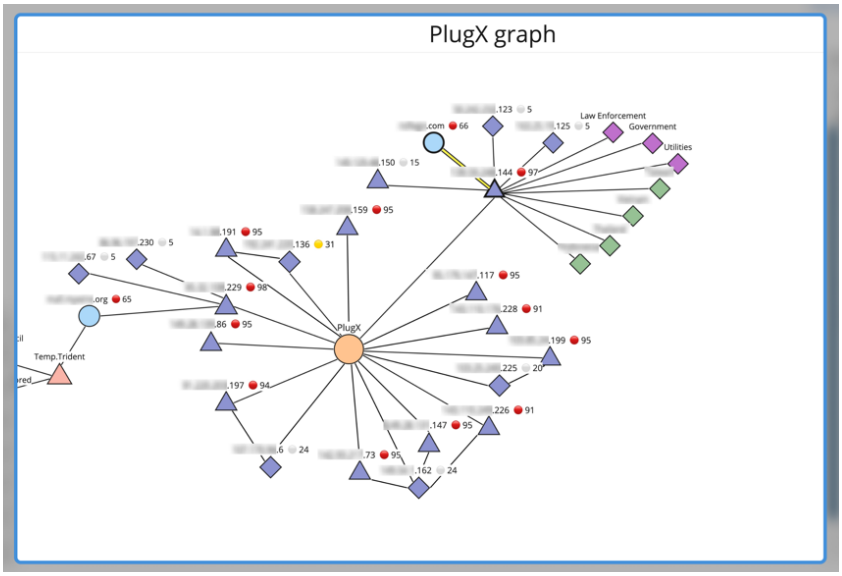


Figure 28: A graph from an investigation into the popular malware PlugX

Another feature that follows both internal and external consistency and standards is the epistemic action of hovering over elements in an interface to access more information, see figure 29. In this Network graph when a user places the cursor over an edge or a node they will access a summary or part of the information available on that element. For an edge that would mean any Insikt notes that established the association and how many other references there are claiming the existence of the association. Here it is important to the users that the sources with the most confidence are prioritized. According to the same heuristic, the users should be able to hover over a node to get information about that entity. This was not evaluated with users but it could be assumed that the top part of an Intelligence Card is expected when hovering over a node. Note that it is simply a suggestion and a feature that needs to be

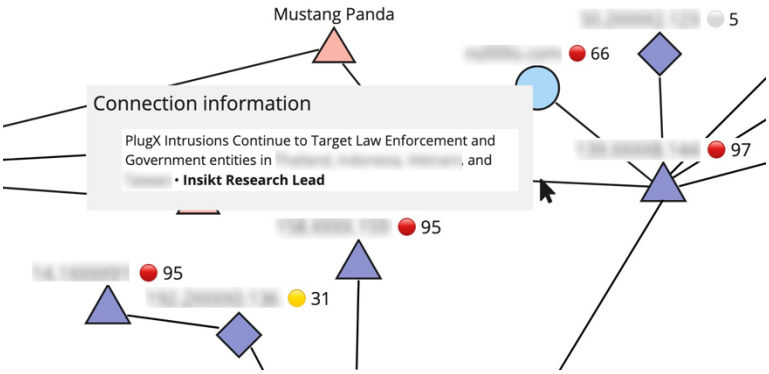


Figure 29: An edge being hovered and showing the Link connecting the two entities

ideated upon and evaluated. Once again according to heuristic four there needs to be an appropriate delay before the information appears. It can also not be too short as that might be distracting and annoying according to the users if moving the cursor over the graph triggers pop-ups of all elements the cursor passes. This is only a summary or a small part of the information available on nodes or edges and the users will have access to more of the information but that will be discussed in the part about the Information panel. As in the rest of the portal, users can also access the full Intelligence Cards of any entity within the graph.

8.2.1 Encoding information into edges and nodes

The actual graph will consist of two main elements, nodes, and edges. As there will be multiple types of each of these, there is a need for users to be able to distinguish them from each other. Without distinctions, the graph will only be presenting a big spider web and users will not be able to interpret the information into actionable intelligence. The encodings of these two elements were not under evaluation during any of the interviews as it was decided that the main functionality, being able to distinguish them from each other was enough at this stage. The encoding was designed based on the theories of encoding them using different channels and aiming to allow preattentive processing when applicable, two theories presented in the Design Theory. Further investigations in different options for the encoding of these elements can be more useful to carry through when having an interactive prototype. The encodings of each of these elements will be described separately below.

The encoding of nodes presented in the concept consists of five components that make each individual node distinctive from the others. Depending on the current zoom level which affects the size of each node one of these components, the icon, will be discarded when the node is too small for the icon to be able to be distinguished. Each entity type will have one assigned background color and icon which distinguishes them, these are placed within the shape of the node. In the concept, the shape of the node indicates its attributed role, this as the shape is processed as a different channel and by that can be used to represent another attribute. There are three types of roles; adversary, victim, and neutral. Different versions of encoding nodes can be seen in figure 30. The colors, icons, and roles will always have the same attribute tied to them in all graphs allowing users to learn and remember what each of these components represents.

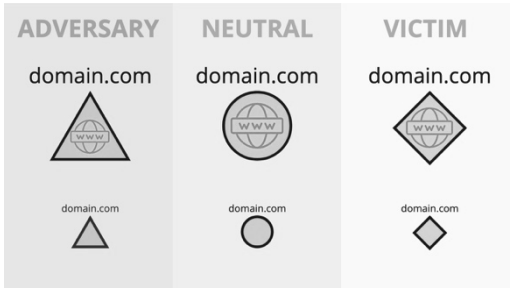


Figure 30: Different ways of encoding nodes

As described earlier these encodings were not the subject during the evaluations but they did receive some feedback. Users requested a way to distinguish the different role attributes but visualizing them through the shape of the node got mixed reviews. Some users expressed a nervousness regarding the risk of them being hard to distinguish as the graph grows bigger. Having the icons is something all interviewees were positive towards, many expressed that it

is used in other tools they use and that it would be useful in this as well. There were some concerns regarding the design of these icons. It is not always clear, to begin with, what it represents but on the other hand, users can learn and get used to them rather quickly. Above the node, the entity's name will be displayed together with its current risk score. The decision to show the risk score of each entity was decided between the two iterations since it was requested by the users and they expressed that it would speed up their process to have it constantly visible. The risk score is one of the main indicators used by the analyst when deciding how to proceed in their investigations and was considered critical to be shown already at the overview of the graph.

> *"Just so that it immediately flags to the user what entities are important and they are immediately able to pivot to the ones who have a red circle."*
> *- Analyst 3*

The encoding of edges relied on the assumption made to divide the confidence levels of associations into just two distinct options, Links and connections. This is the only encoding applied to the concept and it is indicated by the type of line representing the edge. High confidence associations, Links, are represented by regular lines while the lower confidence connections are represented by dashed lines, visualized in figure 31. This type of distinction was discussed with users during the second evaluation phase, they explored some concerns that these types might be harder to distinguish as the graph grows but they expressed a clear need to be able to distinguish them. A suggestion from one of the users was to allow toggling whether the low confidence connections should be visible or not. Such functionality would be useful to declutter and clarify differences between different parts of the graph even at brief overviews.

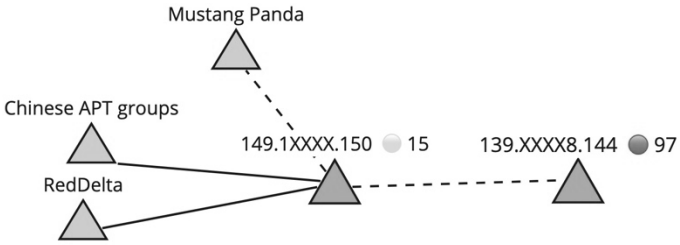


Figure 31: A way of encoding binary information into edges

8.2.2 Indicating changes

The main interaction a user will do within a graph is to add more nodes to the graph and how they can do that will be discussed in the next section. Some of the effects those actions will have are more appropriate to cover here, however. The first is the visibility of new nodes and connections after a pivot has been made or any new entity has been brought into the graph. Whenever that happens any new nodes and edges need to be distinguished from the previously existing ones. This is especially important as the graph grows since then it will be harder to keep track of changes. This feature is missing on other similar tools according to the users and makes identifying new nodes and edges more difficult and time-consuming. The purpose of this is to ensure that the users can be efficient and not spend time visually searching the interface to find the new connections instead of looking into them. Using encodings such as

highlighting in the form of contrasts, which comply with preattentive processes will guide the user's attention to the newly discovered nodes.

This indication feature also becomes very important since every new node might not only be connected to the one the user pivoted from. Within the graph there could, or will, be associations drawn from new entities in the graph to already existing ones. This means that whenever a node is brought into the graph the system checks for connections to any other node already existing in the graph. This feature could be limited by the confidence level to make it faster. This automatic check-up will be very useful to the users as they receive answers to questions they did not even ask yet. However, these connections could go to any entity within the graph, and to make it easily visible they need to be indicated where they have appeared in accordance with the first usability heuristic.

8.3 Action Menu

Figure 32: The Action menu in focus. It might be small but it is powerful!

Interacting with the graph and performing the most common task of an investigation, pivoting, will be accessible through a pop-up menu, as seen in figure 32, in this project called the Action menu. The Action menu is accessed by right-clicking in the graph window. The Action menu is the main interaction point for manipulating the information shown in the graph. It could allow numerous functions but this project has focused on a selected amount which was considered critical for the defined General Use Case. The Action menu is shown as a pop-up and the content of the pop-up will

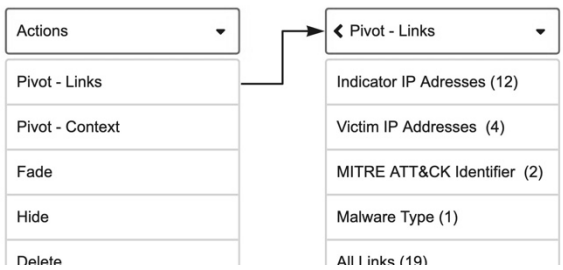


Figure 33: The Action menu before and after selecting Pivot – Links. The results of the pivots are indicated in the parenthesis

be adjusted based on what type of element the user has clicked. When active, the Action menu will update its internal content based on the previous action, see figure 33.

