Towards Designing Effective Visualizations for DNS-based Network Threat Analysis

Rosa Romero-Gomez, Yacin Nadji, and Manos Antonakakis
School of Electrical Engineering and Computing
Georgia Institute of Technology
{rgomez30,yacin,manos}@gatech.edu

ABSTRACT

As threat detection systems become critical for protecting modern organizations, visualization has emerged as an essential tool for security analysts to understand network threats. However, there is currently little research in designing and evaluating effective network threat analysis visualizations. To address this problem, we take a *user-centered approach*, starting with designing an open source threat analysis console for DNS-based network threat analysis grounded in both an understanding of analysts' needs and tasks and security visualization best practices. The proposed open source threat analysis console, called THACO (THreat Analysis COnsole), leverages open DNS datasets, domain WHOIS records, and both public malware and domain blacklists. It also uses a visually scalable visualization technique, a *multi-grouping, zoomable treemap*, to adapt to DNS-based network threat analysis needs. Then, we conduct a user study with 7 in-situ and 31 online IT security practitioners. Our code for THACO and THACO itself will be opened to the community in order to further improve the ability of analysts to perform network threat analysis and better secure their networks.

Index Terms: H.5.2 [Information Interfaces & Presentations]: User Interfaces - Graphical User Interfaces (GUI); C.2.0 [Computer-Communication Networks]: GeneralSecurity and Protection; I.3.8 [Computer Graphics]: Applications

1 INTRODUCTION

Automatically detecting malicious activity is a fundamental part of network security. It ranges from detecting intrusions with signatures [22, 35] to detecting malicious behavior with machine learning approaches [12, 34]. While these systems generate alerts automatically, the presence of false positives and the severity of the base-rate fallacy in detection systems [14] still require a human to be "in the loop" to ensure detection errors do not interrupt normal activity. Accordingly, known best practice [7] dictates tiers of security analysts (hereafter, analysts) to review alerts, where additional context outside of the detection system is used to draw inferences about the actual enemy identities, motives, and sponsorship, which has been conceptualized as *network threat analysis* [20].

Fundamental to network threat analysis is reviewing data from network threat intelligence sources such as domain WHOIS records, DNS (Domain Name System) datasets, and many public and private tools and websites in order to discern attack patterns at the security community level. These intelligence sources must be accessed manually through separate tools and procedures, collated, and correlated back to the original data, which makes network threat analysis more challenging. To overcome these issues, analysts use their knowledge and personal experience with similar security events to establish the severity of the security event [26, 16, 41, 40]. Nevertheless, relying on the experience and knowledge of analysts sacrifices efficiency, effectiveness, scalability, consistency, and visibility. As networks grow and security threats increase, organizations are hard pressed to find analysts that posses the requisite expertise to immediately accomplish effective network threat analysis.

One key approach to scale network threat analysis is visualization, which has been characterized as key due to its ability to highlight patterns and anomalies in large amounts of data [23, 21, 19, 40]. Moreover, the support of visualization tools can foster expertise in novices, help experts to generate ground truth that can be shared with the security community, and used to train statistical models. In this paper, we focus on assisting analysts to perform network threat analysis through visualization of multiple heterogeneous network threat intelligence sources. Particularly, we accomplish this by leveraging open DNS datasets, domain WHOIS records, and both public malware and domain blacklists.

In summary, this paper makes the following contributions: **1)** We develop THACO, an open source threat analysis console[1] focused on assisting analysts performing DNS-based network threat analysis; **2)** We evaluate both its utility and usability with real-world analysts. We show that the threat analysis experience of participants affects neither task completion rates nor task completion times. We also show that experienced analysts satisfaction garnered our tool an "A" grade based on our usability surveys, which means they found it very easy and pleasant to use; and **3)** We open the code of THACO and THACO itself to improve and further develop the security community's ability to perform DNS-based network threat analysis. The code used in this paper is publicly available at the Active DNS project website[2]. The data are publicly available at this website under request.

2 RELATED WORKS

Tools for supporting DNS-based network threat analysis can be divided into two different groups according to their use or not of visualization techniques. Some relevant freely available examples of the first group are ThreatCrowd [3] and DNSViz [10]. These tools mainly use non-interactive node-link graphs as a visualization technique to support investigations between network threat indicators such as IP addresses and domain names from open source feeds. Similarly, Maltego [4] and PassiveTotal [11], commercially available tools (although they also provide limited community editions), make use of node-link graphs and calendar visualizations for visualizing DNS querying behavior. Nevertheless, as stated in previous research [25], the readability of node-link graphs deteriorates when the size of the graph is bigger than twenty vertices, hampering the scalability of this technique for network threat analysis. As examples of the second group, ThreatMiner [5] is a free data aggregator that relies on data tables to support both pivoting and data en-

[1] http://ipviz.gtisc.gatech.edu/. Disclaimer: The authors asked for a demo of the open source threat analysis console, THACO, to be hosted on the Active DNS Project website to support the exploration of its features by the security community.

