Visual Exploration of Network Hostile Behavior

Jorge Guerra
ITIC - UNCuyo
Mendoza, Argentina
jguerra@uncu.edu.ar

Carlos Catania
ITIC - UNCuyo
Mendoza, Argentina
ccatania@itu.uncu.edu.ar

Eduardo Veas
KTI - TUGraz
Graz, Austria
eveas@know-center.at

ABSTRACT

This paper presents a graphical interface to identify hostile behavior in network logs. The problem of identifying and labeling hostile behavior is well known in the network security community. There is a lack of labeled datasets, which make it difficult to deploy automated methods or to test the performance of manual ones. We describe the process of searching and identifying hostile behavior with a graphical tool derived from an open source Intrusion Prevention System, which graphically encodes features of network connections from a log-file. A design study with two network security experts illustrates the workflow of searching for patterns descriptive of unwanted behavior and labeling occurrences therewith.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces – Graphical user interfaces (GUI), Interaction styles, Evaluation/methodology; H.3.3. : Information Search and Retrieval – Information filtering, Search process

Author Keywords

Network security, visual analytics, intrusion detection

INTRODUCTION

Network security is a challenging field of research. It builds on data-driven methods to develop techniques to identify threats, for example, building predictive models using machine learning or statistical methods [1]. Botnet malware is one kind of threat of particular interest in network security. It is extremely hard to detect and can be used as starting point for different kinds of attacks: key logging, denegation of service and SPAM are some of them [2]. Despite the growing needs of sample data and the community best efforts, there is a lack of datasets and labels are scarce when even available [4]. This paper introduces a visual tool to find and label network behavior, identifying hostile and normal traffic with two goals: to observe how experts analyze network traffic and to motivate researchers to complete and exchange ground truth data.

ESIDA'17, March 13 2017, Limassol, Cyprus
© 2017 ACM. ISBN 978-1-4503-4903-1/17/03…$15.00
DOI: http://dx.doi.org/10.1145/3038462.3038466

A recent approach used for encoding network behavior has been proposed by the Stratosphere Intrusion Prevention System (IPS) [3], which is a large effort for offering a state of art IPS for Non Governmental Organizations (NGO).

The approach used by Stratosphere IPS to encode the behavior of a connection, starts by aggregating the flows according to a 4-tuple composed of: the source IP address, the destination IP address, the destination port and the protocol. All the flows that match a tuple are aggregated together and referred as a *Stratosphere connection*. From a traffic capture several of these Stratosphere connections are created. Each one of the these Stratosphere connections contains a group of flows. The behavior of a connection is computed as follows:

1. Extract three features of each flow: size, duration and periodicity.

2. Assign to each flow a *state* symbol according to the features extracted and the assignment strategy shown in Table 1.

3. After the assignment, each *connection* has its own string of symbols that represents its behavior in the network.

	Size Small			**Size Medium**			**Size Large**		
	Dur. Short	**Dur. Med**	**Dur. Long**	**Dur. Short**	**Dur. Med**	**Dur. Long**	**Dur. Short**	**Dur. Med**	**Dur. Long**
Strong Per	a	b	c	d	e	f	g	h	i
Weak Per.	A	B	C	D	E	F	G	H	I
Weak Non-Per.	r	s	t	u	v	w	x	y	z
Strong Non-Per	R	S	T	U	V	W	X	Y	Z
No Data	1	2	3	4	5	6	7	8	9

Symbols for time difference	
Between 0 and 5 seconds:	.
Between 5 and 60 seconds:	,
Between 60 and 5 mins:	+
Between 5 mins and 1 hour	*
Timeout of 1 hour:	0

Table 1. Symbol assignment strategy to encode network behavior.

A sample **behavioral encoding** is shown in Fig. 1. The figure shows the symbols representing all the flows for a Stratosphere connection based on UDP protocol from IP address 10.0.2.103 to port 53 of IP address 8.8.8.8.

2.4.2*4.R.R*a*b*a*a*b*b*a*R.R*R.R*a*a*b*a*a*a*a*

Figure 1. An example behavioral encoding of connection from IP address 10.0.2.103 to destination port 53 at IP address 8.8.8.8 using UDP.

Experts label botnet connections using a combination of command line tools developed to work with the symbol encoding.