8.3.1 Expanding the investigation

The user can expand the graph through two types of actions found within the Action menu, adding a node manually or through pivoting. Adding a node manually is useful when investigating a certain hypothesis since it allows the user to add a node without knowing whether it has any associations to existing nodes. When adding a node manually the graph will cross-reference it to all existing nodes in the graph and draw out any associations found as edges. If there are no existing associations, the node will still be added but will not be connected through any edges to the rest of the graph, at least not at that stage of the investigation. To add a node manually the user opens the Action menu by clicking an empty space in the graph, and chooses that option in the Action menu. This will allow the user to perform, what the portal refers to as, a "quick search". Incorporating the quick search aids users' mental models through internal consistency as it is a functionality already existing in the portal. Utilizing the functionality of the quick search results in specific entities rather than a bunch of events or references as in the case of doing an advanced query. At other places in the portal, performing a quick search opens the Intelligence Card of the specific entity.

The second action to expand the graph is through pivoting, this is a crucial function as it is the most common task in an investigation and where users spend most of their time. In the same manner, as for adding nodes manually, the pivoting action is also available in the Action menu. The pivoting action has been decided to be split into two types of actions. This based on feedback from the users telling that the type of associations is critical in investigations and thereby dividing pivoting based on the confidence level of associations. The user can choose to pivot just for high confidence association, Links, or to pivot for any type of connection. These are accessed through their own options in the Action menu and were distinguished based on the pre-consumption on dividing associations binary using the division suggested in this project.

Depending on the entity type of the selected node the sortings of the available entity-types to pivot to will be arranged differently. This as some types of paths are more common and since that should be easier to access by presenting them first. As there are a lot of entity types to pivot for, a suggestion is to also allow the user to search for types instead of having to scroll a long list. This could especially be helpful when pivoting for less common types as it would speed up their workflow instead of scrolling and manually searching the list. The users should also be able to pivot for all Links or all connections in one single action which would be a way to skip manual labor and quickly expand the graph.

When shown the list of entity types available for pivoting, the list will contain an indication of how many new nodes each type of pivot-action will result in. This indicator was considered really helpful by users as it gives them a sneak peek into what a pivot will result in and could

help them in choosing the path of their investigation. This indicator can also be seen as a sort of warning as it would indicate both if the pivoting action would result in an empty pivot or if it would result in a very high amount of new nodes. If the indicator shows there are not any associations of a specific entity type that option would be disabled but still visible in the Action menu. On the other hand, a high indicator could imply that it would risk cluttering the graph to pivot for those. For this, it is suggested to allow the user to apply filters to their pivots before implementing them. This option has been discussed with users but has not been visually represented how it should be accessed. All interviewees expressed positive feelings towards such a feature as long as it is seen as an extended interaction and that they are not always forced to go through those extra steps when pivoting. Some suggested filters that could be applied according to users are risk scores, source types, and locations. How these filters would act can also differ and need to be decided before implementing them into the graph. They could be seen as purely adding the entities which pass the filters or that it shows all entities but the ones who do not pass the filters are clustered. This way the user can still access them while the cognitive load is reduced and the attention is directed to the nodes of higher importance.

"I like the idea of giving the upfront count as well, whether the engineers will like you for that is something else. I think it's massively helpful!" - Analyst 1

"I like being able to see how many of each type of indicator would come up if you would click on them." - Analyst 2

8.3.2 Discarding irrelevancies

As described by the General Use Case it is equally important to not only be able to expand an investigation but also to discard irrelevant information. This type of functionality can also be accessed through the Action Menu in the same way as for pivoting. Accessing the Action Menu from a node will allow the user to perform two types of decreasing actions. It allows the user to remove nodes and hide nodes. As the naming imply, removing a node will permanently discard it from the investigation, it will be removed from the Network graph and from the list in the Legend Panel. Hiding a node will instead visually remove it from the graph, it will no longer be visible or accessible within the Network graph window but in contrast to removing it, the entity will remain in the legend list. The legend will allow the user to make the node visible again through its show button. These functionalities are critical to reduce the number of irrelevant indicators, both to lower cognitive load and to allow users to focus their attention on indicators that are considered suspicious or in other ways more relevant to the investigation. When pivoting, new nodes added will still be queried for any connections to hidden nodes as well. This could result in a previously considered less relevant node being seen as more relevant based on the newly discovered associations. Based on this it is suggested that associations to hidden nodes should be indicated to the user when discovered but it has not been visualized or suggested how to implement such a feature.

8.3.3 Forming actionable intelligence

In addition to pivoting and discarding actions, it is also considered important to be able to perform epistemic actions to display more specific information or to focus attention. Two features for this that are concluded to be useful are a fade feature and a note- or comment feature.

The fade feature can be seen as a semi-version of the hide feature as it allows users to fade out nodes and edges which they consider to be of less importance as in figure 34. As described by users it could be useful especially when conducting research over a longer time period since it then could indicate to themselves which paths to research further or not. A faded branch of the graph could be used to indicate that it has been investigated but decided that it currently does not lead further. This feature could also be useful when sharing graphs as it can direct the attention of the recipient while still showing other paths that have been researched as well.

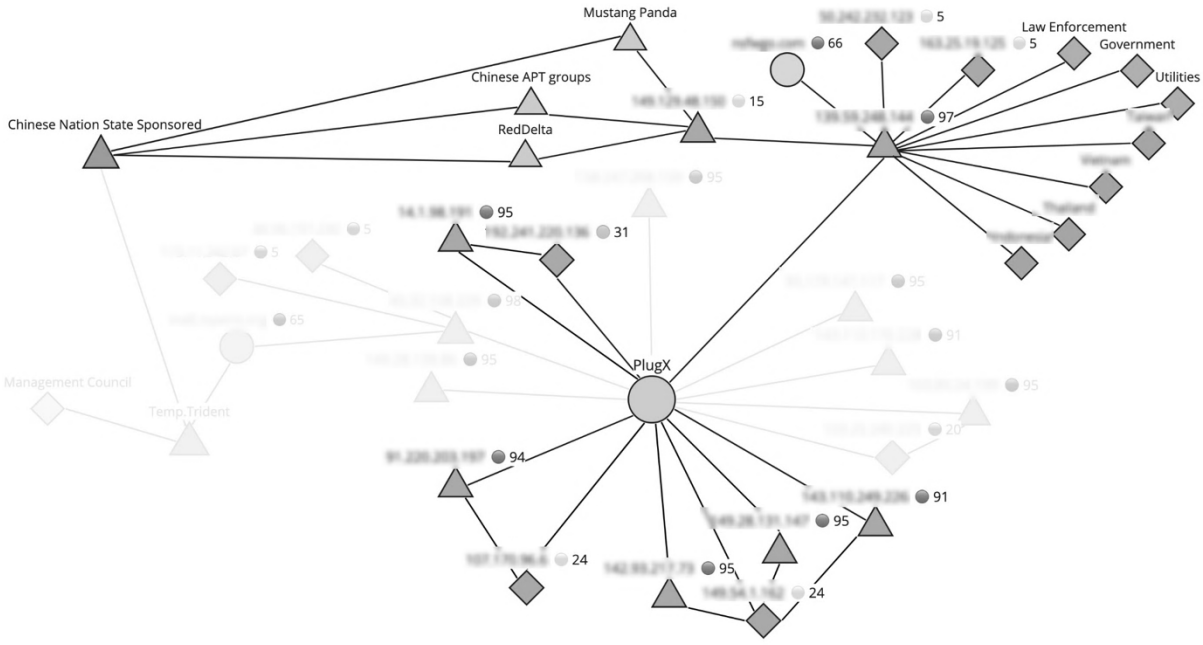


Figure 34: Image showing the fade functionality

Having a comment-, or note-feature is also useful according to the user for approximately the same reason, and also to add extra context to the current investigation or thought process behind it. This feature is also included in some other parts of the portal and allowing the same functionality in the graph complies with internal consistency. Adding notes or comments would also be useful for the users as it helps them remember rather than having to recall why they performed certain actions. Incorporating such notes could also be helpful for the use case of exporting images of the Network graph. Including images in TIA reports is commonly used to represent findings with visual components as well as written and notes could help the receiver to better understand the content of the image.

8.4 Information panel

Figure 35: Image with the information panel in focus

As mentioned earlier in the part about the Network graph the users will be able to access a short summary of information concerning an edge or a node by hovering the element in the graph. During the interviews, it was clear that the TIAs expressed a need to explore as much information and data as they deem necessary to create their intelligence. This is accommodated in a collapsible Information panel on the right-hand side of the graph as shown in figure 35. The panel will initially be open and show information regarding the currently selected node or edge. Whenever a new node or edge is selected by left-clicking it, the panel will be updated with information of what was selected. If the user decides to close the panel it will stay closed until the user decides to re-open it. At this point, this is achieved by clicking the arrow button at the top of the panel but there are multiple possible actions that could be incorporated. For example, a hidden panel could be brought out by double-clicking on an edge or a node, these are interactions that have not been evaluated but are possible future suggestions.

8.4.1 Using the panel to drive investigations

The information contained within the Information panel will be a selection of the information available within an Intelligence Card (figure 36). The panel allows another possibility to pivot within the graph and as such, parts including other entities should be prioritized within the panel. With that in mind, the most obvious parts to show when an entity is selected are the Links, Insikt notes mentioning the entity, and the context section found in its Intelligence Card. Where there are other entities within the Links or other parts in the information panel the users should be able to add these entities to the graph. The new entities should be possible to add both as single entities, and as groups based on the grouping of information in the panel. When evaluating the Information panel the users all agreed that they would still need access to the full Intelligence Card of an entity as they would presumably want to investigate some entities more thoroughly. This will be available in the same manner as everywhere else in the portal where an Intelligence Card is opened as a modal overlay. From an open Intelligence Card, it should also be possible to add entities to the graph similar to adding them via the Information panel. The action should not close the Intelligence Card making it possible to perform multiple actions from the same card. Similar to when elements are added by pivots, new entities and their connections added from Intelligence Cards or the Information Panel will be distinguished, once again according to Nielsen's first and fourth heuristic.