[2] http://www.activednsproject.org/

978-1-5386-2693-1/17/$31.00

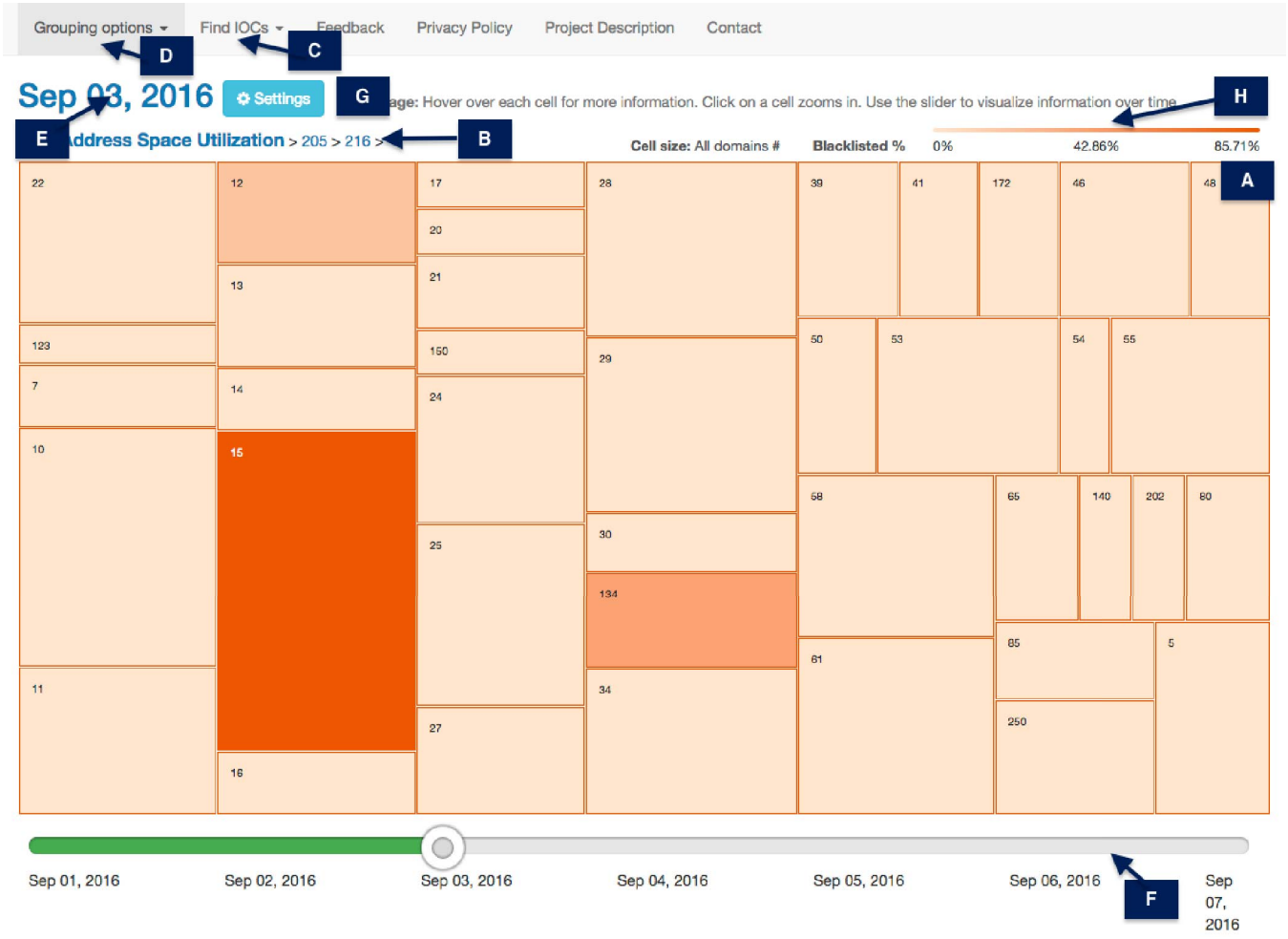

Figure 1: The user interface of THACO displaying networks under 205.216.0.0/16 seen on September 3th, 2016. 205.216.15.0/24 contains the highest percentage of blacklisted domain names according to the cell color and cell size.

richment around threat indicators such as IP addresses and domain names. Similarly, AlienVault OTX [1] uses data tables to display community-generated threat data. Data tables are essential for reviewing the raw data; however, they do not facilitate the discovery of more general patterns, which prevent analysts from taking action.

THACO differs primarily from the tools described above in that it is: more visually scalable, designed for heterogeneous data, and developed for experts based on input from real-world analysts. Our design uses both interviews with experts and a user study with real-world analysts.

3 THACO: Open Source THreat Analysis COnsole for DNS-based Network Threat Analysis

3.1 Domain Problem and Data Characterization

The first step to design suitable visualizations for DNS-based network threat analysis must be to understand analysts' goals and needs. Aiming at drawing relevant quotes from experts referring to both tasks and data needs, we first conducted diverse informal interviews with two domain experts in DNS-based network threat analysis over the period of two months. During this period, diverse prototypes were also tested iteratively by these experts to understand the requirements. We complemented this information with a review of previous work focused on both studying the role of security analysts [26, 28, 20, 18, 30] and identifying malicious uses of DNS [12, 13, 42]. Based on the acquired understanding of the DNS-based network threat analysis practice, we identified a set of data requirements and tasks to be assisted by our threat console. These data requirements and tasks, are divided in two main high-level categories: *bottom-up threat analysis* and *top-down threat analysis*.

1. **Bottom-up threat analysis**. With a more reactive nature, bottom-up threat analysis can be characterized as the most traditional way to perform DNS-based network threat analysis. The goal of this analysis is to research the context of a security event related to an IP address or a domain name triggered by a network detector. Our research revealed nine important data requirements: 1) the total number of IPs historically associated with a domain name; 2) the total number of domain names historically associated with an IP; 3) the geographical location of an IP address; 4) the Autonomous System (AS) associated with an IP, which is a connected group of one or more IP prefixes run by one or more network operators; 5) the number of distinct malware samples that connected to an IP; 6) the number of new domains that connected to an IP from the previous day; 7) the number of domain names that are listed in public blacklists; 8) the registration dates for a domain name; and 9) the distinct registrars associated with a domain name.

2. **Top-down threat analysis**. Modern threats are persistent, flexible, and often evade network defenses. Rather than wait for security events to be triggered by a network detector, analysts actively search for unknown threats to prevent or minimize damage. This analysis, therefore, relies on data coverage of the threat, and the ability to navigate through multiple dimensions of data easily. In other words, analysts need to get an overview of the activity in the network from multiple perspectives in order to discover potential threats. Data requirements this category include: 1) overview of the number of blacklisted domain names by IP prefix, which is a range of IP addresses that can correspond to one or more networks; 2) overview of the the number of blacklisted domain names by ASs; and 3) overview of the number of malicious domain names by geographical location.

3.2 Datasets

To address the data requirements presented above, THACO uses active DNS datasets[2] in combination with public domain blacklists such as Abuse.ch [8], malware blacklists such as Malware DL [9], and domain WHOIS records from VirusTotal [6]. Due to passive DNS datasets are challenging to collect and often rquire restrictive legal agreements, we decided to use open, freely available DNS datasets that can help increase the situational awareness around modern threats. Further details regarding active DNS datasets are described in [29].