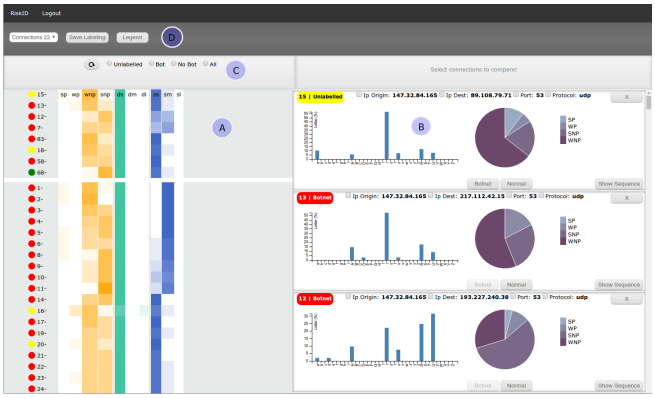

Figure 2. RiskID User Interface. The left panel collects descriptors for all connections with heatmaps of different colors for periodicity, size and duration. The right panel shows details for up to three connections.

They look at periodicity to determine if the connection contains flows occurring at periodic intervals. They check for blacklisted IPs and use databases to determine services.

THE RISKID TOOL

RiskID is a visual analytics tool that combines visualization with clustering techniques to assist the user in the process of labeling connections. The interaction between users and the automatic component is designed to illustrate, in a comprehensive manner, the connections behavior, and to group them both visually and with clustering algorithms to facilitate Botnet detection. The process is as follows:

1. RiskID receives a JSON file that contains for every Stratosphere connection (from here referred simply as connection) some basic network information, such as IP addresses and Ports, together with its corresponding behavioral encoding.

2. The feature extraction module analyzes all connection behavioral encodings and creates for every connection a new vector summarizing the information in terms of periodicity, duration and size.

3. The cluster composition module analyzes the feature vectors and groups them according to a standard similarity measure.

4. The UI represents the list of connections with a heatmap of the feature vectors, using different colors for each type of feature.

5. The user explores the connection list to discover common patterns through color similarity. During this process she can select potentially similar connections.

6. Upon selection, details about the connection composition are shown in order to facilitate comparison.

7. Eventually, when the user finds a high coincidence between selected connections she can proceed to label them as "Botnet" or "Normal".

It is worth noting that a correct labeling process mainly depends on the user selection strategy.

User Interface

RiskID's UI layout has two main blocks that display connection information at different levels (Figure 2). The first block shows the Connection Overview through a list of connections (Figure 2 left). The application displays general information about their composition. The clustering information of the different connections is represented with a different background color in the connection list. The second block shows a Detailed Connection View (Figure 2 right). For each connection selected in the list, the connection viewer displays detailed information about the connection including its current label and symbol frequencies.

Interactions and Visual Design

The application design focuses on facilitating to the user a set of visual tools to analyze the connection composition and the dataset being labeled. The user can interact with the different components of the application to obtain insights for improving the precision and confidence of the labeling process.

The Connection List

The Connection List shows the connections grouped according to the similarity in their encoding behavior. A clustering algorithm is executed to form groups. The current version of the application uses a k-means algorithm based on L2 distance to form the groups. The optimal number of groups is selected by the Elbow method, which consists of increasing the number of clusters until the marginal gain of the variance explained by the model is negligible. The clustering process helps the user get a first approximation of similar connections.

Each connection is converted from its original encoding describing its behavior to a 10-dimensional numerical vector (denoted as feature vector) where the first four dimensions represent periodicity (strong periodicity, weak periodicity, weak non periodicity and strong non periodicity respectively), the other three refer to duration (duration short, duration medium and duration large respectively) and the last three represent the size (size short, size medium, size large).

The feature vector for a given connection is generated considering, for the complete symbol sequence, the cumulative frequency of the corresponding values associated to the behavioral encoding shown in Table 1. At the end of the sequence a percent of each feature is calculated and normalized between the values [0,1].

A Heatmap is used to represent the feature vector of the connection. Different color were used to differentiate feature types. The intensity in the color scale indicates a given feature is predominant over the rest. With the heatmap, it is intuitive to recognize the predominant features of each connection and, more importantly, relate connections with similar features. The user can customize the list of connections using a set of filtering options that appear at the top of the list (Figure 2 left), e.g., filtering by label.

The connection label is shown with a circle in the left in a traffic light metaphor: red circle means "Botnet", green circle means "Normal" and yellow circle means "Unlabeled". As the user labels a new connection, the color of the circle changes

accordingly. The position of the circle and its color facilitates the analysis of groups of connections with same tags. It also helps the user find potential connections to be labeled.

The Detailed Connection View

The Detailed Connection View, located on the right panel of the application, displays detailed information about selected connections, including: origin and destination IP addresses, destination port and protocol. This network information can be used for filtering the connection list,e.g., by destination IP. Hereby, the user can inspect a particular subset of connections that share the selected network features.

The Detailed Connection View includes also a bar chart describing the frequency of occurrence for each letter in the behavioral encoding. Looking at the bar chart, the user can easily observe the difference between the letter distribution along different connections.