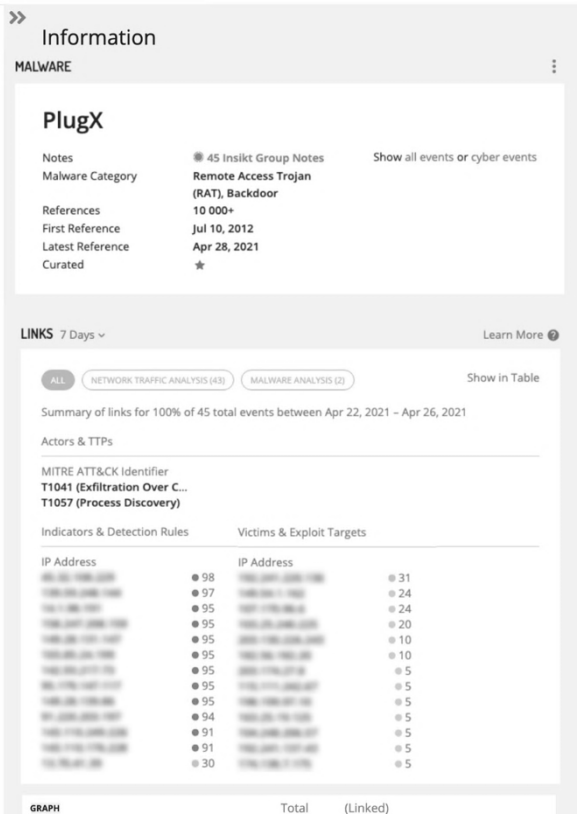


Figure 36: The Information panel with an entity selected showing Risk Score and recent Links

> *"I think it is extremely important that the user can access the entire Intelligence Card rather than only accessing a snippet." - Analyst 3*

> *"The main thing I think about is the ability to understand what is driving a Link between two entities and being able to click an edge to see what's behind that Link. It also helps to drive further pivoting if we can understand what type of connection it is" - Analyst 1*

8.4.2 Limited screen estate

A decision needs to be made whether the panel should be scrollable as that will limit the amount of information that can fit within it. To aid that decision one needs to take into account the information shown when selecting an edge. As mentioned earlier the source with the most confidence should be prioritized and the most confident source at the moment is the Links,

provided as Insikt research notes or network traffic analysis. The network traffic analysis does not require a large amount of space but the Insikt notes will in many cases require more space than what is available with the layout of an Insikt note as it is today (figure 37). As the users will want to explore the connections they deem important, access to the references should be convenient within the information panel. It is likely that where there are Insikt notes there will also be other references connecting the entities and they will be shown as “X additional references” and be available to open in a table like elsewhere in the portal.

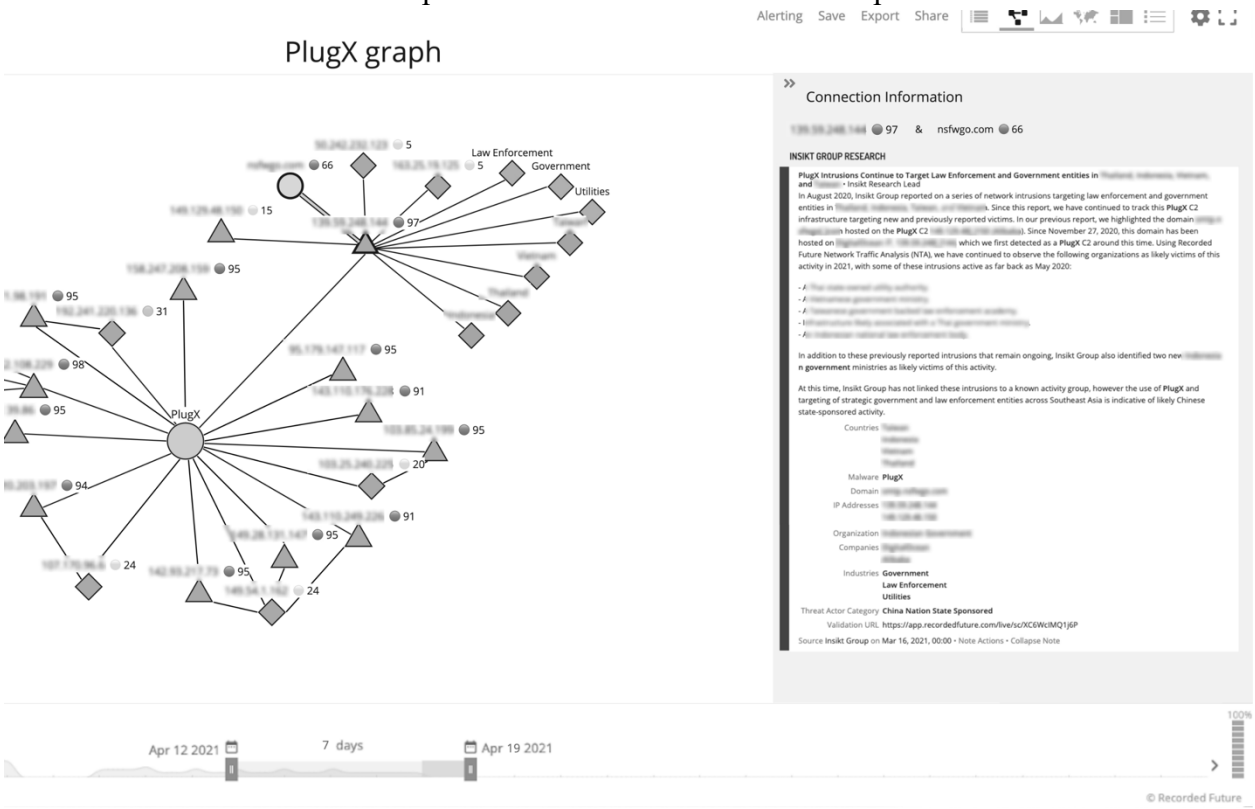


Figure 37: Image showing a selected edge and the Information panel

There is an issue that has not been designed for concerning the Information panel and that is how it behaves when there are multiple entities or edges selected. Possible ways to allow access to information when selecting multiple elements in the graph include stacking them on top of each other in the panel, tabs, or a list within the panel. None of these options has been evaluated with users and requires usability testing to draw any conclusions regarding their usefulness.

8.5 Legend

Figure 38: Image with the Legend in focus

This panel is placed at the left-hand side of the interface, above the History panel. The main element and base of this panel is the legend connected to the Network graph as seen in figure 38. The Legend contains all entity types currently populating the graph and the respective colors they are encoded within the graph. As new nodes and edges are added to the graph through pivoting the legend updates with every step. Updating the indicator of how many nodes there are of each type and adding new headlines if a new entity-type would be added. By updating the legend in accordance with changes in the graph users have a clear overview of the current system status. The same applies when using any of the interaction points within the legend, the feedback of a certain action must be communicated to the users, this in accordance with the first usability heuristic. Having a legend is very common for graphs and diagrams, it aids the users since they can recognize the encoded colors rather than having to remember the meaning of each one. Directly to the right of each entity-type header, there is a number within parentheses indicating the number of entities of that type currently populating the graph. The entity-types are represented as headlines in the list, on their left side, there is a button for expanding the list. When expanded, all entities of said entity type are listed, showing their name and if applicable their risk score. The specific entities are sorted based on multiple factors, firstly their role; adversary, victim, or neutral, if this type of attribution information is available. Then they are sorted high to low based on their risk score, and lastly in alphabetical order if the risk scores are the same or not applicable.

8.5.1 Visual searches

As explained earlier the entity-types are represented as headers in the legend-list, on their left side, there is a button for expanding the list of each type. The button contains a plus- or minus

symbol depending on the current state seen in figure 39. This was decided since it is a common way of designing hierarchical lists and since it follows external consistency.

Clicking a header and thus selecting that specific entity type will highlight all entities of that type in the graph by fading out all other entity nodes. This functionality is used in other elements of the portal and thus aids the internal consistency of the overall product. The state of the graph can be altered by either selecting another header to change which type is highlighted or by clicking in an empty area to de-select, restoring the graph to its normal state. The same functionality can also be achieved through hovering a header but in that case, the graph returns to its normal state as soon as the cursor is moved from the header. This functionality can also be used on specific entities through the same types of interactions. This feature allows epistemic action to help users to locate a certain entity which is especially useful when the graph has grown in size. Highlighting specific entity types can aid users to get an overview of how different types of indicators are more or less associated with different parts of the graph and therefore also to different parts of the investigation.

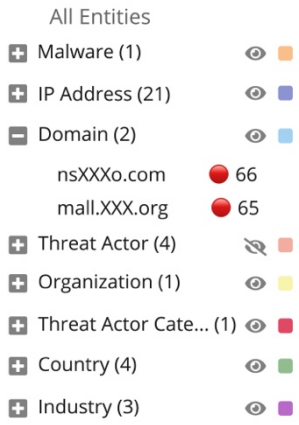


Figure 39: The Legend with the Domain list expanded and the Threat Actor category hidden

8.5.2 Limiting the graph

Aligned to the right-hand side edge of the legend, for each listed item there is a toggle button for hiding or showing said item. Since the choices are binary, to show, or hide, a toggle button is most reasonable and also aligns with external consistency to other types of similar software or use cases. The toggle button is represented by an eye symbol that is either open or closed, this also complying with external consistencies as well as matching users' mental models. This type of button is available both for each specific item but also at the header level which if used applies to all entities of that specific type. Hiding nodes will make them disappear from the graph and therefore unavailable to interact with until shown again. As explained by users in interviews, different types of entities will be more or less interesting to pivot for depending on the starting entity type. This hide-functionality can be useful when investigating a certain path to temporarily reduce the number of nodes, allowing the user to focus on the other remaining nodes and reduce their cognitive load.

8.6 History panel

Figure 40: Image with the History panel in focus

This combination of features and functionalities is arranged as a panel and found in the bottom left corner of the interface, below the legend panel (figure 40). As the name implies, the History panel contains the history of actions taken by the user. All actions leading to the adding or removing of nodes or edges will be represented by a new action-row in the panel. This type of functionality is useful since it allows the user to get an overview of their own or someone else's workflow and can trace their steps in the investigation. The panel also allows the user to regret one or several steps, or to indicate what consequence a certain action had on the graph.

8.6.1 Oh no, go back!

The panel header contains buttons allowing the user to undo the last actions taken and also if undo has been used, the ability to redo actions. This is useful if the user would have used the undo-functionality a step too far. This feature was chosen since it was discovered to be a highly valued functionality mentioned by three out of four users in the interviews. This type of functionality also complies with the third usability heuristic, User control and freedom, by Nielsen as it allows users to undo mistakes. To comply with external consistency it could also be implemented to allow the common keyboard shortcuts to undo and redo which might be useful for users to speed up their workflow. This type of keyboard shortcuts was not investigated during any of the user interviews but is considered useful since time is of high importance for this type of user.