3.3 Visualization and Interaction Design

According to previous design guidelines, frameworks, and recommendations for security visualization [32, 27, 39], monitoring tools typically require low interaction, while tools designed for analytical tasks require significantly more interactive activity. It particularly requires the ability for analysts to have multiple views of the same or related data, as well as several levels of detail.

Following these recommendations and aiming to assist the aforementioned two modes of DNS-based network threat analysis, in THACO, we choose to use a *multi-grouping, zoomable treemap* (see Figure 1.A) over other visualization methods for several reasons. Traditional treemaps allow the user to navigate the data in a hierarchical manner, which enables fast and scalable exploration of large data sets [31]. Typically, however, treemaps draw all levels of the hierarchy at once, inhibiting readability and scalability for extremely large data sets. Our treemap uses interaction to support top-down threat analysis and reveals the information incrementally: clicking on a cell zooms in. For our threat console, readability was a secondary concern to navigation, hence a treemap is preferable to alternatives such as wrapped bar charts.

This treemap is also supported by the component *breadcrumbs* (see Figure 1.B), which summarize navigation paths and, in the case of the IP hierarchy, allow analysts to explore IP neighbors. On the other hand, in order to support bottom-up threat analysis tasks, THACO also provides a search feature for specific IP addresses and domain names included on the top-level menu of the user interface and called "Find IOCs"(see Figure 1.C). Analysts can also take advantage of this search feature for top-down analyses in order to access to specific levels of detail in the treemap such as a specific IP prefix (e.g. 104.16.106.0/24). The treemap also works well because its ability to effectively make use of both size and color for encoding different data attributes. By default, the size of each rectangle in the treemap is proportional to the total number of domain names resolving to a specific IP prefix; and the color represents the percentage of blacklisted domain names that resolve to that IP prefix. However, in order to fully support the data requirements for top-down analysis tasks, we also provide two additional grouping options for the treemap accessible through the top-level menu of the user interface (see Figure 1.D) that allow to visualize both an overview of the number of blacklisted domain names by ASs and an overview of the number of blacklisted domain names by geographical location. To navigate across days, users can either click on the date label on top of the treemap (see Figure 1.E) or use the time slider located below the treemap (see Figure 1.F). This time slider was implemented as a result of participants' feedback collected during the user study further described in Sections 4 and 5. Using an adjustable panel that can be accessed through the "Settings" button on the user interface (see Figure 1.G), analysts are also able to set up three different visual encodings for the treemap such as *cell size*, *cell color*, and *cell size scale*. The color scales provided are created using ColorBrewer [2] to follow visualization best practices by including a colorblind-safe version. Finally, in order to guide analysts to the cells with highest and lowest percentages of blacklisted domain names, we also display a color legend on the top-right side of the user interface (see Figure 1.H).

3.4 Use Cases

Next, we present two example use-cases that were carried out by two fictional analysts, Dirk and Kate. The key goal with this section is to illustrate how analysts would perform bottom-up and top-down threat analysis tasks on THACO. Dirk's task is to research the context of a security event triggered by a network detector. Kate will "hunt" [30] for undetected threats that use her company's hosting for malicious intent, using the console as a guide.

3.5 Bottom-up Threat Analysis

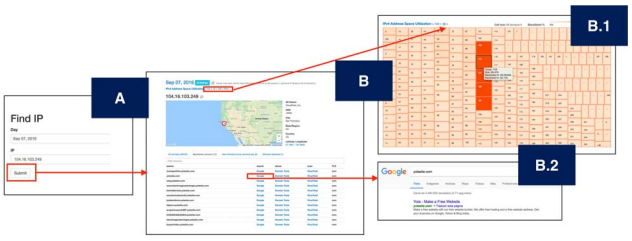

Figure 2: Screenshots of THACO showing the steps taken by Dirk to identify geographical, infrastructure, and reputation-related information of 104.16.103.249/32 on September 7, 2016.

On September 7, 2016 Dirk receives an alert from a reliable detector that malware was downloaded to a machine on his network from 104.16.103.249 resolved by the domain "$otoy.yolasite[dot]com$." In Figure 2.A, he enters the IP address and the threat console displays relevant information that includes: geographic information such as origin city, state, and country; infrastructure information such as related historical domain names; and reputation information such as blacklisted, or newly created, domains and malware lookups (see Figure 2.B). Dirk notices that 38,124 unique domains that have pointed to this IP, 12 domains are listed on public blacklists, and one malware sample has contacted this IP. Next, he sees that all of the blacklisted domains are child labels of the zone $yolasite[dot]com$ and appear to have unrelated third-level labels. Dirk then clicks on the automatically generated Google Search for $yolasite[dot]com$ to get some background information on this domain. He notices that Google describes the domain as a "free website builder", and using his earlier finding, guesses that users' websites are distinguished by child labels domains. Dirk verifies that this creates a unique child label under the $yolasite[dot]com$ zone. Due to the previous findings, Dirk concludes that a malicious free website could have delivered malware to an internal machine, but the whole zone is unlikely to be entirely malicious, so he removes $yolasite[dot]com$ from the blacklist, leaving only the malicious child domains.

Dirk then wants to expand his knowledge to identify other threats hosted by $yolasite[dot]com$. From the same view, THACO also shows him an overview of related infrastructure using a zoomable treemap accessed via breadcrumbs on the top-left of the user interface (see Figure 2.B.1). Breadcrumbs summarize the navigation paths taken by the user. In this case, Dirk navigates through the IP hierarchy represented by each octet of the remote IPv4 address as a sequence, i.e., networks of size /8, /16, and /24 in Classless Inter-Domain Routing (CIDR) notation.