García emphasized the importance of periodicity for recognizing Botnet behavior [3]. Hence, the detailed connection view includes a pie chart of periodicity distribution inside each connection (Strong Periodicity, Weak Periodicity, Weak Non-Periodicity and Strong Non-Periodicity). The user can access the original symbol-based behavioral encoding by clicking a button.

Connection comparison

One important advantage of RiskId is the possibility to compare two or more connections. Every time the user select a new connection, it moves to the top of the list and remains atop as the user scrolls the list. Hereby, the user can visually search similar connections for detailed comparison. Each newly selected connection is placed under the previous selected one, and the details are stacked in the Detailed Connection View. Thus, the user can start a detailed comparison.

The general idea behind the comparison feature is for the user to select several labeled connections visually similar to an unlabeled connection. A connection that upon inspection visually resembles the selected group of labeled connections has a high probability of having the same label of such group.

Implementation Details

RiskId is a Web-based tool mainly implemented in JavaScript using NodeJS. We made use of libraries like JQuery and d3.js for the UI. The cluster construction and feature vector generation are performed entirely on the client side.

CASE STUDY: WORKFLOW ANALYSIS

A study was conducted to observe experts while identifying botnet behavior with our tool. The goal was to observe the strategies the experts employ when using a visual tool.

Data

The dataset for the study was derived from three already labeled datasets coming from network traffic captures taken from CVUT university campus networks. Datasets are publicly available as part of the Malware Capture Facility Project (**MCFP**) [3].

Table 3.1 provides brief information about each of the three datasets. The first two columns show the ID used for referencing the dataset and a brief description of the malware included in such group. Then, in column three and four, the number of connections labeled as Botnet and Normal. Finally, the last column shows the ID of the dataset in MCFP.

ID	Desc.	Botnet Conn.	Normal Conn.	MCFP IDs
A	**Bonet Neris**	2101	713	CTU13-42
B	**Bonet Neris**	1684	128	CTU13-43
C	**Bonet DonBot**	188	300	CTU13-46

Table 2. General information about datasets

For the purpose of the study, the three datasets were merged. The original labels are kept on on a randomly selected 75% percent of the merged dataset, leaving the remaining 25% as unlabeled.

Participants

Two participants (1 Female, 1 Male) took part in the study. They are experts in network security and have expertise working with the Stratosphere IPS to label network behavior. Each participant had to fill a pre-questionnaire prior to the study, assessing general knowledge and familiarity with the domain of network security, analysis tools for network logs, as well as visual tools for data analysis. Both participants had several years (4 and over 10) of experience in the field of network security. They have created a datasets to analyze network behavior (4 and over 10), and had partially or fully labeled bot activity (1 and 4 datasets). Participants have not used visual assistance to label datasets before, nor are they aware of visual tools to support the task.

Methodology

Participants were introduced to the interface using a written tutorial. The meaning of the color scale for the heatmap overview was explained. Furthermore, the introduction covered how to select / de-select connections, the detail view and the raw (drill down) connection view. Participants were then asked to work with the tool for approx. 40m to identify normal or hostile behaviors. This implied basically finding "unlabeled" connections and labeling them with "Normal or Botnet" labels. The session was logged. The start time and the time for each label event were logged as well as UI actions such as selecting a connection, opening details for a connection, etc. At the end of the session, participants had to fill a NASA TLX questionnaire and a post-questionnaire asking about the interpretation of the interface visual features and the workflow followed during labeling.

Outcomes

This study aimed to analyze the workflow and decisions taken while looking for undesired behavior in network logs. We performed analyses of performance, effort and time spent, workflow analysis, and the personal experience. We took the first label event as an indication of progress, so the activities and time until the first label were specially considered.

Performance. Participants labeled a total of 24 connections, 9 of which were correctly labeled (37,5%). There were 8 Botnet

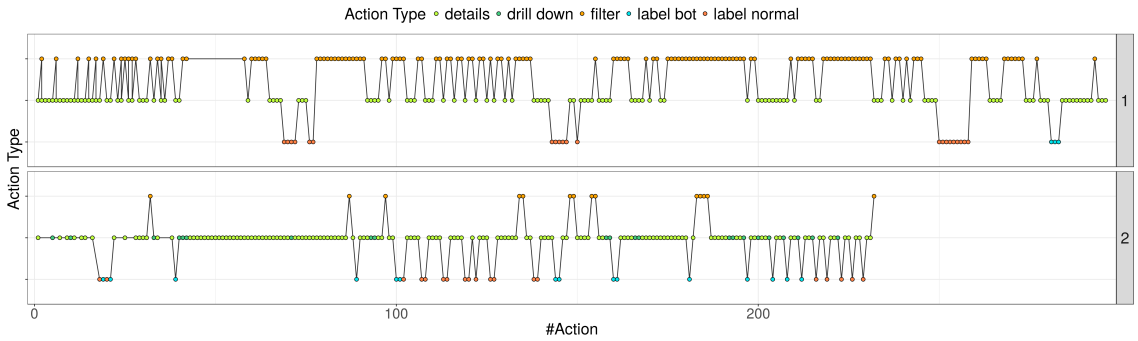

Figure 3. Workflows. Actions are distributed from top to bottom in three levels: filter, details, label. Participants followed two different strategies, E1 concentrated on filtering while E2 used details and comparison.