> *I've found sometimes when I'm doing this type of pivoting. I've clicked on something and a ton of stuff pops up and I'm like "oh no, now it's really messy, I wanna go back to where I was" - Analyst 2*

8.6.2 Step-by-Step information

Each action-row, seen in figure 41, contains information about which node it was based on, the type of action performed, what the action resulted in, and lastly, when it occurred. This information was assumed to best represent the action while still limiting it to reduce the risk of cluttering the panel. The amount of information in each row seemed reasonable according to feedback from the interviews. It was not the subject of interest in any specific question, however, the content did not receive any negative feedback from users.

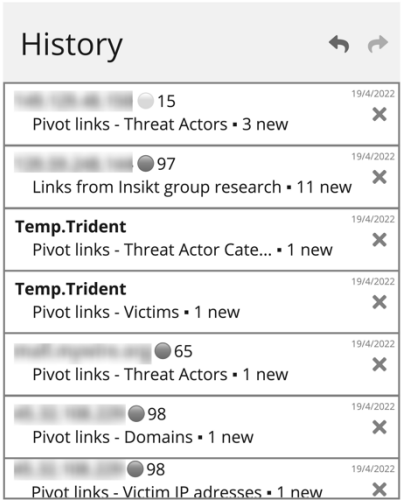


Figure 41: Image showing only the History panel

Each action-row contains multiple functionalities, one of which is the ability to remove an action taken several steps ago without having to manually undo all actions back to that specific one. This is accomplished by the x-button aligned to the right-hand side of the action-row. Having the opportunity to undo or regret actions was a clear demand and need from the users but having the opportunity to back even further or to remove a step in the middle of the chain did not get any clear conclusions during the interviews. Two different options for this functionality were discussed, the first option suggested that removing a step in the middle of the chain would remove just that specific action. This came with some follow-up questions that needed to be answered, what would happen to actions that were taken on nodes resulting from that removed action? Would they also be removed as they were only connected to the removed nodes or would they remain in the graph without having any connection to the other remaining nodes? The other option was that removing a step in the middle of the chain would instead undo all actions subsequent to that one, returning and starting over from that version of the graph. As no conclusions regarding this functionality could be drawn from the user interviews it had to be discussed internally in the project group. The conclusion from the discussion was to keep the functionality and currently suggesting that making such an action should result in the second option described above. Pressing the x-button will send the user back to the version of the graph as it was before that action and undo all actions taken after that point. This was suggested as it was considered to have the most external consistency and to be the easiest to understand. To aid the user in understanding the functionality and help with error prevention, a modal could be used. The modal could indicate to the user that all steps from there on also will be removed and prompt the user to accept this before implementing the action. To aid the user understanding and indicate the system status, nodes and edges which would be removed by such an action could be faded in the graph while the modal is shown. This would indicate what the result of stepping back to that point will result in, aiding the user in the decision-making.

By clicking and thereby selecting an action-row, as in figure 42, the specific action and its result will be highlighted in the graph by fading all non-related nodes and edges. This

highlighting of specific parts of the graph will remain active until de-selected by clicking in an empty area or by selecting something else in the interface. Although some components in the graph are faded they are still accessible and available for selection, thus de-selecting the previous selection. The selected action-row in the history panel is distinguished by color. The same functionality applies when hovering an action-row but then the graph then returns to its normal state as soon as the cursor is moved out of the action-row. This hovering of an action-row is represented by another color variation different from the color for a selected action-row. These color choices for selected respectively hovered items should be the same as in the rest of the portal to keep internal consistency, in accordance with users' mental models and as suggested by the fourth of Nielsen's usability heuristics.

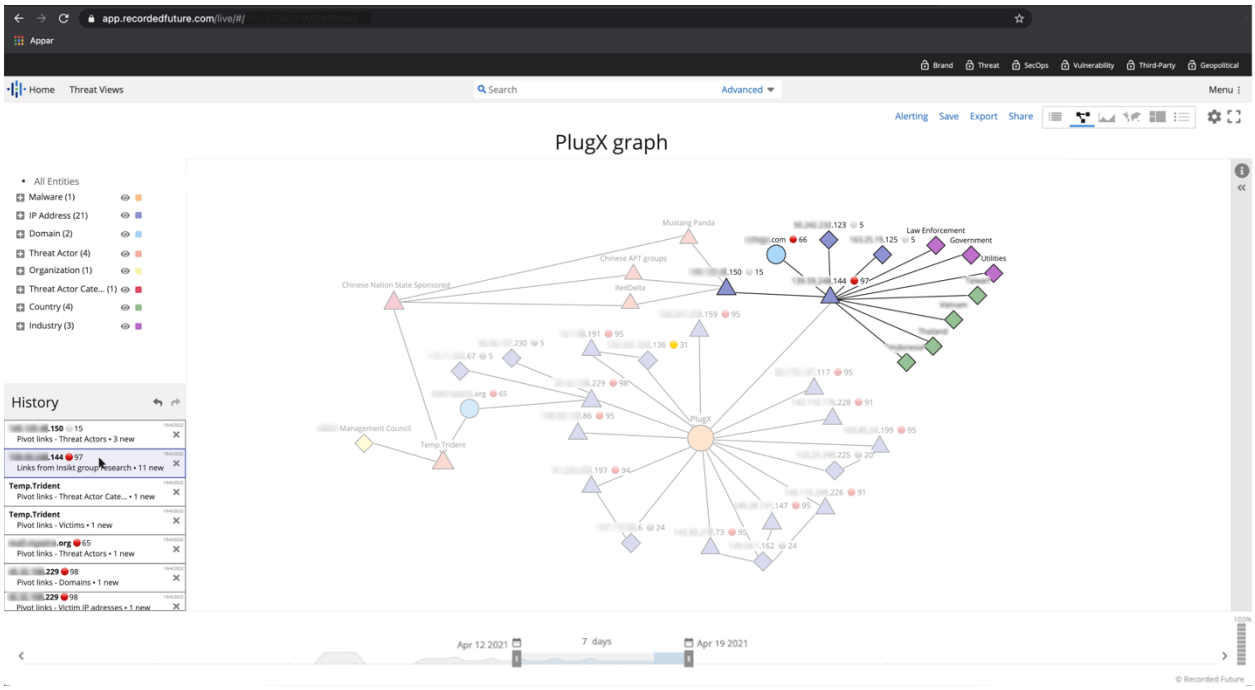


Figure 42: A selected action row with the corresponding pivot highlighted

8.7 Design conclusions

The designs presented in the sections above presented multiple features which were designed to fulfill the project aims. The components held several features, each aiding the overall user experience and raising the ability of the concept to be useful in threat intelligence investigations.

Each of the above-presented components focused on different functionalities and aimed to fulfill the set of requirements defined during the process. The connection between component, functionality, and requirement is presented in the following extended version of the table presented in the second create phase, see table 2.

Component	Functionality	Requirement
Overall design	B. Handling of the AQB within the Network graph H. Time Filtering I. Saving graphs	VII VIII IV, IX
Network graph	A. Visualizing entities and associations E. Visualized attributions for entities & associations	II, III V, VI, VII
Action Menu	C. Manual pivoting between entities	I, IV, VI
Information Panel	D. Access to underlying entity & association information	V, VI, VII, X
Legend	A. Visualizing entities and associations E. Visualized attributions for entities & associations	II, III V, VI, VII
History Panel	F. Ability to undo and discard actions G. Tracking of work path	IV, VIII IV, VIII, IX

TABLE 2, The different parts of the tool, the functions and the corresponding requirements

As described in table 2, each component focused on different functionalities and by that fulfilled different requirements and needs of the users. The overall design contained features enabling user flow between the Network graph and other components of the current system (Figure 43). It inherited the timeline feature from the portal enabling users to limit their investigations by setting a time frame. It also describes how users are allowed to save their progress to provide opportunities to maintain investigations and graphs over longer time periods, as well as to share graphs with others.

The Network graph component suggested how to visualize the two main elements of the graph, the entities in the form of nodes and the associations as edges between them. It also described possibilities in how to visualize attributes assigned to the nodes and edges which could direct user focus and guide their actions. Within the Network graph, the Action menu component could be accessed, this provided the main interaction point for investigation via the pivoting functionality it provides.