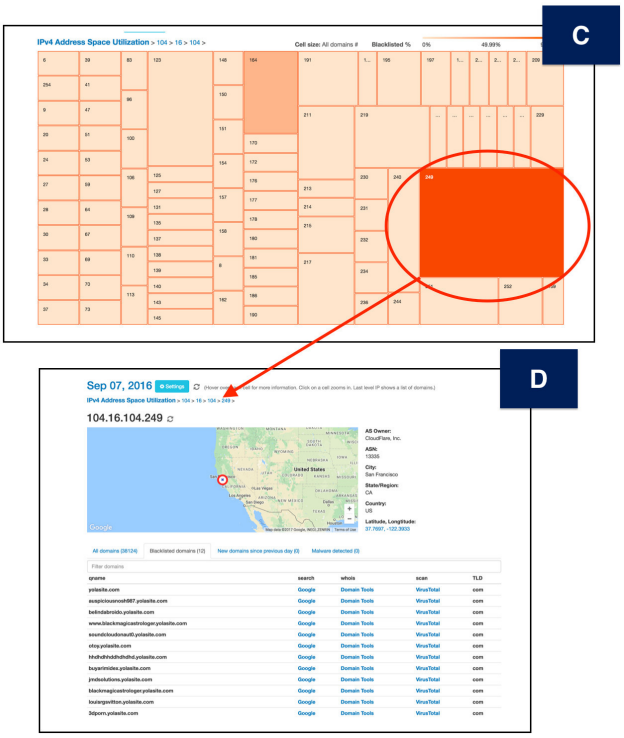

Figure 3: Screenshots of THACO showing the views the user visualizes after following the steps in Figure 2.

Returning to the visualization, the size of each rectangle in the treemap is proportional to the total number of domain names pointing to IPs contained within the network, and the color represents the percentage of blacklisted domain names that resolve into that CIDR. Moving the mouse over any rectangle of the treemap gives a pop-up label displaying the percentage of blacklisted domain names. A relative color legend is also provided at the top-right of the treemap that updates according to the treemap level. Our analyst uses the breadcrumb component to view the network 104.16.0.0/16 as seen on September 7, in order to determine if there are other /24 networks of interest. Dirk's eyes will be naturally drawn to the largest and darkest-colored rectangles in the treemap (see Figure 2.B.1). Dirk realizes that 104.16.104.0/24, 104.16.105.0/24, 104.16.106.0/24, and 104.16.107.0/24 networks also contain high percentages of blacklisted domain names.

Dirk chooses to first focus on the "neighbor" of 104.16.103.0/24, 104.16.104.0/24, so he clicks the corresponding rectangle in order to identify any IP addresses under this /24 with high percentages of blacklisted domain names. As a result of this action, these IPv4 addresses are displayed (see Figure 3.C). Dirk again will be drawn to the largest and darkest-colored rectangle in the treemap, 104.16.104.249/32, revealing a new IP address that should be explored. Clicking on the corresponding cell reveals geographical, infrastructure, and reputation-related information of 104.16.104.249/32 (see Figure 3.D). He again sees many $yolasite[dot]com$ child domains and modifies his detector to alert him on communication to these IP addresses as well. Finally, he creates a ticket for IT to remove the malware, including a link to the malware sample analysis results already discovered.

3.6 Top-down Threat Analysis

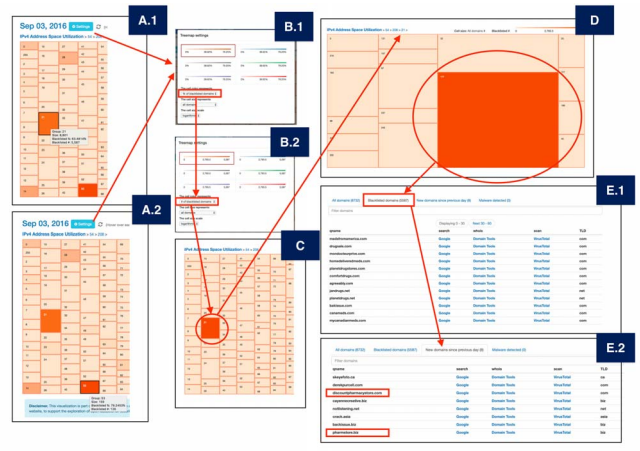

Figure 4: Screenshots of THACO showing the steps taken by Kate to identify networks under 54.208.0.0/16 with the highest percentage of blacklisted domain names on September 3, 2016.

On September 3, 2016 Kate investigates her hosting provider's network prefix, 54.208.0.0/16, to hunt for potentially malicious activity. In Figure 4.A.1, she notices that 54.208.21.0/24 and 54.208.53.0/24 contain similar percentages of blacklisted domain names. In order to better distinguish them, she mouses over each corresponding cell (Figures 4.A.1 and 4.A.2). Kate discovers that 54.208.53.0/24 contains a higher percentage of blacklisted domain names; however, 54.208.21.0/24 is associated with a higher number of blacklisted domain names.

Guided by this insight, Kate clicks on the settings panel located at the top-left of the user interface in order to adjust the default cell color to the number of blacklisted domain names (Figures 4.B.1 and 4.B.2). As a result, the treemap updates the displayed information according to the new configuration. Kate now confirms that 54.208.21.0/24 contains a greater number of blacklisted domain names and chooses to investigate this network first. She then clicks on the corresponding cell to see that the IP addresses under 54.208.21.0/24 also contain high numbers of blacklisted domain names (see Figure 4.D). Kate identifies that 54.208.21.117/32 contains 5,587 blacklisted domain names.

In order to get further detail on 54.208.21.117/32, she clicks on the corresponding cell. Consequently, geographical, infrastructure, and reputation-related information of 54.208.21.117/32 is shown. Kate realizes that the blacklisted domains are likely related to illegal online pharmaceutical sales based on the domain strings, e.g., $medsfromamerica[dot]com$, $drugsale[dot]com$, etc. (see Figure 4.E.1). Kate creates a ticket for her company's IT team to revoke the accounts associated with these malicious domains. Moreover, clicking the tab for new domains (see Figure 4.E.2) reveals that 54.208.21.117/32 is also resolved by eight new domain names. Kate hypothesizes that two of these new domain names are likely related to this campaign, based on their strings that also suggest illicit drug activity, specifically, $discountpharmacystore[dot]com$ and $pharmstore[dot]biz$. At this point, she infers that these new domain names should also be removed and adds them to the ticket to IT she created previously.