connections correctly labeled out of 11 (73%), and 1 Normal connection correctly labeled out of 13 (7%). Expert 1 (E1) labeled 4 connections, 1 Botnet (with 0% correctly labeled) and 3 Normal (33% correctly labeled). Expert 2 (E2) labeled 20 connections, 10 Botnet (80%) and 10 Normal (0%).
E1 took 30m until the first label, and 44.5m until the first correct (Normal) label. E2 started right away. The first label occurred at 2m10s and was also a correct (Botnet) label. Connections were not just labeled and forgotten, in some cases a label was removed or edited immediately after or at a later point in time.

Subjective Performance. Subjective results are reported by participant. Workload was calculated with R-TLX (the average of the TLX scales), TLX was measured on the range 1-10. Workload for E1 was 3.83 and for E2 2.33. This indicates that participants did not find it altogether stressful to perform the task. While E1 reported a high (2) self assessed performance, effort (6), frustration (6) where above the center of the range, and mental (3) demand was low. In contrast, E2 reported a medium performance (5), but felt less effort (3), frustration (1) and mental demand (1). A system usability scale (SUS) reported high scores (E1=72, E2=78) in the range of a B^+ grade. In general, the system was usable to the extent that participants felt confident with their performance.

Workflow. We performed action analysis of logged activity. To this end, we categorized actions in three types (filtering, details, and labeling). These actions are in-line with known visualization workflow (overview, filtering, details), since the overview in the case of RiskID is always visible. All filtering actions (filter by IP, filter by port, protocol, etc.) fell under the filtering category. We distinguish two labeling actions (label botnet, label normal) and two detail actions (connection details=details, connection symbol sequence=drill down). Figure 3 illustrates the two different strategies each expert followed to accomplish the task. E1 favored filtering as a means to find unlabeled connections that shared characteristics with labeled ones. E2 relied mostly on multiple comparisons.

DISCUSSION AND OUTLOOK

Accuracy in the labeling task was low. It is interesting to note that more errors were made in wrongly labeling normal connections. E2 had relatively good accuracy in labeling botnet connections (80%). The two experts followed rather different strategies, E1 used filtering and E2 used multiple comparisons as a basis for decision. E2 also inspected the symbol sequence (drill down). Visual and interaction design were well received. From the usability and workload, RiskID was relatively simple to learn and usable for both participants. They reported low effort in using the tool. In spite of the visual design and interactive features, the accuracy was low. We believe the task of labeling requires complementary information. For example, one activity that is done frequently is checking blacklisted IPs. In designing RiskID, we tried to keep with the encoding of Stratosphere IPS. We believe that the visual design could overcome its drawbacks by generalizing on features (periodicity, size, duration). In the future we will rework RiskID to generalize features and include missing information. We will also carry out a follow up study with more experts. Finally, we will investigate what patterns are shared by botnet connections to highlight potentially hostile behavior in unlabeled connections and recommend actions to users.

ACKNOWLEDGMENTS

The authors would like to thank the financial support received by Argentinean ANPCyT- MINCYT through the project PICT 1435-2015 as well as the support received by Stratosphere IPS project.

REFERENCES

1. Sebastian Abt and Harald Baier. 2013. Are We Missing Labels ? A Study of the Availability of Ground-Truth in Network Security Research. *Badgers 2014* (2013).
2. Maria Jose Erquiaga, Carlos Catania, and Carlos García Garino. 2012. An analysis of network traffic characteristics for Botnet detection. *CACIC 2012, XVIII Argentinean Congress of Computer Science* (2012).
3. Sebastian Garcia. 2014. *Identifying, Modeling and Detecting Botnet Behaviors in the Network*. Ph.D. Dissertation. UNICEN University. DOI: http://dx.doi.org/10.13140/2.1.3488.8006
4. Benjamin Sangster, Thomas Cook, Robert Fanelli, Erik Dean, William J Adams, Chris Morrell, and Gregory Conti. 2009. Toward Instrumenting Network Warfare Competitions to Generate Labeled Datasets. *USENIX Security's Workshop on Cyber Security Experimentation and Test (CSET)* (2009).