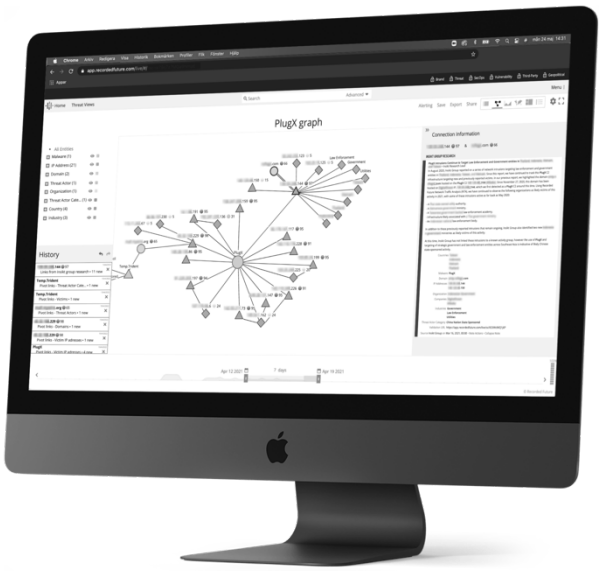


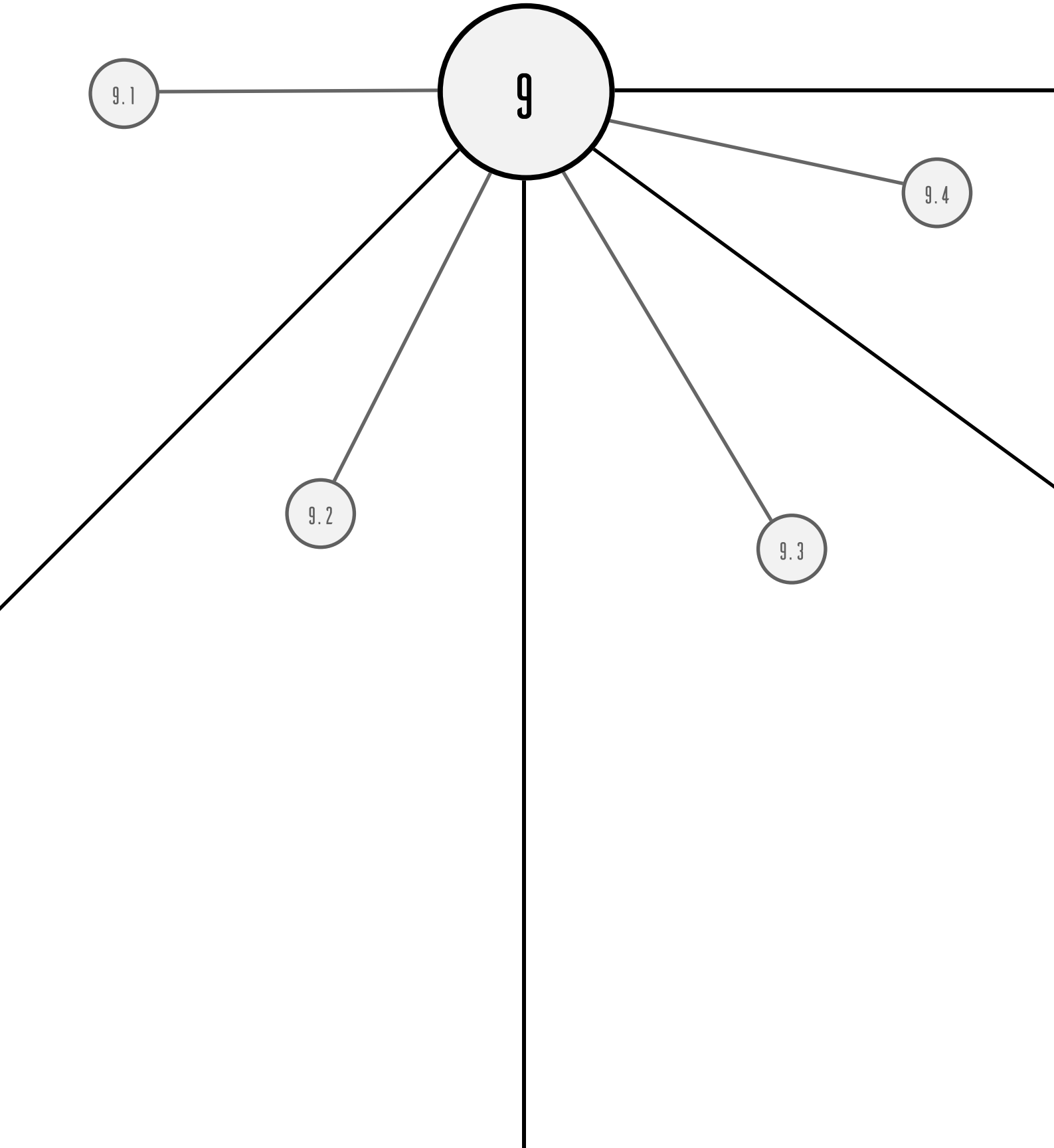
Figure 43: Network graph mockup

Through the Information panel, one of the major insights from user interviews was incorporated, allowing them to inspect what is known about every node or edge currently found in the graph.

The information shown in the panel describes the background to its attributes in the graph and also allows the user to pivot to new associations. It also has a clear relation to crucial platform elements such as the Intelligence Card and allows the user to proceed in whatever part of the portal they prefer.

Having a clear correlation to the Network graph, the Legend component aids the user to interpret the information shown in the graph as well as allowing them to control certain aspects of what is shown. Lastly, the History-panel gives the user a clear overview of the steps and progress of their investigation, as well as allowing them to discard actions by stepping backwards.

RECOMMENDATIONS



9. Recommendations

During the project, multiple design suggestions were constructed and evaluated while it is still clear that some features had to be prioritized lower than others. This section provides the design recommendations for future iterations of the Network graph.

9.1 Beware of clutter

"I can see it both being beneficial but also a little noisy." - Analyst 4

The risk of clutter becoming an issue was always present in every phase of the project. The data contained in the Recorded Future Intelligence Graph is vast and the choice of not limiting what to include will come with a cost. Even though the data might be useful, including it in the graph will increase the amount of information thus increasing the risk of clutter becoming an issue. How to include data in the graph is an interesting design problem as there are really only two main elements in a network graph, nodes, and edges. It is possible to encode information into both nodes and edges but there are limits to how much information can be encoded before it becomes very complex. This leaves the option to put information in text floating in the graph which will inevitably lead to clutter. In the interviews, it was clear that the users wanted to avoid clutter from fearing the graph becoming too noisy.

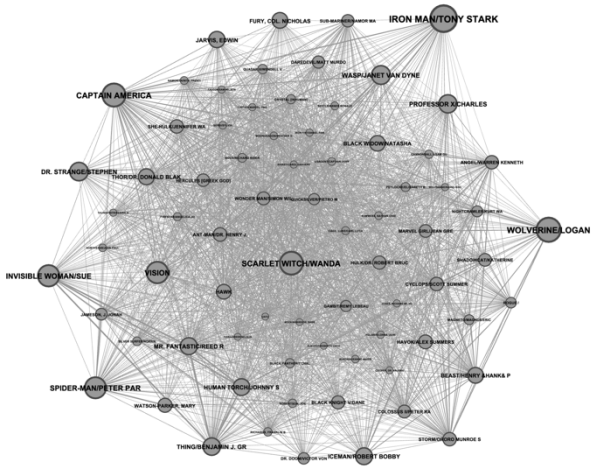


Figure 44: Showing what a cluttered graph could look like using characters from the Marvel universe and their connections. Data retrieved from: https://syntagmatic.github.io/marvel/

When designing the tool careful consideration needs to be put into what data to include and its benefits need to be evaluated and compared against the drawbacks of including it. Possibilities to investigate further include creating different interfaces based on specific use cases or users or the option to customize the graph and let the users decide for themselves what information to include in the graph. Another way to avoid or at least control the amount of clutter is how the graph changes based on its scale. Depending on how far zoomed in or out the user is in the graph the information should adapt which could reduce cluttering. Not only by changing whether icons or text is shown or not, but perhaps by limiting what connections are shown, or by clustering nodes when zoomed out. One option would be to only show low confidence connections when zoomed in below a certain threshold. Whatever direction is chosen when the concept is developed further, clutter and noise in the graph will always require much consideration.

9.2 Begin building

"Code does what you tell it to do" - Senior Architect

A large hurdle when designing was the uncertainty of what was considered possible or not. While it might be true that code will do what it is told to do, it is difficult or impossible to imagine exactly how it will work in reality. To actually see how the Network graph behaves, what data is received when pivoting, and exactly how one can pivot will have a large impact on how the interface should be designed.

Creating the possible interactions for the user to apply filters to the graph will also benefit from closer cooperation with developers. The team will have to explore the multiple ways it is possible to incorporate filters and evaluate their usability. This evaluation should consider both applying one filtering at a time or a possible scenario where there are multiple different filters active simultaneously. As mentioned earlier, different types of entities benefit from different kinds of filters, IP addresses will benefit by a tight time frame, while Threat Actors need a broader. However, there are clear usability issues with such a functionality. When the graph becomes larger the complexity of having multiple different filters active on each pivot becomes difficult to keep track of for the user. There seems to be an advantage of being able to prototype the tool with real data while designing the more complex features.

9.3 The Advanced Query Builder

"I would say that the query builder is a must, most likely." - Analyst 3

Despite being a feature of the current portal that the users thought would, or could, be essential in the graph the AQB was left pretty much on the drawing board after Evaluate 1. This was a conclusion of multiple factors but the major one was the risk of instantly overpopulating the graph as a result of a too broad query. The best way to incorporate the AQB in the current concept is to allow users to import individual results, references or events, from a query into a graph. This means that the users get to decide what events they deem important from the AQB result and to continue investigating the entities in that event.

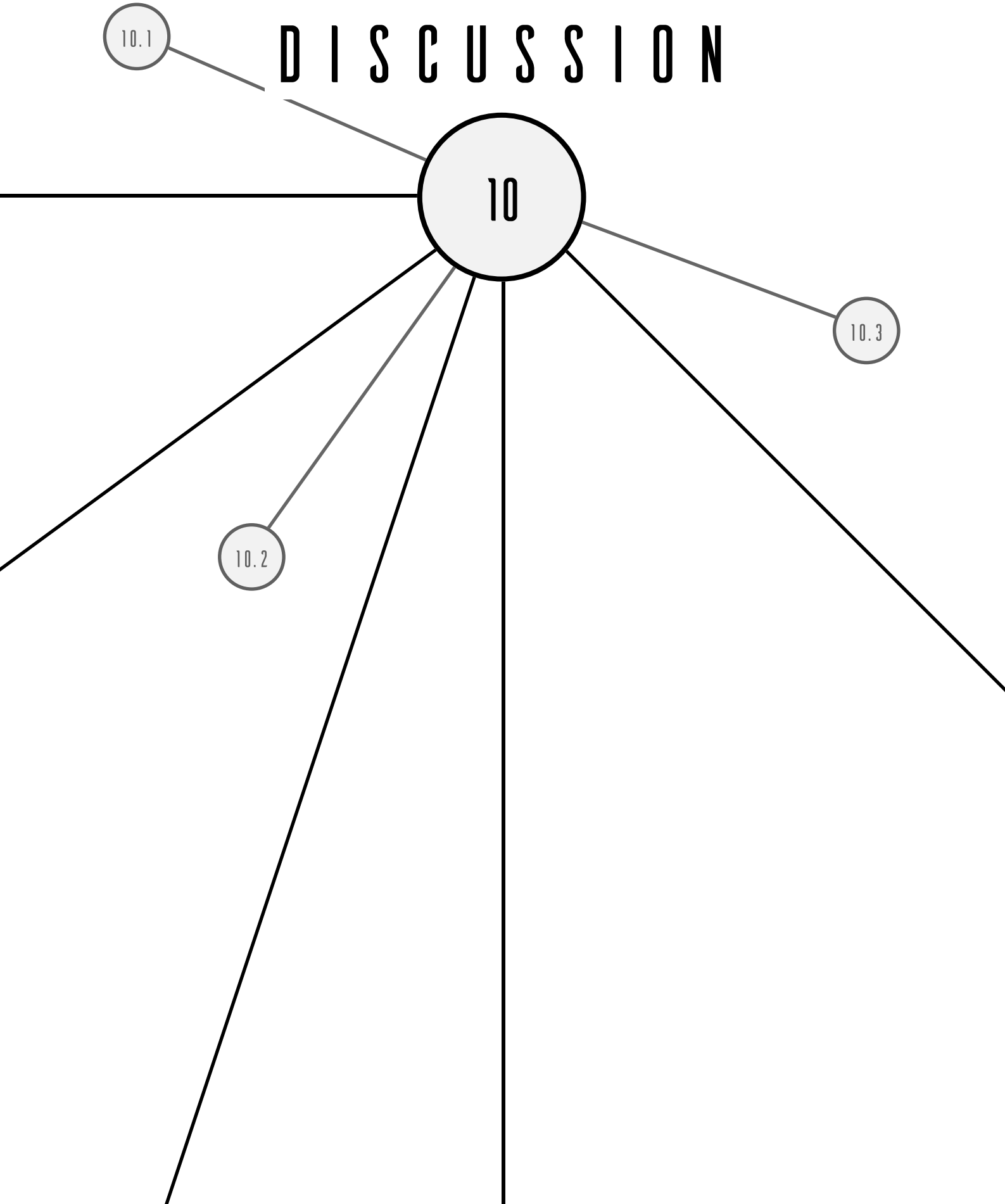
To include access to the AQB within the Network graph it needs to be further evaluated how the users can limit their results, without the risk of losing important information, or how to quickly assess and discard results that are not relevant. Discarding the possibility of incorporating the AQB within the Network graph is not only against requirement VII, providing seamless experience between graph and other portal features. It could also risk a loss of utility for users who are skilled in using the AQB. If incorporated well the ABQ together with the Network graph could form an incredibly powerful tool for analysts both at the beginning of an investigation and during investigations.