4 User Study

While THACO is grounded in security visualization best practices and addresses both data and tasks needs for DNS-based network threat analysis, its use should be evaluated by real-world analysts. Our goal is to answer the following research questions:

1. *Does THACO allow analysts to perform DNS-based network threat analysis tasks?* To respond to this research question, we asked participants to conduct different tasks scenarios using THACO. For each task scenario, we measured both **task completion rates** and **task completion times**.

2. *Can THACO also be used by novice analysts in an effective way?* In order to measure the influence of experience in the use of THACO, we conducted a "between-subjects" t-test using both **task completion rates** and **task completion times** for each task scenario.

3. *Is THACO pleasant and easy to use for analysts?* According to Nurse et al. [33], as much as is feasible, security tools should aim to provide users with a positive and satisfactory experience. Following previous work in user evaluation in visualization [24, 39], we evaluate **overall user experience** of using THACO in terms of how easy and pleasing it is to use, as well as in terms of both **spatial organization** and **navigation support** to locate relevant information.

4.1 Participants

Recruitment of a representative sample for this user study was one of the major challenges. Unlike user evaluations that can be performed with participants from the general population, our user study requires participants with a specific background.

In an effort to overcome these issues, and allow a diverse range of participants, we recruited both in-situ and online participants. Specifically, 7 in-situ and 31 online IT security practitioners, from both academia and industry, with years of experience in network threat analysis ranging from less than a year to more than 10 years, participated in this study. Apart from the level of experience of participants and gender, no further personal information was collected. In-situ participants visiting our laboratory were recruited only by e-mail, through professional contacts of the research team. To recruit online participants, we also used social media platforms like Twitter. IRB (Institutional Board Review) approval was obtained to enroll participants in this study and all responses were only collected after written consent was obtained. Participants were not compensated for their time.

4.2 Procedure

In this section, we describe the procedure employed in the user study of THACO according to the type of participation. While for the in-situ evaluation participants conducted *tasks scenarios* and *semi-structured interviews*, for the online evaluation participants conducted a web-based survey including both *System Usability Scale (SUS)* and *closed and open-ended questions*. By using diverse quantitative and qualitative methods, data from multiple perspectives helps to mitigate the effects that the limitations of any one particular technique.

4.2.1 In-situ Evaluation

In our study, we invited 7 IT security practitioners to visit our research laboratory, with experience in DNS-based network threat analysis ranging from one year for three of the participants to more than 10 years for four of them. After giving them a brief ten minute oral introduction of THACO, they were instructed to complete five main tasks scenarios using THACO deployed as a website. The first participant with more than 10 years of experience was used to test the procedure and identify potential issues in our user study. The results obtained from this participant were discarded and they were not included in the final analysis.

Task scenario selection was guided by three main objectives: (1) to provide tasks related to both bottom-up and top-down threat analysis; (2) to provide tasks where we predict users will have difficulties, such as Task Scenarios 2 and 4; and (3) to provide tasks that enable a more thorough examination of the features provided by THACO, such as Task Scenarios 1 and 3.

- Task Scenario 1: You have received a high-level alert from a network detector around malicious activity in your company's network related to 104.0.0.0/8 that was triggered on September 3, 2016. Use the threat console to find IP blocks with 80% or higher of blacklisted domain names across the hierarchy (/16,/24,/32). Taking advantage of as many mechanisms provided by the tool as you need, identify these IP blocks and explain the reasoning behind your answer.

- Task Scenario 2: You have received a high-level alert from a network detector related to 54.230.205.0/24 that was triggered on September 1, 2016. Use the threat console to identify IP addresses with the highest percentage of blacklisted domain names per day from September 1 to September 4. Taking advantage of as many mechanisms provided by the tool as you need, try to characterize these IP addresses in terms of number of blacklisted domains, new domain names, and related malware. Explain your reasoning behind your answer.

- Task Scenario 3: You have received a high-level alert that was triggered on September 4 2016 related to some IP addresses located in Europe and Africa. Use the threat console to identify countries in both Europe and Africa with 50% or higher of blacklisted domain names. Explain your reasoning behind your answer.

- Task Scenario 4: You have been requested to research autonomous systems with 50% or higher of blacklisted domain names from September 1 to September 7. Use the threat console to identify them and explain your reasoning behind your answer.

- Task Scenario 5: You have received an alert from a network detector related to 104.16.103.249/32 on September 3 2016. Use the threat console to understand the context of this IP address in terms of geographical location, ownership, number of blacklisted domain names, new domain names, and domain names associated with malware. Explain the reasoning behind your answer.

In our study, an observer sat with each participant throughout the session, recording observations, noting any difficulties, and any comments made by the participant while he/she completed each task scenario. Finally, after completing the tasks scenarios, we conducted a semi-structured interview for each participant. Each interview lasted approximately thirty to forty minutes and was subsequently transcribed and sanitized to preserve the participants' anonymity. Participants were in the laboratory for no more than two hours. As is typically the case with semi-structured interviews, not all participants were asked the same questions, and not all discussed topics were directly relevant to our research questions.

4.2.2 Online Evaluation

For the online evaluation, we used the System Usability Scale (SUS) [17] to measure the overall user satisfaction of THACO. This type of survey is composed of only ten statements, therefore it is relatively quick and easy for participants to complete, and for researchers or administrators to score. The result of the survey is a

single score, ranging from 0 to 100, and is relatively easy to understand. Furthermore, previous literature in usability [15] has found that SUS is able to provide the most reliable results across a wide range of sample sizes. SUS was proctored through a website and took approximately twenty minutes to complete.

SUS was complemented with six closed-ended questions and two open-ended questions (see Table 1) in order to overcome possible limitations of the SUS scale and obtaining feedback on specific visual representation features, such as spatial organization, provided orientation, and help. For closed-ended questions, a five-value Likert scale was used to collect the opinion of participants ranging from (1) strongly disagree to (5) strongly agree).

Closed-ended Questions
q1. I found the threat console useful for network threat analysis.
q2. I found the organization of the information in the threat console very confusing.
q3. I discovered unexpected relationships among information elements through the threat console (e.g. I found new unexpected blacklisted domain names using the tool).
q4. I was able to locate relevant information elements using the threat console (e.g. IP blocks with high percentage of malicious domain names).
q5. I think this threat console may help me in achieving a better performance in my daily tasks.
q6. I think the organization of the information in the threat console is very clear.
Open-ended Questions
q1. Which features of the visualization do you like the most and the least?
q2. Briefly explain why you like or dislike these features.

Table 1: Closed and open-ended questions included in our web-based survey. For closed-ended questions, we used a five-value Likert scale.

5 Analysis and Results

5.1 Task Completion Rates

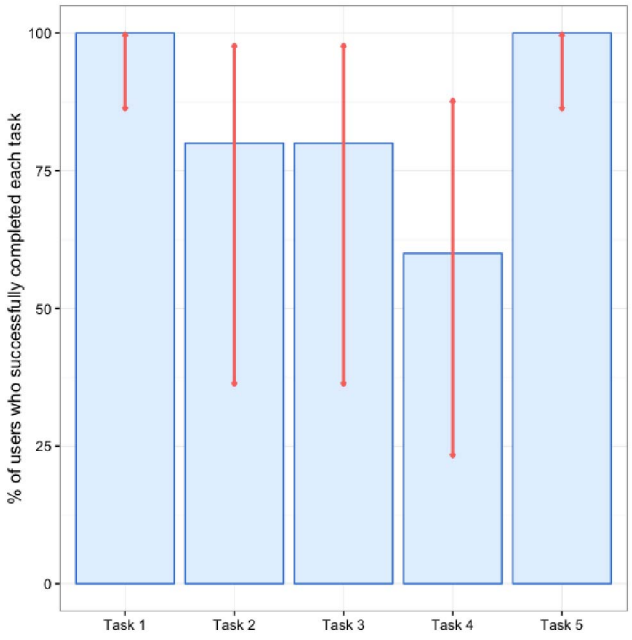

Figure 5: Task completion rates represented as bars, plus the 95% confidence interval for each, represented as error bars.

Figure 5 shows that Tasks 1 and 5 had significantly higher completion rates than Tasks 2, 3, and 4. An average completion rate of 100% for Task 1 means that all participants were able to navigate through the entire IP hierarchy using the treemap to identify IP blocks with high percentages of blacklisted domain names. Similarly, an average completion rate of 100% for Task 5 means that all participants were able to use the threat console to understand the context of a specific IP address. On the other hand, Tasks 2 and 4 correspond to tasks we hypothesized participants would have some difficulties to complete. They were required to keep track of different pieces of information over time. Similarly, Task Scenario 3 is related to the use of the geographical grouping option for the treemap.

Given the diversity of experience of our participants, we were also interested in understanding its influence on these percentages. Accordingly, we conducted a "between-subjects" t-test in order to decide if there were significant differences among the percent of completed tasks from experienced and inexperienced participants. The result obtained from the t-test, (p = 0.74), supports that there was no influence of the experience of participants on these percentages of completed tasks. In other words, experienced and inexperienced participants were able to complete the task scenarios with similar completion rates using THACO.

Finally, we computed the total average task completion rate for THACO. In order to interpret the obtained score of 84%, we used an analysis of about 1200 usability tasks across different products [37], which reported an average task completion rate of 78%. In accordance to this analysis, an overall task completion rate of 84% for THACO is better than 61.3% of all usability tasks analyzed.

5.2 Task Completion Times

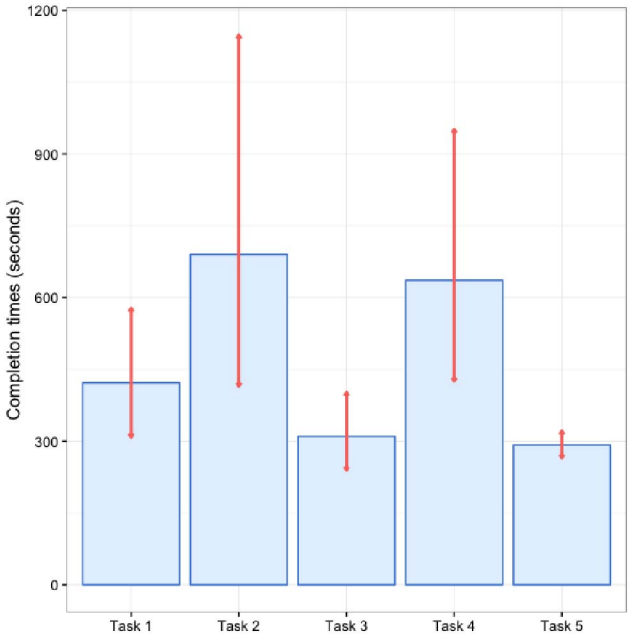

Figure 6: Task completion times represented as bars, plus the 95% confidence interval for each, represented as error bars.

Figure 6 shows the completion times per task, along with a 95% confidence interval for each. The 95% confidence intervals for each task was calculated using the natural logarithm [38] to approximate a normal distribution and plotted as error bars. As shown in this figure, Tasks 2 and 4 had significantly higher average completion times than Tasks 1, 3, and 5. These task completion times complement, thus, previous results displayed in Section 5.1 with the exception of Task 3, which is related to the usage of the geographical grouping option for the treemap. Tasks 2 and 4 were indeed more difficult and took longer to complete, as we had predicted. Con-

versely, Tasks 1 and 5 were completed faster and all participants were successful.

Finally, we conducted a "between-subjects" t-test in order to decide if there were significant differences between the completion times from experienced and inexperienced participants. The result obtained from the t-test, ($p = 0.62$), supports that the experience of participants did not significantly influence task completion times. In other words, experienced and inexperienced participants were able to complete task scenarios in a similar amount of time using THACO.

5.3 User Satisfaction

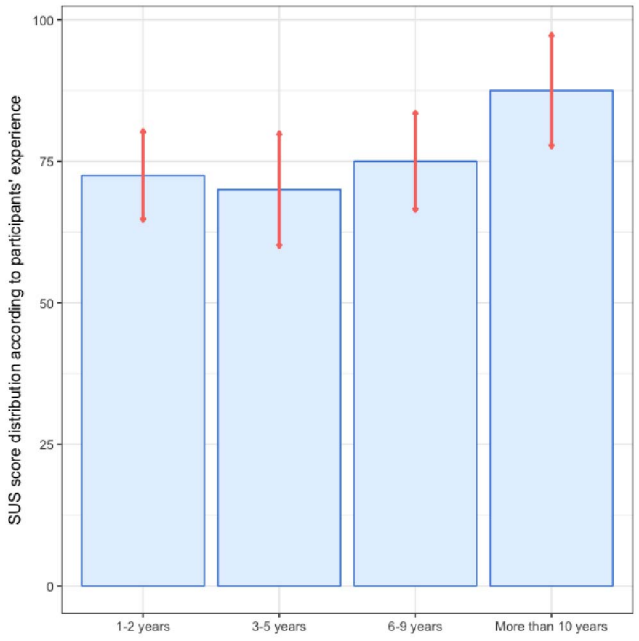

Figure 7: Distribution of SUS scores according to participants' years of experience in network threat analysis.

To measure the overall user experience of THACO, we calculated the SUS score following the procedure proposed by Brooke [17]. To calculate the SUS score, we first needed to sum the score contributions from each item. Each item's score contribution will range from 0–4. For items 1, 3, 5, 7, and 9 the score contribution is the scale position minus 1. For items 2, 4, 6, 8, and 10 the contribution is 5 minus the scale position. Finally, we needed to multiply the sum of the scores by 2.5 to obtain the overall value of SUS.

Figure 7 shows the distribution of the SUS scores according to the years of experience of participants, along with a 95% confidence interval for each range. From this figure, the highest score of 85.5 was obtained for participants with ten or more years of experience. However, in order to interpret these SUS scores, we needed to convert them to a percentile rank through a normalization process. Consequently, scoring at the overall mean score of 76.5 means that the threat console in terms of usability has higher perceived usability than 70% of all products tested. Similarly, scoring at the mean of 85.5 for the most experienced participants can be interpreted as better than 80% of all products tested. It is worth noting that among the most experienced participants that had 10 years or more experience in the threat analysis field, the threat console achieved a SUS score of 85.5 over 100. In other words, we achieve "A' as a grade among the most experienced participants, following the letter grades proposed in [36]. This is a very promising result, as we can further improve an already useful threat analysis console that the most senior threat analysts felt easy and pleasant to use.

In order to both overcome SUS limitations for identifying potential symptoms of confusion by participants and obtaining specific feedback on visual representation features, we complemented the obtained SUS score with closed and open-ended questions included in our web-based survey. In order to statistically interpret these results, we computed the median for each question and the Inter-Quartile Range (IQR) of each item in the five-value Likert scale. As shown in Table 2, participants expressed strong agreement on four main questions: (1) the utility of THACO for network threat analysis (q1: Mdn = 4, IQR = 0); (2) the adequate spatial organization of the information provided by THACO (q2: Mdn = 2, IQR = 1, q6: Mdn = 4, IQR = 2); (3) the support of THACO to achieve a better performance on threat analysis daily tasks (q5: Mdn = 4, IQR = 1); and (4) the orientation provided by THACO to find relevant information around the context of a threat indicator (q4: Mdn = 4, IQR = 1). Nevertheless, opinion seems to be divided with regard to the support of the tool to find unexpected relationships among pieces of information (q3: Mdn = 3, IQR = 2).

	q1	q2	q3	q4	q5	q6
Median	4	2	3	4	4	4
IQR	0	1	2	1	1	2

Table 2: Median and IQR for closed-ended questions from Table 1.

5.4 Discussion

Generalizing from the tasks scenarios, this study shows that THACO better assists DNS-based network threat analysis tasks focused on a specific day. In particular, the completion rates for Task 1 and Task 5 (see Figure 5) were always 100 per cent. Likewise, they were on average performed more quickly in these same tasks scenarios (see Figure 6). The success rates for Tasks 2 and 3, which involve keeping track of different pieces of information over time registered the worst completion rates (see Figure 5) and times (see Figure 6). To carry out these tasks, we observed that some participants were taking notes on paper of the information observed per day and then, they correlated it with information from previous days. As potential solutions to temporal issues of THACO, some participants proposed the inclusion of time sliders to be able to easily navigate across days. This suggestion has been already implemented in the future version of THACO. Both our experiments demonstrate that THACO is easy to use regardless of the experience of participants..

6 Conclusions

We have designed an open source threat analysis console, called THACO (THreat Analysis COnsole), which leverages open DNS datasets, domain WHOIS records, and both public malware and domain blacklists to adapt to DNS-based network threat analysis needs. In order to demonstrate both its utility and usability, we have conducted a user study utilizing 7 in-situ and 31 online IT security practitioners, from both academia and industry, with years of experience in network threat analysis ranging from less than a year to more than 10 years. Our study demonstrates that THACO is usable and useful to analysts of all experience levels. Further work should be oriented towards compare THACO with other existing network threat analysis tools such as Maltego or ThreatCrowd.

Acknowledgements

This material is based upon work supported in part by the US Department of Commerce grant no. 2106DEK, National Science

Foundation (NSF) grant no. 2106DGX and Air Force Research Laboratory/Defense Advanced Research Projects Agency grant no. 2106DTX. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the US Department of Commerce, National Science Foundation, Air Force Research Laboratory nor Defense Advanced Research Projects Agency.

REFERENCES

[1] Alienvault otx. https://otx.alienvault.com/. Accessed: 2017.

[2] Color brewer 2.0. http://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3%. Accessed: 2017.

[3] Threatminer. https://www.threatcrowd.org/. Accessed: 2017.

[4] Threatminer. https://www.paterva.com/web7/buy/maltego-clients.php. Accessed: 2017.

[5] Threatminer. https://www.threatminer.org/index.php. Accessed: 2017.

[6] Virus total. https://www.virustotal.com/. Accessed: 2017.

[7] Building a world-class security operations center: A roadmap. https://www.sans.org/reading-room/whitepapers/analyst/building-world-class-security-operations-center-roadmap-35907, 2015.

[8] Abuse.ch domain blacklist. http://www.abuse.ch/, 2016.

[9] Malware domain list. https://www.malwaredomainlist.com/, 2016.

[10] Sandia national laboratories, dnsviz. [online]. http://dnsviz.net, 2016.

[11] Passivetotal. https://www.passivetotal.org/, 2017.

[12] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster. Building a dynamic reputation system for dns. In *USENIX security symposium*, pages 273–290, 2010.

[13] M. Antonakakis, R. Perdisci, W. Lee, N. Vasiloglou, and D. Dagon. Detecting malware domains at the upper dns hierarchy. In *USENIX security symposium*, volume 11, pages 1–16, 2011.

[14] S. Axelsson. The base-rate fallacy and the difficulty of intrusion detection. *ACM Transactions on Information and System Security (TISSEC)*, 3(3):186–205, 2000.

[15] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.

[16] D. Botta, R. Werlinger, A. Gagné, K. Beznosov, L. Iverson, S. Fels, and B. Fisher. Towards understanding it security professionals and their tools. In *Proceedings of the 3rd symposium on Usable privacy and security*, pages 100–111. ACM, 2007.

[17] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[18] S. Caltagirone, A. Pendergast, and C. Betz. The diamond model of intrusion analysis. Technical report, CENTER FOR CYBER INTELLIGENCE ANALYSIS AND THREAT RESEARCH HANOVER MD, 2013.

[19] B. D. Czejdo, E. M. Ferragut, J. R. Goodall, J. Laska, et al. Network intrusion detection and visualization using aggregations in a cyber security data warehouse. *Int'l J. of Communications, Network and System Sciences*, 5(09):593, 2012.

[20] A. D'Amico, K. Whitley, D. Tesone, B. O'Brien, and E. Roth. Achieving cyber defense situational awareness: A cognitive task analysis of information assurance analysts. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 49, pages 229–233. SAGE Publications Sage CA: Los Angeles, CA, 2005.

[21] A. D. D'Amico, J. R. Goodall, D. R. Tesone, and J. K. Kopylec. Visual discovery in computer network defense. *IEEE Computer Graphics and Applications*, 27(5), 2007.

[22] D. E. Denning. An intrusion-detection model. *IEEE Transactions on software engineering*, (2):222–232, 1987.

[23] R. F. Erbacher, K. L. Walker, and D. A. Frincke. Intrusion and misuse detection in large-scale systems. *IEEE Computer Graphics and Applications*, 22(1):38–47, 2002.

[24] G. A. Fink, C. L. North, A. Endert, and S. Rose. Visualizing cyber security: Usable workspaces. In *Visualization for Cyber Security, 2009. VizSec 2009. 6th International Workshop on*, pages 45–56. IEEE, 2009.

[25] M. Ghoniem, J.-D. Fekete, and P. Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 17–24. Ieee, 2004.

[26] J. Goodall, W. Lutters, and A. Komlodi. The work of intrusion detection: rethinking the role of security analysts. *AMCIS 2004 Proceedings*, page 179, 2004.

[27] J. Jacobs and B. Rudis. *Data-Driven Security: Analysis, Visualization and Dashboards*. John Wiley & Sons, 2014.

[28] A. Komlodi, J. R. Goodall, and W. G. Lutters. An information visualization framework for intrusion detection. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, page 1743. ACM, 2004.

[29] A. Kountouras, P. Kintis, C. Lever, Y. Chen, Y. Nadji, D. Dagon, M. Antonakakis, and R. Joffe. Enabling network security through active dns datasets. In *International Symposium on Research in Attacks, Intrusions, and Defenses*, pages 188–208. Springer, 2016.

[30] R. M. Lee and R. Lee. The who, what, where, when, why and how of effective threat hunting. https://www.sans.org/reading-room/whitepapers/analyst/who-what-where-when-effective-threat-hunting-36785, 2016. Accessed: 2017.

[31] F. Mansmann, F. Fischer, D. A. Keim, and S. C. North. Visualizing large-scale ip traffic flows. In *Vision, Modeling, and Visualization*, 2007.

[32] R. Marty. *Applied security visualization*. Addison-Wesley Upper Saddle River, 2009.

[33] J. R. Nurse, S. Creese, M. Goldsmith, and K. Lamberts. Guidelines for usable cybersecurity: Past and present. In *Cyberspace Safety and Security (CSS), 2011 Third International Workshop on*, pages 21–26. IEEE, 2011.

[34] R. Perdisci, W. Lee, and N. Feamster. Behavioral clustering of http-based malware and signature generation using malicious network traces. In *NSDI*, pages 391–404, 2010.

[35] M. Roesch et al. Snort: Lightweight intrusion detection for networks. In *LISA*, volume 99, pages 229–238, 1999.

[36] J. Sauro. Measuring usability with the system usability scale (sus). http://www.measuringu.com/sus.php. Accessed: 2017.

[37] J. Sauro. What is a good task-completion rate? http://www.measuringu.com/blog/task-completion.php. Accessed: 2017.

[38] J. Sauro and J. R. Lewis. Estimating completion rates from small samples using binomial confidence intervals: comparisons and recommendations. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 49, pages 2100–2103. SAGE Publications, 2005.

[39] D. Staheli, T. Yu, R. J. Crouser, S. Damodaran, K. Nam, D. O'Gwynn, S. McKenna, and L. Harrison. Visualization evaluation for cyber security: Trends and future directions. In *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, pages 49–56. ACM, 2014.

[40] S. Walton, E. Maguire, and M. Chen. A visual analytics loop for supporting model development. In *Visualization for Cyber Security (VizSec), 2015 IEEE Symposium on*, pages 1–8. IEEE, 2015.

[41] R. Werlinger, K. Hawkey, and K. Beznosov. Human, organizational and technological challenges of implementing it security in organizations. *HAISA*, 8:35–48, 2008.

[42] S. Yadav, A. K. K. Reddy, A. Reddy, and S. Ranjan. Detecting algorithmically generated malicious domain names. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pages 48–61. ACM, 2010.