9.4 Strategies

Another possible feature of future iterations brought up during the process by technical experts is the idea of Strategies. The concept is that Recorded Future identifies steps or procedures a user will want to perform in all or most investigations or certain sequences of pivots which are common in certain situations. There could be one or multiple strategies able to execute from any entity type to be used in its specific scenario. At first, the "most obvious" strategies can be developed as it would simply be the quickest way to bring more information from the specific entity into the graph. This could be information that will always be available in some quantity and be more or less interesting. With time and usage data, more strategies can be developed for specific kinds of investigations and use cases.

The strategies and how to use them were never the point of focus in the project as they will presumably not be useful in isolation. In order for them to be useful the rest of the tool needs to be there and the users need to be able to conduct the steps themselves as well. In the interviews, the users were somewhat skeptical towards including the strategies. This skepticism originated from the need the TIAs expressed in having to research every reference themselves and make an assessment if they believed in the statement or not. They expressed a fear that a strategy intended for a use case risks missing things or executing pivots to indicators they do not care about. While being mainly skeptical, they could also see the usefulness of the strategies and that they could speed up the process of their work when they are implemented well enough. They also expressed interest in being able to define and set up their own strategies. Something to keep in mind is that the users interviewed were only a small set of users who are already experts in using the platform and expected to quickly become expert users of the Network graph as well. While they might not have the most use of the strategies they would presumably be taking part in developing such strategies to enable other, less skilled users. With time and cooperation, the strategies could become one of the most powerful tools when exploring the Recorded Future Intelligence Graph as it would combine the powers of humans and computer automation.

DISCUSSION



10. Discussion

To understand whether the project achieved its aim and aided towards fulfilling its purpose a discussion will be presented on the subject. This chapter will also explore the process and applications of methods together with the larger impacts the project and the developed Network graph will have in a larger context.

10.1 The design concept

The aim of the project was to create a conceptual design solution to visualize the Recorded Future Intelligence Graph to aid TIAs with their process and decision-making. During the project, the usefulness of such a tool became clear, as well as its potential to aid the users interviewed during the process. The main reason for this conclusion is the General Use Case as defined after the Exploration. The possibility to create a tool that aids the users in one, or most, parts of the presented use case with a quick continuous workflow and access to the Recorded Future Intelligence Graph is apparent. However, the applicability of the General Use Case has to be questioned as it was developed based on the thoughts of only four actual analysts. If it can be applied to the general TIA and not only the ones within Insikt remains to be confirmed. It is not unlikely that the General Use Case might be too general and needs to be more specified to reach its full potential and be as useful as possible. To ensure that the aim is fulfilled the development of the General Use Case needs to continue together with the development of the Network Graph.

The scope of the project made it necessary to, at certain points, simply make design decisions for some parts or features without giving them as much consideration or evaluation to be able to focus on other aspects. One example of such decisions was the encoding of nodes and edges. It was considered necessary to encode information into the nodes in some way but an attempt to find the best way was never made. The encoding of nodes could have been a project in itself and the decision to not spend time evaluating different options allowed the project to proceed further within other aspects. Presenting one possible way of the encoding of nodes was necessary to convey both the overall functionality of the concept as well as the functionality of other more specific features. Without these types of decisions, the project would have stalled with too many options to choose from in several different features. Making decisions like these resulted in some parts of the concept being less developed and evaluated while it allowed further exploration of other parts. The choices of which parts to focus on were based on the fact that they were considered more valuable to evaluate as they could have altered the workflow while other parts such as the encoding of nodes were seen as merely having different visual options which could be postponed to future iterations.

Presented in the Results: Design Solutions chapter is the design concept of how to visualize the Recorded Future Intelligence Graph. While the aim indeed never was to develop more than a concept it is clear that there is much development left before the tool can become a reality. In its current state, it is highly adapted to the preferences of a few users and many of its parts are not yet thoroughly explored nor evaluated. Since no such tool as this is available as of now

in the Recorded Future platform this project aimed to explore possibilities and to form a general standpoint. In regards to the aim, the result achieved is at a level that was both expected and intended and it is considered to lay a firm foundation for future iterations and further development. Further development projects could then be more focused on certain features allowing it to dig deeper into those needs and functionalities and come to more specific conclusions.

10.2 Process disposition and application of methods

A major factor in the result of the project was the time needed and taken for project members to gain an understanding of the Cyber Security and Threat Intelligence domain. From the start, it was understood that it would require time to get the basic knowledge required to interact with and understand the potential users. A considerable amount of time was spent on reading as much as possible to raise the level of understanding. This is considered to have resulted in better interviews as it appeared that the interviewees rarely tried to adapt their answers to an assumed lower level of understanding. Without the time spent to learn the domain this adaptation would absolutely have been necessary. This did however come at the cost of the time spent ideating solutions and designing the concept. It is pointless to speculate on where the project could have ended up if less time was spent understanding and more spent in designing. What can be assumed is that the result would have been different. More time could also have been spent on evaluating the ideas which most considerably would have led to more well-founded conclusions on even more aspects of the final concept. This is something that definitely would be useful to allocate time for if the project would be continued.

10.2.1 User group

During the whole process, only four TIAs were interviewed and they were all employed by Recorded Future. In addition, the four of them were already positive towards the possible implementation of a network graph. They were recruited based on the fact that they expressed interest in the project and in providing input for developing the requirements of such a solution. This will have had a large impact on the result of the project both in how they evaluated the concept but also that they might be a very specific subgroup within TIAs conducting certain types of investigations. It is possible that the General Use Case, presented as one of the major results of this project, is accurate to how the analysts in Insikt work but not representative of a general Threat Intelligence Analyst. While the lack of users to interview was a result of it simply being difficult to get access to users during the project there is certainly a need to do further user research when developing and implementing the Network Graph. One advantage of conducting several consecutive interviews with the same users throughout the project was the fact that less time had to be spent explaining the aim of the project and concept and more time could be spent focusing on the actual features and interactions.

10.2.2 Non-interactive prototypes

During all evaluations of the concepts with users, the workflow and interactions were presented by static wireframes shown in a presentation. The main reason for this choice was limited time,

producing static images would take considerably less time than an interactive prototype. While there are tools that allow designers to develop advanced interactive prototypes they were not suitable for this project as they require large amounts of manual labor. For a prototype of a tool with fewer options and where an assumption can be made that the users would work in a similar or the same manner, it would be preferable. In a product like this Network graph, the assumption was made that each user would approach an investigation in their own way. While there would be similarities, the number of possible paths required would be massive. To construct a prototype simply allowing one pre-decided path of an investigation would by that be a waste of time since it would take longer than producing static frames yet provide the same results. At the same time, producing a prototype allowing all these possible paths was considered to be highly time-consuming which then would have limited the number of iterations. With the decision of just producing static wireframes, it was possible to conduct more evaluations and iterations but it most likely affected the results. If interactions with a prototype could have been observed, assumptions the users make and what actions they would want to execute in a given situation could have been noticed. Now the result of the project is based only on discussions and assumptions based on the work path decided by the project members and presented through the static wireframes. This is one factor to why one of the recommendations for future development is to start building as soon as possible. This would allow more evaluations and observations to gain an in-depth understanding of users' workflow in this type of tool.

10.3 The impact

As one tool of many, the impact of the Network Graph might be small when looking at the Cyber Security field as a whole. Similar to writing code or creating art, there is rarely just one correct answer or a single correct way for TIAs to conduct investigations. Every analyst has their own preferences or has been taught in different ways, impacting what tools they prefer to use. What tools an organization decides to invest in will have an equally large impact. However, the continued development of tools for analysts, SOCs, or any other function within Cyber Security operations will have a large impact when it comes to protecting both organizations and individuals.

As stated in the introduction there is a lack of cyber security professionals globally and this lack needs to be covered in some way. When society's increasing reliance on digital systems the risk connected to these systems increases as well. Any entity, from governments and companies with considerable resources, to start-ups and individuals with far fewer resources, is a potential target of cybercrime. As discussed in the introduction the cybersecurity industry is a multi-billion dollar industry in a constant fight against cyber criminals where both try to find vulnerabilities, one to exploit them and one to minimize them. With enough motive, the threats will continue to innovate to find opportunities to compromise systems and achieve their goals, monetary, ideological, or other. The greatest indicator that there will always be a need for improved cyber security is the fact that many APTs are state-sponsored actors. Their actions are mandated by the motives that the government funding them has. They are not driven by the challenge the hacking itself presents or the need for personal monetary gain, they also risk less

since they are protected by their employing government. As the amount of threats grows and the complexity of challenges increases, the access to great solutions that aid the people defending assets needs to increase in order to stay ahead of these constantly evolving adversaries. On the contrary there is a large risk in making data like Recorded Future's easier to analyze. If threat actors manage to get access to the data, the damage they can cause is massive, especially with the amount of personally identifiable information within the portal. This is the case for regular users as well, as they too get access to this sensitive data both regarding their own organization and others and must therefore act with care and not exploit the information presented to them. While security companies aim to protect others they have to be incredibly well protected themselves in order to not be used as a weapon instead of as a shield.

For a government to protect its citizens or a company to protect its customers and users there is a need for tools that enable them. Ideally, a company could employ enough talent to efficiently combat the threat but in reality, there is a workforce skill gap. Part of the need for a larger workforce is that there are simply too many tasks to complete, too many alerts to research and discard or that demands some sort of action. This means that there is a need for solutions that can provide users with more actionable intelligence quicker to improve the efficiency of the cybersecurity workforce. While the impact of just a single tool might not be as large as one might hope, it is the availability of these tools that allows and aids TIAs to do their work. Software-as-a-Service solutions like the Recorded Future Platform, provides information and intelligence on a scale that few individual organizations are able to conclude on their own. By integrating multiple ways of interacting and presenting this intelligence even more users will be able to benefit from it. These services and solutions offer opportunities enabling smaller companies that cannot host the competence internally to still keep track of their threat landscape and protect themselves. As all of this keeps developing the Recorded Future Intelligence Graph will improve and the Network graph will be improving along with it. In the end the development of the Network graph will aid TIAs in making the world a little bit safer.

CONCLUSION

11

11. Conclusions

Threat Intelligence Analysts need to analyze and communicate complex information to ensure or improve the security of their organizations or clients. Cyber security is a constant race between cyber security teams and malicious actors. For a security team to be able to stay ahead of their malicious adversaries they need tools that aid their decision-making process. For the TIA those tools provide them with data and information that they can turn into actionable intelligence. The purpose and aim of the project were to create a conceptual design for a solution that would aid TIAs in their investigations by visualizing the relationships between entities.

When creating a concept in the early development of a tool it was beneficial to talk to the same users throughout the process. In this stage of development it was useful that they were both in the loop and invested in the project. This allowed comparisons between the evaluations and kept the project moving forward while exploring a fair amount of opportunities. The result of the thesis consists of a defined use case generalizing the workflow of a TIA which then was supported by the design concept. Although every investigation differs, according to the users they all go through the five steps defined in the general use case. Defining a use case was greatly beneficial for the project as it specified the needs and tasks the solution should aid the user to complete. The developed design concept constitutes a foundation and guidance to build a tool that will aid TIAs by allowing visual exploration of the Recorded Future Intelligence Graph. The concept complements existing tools by providing a visual aid for analyzing the attributions and relationships between entities. For both analysis and communication, this concept can speed up and improve the work of a Threat Intelligence Analyst aiding them in the process of interpreting, forming and communicating actionable intelligence.

As stated by Ben Shneiderman, (2014, 730):
"...visual strategies for exploring complex data lead to more potent and meaningful insights."
Within the context of cyber security more potent and meaningful insights means keeping more assets protected, less money lost and ultimately more lives saved.

References

Bianco, D. J. (2013, 3 1). *The Pyramid of Pain*. detect-respond. Retrieved 2, 2021, from http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html

British Design Council. (2005). *A study of the design process*. Retrieved 2, 2021, from https://www.designcouncil.org.uk/sites/default/files/asset/document/ElevenLessons_Design_Council%20(2).pdf

Budiu, R. (2014, 7 6). *Memory Recognition and Recall in User Interfaces*. nngroup. Retrieved 3, 2021, from https://www.nngroup.com/articles/recognition-and-recall/

Caltagirone, S., Pendergast, A., & Betz, C. (2013, 7 5). *The Diamond Model of Intrusion Analysis*. US Department of Defense. Retrieved 2, 2021, from https://apps.dtic.mil/sti/pdfs/ADA586960.pdf

Check Point. (N/D). *What Is Sandboxing?* Checkpoint. Retrieved 5, 2021, from https://www.checkpoint.com/cyber-hub/threat-prevention/what-is-sandboxing/

Cimpanu, C. (2021, 5 18). Darkside gang estimated to have made over $90 million from ransomware attacks. *The Record*. https://therecord.media/darkside-gang-estimated-to-have-made-over-90-million-from-ransomware-attacks/

Cooper, A., Reimann, R., Cronin, D., & Noessel, C. (2014). *About Face : The Essentials of Interaction Design* (4th ed.). John Wiley & Sons, Incorporated. 9781118766576

De Groot, J. (2020, 11 25). *What is a Security Operations Center (SOC)?* Digitalguardian. Retrieved 5, 2021, from https://digitalguardian.com/blog/what-security-operations-center-soc

DSDM Consortium. (2003). *DSDM: Business Focused Development* (J. Stapleton, Ed.; 2nd ed.). Pearson education limited.

Dunne, C., & Shneiderman, B. (2013, 4). *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Retrieved 1, 2021, from https://dl.acm.org/doi/10.1145/2470654.2466444

Fargnoli, M., Rovida, E., & Troisi, R. (2006). *THE MORPHOLOGICAL MATRIX: TOOL FOR THE DEVELOPMENT OF INNOVATIVE DESIGN SOLUTIONS.* Retrieved 5, 2021, from https://axiomaticdesign.com/technology/icad/icad2006/icad2006_21.pdf

Friis Dam, R., & Yu Siang, T. (2021, 1). *How to Select the Best Idea by the end of an Ideation Session*. interaction-design. Retrieved 4, 2021, from https://www.interaction-design.org/literature/article/how-to-select-the-best-idea-by-the-end-of-an-ideation-session

Gartner. (2013, 5 16). *Definition: Threat Intelligence*. Gartner. Retrieved 1, 2021, from https://www.gartner.com/en/documents/2487216/definition-threat-intelligence

Gibbons, S. (2016, 7 31). *Design Thinking 101*. nngroup. Retrieved 2, 2021, from https://www.nngroup.com/articles/design-thinking/

Google LLC. (N/A). *Solution Sketch*. https://designsprintkit.withgoogle.com/methodology/phase3-sketch/solution-sketch. Retrieved 3, 2021, from https://designsprintkit.withgoogle.com/methodology/phase3-sketch/solution-sketch

Gundert, L. (2020). *The Risk Business*. CyberEdge. 978-1-948939-13-3

Harley, A. (2018, 7 3). *Visibility of System Status (Usability Heuristic #1)*. nngroup. Retrieved 3, 2021, from https://www.nngroup.com/articles/visibility-system-status/

Hutchins, E. M., Cloppert, M. J., & Amin, R. M. (N/D). *Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains*. Lockheed Martin. Retrieved 2, 2021, from https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf

Interaction Design Foundation. (N/A). *What is Wireframing?* interaction-design. Retrieved 3, 2021, from https://www.interaction-design.org/literature/topics/wireframing

ISC2. (2020). *Cybersecurity Professionals Stand Up to a Pandemic* [Cybersecurity Workforce Study].

Kirsh, D., & Maglio, P. (1994, 10). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, *18*(4), 513-549. https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog1804_1

Krause, R. (2021, 1 10). *Maintain Consistency and Adhere to Standards (Usability Heuristic #4)*. nngroup. Retrieved 3, 2021, from https://www.nngroup.com/articles/consistency-and-standards/

Laubheimer, P. (2015, 3 23). *Preventing User Errors: Avoiding Unconscious Slips*. nngroup. Retrieved 3, 2021, from https://www.nngroup.com/articles/slips/

Lee, R. M. (2017). *Threat Intelligence and Me*. N/D. 978-1541148819
MacGregor, R. (2015, 5 29). *Diamonds or chains*. pwc.blogs. Retrieved 4, 2021, from https://pwc.blogs.com/cyber_security_updates/2015/05/index.html

Nielsen, J. (1994, 4 24). *10 Usability Heuristics for User Interface Design*. nngroup. Retrieved 4, 2021, from https://www.nngroup.com/articles/ten-usability-heuristics/

Nowell, L., Schulman, R., & Hix, D. (N/A). *Graphical Encoding for Information Visualization: An Empirical Study*. Retrieved 3, 2021, from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1173146&tag=1

Porkorny, Z., Barysevich, A., Gundert, L., Liska, A., McDaniel, M., & Wetzel, J. (2019). *The Threat Intelligence Handbook* (2nd ed.). CyberEdge Group.

Purplesec. (2021). *2021 Ransomware Statistics, Data, & Trends*. purplesec. Retrieved 4, 2021, from https://purplesec.us/resources/cyber-security-statistics/ransomware/

Recorded Future. (2017, 3 8). *Threat Intelligence, Information, and Data: What Is the Difference?* RecordedFuture. Retrieved 01, 2021, from https://www.recordedfuture.com/threat-intelligence-data/

Recorded Future. (2021). *Platform*. RecordedFuture. https://www.recordedfuture.com/platform/

The Recorded Future Team. (2020, 7 9). *How Analytic Frameworks Lay the Groundwork for Applying Security Intelligence*. RecordedFuture. Retrieved 3, 2021, from https://www.recordedfuture.com/analytical-threat-intelligence-frameworks/

Rosala, M. (2020, 11 29). *User Control and Freedom (Usability Heuristic #3)*. nngroup. Retrieved 3, 2021, from https://www.nngroup.com/articles/user-control-and-freedom/

Rynes, A., & Bjornard, T. (2011, 7). *Intent, Capability, and Opportunity: A Holistic Approach to Addressing Proliferation as a Risk Management Issue*. Idaho National Library. Retrieved 3, 2021, from https://inldigitallibrary.inl.gov/sites/sti/sti/5223019.pdf
Salama, V., Marquardt, A., Mascarenhas, L., & Cohen, Z. (2020, 10 29). Several hospitals targeted in new wave of ransomware attacks. *CNN*.

Sharp, H., Preece, J., & Rogers, Y. (2019). *Read Online Download Book Add to Bookshelf Share Link to Book Cite Book Interaction Design : Beyond Human-Computer Interaction* (5th ed.). John Wiley & Sons, Incorporated. 9781119547259

Shneiderman, B. (2014, 2 14). The Big Picture for Big Data: Visualization. *Science*, *343*(6172), 730. 10.1126/science.343.6172.730-a

Shneiderman, B., & Plaisant, C. (2005). *Designing the User Interface* (4th ed.). Person Education. 0-321-19786-0

Shneiderman, B., & Plaisant, C. (2015, 5 22). Sharpening Analytic Focus to Cope with Big Data Volume and Variety. *IEEE Computer Graphics and Applications*, *35*(3), 10-14. 10.1109/MCG.2015.64

Slowik, J. (2020, 11 20). *Current Events to Widespread Campaigns: Pivoting from Samples to Identify Activity*. Domaintools. Retrieved 4, 2021, from https://www.domaintools.com/resources/blog/current-events-to-widespread-campaigns-pivoting-from-samples-to-identify

Threat Analysis Group. (N/A). *Threat, vulnerability, risk – commonly mixed up terms*. Threatanalysis. Retrieved 03, 2021, from https://www.threatanalysis.com/2010/05/03/threat-vulnerability-risk-commonly-mixed-up-terms/

Treffinger, D. J., Isaksen, S. G., & Stead-Dorval, K. B. (2006). *Creative problem solving : an introduction* (4th ed.). Waco, Tex. : Prufrock Press.

Treisman, A. (1985, 8). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*(2), 156-177. https://www.sciencedirect.com/science/article/pii/S0734189X85800049

Ware, C. (2013). *Information Visualization* (3rd ed.). Elsevier.

Waters, K. (2009). *Prioritization using MoSCoW*. Retrieved 4, 2021, from https://cs.anu.edu.au/courses/comp3120/local_docs/readings/Prioritization_using_MoSCoW_AllAboutAgile.pdf

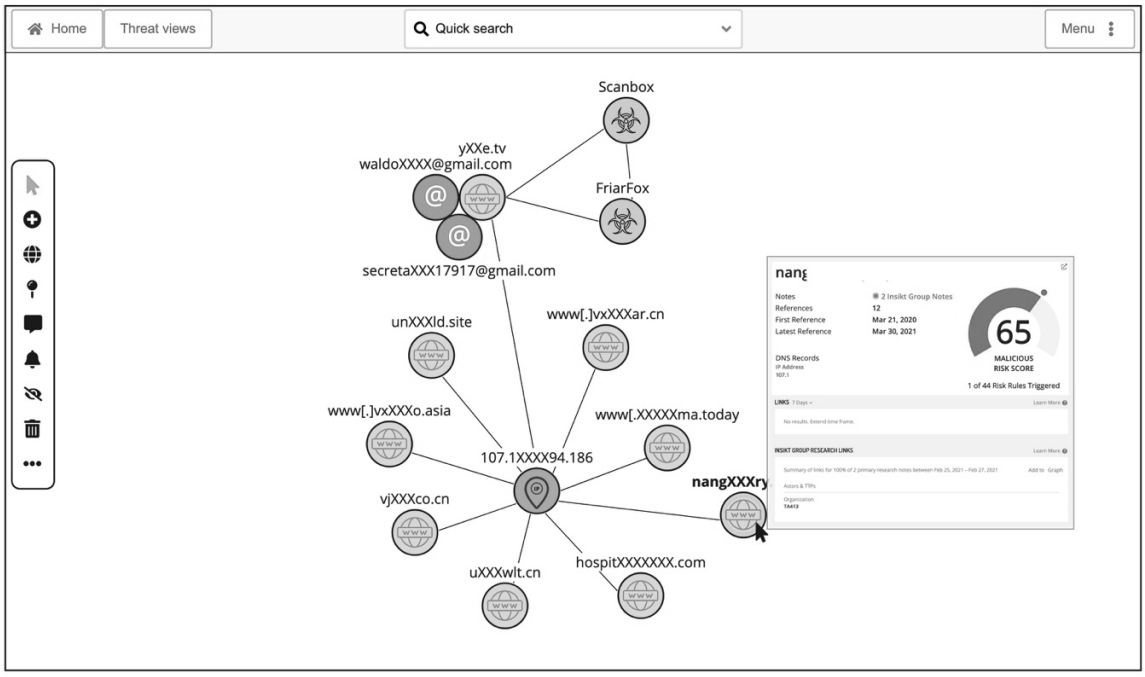
Whitenton, K. (2018, 3 11). *The Two UX Gulfs: Evaluation and Execution*. nngroup. Retrieved 3, 2021, from https://www.nngroup.com/articles/two-ux-gulfs-evaluation-execution/

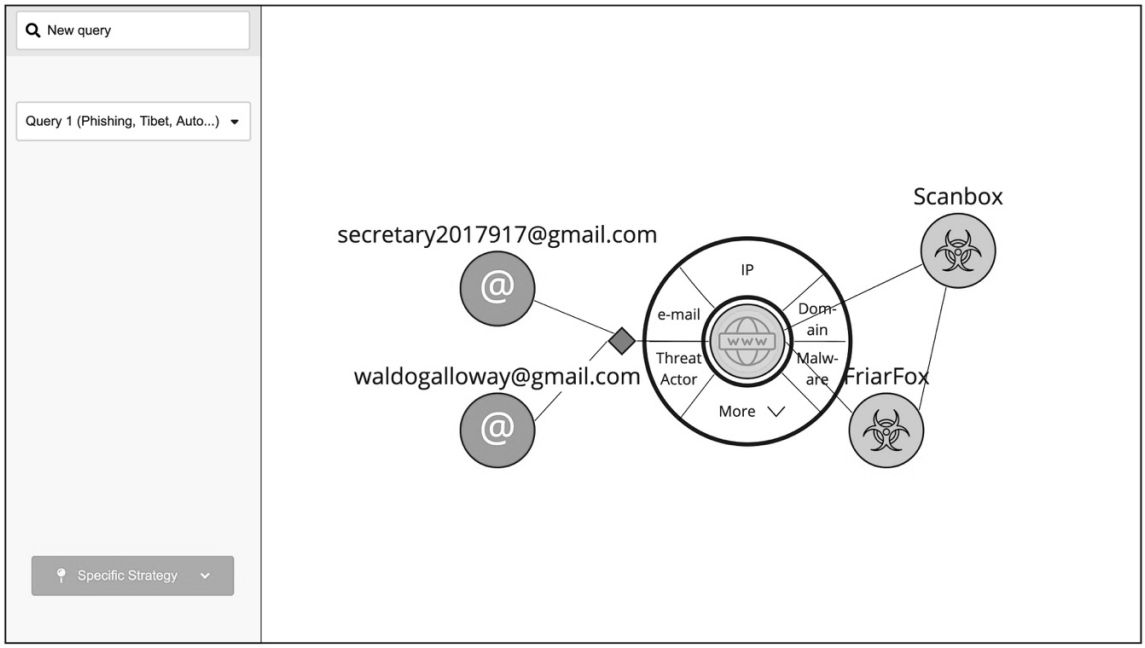
Wilson, C. (2013). *Brainstorming and Beyond: A User-Centered Design Method*. Elsevier Inc.

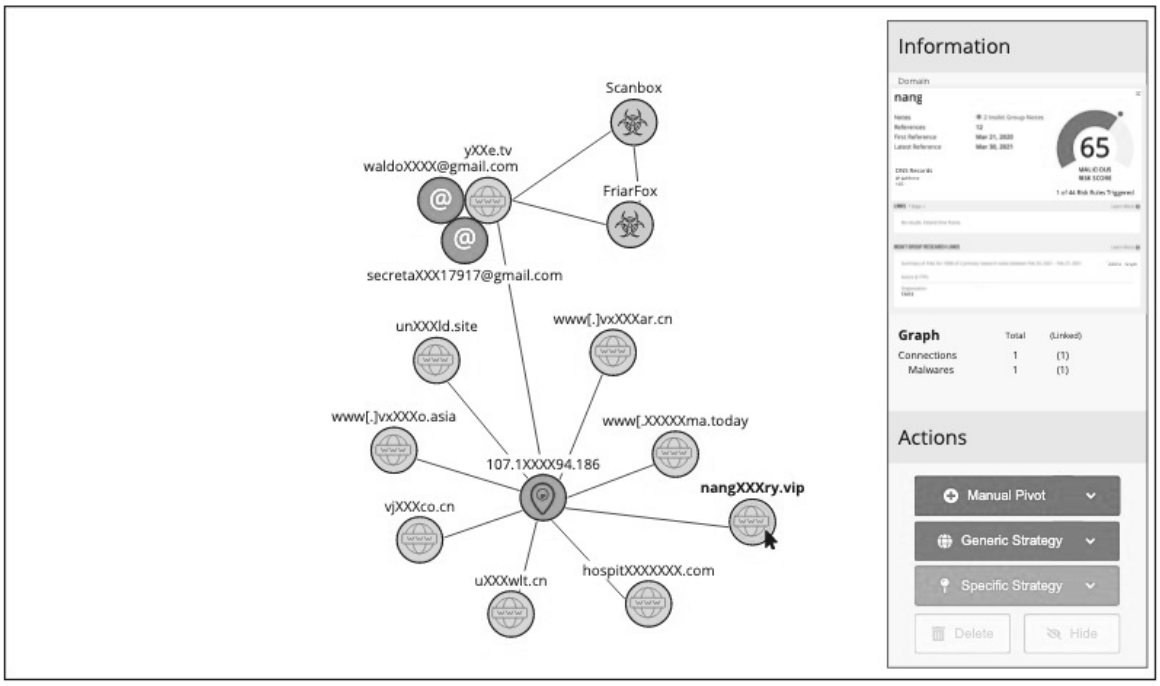
Wood, L. E. (1997, 3). Semi-structured interviewing for user-centered design. *Interactions*.

Appendix

Appendix 1






Scanbox
yXXe.tv
waldoXXXX@gmail.com
FriarFox
secretaXXX17917@gmail.com
unXXXld.site
www[.]vxXXXar.cn
www[.]vxXXXo.asia
www[.XXXXXma.today
107.1XXXX94.186
nangXXXry.vip
vjXXXco.cn
uXXXwlt.cn
hospitXXXXXXX.com
Information
Domain
nang
65
Graph
Total
(Linked)
Connections
1
(1)
Malwares
1
(1)
Actions
Manual Pivot
Generic Strategy
Specific Strategy
Delete
Hide

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY