Scientific Research Publishing

Journal of Information Security, 2022, 13, 210-243
https://www.scirp.org/journal/jis
ISSN Online: 2153-1242
ISSN Print: 2153-1234

# Systematic Review of Graphical Visual Methods in Honeypot Attack Data Analysis

**Gbenga Ikuomenisan*, Yasser Morgan**

Faculty of Engineering & Applied Science, University of Regina, Regina, Canada
Email: *gti002@uregina.ca, Yasser.Morgan@uregina.ca

**How to cite this paper:** Ikuomenisan, G. and Morgan, Y. (2022) Systematic Review of Graphical Visual Methods in Honeypot Attack Data Analysis. *Journal of Information Security*, **13**, 210-243.
https://doi.org/10.4236/jis.2022.134012

**Received:** May 4, 2022
**Accepted:** August 20, 2022
**Published:** August 23, 2022



## Abstract

Mitigating increasing cyberattack incidents may require strategies such as reinforcing organizations' networks with Honeypots and effectively analyzing attack traffic for detection of zero-day attacks and vulnerabilities. To effectively detect and mitigate cyberattacks, both computerized and visual analyses are typically required. However, most security analysts are not adequately trained in visualization principles and/or methods, which is required for effective visual perception of useful attack information hidden in attack data. Additionally, Honeypot has proven useful in cyberattack research, but no studies have comprehensively investigated visualization practices in the field. In this paper, we reviewed visualization practices and methods commonly used in the discovery and communication of attack patterns based on Honeypot network traffic data. Using the PRISMA methodology, we identified and screened 218 papers and evaluated only 37 papers having a high impact. Most Honeypot papers conducted summary statistics of Honeypot data based on static data metrics such as IP address, port, and packet size. They visually analyzed Honeypot attack data using simple graphical methods (such as line, bar, and pie charts) that tend to hide useful attack information. Furthermore, only a few papers conducted extended attack analysis, and commonly visualized attack data using scatter and linear plots. Papers rarely included simple yet sophisticated graphical methods, such as box plots and histograms, which allow for critical evaluation of analysis results. While a significant number of automated visualization tools have incorporated visualization standards by default, the construction of effective and expressive graphical methods for easy pattern discovery and explainable insights still requires applied knowledge and skill of visualization principles and tools, and occasionally, an interdisciplinary collaboration with peers. We, therefore, suggest the need, going forward, for non-classical graphical methods for visualizing attack patterns and communicating analysis results. We also recommend training investigators in

visualization principles and standards for effective visual perception and presentation.

## Keywords

Honeypot Data Analysis, Network Intrusion Detection, Visualization and Visual Analysis, Graphical Methods and Perception, Systematic Literature Review

## 1. Introduction

Information visualization can be described as the study and development of interactive visual representations of complex, abstract data for the purpose of revealing patterns and gaining insights and actionable knowledge for informed human cognitive decision-making [1] [2]. As an established scientific field [2], its application in the fight against cybercrime involving the detection and identification of intrusive attacks from network traffic data is of immense benefits [3] [4] [5].

However, cybercriminals have been observed to still infiltrate networks undiscovered, typically exfiltrating and thereafter encrypting valuable user and/or business data mainly for financial gains [6]. Since traffic data generated in today's networks are typically complex and massive making a easy discovery of useful attack information a challenging task [7], we arguably attribute a typical cause of the aforementioned phenomenon (*i.e.* undetected cyberattacks) to inappropriate construction of graphical methods by the security analyst for information visualization of attack data. We observed that most cyber security researchers are perhaps not abreast of established visualization principles and appropriate visualization methods, and therefore are not able to exhume patterns from data.

Data visualization principles, for all intent and purpose, are proven and established visualization best practices that can be used for effective visual communication and early discovery of patterns. In this study, we aim to investigate the visualization practice of security researchers during the analysis of network intrusion detection data. Our purpose is to identify the typical graphical methods used for visualization and communication of network intrusion detection data, and to evaluate researchers' understanding of basic visualization principles. We limit the focus of our studies to Honeypot server intrusion detection since the field of network intrusion detection is vast. In addition, the Honeypot server is a fast-evolving deception-based network security defense mechanism which has been observed to be applied in recent times in numerous security types of research for the collection and analysis of network intrusive attack data [8] [9].

To the best of our knowledge, we have not found the systematic survey and/or review articles evaluating visualization metrics, methods, and practices relating to attacks and intrusion detection in Honeypot cybersecurity research. Most re-

lated survey and/or review articles in the field of Honeypot did not follow systematic reviewing and reporting methods [10]. Most related literature surveys, particularly in fields such as information visualization [2] [11], network anomaly detection [12], data preprocessing, and dimensionality reduction [13] [14] have not presented findings on how visualization methods are used in relation to Honeypot data analysis. Finally, there have been few Honeypot research contributions in the areas of Honeypot data analysis and visualization [10] and this paper hopes to fill this gap.

In this study, we will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology [15] in assessing Honeypot empirical papers in scientific literature focusing on exploration and visual analysis of Honeypot intrusive attack traffic data. Based on our inclusion and exclusion criteria, relevant papers will be searched for, screened for eligibility, and included in the study. We will extract and evaluate information about the included papers and the graphical figures used in the display of analysis results as presented in the included papers. We reason that information about graphical figures in these papers is an indication of the knowledge and expertise of the security researcher in data visualization since it is used to communicate important research discoveries. We believe that statistical analysis of the extracted data will give insights into the practice of graphical method construction and the use of graphical visualization methods by the security researcher during the visualization of intrusion detection data both for pattern discovery and information communication.

In order to give useful background information to the reader preceding analysis and discussions, in Section 2, we presented an overview of: Honeypot definitions, classifications, and log collection; visualization and visual analytics pipelines; and visual perceptual elements, and common graphical methods, their use, and limitations. Section 3 gave a summary of related survey studies in the areas of information visualization, network anomaly visualization, and data dimensionality reduction. Section 4 detailed the methodology adopted by this. The research results, analysis, and discussion were presented in Section 5, and in Section 6, the conclusion in addition to common visualization principles, best practices, and recommendations were presented.

## 2. Background

### 2.1. Honeypot Types and Log Management

A Honeypot server is a bait technology designed to spuriously engage only attackers, deriving intended value only when it has been effectively probed, attacked, and compromised [16]. The first instance of a Honeypot was Fred Cohen's Deception Toolkit and examples of Honeypot types are shadow Honeypots (deployed with network intrusion detection systems), honeynets (networks of Honeypots), honeywall (a type of Honeypot firewall), and Honeypot deployment frameworks [17] [18].

Honeypots are generally classified into two types based on the level of interactivity with attackers and how well it is able to log data about attacker interactivity [19]:

- Low interaction Honeypots deployed as simulators of real operating systems; easily detectable by attackers; may not be able to engage attackers for longer periods; generally used to investigate simple attack patterns and trends; can collect a large amount of data about attacker activities; are generally simple to deploy and manage; and are less resource intensive
- High interaction Honeypots deployed on real operating systems; not easily detected by attackers; can engage attackers for longer periods; can investigate a wider scope of attacker activities; can collect a relatively larger amount of data about attacker activities and malwares and other malicious artifacts; can help detect zero-day vulnerabilities; can be difficult to deploy and manage; and are more resource intensive.

Other Honeypot classifications are: production Honeypots, research Honeypots, physical Honeypots, virtual Honeypots, server Honeypots and client Honeypots [20].

Usually, Honeypot servers are configured with open and vulnerable operating systems services and applications, and are deployed to different geolocations to be attacked (as shown in **Figure 1**). Network traffic data to and from the Honeypot traps or sensors are collated, processed, visually analyzed and inspected interactively in order to discover and identify interesting attack patterns, and also gain insights and actionable knowledge needed to improve overall network security [21] [22].

## 2.2. Visualization and Visual Analysis Pipelines

Data visualization is the transformation of raw data into simple, intuitive and interactive visual data abstractions for the purpose of conveying meaning and augmenting human cognition [23]. It can be regarded as a data flow model which involves a series of functional data abstraction modules which when executed as a pipeline processes and transforms raw input data into visual components

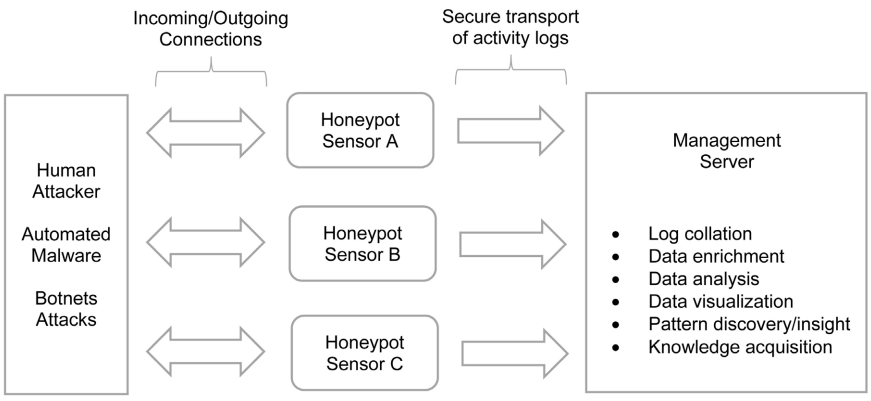


**Figure 1.** Generic Honeypot deployment, log collection, and management for intrusion detection.

such as charts and maps for further analysis. Raw data are abstractions of the universe and can be of type numeric, non-numeric, or both. DV can be largely categorized into two: scientific visualization, which involves processing of multidimensional scientific data; and information visualization, which involves processing of multivariate business data.

Scientific visualization [7] commonly involves visual transformation of scientific data for visual investigation of a scientific entity, for example, the visualization of human cell DNA structure [24] in antibodies studies. In scientific visualization the display dimension is mostly the same as the dimension in which the scientific data originally exists. Consequently, the display dimension is mostly known (*i.e.* already given) and the visual transformation process is mostly not regarded as a dimension scaling problem. Common scientific visualization dimensions are 1D, 2D, and 3D [7].

Information visualization [2] commonly involves visualization of tabular multi-variable business (or non-scientific) data such as the one generated from within financial institutions or computer networks. It is the study of transformation of non-scientific data into corresponding visual representations and the layout of visual data abstractions on views.

In information visualization, raw business data is typically characterized with large number of correlated input variables consequently requiring display dimension reduction for easy visual inspection [7]. It is the responsibility of the human analyst to choose an optimal display dimension for representative visual transformation, easy visual interaction, and pattern discovery and knowledge gain.

Numerous visualization pipelines and/or frameworks have been proposed in literature for guided data analysis, visualization, and knowledge extraction [2] [14] [25]. **Figure 2** shows a 3-step generic information visualization framework for forensic Honeypot data analysis. First, collected network traffic data is processed through a data abstraction module where raw data is cleaned and transformed into an appropriate form using existing data analysis techniques

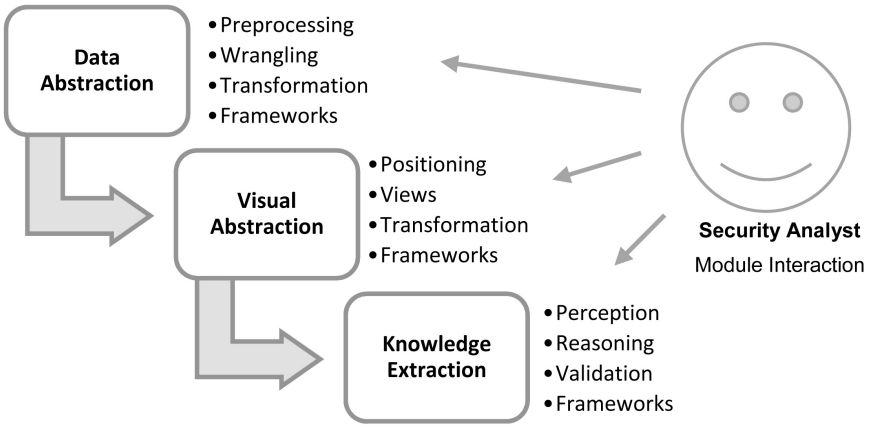


**Figure 2.** Generic information visualization process flow for forensic analysis of Honeypot data.

and methods (e.g. scaling, dimension reduction). Second, visual components such as such charts, plots, maps and diagrams are then created from the transformed data through the visual abstraction module. Finally, knowledge extraction module organizes created visual components on views, supporting the security analyst in interactive visual inspection of the network traffic data.

Card *et al.* [1] proposed a generic information visualization pipeline well- referenced in data visualization research. It describes the functional steps required for knowledge crystallization—a process of achieving the most intuitive visual representation of data for a cognitive task. In the first functional step, raw multivariate data (typically with imperfections such as missing values, outliers, etc.) is preprocessed and transformed into a compressed qualitative data representation in low-dimension space. The second functional module creates corresponding visual elements such as charts and maps using the transformed data as input. In the third functional module, the mapped visual elements are transformed into views for visual inspection. The human analyst can interact with every step in the information visualization pipeline to improve the results of the visualization task. The quality of results though largely depends on the knowledge of the task's domain and technical expertise of the analyst [14].

The data model proposed by Card *et al.* [1] is regarded as a basic information visualization pipeline and there are a number of notable extensions in this area. For example, Chi [26] extended the work of Card *et al.* [1] to include data abstraction steps with which a user can interact. Chi [26] reveals that the state of the data keeps changing as it goes through abstraction functions in a visualization pipeline. The information visualization data flow model proposed by Haber *et al.* [27] was extended in the work of Santos *et al.* [7] for both scientific (multidimensional) and information (multivariate) visualization. In addition, Moreland's [11] work basically explores considerations in pipeline designs in which a simple visualization pipeline is presented and includes a sequential read, filter, and render functional modules which represent the data input, data transformation, and data output components.

Classical information visualization has evolved to include visual analytics [2]. While information visualization concerns the visual representation and presentation of abstract multivariate data mostly in low-dimension space, visual analytics concerns making sense of visual data through visual inspection, and hypothesis formulation and validation for human cognition tasks [2]. Thus, a generic visual analytics pipeline can be regarded as a series of connected steps in a data model required for visualizing and inspecting multivariate data in order to discover patterns and gain insights and knowledge to augment human cognition [1]. The basis of visual analytics is formed from the idea of human-computer interaction for problem-solving [2]. Since the knowledge of the security analyst about the current state of the network is limited, visual analytics helps in pattern discovery from which hypotheses can be formed and proven leading to insights and knowledge gains. Generally, information visualization and visual analytics data models convert raw data into visual objects which can be combined and

presented to the analyst as views [2] for inspection.

The visual analytics pipeline of Keim *et al.* [28] is well-known in literature and focused on knowledge generation from observable visual elements displayed on views. The authors introduced functional user interaction steps in an iterative fashion for the analysis and knowledge generation tasks. Ltifi *et al.* [29] and Sacha *et al.* [30] are also well-known visual analytics pipeline extensions of Keim *et al.* [28]. The Keim *et al.* [28] visual analytics model involves two phases: data mining reinforced with visual analysis. It includes all functional steps in a generic information visualization pipeline proposed by Card *et al.* [1]. It includes four modules of data processing, information visualization, pattern discovery, and knowledge generation as executables for automated data analysis which is reinforced with visual data exploration for interactive pattern building and visualization.

Alonso *et al.* [31] describe visual analysis and inspection of Honeypot data as an alternate network monitoring tool. Essentially, based on the reference visualization frameworks discussed in this section, a typical Honeypot information visual analytics model as shown in **Figure 2** for inspection of distributed Honeypot multivariate network attack data will require that raw data be aggregated, cleaned, compressed, filtered, mapped, rendered, and interactively analyzed with continuous qualitative feedbacks in order to discover patterns and gain explainable insights and knowledge into Honeypot intrusion detection data for cognitive action by the security analyst.

## 2.3. Graphical Perception and Methods

The discovery of attack patterns hidden in raw network traffic data and the generation of explainable insights by the human security analyst is strongly connected to how the human visual system perceives and reasons visual data. Hence, for effective knowledge generation from visual displays the knowledge in established visualization principles and best practices, and the expertise in the thoughtful selection of graphical methods and application of visual channels (**Figure 3**) for visual analysis tasks is paramount. In this section, we present a quick background to common graphical elements and methods in the context of common visualization principles and best practices.

### 2.3.1. Common Visualization Channels

In order to extract information from a graphical display, the human visual system (through the brain) typically performs basic perceptual tasks such as identifying the position of a point and estimating the difference in lengths of two lines by using graphical channels. The information extraction metrics (e.g. position, length) used in strategically encoding data (quantitative or categorical) are called visual channels. **Figure 3** shows a list of visual channels [32] [33] ranked (from most effective to least effective) in graphical methods for encoding of quantitative or categorical data. Typically, it continues to get more difficult to perceive small quantitative changes as one goes down the list. The authors revealed that

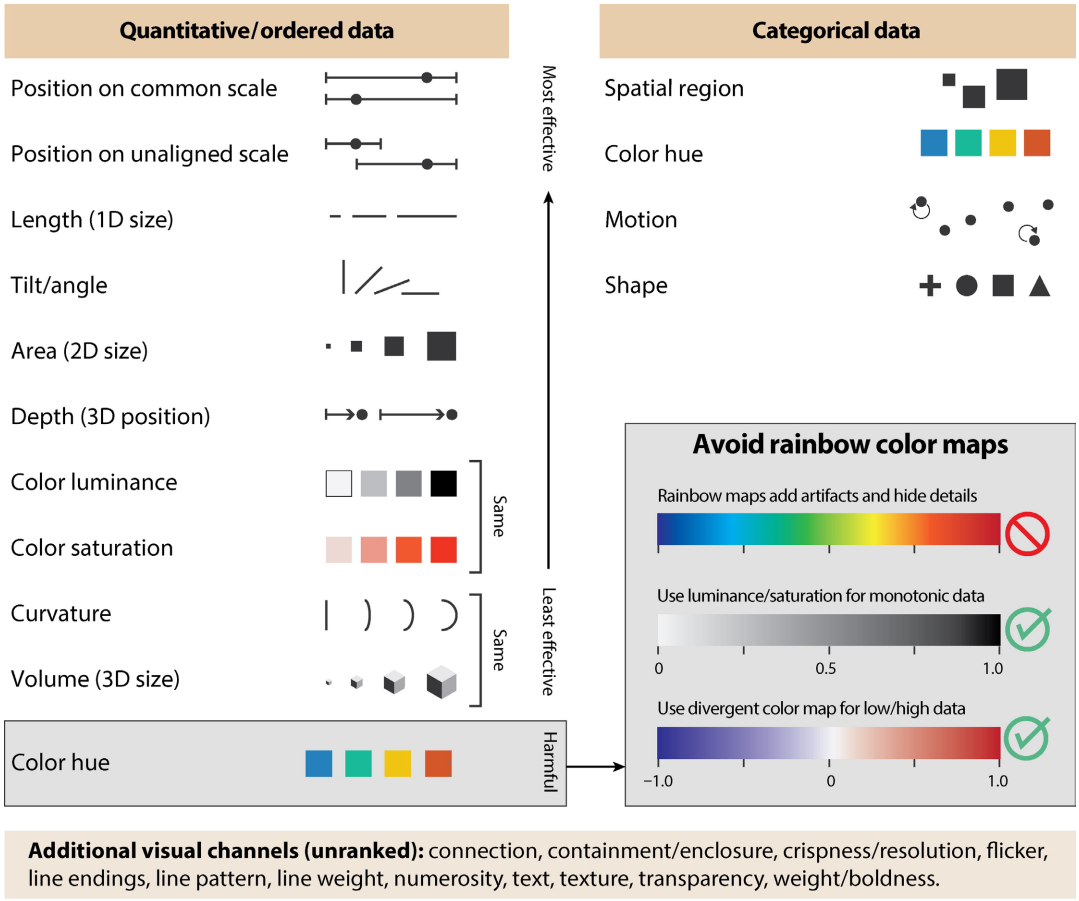


**Figure 3.** Visual channel ranking [32]. This figure shows the visual channels which are most and least effective for encoding and decoding of quantitative and ordered/categorical data. The most effective is position typically used in line/bar charts; area/curvature is typically used in pie charts and are less preferred (even discourage) compared to bar charts; luminance/saturation is also encouraged against hue when using color to encode quantitative data; color hue can add visual artifacts to visual representations of quantitative data and as such may obscure findings, however are very useful in representation of categorical data. These rankings are very useful and can serve as a guide for construction of effective graphical visualization methods.

the human visual systems can better perceive small changes (or differences) in quantitative data using length as the comparison channel (e.g. in line or bar chart). On the other hand, visual channels such as area, angle, direction, volume, curvature, shading or color proves relatively more difficult for the brain to process thus resulting in less effective graphical visual perception.

Color is regarded an essential part of visual information communication and is actively used in information encoding in visualization tasks [32]. When applied thoughtfully it can create aesthetics in graphical methods promoting visual perception. Color is mostly used to draw the viewer's attention to important aspects of a graphical visual display. However, the human visual system can have defects (e.g. color-blindness) making color to be perceived differently amongst viewers. Additionally, the use of color hue for encoding quantitative data have been scientifically discouraged in visualization research due to the possibility inclusion of obfuscating visual artifacts possibly misrepresenting data [32]. To ac-

commodate known visual defects, perception tailored color palettes such as Cividis and Viridis have been developed. In addition, it is scientifically encouraged to present color as shades of gray, which may not be visually appealing, but has been proven to communicate information in a clear and intuitive fashion.

Furthermore, values encoded with color can be surprisingly read inaccurately by the human visual system. For example, the perception of color can be influenced by other neighboring colors resulting in an optical illusion. This unintended visual effect can occur if different colors or variations of the same color are not thoughtfully applied in graphical method construction leading to possible information misrepresenting. In essence, when displaying quantitative data and small differences in measurements between entities is important, it is recommended to encode values with position, line, or size rather than lightness, saturation, or color hue. The visual channel rankings can serve as a guide for effective graphical construction during information visualization tasks.

#### 2.3.2. Common Visualization Methods

Line charts [34] are typically used with time-series data for revealing patterns and trends.

Pie charts [35] are mostly used to compare slices of a whole pie or show contributions of individual pie slices to the whole pie. As shown in **Figure 3**, area or angle are the visualization channels which can be used for comparison and as previously highlighted have been found to be less effective for discerning of quantitative values as compared to position or length used in line or batch charts. In addition, since pie chart is used to reveal or communicate proportion typically from a series data, it is encouraged to only visualize data in 2-dimensions. Displaying pipe chart in 3-dimensions may distort visual perception. Hence, pie charts are effective in presetting estimated contributions of parts to a whole, but the extra work performed by the brain during perception makes them less attractive for visualization tasks.

Just like in pie chart, a bar chart [36] is a simple to plot and intuitive graphical chart. It is mostly used for displaying counts or proportion and comparison of discreet classifications in data. Notwithstanding, graphical charts have peculiar strengths and limitations which should be known and understood. For instance, simplicity does not always translate into effective and unbiased visual data decoding. This is so since different datasets can lead to the same bar chart representations. Thus, bar charts may not reveal relationships, patterns, or actual distributions in data (e.g. outliers and clusters) which may mislead human thoughts or reasoning. Also, it is ineffective for visual communication of continuous data [37].

A number of graphical charts (e.g. scatter plot [38], box plot [39], histogram [40]) have been found to proof effective (and better than bar charts) for intuitive display and perception of continuous data [37]. These charts can show distribution in data and are mostly used for this purpose. As an example, using the scatter plot enables the security analyst to perceive data sample size, outliers, and

community clusters at a glance which could possibly be hidden when displayed as bar charts. A better visualization approach is to use a scatterplot matrix which can be used to show every pairwise combination of data features as bivariate scatter plots arranged in a matrix form, consequently visualizing all possible 2-dimensional correlations in the data.

A graphical method typically affected by optical illusion is the heatmap [41]. It is mostly used to visualize patterns in massive dataset at a glance. A classical heatmap maps shades of different colors as data representations typically in 2-dimensions using brightness or hue level to differentiate between magnitudes in data values. Larger and smaller magnitudes are color-coded as darker and lighter shades respectively [41]. A limitation of this visual approach (like every other) concerns the possibility of creating unintended color artifacts (such as illusions or gradients) which may obscure actual patterns or introduce bias in how information is visually perceived. As the numbers of columns and rows of the heatmap increases or as the size of the cell reduces, this unwanted effect can increase, making it unrealistic to visualize large amount of data as a large heatmap. To limit this effect, it is scientifically encouraged to use grayscale palette or blackbody color palette to illustrate heatmap (and of course other graphical methods) as opposed to rainbow color palette. In addition, displaying extra withe border around the squared cells can limit the perceivable unintended effect. Other commonly used graphical methods (in literature and practice) are World map and Parallel plot.

## 3. Related Survey Studies

Due to the potency of Honeypot systems in deceptively engaging attack actors and trapping and logging their information, (such as network and system activities, and submitted and downloaded malicious software artifacts) for further analysis work, the Honeypot technology has received great attention from the research community (and practitioners alike) and numerous empirical studies focusing on detection of network intrusive attacks through visual analysis of captured Honeypot traffic data have been conducted [8] [9]. Consequently, there have been articles [20] [42] [43] presenting overviews on the typical Honeypot traffic data analysis metrics and methods.

We consider the work of Nawrocki *et al.* [20] as the only, current, and by far the most detailed survey work focusing on Honeypot data analysis methods in Honeypot research. The authors presented a broad overview of Honeypot software, active software deployment platforms, and Honeypot data analysis metrics and approaches. They identified and described over 60 Honeypot data features which could be used to characterize intrusive attacks, all categorized according to a set of predefined problem statements and/or analysis questions. Notwithstanding, the authors did not consider, in-depth, visualization techniques and tools used during analysis of the security logs generated from Honeypot sensors.

Furthermore, we identified related survey papers published in other research

fields, such as in visualization [2] [11] [44] [45] [46], network anomaly detection [12] [47] [48] [49], and data preprocessing and dimensionality reduction [13] [14] [25] [50]. However, these papers and their likes generally present a broad overview of visualization and data analysis topics such as tools, methods, techniques, pipelines, frameworks, and applications, and are therefore not specific to Honeypot data analysis and visualization.

For example, Moreland [11] presented a review of features of visualization pipelines from the view of visualization pipeline designs. The authors defined a simple pipeline design as a connected data flow model having three executable modules: 1) sources, which generates and/or reads data from inputs to outputs, such as file readers and synthetic data generators; 2) filters, which selects data and transforms data between different forms and/or formats, such as machine and deep learning dimensionality reduction methods [13] [49]; and 3) sinks, which operates on and/or writes data from inputs to outputs, such as file writers and image renderers. Each executable module was conceptualized as possibly having one or more data inputs and outputs connected in many-to-many relations.

Similarly, Ltifi *et al.* [2] presented a review visualization pipelines, models, and methods in literature. The authors reviewed various information visualization, visual analytics, and knowledge extraction data pipeline abstraction models such as: the generic visualization reference model of Card *et al.* [1]; the data state reference model of Chi [26]; the temporal and structural visualization models of Daassi Chaouki and Nigay [51]; the software design reference models of Heer *et al.* [52]; the visualization design and validation nested model of Munzner [53]; the semantic interaction model of Endert *et al.* [54]; the visual analytics model of Keim *et al.* [28]; and the visual analytics and knowledge generation model of Sacha *et al.* [30]. The authors also proposed a comprehensive visualization, visual analytics, and knowledge generation pipeline a previous study [29].

Zhang *et al.* [47] presented a review of common network data types and features, typical visual analysis tasks and applications, and classification of existing works on network anomaly visualization. In line with the suggestions of Fernandes *et al.* [12], Zhang *et al.* [47] suggested that visualization tools are used in networks for forensic log analysis, real-time network stream analysis for anomaly or attack detection. Additionally, the authors agreed that network data are of several types such as network alerts, anomalous traffic, attack patterns, etc. [49] and have properties such as temporal, tabular, topological, spatial, high-dimensional, tree, etc. [12] [46] and can be classified into three applied visualization tasks namely, detection and identification, correlation and classification, and awareness and assessment. Network alerts are typically generated to the security analyst in the form of messages in response to detected network attack patterns and/or anomalous traffic [55]. Authors further suggested several sources generated network traffic data for analysis such as alerts generated from security events, direct traffic packets or network flows derived as metadata of actual network traffic data [12].

Fernandez *et al.* [44], in line with Zhang *et al.* [47], acknowledged that visualization generally helps to transform data into simple and compressive perceptual components. The authors agreed that visualization proves useful in revealing patterns inherent in large amount of data [12] and presented a review of common interactive visualization techniques necessary for improving visual analysis tasks and cognition for the analyst. The authors suggested steps for transforming raw data into visual objects such as: mapping to a geometric shape, laying out of visual objects, conveying all data attributes visually, addition of interactive options, and rendering considerations of visual objects on view [2] [11]. Authors further suggested representative perceptual visualizations techniques as use of: different shapes, to information about convey different objects, different object sizes to indicate level of importance or value in relation to others, different colors in leu of different shapes, different depths to illustrate rising or reducing rate, addition of textures, using different opacities, and the use of appropriate labelling of visual objects [46]. Authors also highlighted different sets of 2D and 3D interactive techniques such as filtering zooming, glossing, panning, rotation, scaling and lightening [2].

Yan *et al.* [48] investigated machine learning and deep learning-based methods of visual anomaly detection particularly in image data. The authors suggested image anomaly detection, like raw network traffic data, has promising applications, for example in medical image analysis for detection of lesions, and in intelligent manufacturing, for detection of defects. Authors identified and classified unsupervised visual anomaly detection methods based on two themes: granularity (image and pixel level) and history (pre and post deep learning). Image level detection methods such as Gaussian mixture model [56] and one-class support vector machines [50], and the image reconstruction autoencoders [57], investigates if an image is anomalous, while the pixel level methods such as the deep convolution and generative autoencoders [58] [59], sparse coding [60], Kmeans clustering [61], Gaussian mixed model, and other machine learning ensemble models [62] [63], focuses on detections of anomalous image regions.

Similar to the work of Zhang *et al.* [47], Fernandez *et al.* [44], and Yan *et al.* [48], Soo-Yeon *et al.* [49], presented a review of common visualization techniques for pattern discovery particularly during network traffic data analysis. The authors suggested that in network traffic data analysis, the effectiveness of perceptual visual methods have not been fully explored and that several visualization methods have been proposed and developed to aid effective detection of unknown intrusive attacks. The authors suggested four major categories of approaches to effective network traffic visualization as found in literature are: 1) data filtration and transformation, in which the right amount of data is first selected and then transformed from a high dimension to a low dimension for easy visualization [13]; 2) graph based data representation, in which data properties is captured using a complex graph structure visualized as nodes and edges [64]; 3) pixel-based, in which the entire view of the network data is visualize using vari-

ous colors [65] [66]; 4) multiple view based, in which multiple coordinated views are used together for better understand and insights into network traffic data [5]. Examples of common visualization techniques identified by authors are pie charts, line graph, bar graph, scatter plot, radial and glyph visualization, heat map, parallel coordinate, link view and node-link diagram [46]. The node-link diagrams are graph based data representation methods that are commonly used to reveal connectivity and communication behaviors between source and target systems in a network.

Additionally, several data filtration and transformation methods were identified by Soo-Yeon *et al.* [49], such as intent selection, feature selection, sample selection, data aggregation, linear discriminant analysis (LDA), principal component analysis (PCA), singular value decomposition (SVD), Multi-Dimensional Scaling (MDS), and Self-Organizing Maps (SOM). Similarly, Sorzano *et al.* [13] presented a review and categorization of various high dimensional data reduction techniques commonly used for easy and effective data visualization such as: wavelet and empirical decompositions methods, vector quantization methods (e.g. Kmeans), PCA and variants (e.g. incremental, stream, online, nonlinear, sparse, rotational, robust, and kernel), Manifolds and variations, SOM and variants, Factor analysis, dictionary based methods, and methods based on projections such as MDS and Locally linear embedding (LLE). The authors acknowledge that while the most used techniques are the component analysis-based methods such as the PCA and the projection and dictionary-based dimensionality reduction methods are becoming prominent.

## 4. Methodology

This study follows the PRISMA (Preferred Reporting Items of Systematic Reviews and Meta-Analysis) version 2020 methodology [15] [67] for selection and evaluation of original research papers studying networks threats and attacks captured in Honeypot traffic data. The methodology can be used for synthesizing existing scientific literature and uncover findings from results. It basically involves two phases: planning—where necessary research preparations are made and study focus are defined; and execution—where papers in literature are searched, identified, screened, reviewed, and reporting. The PRISMA systematic approach is beneficial as it guarantees consistency in preparation and execution of all types of systematic reviews and meta-analysis studies therefore enabling research reproducibility. In addition, it ensures that the overall study outcome is free of bias.

### 4.1. Eligibility Criteria

Empirical studies published in English language by reputable research-oriented publishers, and focusing on Honeypot traffic data analysis and visualization, from inception to year 2021 were considered. Identified papers pulled from search results which were not relevant to our study were excluded, for example,

surveys and/or review papers, and other empirical studies which either were not related to Honeypot in any way or did not include graphical figures such as plots, maps, or charts showing data/information supporting key findings relating to discovery of intrusive attack patterns.

To be selected, empirical studies needed to visually analyze Honeypot data and present analysis results and findings as graphical charts, maps, and/or plots in figures. Analysis data must be directly related to Honeypot, firewall, and other network intrusion detection technologies. To be included in this study, selected papers needed to be highly relevant and as such should be published by a reputable research-oriented organization in a high impact peer-reviewed conference proceeding or journal. We reason that articles satisfying these guidelines are of high relevance to the research community and were authored by researchers with considerable cybersecurity expertise. Thus, we would be able to evaluate and model the common visualization practices of the typical knowledgeable security analysist, thereby limiting bias in the overall outcome of the study.

### 4.2. Information Sources and Search Strategy

To identify original research with focus on visual analysis of Honeypot traffic data, an initial Google Scholar search was conducted on the 10th of January, 2022 using an automated python script [68] executed with default parameters through the Google Collaboratory web interface. The automated python script was configured to rank publications by number of citations (per year) with the aim of revealing the most relevant empirical papers in the field. As shown in Table 1, the authors combined the following keywords as a search string in separate searches:

- "Honeypot" and "analysis"—as a search string
- "Honeypot" and "visualization"—as a search string

We saved the search results as a single CSV file and uploaded to Google Docs as a spreadsheet search database for further processing. Records of articles in the search database were found to have been indexed (by the Google Scholar search engine) from several online bibliographic databases. Most notable and of interest are: SpringerLink [69], ScienceDirect [70], IEEE Xplore [71], and ACM Digital Library [72].

Table 1. Characteristics of the paper search process: This table displays the search engine used, the search tool/interface, the date of coverage for each search, the search strings used, and the number of returned articles by database. Two Google Scholar searches were conducted on the 10th of January, 2022 using an automated python script [68] executed with default parameters through the Google Collaboratory web interface.

| Search Engine | Search Interface | Coverage | Search String | Sources | # of Records |
|---|---|---|---|---|---|
| Google Scholar | Google Colab with automated python script for citation ranking | 2003-2021 | 1st search string: Honeypot AND analysis<br>2nd search string: Honeypot AND visualization | SpringerLink [69]<br>ScienceDirect [70]<br>IEEE Xplore [71]<br>ACMDigitalLibrary [72]<br>Others | 25<br>2<br>68<br>9<br>114 |

Table 1 shows the attributes that characterizes the search operation such as: search engine type, search tool/interface type, date of coverage, search keyword and string, and source name.

### 4.3. Selection of Sources

This study follows a three-step process (data cleaning, first screening, second screening, and article rating) for selection of sources based in the aforementioned inclusion and exclusion requirements.

#### 4.3.1. Data Cleaning

The search database was first preprocessed as follows:

- Records were sorted by paper title and author;
- Records with no author(s) and/or title were identified and removed;
- Direct web URL to articles were retrieved (*i.e.* for records without links); and
- The articles were again sorted and grouped by bibliographic database.

Articles from SpringerLink [69], ScienceDirect [70], IEEE Xplore [71], and ACM Digital Library [72] were selected for screening in an attempt to include only qualitative empirical papers.

#### 4.3.2. First Screening

In this step, the title and abstract of each article were independently studied and the contents were traversed quickly in search of any form of graphical figures. Papers with the following characteristics were excluded:

- Duplicates records were removed;
- Papers not presented in English language;
- Papers not presenting original research (e.g. review, survey, etc.);
- Papers not related to Honeypot;
- Papers not focusing on analysis of Honeypot traffic data; and
- Papers not presenting graphical figures.

#### 4.3.3. Second Screening

After the first screening, the full literature content of each article retained were independently reviewed and analyzed. The authors discussed the results of the screening process and agreed on the most relevant articles to be selected. In this step, papers not relevant to our study were further excluded, in particular:

- Papers not directly related to Honeypot; and
- Papers not specifically presenting analysis results visually using graphical charts, plots, maps, diagrams (e.g. bar, pie, scatter plots) etc. were further identified and excluded.

Only papers presenting graphical figures highlighting important discoveries (such as attack frequency, temporal trends, and spatial and topological patterns) in Honeypot data analysis were selected. Graphical figures are important in showing the visual data that support key findings/discoveries as outcomes of data analysis tasks. Hence, graphical figures presented in papers can serve as an

evaluation metric for determining the level of knowledge and expertise in data visualization.

Selected papers after the second screening were regarded as eligible and included in this study for further review and analysis.

### 4.4. Data Collection Process

In order to categorize, synthesize, and analyze each paper included in the study, we collected data items characterizing the paper (Table 2) and the study.

#### 4.4.1. Data on the Paper

- Author—name(s) of the authors and reference to as stated in our search database.
- Citation—this is the number of citations as stated in our search database.
- Year—this is the year of publication as stated in our search database.
- Type—this is the of publication type as stated in our search database (e.g. journals, conferences, book series).

#### 4.4.2. Data on the Study

We studied the included Honeypot data analysis papers independently to identify typical visualization techniques used in the display and discovery of intrusive attack patterns inherent in collected Honeypot attack data.

- Graphical visualization methods—selected charts, plots, maps, etc.

Additionally, we extracted the following data items from the graphical figures in order to evaluate and model visualization practices of a presumed skilled cybersecurity researcher. We aim to understand how the following factors influence what visualization methods are selected.

- Security metric combinations—data features selected for analysis.
- Analysis method—the analysis depth/breath e.g. basic or extended.

### 4.5. Synthesis of Results

We tabulated extracted data items characterizing the included paper for quick reference and statistical analysis. We also tabulated and charted the extracted data items characterizing our study, and finally synthesized and discussed our findings.

## 5. Results and Discussion

In this section, we present our results, analysis, findings on the characteristics of the included papers and the characteristics of the extracted graphical figures.

### 5.1. Paper Characteristics

#### 5.1.1. Results

The PRISMA flow diagram of the source selection process is shown in Figure 4. It summarizes how the most relevant Honeypot related papers were identified, screened, and selected from our search database.

**Table 2.** Paper characteristics: This table shows the Honeypot emperical papers included in this work and their physical features. The number of citations per year, helps to show the relevance of paper irrespective of the publication year. It also shows the observed graphical methods generally used in Honeypot data analysis for visualization and presentation of attack partterns.

| Paper | # of Citations | # of Citations per Yr. | Year | Type | Publisher | Graphical Method |
|---|---|---|---|---|---|---|
| Chuvakin [73] | 25 | 1 | 2003 | Journal | Elsevier | Bar |
| Baykara [74] | 39 | 10 | 2018 | Journal | Elsevier | Bar, Pie |
| Vasilomanolakis [75] | 59 | 8 | 2015 | Conference | ACM | Line |
| Kobayashi [76] | 1 | 0 | 2019 | Conference | ACM | Bar, Hilbert-curve |
| Belqruch [77] | 9 | 3 | 2019 | Conference | ACM | Pie, Bar |
| Song [78] | 251 | 23 | 2011 | Conference | ACM | Line, Area, Pie |
| Singh [79] | 24 | 2 | 2011 | Conference | IEEE | Pie |
| More [80] | 7 | 1 | 2013 | Conference | IEEE | Line |
| Almotairi [81] | 43 | 3 | 2009 | Conference | IEEE | Line, Scatter |
| Koniaris [82] | 42 | 5 | 2013 | Conference | IEEE | Bar, Line, World map |
| Almotairi [83] | 38 | 3 | 2008 | Conference | IEEE | Scatter |
| Chen [84] | 30 | 0 | 2005 | Conference | IEEE | Pie, Bar, Line |
| Kyriakou [85] | 10 | 2 | 2018 | Conference | IEEE | Line, Bar |
| Fraunholz [86] | 15 | 3 | 2017 | Conference | IEEE | Line, Bar |
| Yeh [87] | 18 | 1 | 2008 | Conference | IEEE | Pie |
| Fraunholz [88] | 12 | 2 | 2017 | Conference | IEEE | Pie, Line, Bar, Histogram |
| Visoottiviseth [22] | 17 | 2 | 2011 | Conference | IEEE | World map |
| Pomsathit [89] | 13 | 1 | 2012 | Conference | IEEE | Line |
| Lihet [90] | 4 | 1 | 2018 | Conference | IEEE | Bar |
| Koniaris [91] | 22 | 3 | 2014 | Conference | IEEE | Bar, Pie, Line, World map |
| Wang [92] | 5 | 2 | 2020 | Conference | IEEE | Pie, Bar, Line |
| Sethia [93] | 10 | 3 | 2019 | Conference | IEEE | Bar |
| Krasser [5] | 143 | 8 | 2005 | Conference | IEEE | Parallel plot, Scatter plot |
| Sokol [4] | 6 | 1 | 2015 | Conference | IEEE | Heat map |
| Shyla [94] | 0 | 0 | 2021 | Conference | IEEE | Area chart, World map |
| Dowling [95] | 12 | 2 | 2017 | Conference | IEEE | Line chart, World map |
| Lakh [96] | 0 | 0 | 2019 | Conference | IEEE | Pie chart, Bar chart |
| Djap [8] | 0 | 0 | 2021 | Conference | IEEE | Donut chart |
| Moore [97] | 13 | 2 | 2015 | Conference | Springer | World map, Bar chart |
| Sánchez [98] | 1 | 0 | 2015 | Conference | Springer | Scatter plot, Dendrogram chart |
| Cao [99] | 12 | 2 | 2017 | Conference | Springer | World map |

**Continued**

| | | | | | | |
|---|---|---|---|---|---|---|
| Kemppainen [21] | 5 | 0 | 2018 | Book | Springer | Bar, Pie, Line, World map |
| Abbas [100] | 4 | 0 | 2007 | Conference | Springer | Bar chart |
| Nicomette [101] | 55 | 5 | 2011 | Journal | Springer | Line chart, Scatter plot |
| Agrawal [102] | 12 | 2 | 2017 | Journal | Springer | Line chart |
| Alonso [3] | 9 | 1 | 2010 | Conference | Springer | Scatter plot |
| Zurutuza [103] | 11 | 1 | 2011 | Conference | Springer | Scatter, Point chart |

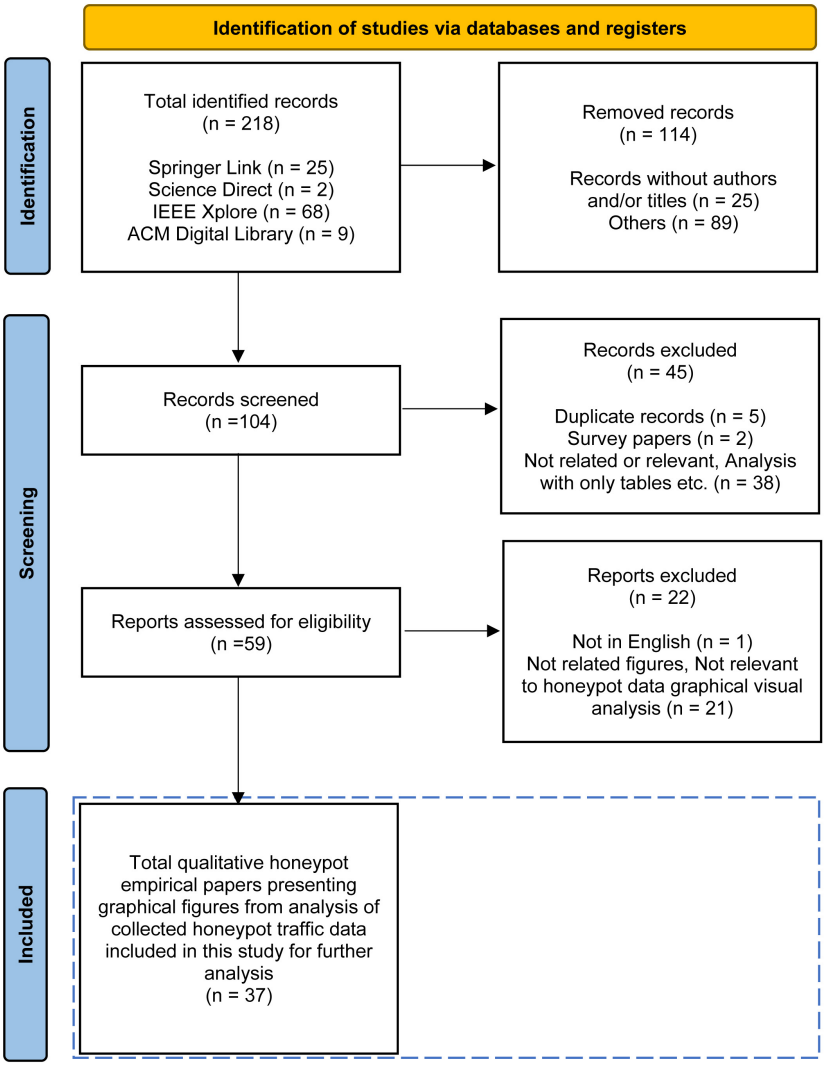


**Figure 4.** The PRISMA flow diagram showing the process of systematic literature review in this research. Papers were identified through Google Scholar searches, filtered only from major bibliographic sources, screened for eligibility, and finally selected and included in this study.

A total of 218 survey papers were initially retrieved: 25 from SpringerLink [69]; 2 from ScienceDirect [70]; 68 from IEEE Xplore [71]; 9 from the ACM Digital Library [72]; and 114 from remaining sources. Before the screening

phase, 25 records without authors and/or titles, and 89 records not belonging to the aforementioned bibliographic databases were removed, remaining 104 records. After the title and abstract screening 45 articles were excluded: 5 duplicates, 2 were survey papers, and 38 were either not original research, directly related to Honeypot technology, or presented Honeypot analysis results using only tables. Thus, we retained 59 papers for further investigation. After the full-text screening, we further excluded 22 papers which were not relevant to this study: 1 paper was not presented in English language and the others were either not related to graphical visual analysis of Honeypot traffic data. Thus, we selected 37 qualitative Honeypot empirical papers, which were included in this study.

#### 5.1.2. Analysis

Analysis of the publication source reveals that about 17% of the total identified papers were selected as related and relevant for this study. **Figure 5** shows the contributions of the major online bibliographic sources selected for initial screening. The dark gray right bars show the number of excluded papers while the left light gray bars show the number of included papers. It can be observed that the highest number of selected (46) and included (22) papers respectively are from IEEE Xplore [71]. In addition, most of the papers selected from both ACM Digital Library [72] and ScienceDirect [70], and about half of the papers selected from SpringerLink [69] were included in the study. Furthermore (as shown in **Table 2**), more than half of the included papers were published by IEEE in conference proceedings between 2005 and 2021, and only four (4) papers were published in journals between year 2003 and 2018.

### 5.2. Figure Characteristics

#### 5.2.1. Common Methods

Graphical visual methods such as charts, plots, diagrams, and maps are typically used for compressing and encoding volumes of textual tabular data for effective information display, visualization, pattern discovery, and information communication.

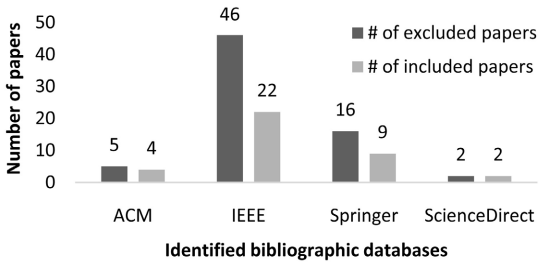


**Figure 5.** The contributions of major online bibliographic sources selected for initial screening. The dark gray left bars show the number of excluded papers while the right light gray bars show the number of included papers. It can be observed that the highest number of selected (46) and included (22) papers respectively are from IEEE Xplore. Also, for effective visual perception, the legends which give useful information about excluded and included papers were brought closer to the bars which were also color-coded using shades of grey and were apropriately labelled.

Based on the number of included papers observed presenting relevant graphical figures (Table 3), the following visualization methods (and related variants) have been identified as the commonly used charts: Bar chart (46%), Line chart (49%), Pie chart (32%), World map (22%), and Scatter plot (19%). Of these, Bar charts was observed to be the most used (37 figures) followed by Line chart (22 figures) and Pie charts (21 figures). Other graphical methods, such as Heatmap, Parallel plot, Dendrogram, Histogram, and Hilbert-curve have been sparingly observed in the included papers.

In essence, about the same types of visualization techniques have been found in literature since year 2003 for visualization of Honeypot data and communication of analysis results. Therefore, it is only logical to infer that similar class of patterns have consistently been unearthed from Honeypot security data and revealed both to the analyst and the readers. We suggest to security analysis that it is high time this visualization practice change. We belief that graphical charts are like different 2-dimensional X-ray lenses only having the capability to reveal and communicate only a part of the full story inherent in the original data. Business data are typically multivariate and live in high dimensional space thus requiring different views (*i.e.* multiple views) from different lenses (*i.e.* different chart types) for comprehensive analysis and storytelling.

**Table 3.** Distribution of graphical methods as identified from figures in the 37 included Honeypot papers. As shown, Bar chart (17 papers), Line chart (18 papers) and its variations, Pie chart (12 papers) and its variations, World map (8 papers), and Scatter plot (7 papers) have been identified as the commonly used graphical methods in visualization of Honeypot traffic data. Of these, Bar chart is observed to be the most used. The first three chart types are typically used for basic statistical analysis, while the others for discoveries of the depth and breadth of intrusive attacks.

| Graphical Method | # (%) of Papers | # of Figures |
|---|---|---|
| **(a) Basic visual methods** | | |
| Bar chart | 17 (46%) | 37 |
| Line/Point/Area chart | 18 (49%) | 22 |
| Pie/Donut chart | 12 (32%) | 21 |
| **(b) Extended visual methods** | | |
| Histogram | 1 (3%) | 2 |
| World map | 8 (22%) | 9 |
| Heatmap | 1 (3%) | 2 |
| Scatter plot | 7 (19%) | 7 |
| Parallel plot | 1 (3%) | 1 |
| Dendrogram | 1 (3%) | 1 |
| Hilbert-curve | 1 (3%) | 1 |

### 5.2.2. Analysis Approach

We observed that Bar charts, Line charts, Pie charts, and other related variants such as Area chart, Point charts, and Donut charts are visualization techniques typically used in basic statistical analysis of Honeypot traffic data as shown in Table 3.

As previously discussed, Line charts (and variants) are typically used in investigating attack trends and interesting temporal patterns [86] [91] [94]; while the Bar and Pie charts are typically used for investigating attack frequencies or proportions in relation to some categorical Honeypot security data features [8] [21] [87] [88]. The other methods are typically used for extensive data analysis such as in attack community detection [83] [98] and geospatial attack correlation [97] [99].

Furthermore (as shown in Figure 6), relevant graphical figures presenting the basic visualization techniques were identified in a total of 30 papers out of the 37 included papers while the extended methods were observed in only 18 papers. Therefore, the basic visual methods (about 81%) were observed in the included papers more than the extended visual methods (about 49%). We attribute this to general understanding that that the basic visualization techniques are familiar, simple to construct, and are easily understood by the analyst. This observation thus supports our argument, that basic analysis of Honeypot traffic data is the most conducted and the extended visual methods such as one which reveal attack connection distribution (e.g. histogram), community and/or correlation patterns (e.g. scatter plots) are seldomly used.

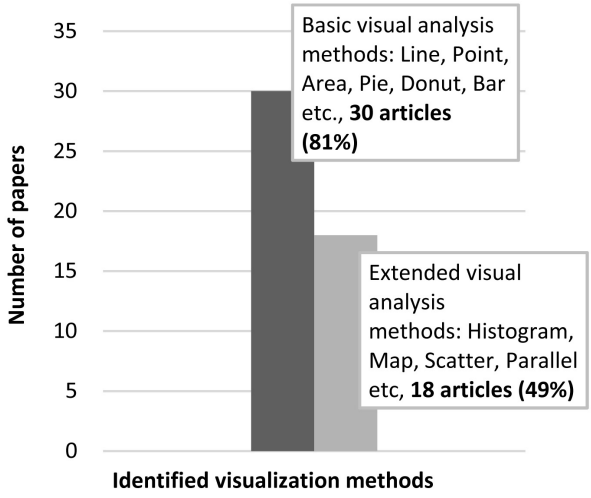


Figure 6. The number of papers with Identified basic or extended visualization methods. Relevant graphical figures presenting the basic visualiza-tion techniques were identified in a total of 30 papers out of the 37 included papers while the extended visualization methods were ob-served in only 18 papers. This further shows that basic visual methods are most used and basic Honeypot data analysis are most con-ducted. Also, for effective visual perception and presentation, the bars were color-coded using shades of grey and were appropriately labelled which give useful information about basic and extended analysis methods at a glance.

Also, for effective visual perception and presentations, the bars shown in **Figure 6** were intentionally color-coded using shades of grey and were appropriately labelled which give useful information about basic and extended analysis methods at a glance.

#### 5.2.3. Visualization Tasks

With further analysis of the visualization tasks carried out during visual analysis of the Honeypot traffic data in the included papers, we identified analysis questions relating to the attacker/attack source (human or malware), target (Honeypot), and network traffic connections leading to the selection of security metrics and associated graphical methods as shown in **Table 4**. As expected, attack frequency/volume (*i.e.* the count or proportion the attackers' incoming connections) are commonly displayed using Line and/or Bar charts with respect to Time Unit (e.g. Hour, Day, Week, and Month) [77] [82] [86] [94]. Also, Pie chart, Bar chart, and World map are the graphical display methods typically used during investigation of the target and/or attacker/attack source (e.g. destination port, source internet protocol, country, malware/shellcode type, operating system, username/password etc.) [78] [96] [97] [99].

We also observed that Heatmap was used solely for temporal analysis using different time units for revealing temporal patterns (e.g. nocturnal and diurnal) in Honeypot data [4]. In detecting attack patterns, researchers typically clean, scale, and transform Honeypot data from high-dimension to low-dimension (e.g. 2D or 3D) to enable easy visualization. In such a case, low-dimension features of the transformed Honeypot data are typically charted using scatter plot (or variations) as the graphical visualization method [5] [81] [83] [103]. As previously discussed, the scatter plot can be used to show distribution in data and as such have been observed to be the most used graphical display for revealing attack cluster and anomaly in Honeypot data.

#### 5.2.4. Visualization Trend

As previously stated, the visualization trend or practice favors frequent and infrequent use of basic and extended visualization methods respectively and the graphical methods type are somewhat evenly distributed over the study period as shown in **Table 3**. Notwithstanding, we observed that researchers are beginning to use uncommon and advanced visualization methods such as Histogram (in 2017) and Hilbert-curve (in 2019) for Honeypot data analysis. Other researchers [5] [75] [104] [105] [106] even created customized visualization techniques for out-of-the-box pattern discovery and perceptual experience. This, is a positive development in our opinion as scientific researchers have already been encouraged [32] [37] to move on from basic visualization techniques to better and sophisticated techniques which have been tested and proven to be more effective in decoding encoded data, and revealing patterns.

### 5.3. Summary and Recommendations

There are numerous information visualization methods which can be used for

Table 4. Common Honeypot visualization tasks: This table shows the commonly investigated attack metrics, features, and/or analysis questions and their associated graphical visualization methods used during Honeypot attack data analysis for attack pattern discovery and reporting.

| Metric | Graphical Method | Ref. |
|---|---|---|
| Distribution of Attacker Source IP Connections per Time Unit | Line | [21] |
| Number of Attacker Session by Time Unit | | [75] |
| Number of Attacker Shellcode Sessions by Time Unit | | [78] |
| Number of Attacker Source IP (Connections) per Ports Sequence (per Sensor) per Time Unit | | [80] |
| Number of Attacker Source IP Connections per Distinct Attacker Source IP | | [82] |
| Number of Attacker Source IP Connections per Distinct Protocol | | [84] |
| Number of Attacker Source IP Connections per Time Unit | | [85] |
| Number of Attacker Source IP over Time Unit | | [86] |
| Number of Attacker SSH Sessions by Time Unit | | [88] |
| Number of Malware Attack Connections per Exposed Destination Port | | [89] |
| Number of Packets per Time Unit | | [91] |
| Number of Packets per Time Unit per IDS Type | | [92] |
| Number of Unique Attacker Source IP per Time Unit | | [94] |
| Number of Unique Exposed Honeypot Ports per Time Unit | | [95] |
| Proportion of Attacker Source IP across Targeted Honeypot Sensors | | [101] |
| | | [102] |
| Distribution of Antivirus Alerts by Antivirus type | Pie | |
| Distribution of Attacker Open Session by Country | | [8] |
| Distribution of Attacker Source IP Connections by Country | | [21] |
| Distribution of Attacker Source IP Connections by Distinct Attacker Source IP | | [74] |
| Distribution of Attacker Source IP Connections by Distinct Protocol | | [77] |
| Distribution of Attacker Source IP Connections by Exposed Destination Ports | | [78] |
| Distribution of Attacker Source IP Connections per Attack Type | | [79] |
| Distribution of Attacker Source IP Connections per Distinct Protocol | | [84] |
| Distribution of Attacker Source IP Connections per Distinct Username/Password/Combination | | [87] |
| Distribution of Attacker Source Packets by Distinct Attacker Source IP | | [88] |
| Distribution of Attacker Source Packets by Distinct Protocol | | [91] |
| Distribution of Distinct Attacker Source IP by Country | | [92] |
| Distribution of Distinct Attacker Source IP Connections by Country | | [96] |
| Distribution of Distinct Malware by Country | | [99] |
| Distribution of Malware Samples by Architecture | | |
| Distribution of Shellcode Alerts by Shellcode type | | |
| Number of Attacker Source IP Connections by Exposed Destination Ports | Bar | [8] |
| Distribution of Attacker Source IP Connections per Subnetwork Class | | [21] |
| Number of Attacker Source IP Connections per Connection Status (Failure, Success) | | [73] |
| Number of Attacker Source IP Connections per Time Unit | | [74] |
| Number of Distinct Connection Commands per Time Unit | | [76] |
| Number of Attack Sessions by Exposed Destination Ports | | [77] |
| Number of Attacker Source IP Connections per Country | | [82] |
| Number of Attacker Source IP Connections per Distinct Operating System Name | | [84] |
| Number of Attacker Source IP per Country | | [85] |
| Number of Attacker Source IP Connections per Connection Type (Attack, Intrusion, Total Traffic) | | [86] |
| Number of Attacker Source IP Connections per Distinct Attacker Source IP + Country Code | | [88] |
| Number of Attacker Source IP Connections per Distinct Protocol | | [90] |
| Number of Malware Samples per Distinct Malware Name | | [91] |
| Number of Malware Attack Connections per Exposed Destination Ports | | [92] |
| Number of Attacker Source IP Connections per Distinct Attacker Source IP | | [93] |
| Number of Attacker Source IP Connections per Distinct Username/Password/Combination | | [96] |
| Number of Attacker Source IP Connections per Attack Type | | [97] |
| Number of Attacker Source IP Connections by Distinct Destination IP | | [100] |

**Continued**

| | | |
|---|---|---|
| Time Unit by Time Unit<br>Exposed Destination Ports by Time Unit | Heatmap | [4] |
| Subnetworks of Scanning Sources | Hilbert-curve | [76] |
| Number of Passwords by Password Length<br>Number of Attacker Unique Source IP by Inefficiency Ratio | Histogram | [88] |
| Attacker Source IP by Exposed Destination Port | Parallel | [5] |
| Number of Attacker Source Packets per Distinct Attacker Source IP + Location (e.g. Country Code)<br>Number of Attacker Source IP Connections per Distinct Attacker Source IP + Location (e.g. Country Code)<br>Number of Attacker Source IP Connections per Distinct Attacker Protocol + Source IP Location (e.g. Country Code)<br>Number of Attacker Source IP Connections per Distinct Attacker Source IP + Location (e.g. Country Name)<br>Destination IP + Location (e.g. Country Code) | World Map | [21]<br>[22]<br>[82]<br>[91]<br>[94]<br>[95]<br>[97]<br>[99] |
| Variance Components<br>Attacker Source Packet Size by Time Unit<br>English Dictionary Words by French Dictionary Words | Scatter, Dendrogram | [3]<br>[5]<br>[81]<br>[83]<br>[98]<br>[101]<br>[103] |

display of intrusion detection data and discovery of hidden attack patterns. Notwithstanding, the security analyst is expected to be adequately skill in data visualization and also encouraged to be thoughtful when selecting graphical methods and using colors as these may influence visual perception. It has been observed that scientific researchers mostly consider data visualization a trivial task and as such do not get adequate training or follow data visualization principles and best practices during visual analysis tasks. This misconception, have been identified as a major cause of ambiguous or misleading information visualization and/or visual communication, limiting intuition and consequently, insights and knowledge gains.

The results of this study suggest that the basic charts (such as line, bar and pie) are still actively used in data visualization of intrusion detection data although visualization research has already suggested a move to charts with higher data density, visual effectiveness, and visual expressiveness [32] [37] capable of effectively revealing and communicating hidden patterns in data. Data density describes the amount of information shown in a visual display as compared to the whole display area while visual effectiveness and expressiveness relates to the correctness, clarity, adequacy, and relevance of display data. In essence, we suggest that these basic charts have been used enough and it is high time researchers develop skill in other visual methods both simple and sophisticated which are capable of revealing different underlaying structures in data and also viewing

data from a different aspect.

In general, it is recommended to plan ahead and consider the visualization task and the information audience (both the researcher and viewers e.g. novice, professional) in advance to determine what and how graphical methods would be used. Furthermore, it is encouraged to gain contemporary knowledge and skill in graph construction, and be thoughtful when selecting graphical methods and applying colors in order to ensure that information is appropriately perceived. The following are essential visualization principles and best practices [32] enabling effective and expressive visualization:

- Show and highlight relevant data (e.g. using different shapes, and colors) in the graph as clearly as possible to ensure patterns are observable and the story being told is well communicated;
- Limit unnecessary information (*i.e.* visual artifacts e.g. 3-D in pie chart, bold grid lines, etc.) introduced in graphs that may cause clutter and/or distortion;
- Ensure the graph is well-constructed with the required information to be able to stand alone and as well as complement the text to tell the story;
- Construct graphs using visual metrics easy for the brain to process and the eyes to perceive such as using position, length, size, shape, orientation, and color;
- Annotate important graph sections to give clearer meaning and understanding (e.g. placing explanatory legends at the relevant sections on the graph as opposed to outside the graph);
- Avoid a single busy graph and break it up into smaller simpler charts that could be viewed together (e.g. single-line charts, scatter plot matrix); alternatively, construct compact visualization with high data density using visual channels which enables high visual effectiveness;
- Where possible, avoid the use of pie charts (and its variants) when communicating quantities, bar charts, tree maps or slope charts are often better;
- Ensure labels are concise, explanatory and well positioned for easy visual perception (e.g. rotate bar chart/Y-axis labels to horizontal position where possible, start the count axis in bar chart at zero);
- Use maps with care, make sure to select the required type, and avoid unnecessary information and clutter (e.g. shades, color, texture), bar chart is often better;
- Use color with care, consider and accommodate visual defects (e.g. color blindness) during graph construction, avoid default fonts and rainbow colors, and use gray color shades or other well-represented color palettes where possible;
- Large data are typically multivariate, reduce data complexity through dimensional reduction (e.g. PCA), visualize all aspects of the data before drawing conclusions (e.g. 2D scatter matrix plot);
- Learn and use uncommon visualization methods for revealing different aspects of data;
- To clearly reveal patterns and communicate information effectively, endea-

vor to construct tailored (custom) visualizations as opposed to using generic visualization methods (e.g. customizing generic methods by adding annotations, and spatial information which can help the brain in easy cognitive reasoning);

- It is noteworthy that contemporary visualization tools and applications have sophisticated and highly configurable visualization methods commonly designed for aesthetics and useful in the creation of artworks. However, these tools may introduce unnecessary clutter in visualized data thus obfuscating patterns and insights. So, care should be taken when using visualization tools.

## 6. Conclusions

Information visualization has proven beneficial in intrusion detection data analysis and research as an effective data exploration and network monitoring framework. It typically relates to the mitigation of network adversarial attacks and intrusions detected in collected intrusion detection traffic data through automated data analysis and human visual perception and inspection of encoded graphical visual methods such as charts, plots, maps and diagrams.

In this work, we systematically surveyed Honeypot research papers focusing on the analysis of gathered intelligence data in order to identify the graphical visualization methods used by security researchers during the analysis of intrusion detection data and to evaluate their knowledge and skill in data visualization principles and best practices. We focused on Honeypot data analysis to further bridge the gap between analysis and visualization subfields in Honeypot research. We extracted useful information from graphical figures presenting important findings observed during data analysis and presented our findings.

We observed that basic graphical charts (e.g. line, bar, and pie) are mostly used in the basic statistical analysis of Honeypot data while extended visual methods (e.g. scatter plots and world maps) are typically used for deeper analysis. It was further observed that a significant number of studies did not follow basic visualization principles and best practices in their use of color. This is evident from the use of colors from a rainbow color map in the included papers. Although it is easy to construct the aforementioned commonly used graphical visualization methods, it is generally recommend in visualization research that security analysts and researchers go father and beyond these visual techniques to enable the discovery of novel patterns inherent in business data. We believe that graphical methods are like 2-D X-ray lenses capable of revealing different hidden structures and relationships in data. Therefore, getting appropriate skills in data visualization and moving beyond well-known visualization methods may prove beneficial in the discovery of new patterns in intrusion detection data, useful in mitigating adversarial attacks in today's networks.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References


[1] Card, S., Mackinlay, J. and Shneiderman, B. (1999) Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann Publishers Inc., San Francisco, 579-581.

[2] Ltifi, H., Kolski, C. and ben Ayed, M. (2020) Survey on Visualization and Visual Analytics Pipeline-Based Models: Conceptual Aspects, Comparative Studies and Challenges. *Computer Science Review*, **36**, Article ID: 100245. https://doi.org/10.1016/j.cosrev.2020.100245

[3] Alonso, Á., *et al.* (2010) Understanding Honeypot Data by an Unsupervised Neural Visualization. In: Herrero, Á., Corchado, E., Redondo, C. and Alonso, Á., Eds., *Computational Intelligence in Security for Information Systems* 2010, Advances in Intelligent and Soft Computing, Vol. 85, Springer, Berlin, 151-160. https://doi.org/10.1007/978-3-642-16626-6_17

[4] Sokol, P., Kleinová, L. and Husák, M. (2015) Study of Attack Using Honeypots and Honeynets Lessons Learned from Time-Oriented Visualization. *IEEE EUROCON— International Conference on Computer as a Tool* (*EUROCON*), Salamanca, 8-11 September 2015, 1-6. https://doi.org/10.1109/EUROCON.2015.7313713

[5] Krasser, S., Conti, G., Grizzard, J., Gribschaw, J. and Owen, H. (2005) Real-Time and Forensic Network Data Analysis Using Animated and Coordinated Visualization. *Proceedings from the* 6*th Annual IEEE Systems, Man and Cybernetics* (*SMC*) *Information Assurance Workshop*, West Point, 15-17 June 2005, 42-49.

[6] Verizon (2021) 2021 DBIR Results & Analysis|Verizon. https://www.verizon.com/business/resources/reports/dbir/2021/results-and-analysis

[7] dos Santos, S. and Brodlie, K. (2004) Gaining Understanding of Multivariate and Multidimensional Data through Visualization. *Computers & Graphics*, **28**, 311-325. https://doi.org/10.1016/j.cag.2004.03.013

[8] Djap, R., Lim, C., Silaen, K.E. and Yusuf, A. (2021) XB-Pot: Revealing Honeypot-Based Attacker's Behaviors. 2021 9*th International Conference on Information and Communication Technology* (*ICoICT*), Yogyakarta, 3-5 August 2021, 550-555. https://doi.org/10.1109/ICoICT52021.2021.9527422

[9] Ghourabi, A., Abbes, T. and Bouhoula, A. (2013) Automatic Analysis of Web Service Honeypot Data Using Machine Learning Techniques. *International Joint Conference CISIS*'12-*ICEUTE*12-*SOCO*12 *Special Sessions*, Ostrava, 5-7 September 2012, 1-11. https://doi.org/10.1007/978-3-642-33018-6_1

[10] Ikuomenisan, G. and Morgan, Y. (2022) Meta-Review of Recent and Landmark Honeypot Research and Surveys. *Journal of Information Security*, **13**, 181-209.

[11] Moreland, K. (2013) A Survey of Visualization Pipelines. *IEEE Transactions on Visualization and Computer Graphics*, **19**, 367-378. https://doi.org/10.1109/TVCG.2012.133

[12] Fernandes, G., Rodrigues, J.J.P.C., Carvalho, L.F., Al-Muhtadi, J.F. and Proença, M.L. (2019) A Comprehensive Survey on Network Anomaly Detection. *Telecommunication Systems*, **70**, 447-489. https://doi.org/10.1007/s11235-018-0475-8

[13] Sorzano, C.O.S., Vargas, J. and Montano, A.P. (2014) A Survey of Dimensionality Reduction Techniques.

[14] Sacha, D., *et al.* (2017) Visual Interaction with Dimensionality Reduction: A Structured Literature Analysis. *IEEE Transactions on Visualization and Computer Graphics*, **23**, 241-250. https://doi.org/10.1109/TVCG.2016.2598495

[15] Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. and Group, T.P. (2009) Preferred

Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, **6**, e1000097. https://doi.org/10.1371/journal.pmed.1000097

[16] Spitzner, L. (2002) Honeypots: Tracking Hackers. Addison-Wesley, Boston.

[17] Silva, D. and Rodriguez, G. (2017) A Review of the Current State of Honeynet Architectures and Tools. *International Journal of Security and Networks*, **12**, 255-272. https://doi.org/10.1504/IJSN.2017.10009165

[18] Zobal, L., Kolář, D. and Fujdiak, R. (2019) Current State of Honeypots and Deception Strategies in Cybersecurity. 2019 11*th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops* (*ICUMT*), Dublin, 28-30 October 2019, 1-9. https://doi.org/10.1109/ICUMT48472.2019.8970921

[19] Fan, W., Du, Z., Fernández, D. and Villagrá, V.A. (2018) Enabling an Anatomic View to Investigate Honeypot Systems: A Survey. *IEEE Systems Journal*, **12**, 3906-3919. https://doi.org/10.1109/JSYST.2017.2762161

[20] Nawrocki, M., Wählisch, M., Schmidt, T.C., Keil, C. and Schönfelder, J. (2016) A Survey on Honeypot Software and Data Analysis. Cornell University Library, Ithaca.

[21] Kemppainen, S. and Kovanen, T. (2018) Honeypot Utilization for Network Intrusion Detection. In: Lehto, M. and Neittaanmäki, P., Eds., *Cyber Security*: *Power and Technology*, Springer International Publishing, Cham, 249-270. https://doi.org/10.1007/978-3-319-75307-2_15

[22] Visoottiviseth, V., Jaralrungroj, U., Phoomrungraungsuk, E. and Kultanon (2011) Distributed Honeypot Log Management and Visualization of Attacker Geographical Distribution. 2011 *Eighth International Joint Conference on Computer Science and Software Engineering* (*JCSSE*), Nakhon Pathom, 11-13 May 2011, 23-28. https://doi.org/10.1109/JCSSE.2011.5930083

[23] Wikipedia Contributors (2022) Data and Information Visualization—Wikipedia, the Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Data_and_information_visualization&oldid=1097071271

[24] Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nature Chemistry*, **5**, 182-186. https://doi.org/10.1038/nchem.1548

[25] Kirchner, K., Zec, J. and Delibašić, B. (2016) Facilitating Data Preprocessing by a Generic Framework: A Proposal for Clustering. *Artificial Intelligence Review*, **45**, 271-297. https://doi.org/10.1007/s10462-015-9446-6

[26] Chi, E.H. (2000) A Taxonomy of Visualization Techniques Using the Data State Reference Model. *IEEE Symposium on Information Visualization* 2000. *INFOVIS* 2000. *Proceedings*, Salt Lake City, 9-10 October 2000, 69-75. https://doi.org/10.1109/INFVIS.2000.885092

[27] Haber, R.B. and McNabb, D.A. (1990) Visualization Idioms: A Conceptual Model for Scientific Visualization Systems. In: *Visualization in Scientific Computing*, IEEE Computer Society Press, London, 74-93.

[28] Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J. and Melançon, G. (2008) Visual Analytics: Definition, Process, and Challenges. In: Kerren, A., Stasko, J.T., Fekete, J.-D. and North, C., Eds., *Information Visualization*, Springer, Berlin, 154-175. https://doi.org/10.1007/978-3-540-70956-5_7

[29] Ltifi, H., Ben Mohamed, E. and ben Ayed, M. (2016) Interactive Visual Knowledge Discovery from Data-Based Temporal Decision Support System. *Information Visualization*, **15**, 31-50. https://doi.org/10.1177/1473871614567794

[30] Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.C., Ellis, G. and Keim, D.A. (2014) Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, **20**, 1604-1613. https://doi.org/10.1109/TVCG.2014.2346481

[31] Alonso, Á., *et al.* (2010) On the Visualization of Honeypot Data through Projection Techniques. *Proceedings of the* 10*th International Conference on Computational and Mathematical Methods in Science and Engineering*, *CMMSE*2010 *Almeria*, Almeria, 27-30 June 2010, 1-12.

[32] O'donoghue, S.I., *et al.* (2018) Visualization of Biomedical Data. *Annual Review of Biomedical Data Science*, **1**, 275-304. https://doi.org/10.1146/annurev-biodatasci-080917-013424

[33] Cleveland, W.S. and McGill, R. (1984) Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, **79**, 531-554. https://doi.org/10.1080/01621459.1984.10478080

[34] Wikipedia Contributors (2021) Line Chart—Wikipedia, the Free Encyclopedia. https://en.wikipedia.org/wiki/Line_chart

[35] Wikipedia Contributors (2021) Pie Chart—Wikipedia, the Free Encyclopedia. https://en.wikipedia.org/wiki/Pie_chart

[36] Wikipedia Contributors (2021) Bar Chart—Wikipedia, the Free Encyclopedia. https://en.wikipedia.org/wiki/Bar_chart

[37] Weissgerber, T.L., Milic, N.M., Winham, S.J. and Garovic, V.D. (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. https://doi.org/10.1371/journal.pbio.1002128

[38] Wikipedia Contributors (2021) Scatter Plot—Wikipedia, the Free Encyclopedia. https://en.wikipedia.org/wiki/Scatter_plot

[39] Wikipedia Contributors (2021) Box Plot—Wikipedia, the Free Encyclopedia. https://en.wikipedia.org/wiki/Box_plot

[40] Wikipedia Contributors (20221) Histogram—Wikipedia, the Free Encyclopedia. https://en.wikipedia.org/wiki/Histogram

[41] Wikipedia Contributors (2021) Heat Map—Wikipedia, the Free Encyclopedia. https://en.wikipedia.org/wiki/Heat_map

[42] Bringer, M.L., Chelmecki, C.A. and Fujinoki, H. (2012) A Survey: Recent Advances and Future Trends in Honeypot Research. *International Journal of Computer Network and Information Security*, **4**, 63-75. https://doi.org/10.5815/ijcnis.2012.10.07

[43] McGrew, R. (2006) Experiences with Honeypot Systems: Development, Deployment, and Analysis. *Proceedings of the* 39*th Annual Hawaii International Conference on System Sciences* (*HICSS*'06), Kauai, 4-7 January 2006, 220a. https://doi.org/10.1109/HICSS.2006.172

[44] Fernandez, R. and Fetais, N. (2017) Survey of Information Visualization Techniques for Enhancing Visual Analysis. 2017 *International Conference on Computer and Applications* (*ICCA*), Doha, 6-7 September 2017, 360-363. https://doi.org/10.1109/COMAPP.2017.8079755

[45] Yang, L., Ma, Z., Zhu, L. and Liu, L. (2019) Research on the Visualization of Spatio-Temporal Data. *IOP Conference Series*: *Earth and Environmental Science*, **234**, Article ID: 012013. https://doi.org/10.1088/1755-1315/234/1/012013

[46] Shen, H., *et al.* (2019) Information Visualisation Methods and Techniques: State-of-the-Art and Future Directions. *Journal of Industrial Information Integration*, **16**, Article

ID: 100102. https://doi.org/10.1016/j.jii.2019.07.003

[47] Zhang, T., Wang, X., Li, Z., Guo, F., Ma, Y. and Chen, W. (2017) A Survey of Network Anomaly Visualization. *Science China. Information Sciences*, **60**, 122-138. https://doi.org/10.1007/s11432-016-0428-2

[48] Yang, J., *et al.* (2021) Visual Anomaly Detection for Images: A Survey.

[49] Soo-Yeon, J., Bong-Keun, J. and Jeong, D.H. (2021) Evaluating Visualization Approaches to Detect Abnormal Activities in Network Traffic Data. *International Journal of Information Security*, **20**, 331-345. https://doi.org/10.1007/s10207-020-00504-9

[50] Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C. (2001) Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, **13**, 1443-1471. https://doi.org/10.1162/089976601750264965

[51] Daassi, C., Nigay, L. and Fauvet, M.-C. (2004) Visualization Process of Temporal Data. in *Database and Expert Systems Applications*, Zaragoza, 30 August-3 September 2004, 914-924. https://doi.org/10.1007/978-3-540-30075-5_88

[52] Heer, J. and Agrawala, M. (2006) Software Design Patterns for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, **12**, 853-860. https://doi.org/10.1109/TVCG.2006.178

[53] Munzner, T. (2009) A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics*, **15**, 921-928. https://doi.org/10.1109/TVCG.2009.111

[54] Endert, A., Fiaux, and North, C. (2012) Semantic Interaction for Sensemaking: Inferring Analytical Reasoning for Model Steering. *IEEE Transactions on Visualization and Computer Graphics*, **18**, 2879-2888. https://doi.org/10.1109/TVCG.2012.260

[55] Falih Badran, M., Sahar, N.M., Sari, S. and Taujuddin, N.S.A.M. (2020) Intrusion-Detection System Based on Hybrid Models: Review Paper. *IOP Conference Series. Materials Science and Engineering*, **917**, Article ID: 012059. https://doi.org/10.1088/1757-899X/917/1/012059

[56] Howard, W.R. (2007) Pattern Recognition and Machine Learning20072Christopher M. Bishop. Pattern Recognition and Machine Learning. Heidelberg, Germany: Springer 2006. i-xx, 740 pp., ISBN: 0-387-31073-8 $74.95 Hardcover. *Kybernetes*, **36**, 275-275. https://doi.org/10.1108/03684920710743466

[57] Hinton, G.E. (1989) Connectionist Learning Procedures. *Artificial Intelligence*, **40**, 185-234. https://doi.org/10.1016/0004-3702(89)90049-0

[58] Ning, X., *et al.* (2020) Nonparametric Topic Modeling with Neural Inference. *Neurocomputing* (*Amsterdam*), **399**, 296-306. https://doi.org/10.1016/j.neucom.2019.12.128

[59] Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D. and Steger, C. (2021) The MVTec Anomaly Detection Dataset: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. *International Journal of Computer Vision*, **129**, 1038-1059. https://doi.org/10.1007/s11263-020-01400-4

[60] Carrera, D., Manganini, F., Boracchi, G. and Lanzarone, E. (2017) Defect Detection in SEM Images of Nanofibrous Materials. *The IEEE Transactions on Industrial Informatics*, **13**, 551-561. https://doi.org/10.1109/TII.2016.2641472

[61] Napoletano, P., Piccoli, F. and Schettini, R. (2018) Anomaly Detection in Nanofibrous Materials by CNN-Based Self-Similarity. *Sensors* (*Basel*), **18**, Article 209. https://doi.org/10.3390/s18010209

[62] Carrera, D., Boracchi, G., Foi, A. and Wohlberg, B. (2016) Scale-Invariant Anomaly Detection with Multiscale Group-Sparse Models. 2016 *IEEE International Conference on Image Processing* (*ICIP*), Phoenix, 25-28 September 2016, 3892-3896. https://doi.org/10.1109/ICIP.2016.7533089

[63] Xie, X.H. and Mirmehdi, M. (2007) TEXEMS: Texture Exemplars for Defect Detection on Random Textured Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**, 1454-1464. https://doi.org/10.1109/TPAMI.2007.1038

[64] Akoglu, L., Tong, H. and Koutra, D. (2015) Graph Based Anomaly Detection and Description: A Survey. *Data Mining and Knowledge Discovery*, **29**, 626-688. https://doi.org/10.1007/s10618-014-0365-y

[65] Wagner, C., Wagener, G., State, R., Dulaunoy, A. and Engel, T. (2010) PeekKernel-Flows: Peeking into IP Flows. *Proceedings of the Seventh International Symposium on Visualization for Cyber Security*, Ottawa, 14 September 2010, 52-57. https://doi.org/10.1145/1850795.1850801

[66] Wagner, C., Wagener, G., State, R. and Engel, T. (2011) Digging into IP Flow Records with a Visual Kernel Method. In: Herrero, Á. and Corchado, E., Eds., *Computational Intelligence in Security for Information Systems*, Springer, Berlin, 41-49. https://doi.org/10.1007/978-3-642-21323-6_6

[67] Page, M.J., *et al.* (2021) The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *Systematic Reviews*, **10**, Article No. 89. https://doi.org/10.1186/s13643-021-01626-4

[68] Wittmann, F.M. (2021) Sort Google Scholar by the Number of Citations V2.0b. https://github.com/WittmannF/sort-google-scholar

[69] SpringerLink, Springer Nature Switzerland AG (2021). http://link.springer.com

[70] Science Director, Elsevier B.V. (2021). https://www.sciencedirect.com

[71] IEEE Xplore, Institute of Electrical and Electronics Engineers (2021). https://ieeexplore.ieee.org

[72] ACM Digital Library, Association for Computing Machinery (2021). https://dl.acm.org

[73] Chuvakin, A. (2003) "Honeynets: High Value Security Data": Analysis of Real Attacks Launched at a Honeypot. *Network Security*, **2003**, 11-15. https://doi.org/10.1016/S1353-4858(03)00808-0

[74] Baykara, M. and Das, R. (2018) A Novel Honeypot Based Security Approach for Real-Time Intrusion Detection and Prevention Systems. *Journal of Information Security and Applications*, **41**, 103-116. https://doi.org/10.1016/j.jisa.2018.06.004

[75] Vasilomanolakis, E., Karuppayah, S., Kikiras and Mühlhäuser, M. (2015) A Honeypot-Driven Cyber Incident Monitor: Lessons Learned and Steps Ahead. *Proceedings of the* 8*th International Conference on Security of Information and Networks*, Sochi, 8-10 September 2015, 158-164. https://doi.org/10.1145/2799979.2799999

[76] Kobayashi, H., Zhang, Z., Ochiai, H. and Esaki, H. (2019) Probing Firewalls of Malware-Infected Networks with Honeypot. *Proceedings of the* 14*th International Conference on Future Internet Technologies*, Phuket, 7-9 August 2019, 1-4. https://doi.org/10.1145/3341188.3341190

[77] Belqruch, A. and Maach, A. (2019) SCADA Security Using SSH Honeypot. *Proceedings of the* 2*nd International Conference on Networking, Information Systems & Security*, Rabat, 27-29 March 2019, 1-5. https://doi.org/10.1145/3320326.3320328

[78] Song, J., Takakura, H., Okabe, Y., Eto, M., Inoue, D. and Nakao, K. (2011) Statistical Analysis of Honeypot Data and Building of Kyoto 2006+ Dataset for NIDS Evalua-

tion. *Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, Salzburg, 10 April 2011, 29-36. https://doi.org/10.1145/1978672.1978676

[79] Singh, A.N. and Joshi, R.C. (2011) A Honeypot System for Efficient Capture and Analysis of Network Attack Traffic. 2011 *International Conference on Signal Processing, Communication, Computing and Networking Technologies*, Kumaracoil, 21-22 July 2011, 514-519. https://doi.org/10.1109/ICSCCN.2011.6024606

[80] More, A. and Tapaswi, S. (2013) A Software Router Based Predictive Honeypot Roaming Scheme for Network Security and Attack Analysis. 2013 9*th International Conference on Innovations in Information Technology* (*IIT*), Abu Dhabi, 17-19 March 2013, 221-226. https://doi.org/10.1109/Innovations.2013.6544422

[81] Almotairi, S., Clark, A., Mohay, G. and Zimmermann, J. (2009) A Technique for Detecting New Attacks in Low-Interaction Honeypot Traffic. 2009 4*th International Conference on Internet Monitoring and Protection*, Venice/Mestre, 24-28 May 2009, 7-13. https://doi.org/10.1109/ICIMP.2009.9

[82] Koniaris, I., Papadimitriou, G. and Nicopolitidis (2013) Analysis and Visualization of SSH Attacks Using Honeypots. *IEEE Eurocon* 2013, Zagreb, 1-4 July 2013, 65-72. https://doi.org/10.1109/EUROCON.2013.6624967

[83] Almotairi, S., Clark, A., Mohay, G. and Zimmermann, J. (2008) Characterization of Attackers' Activities in Honeypot Traffic Using Principal Component Analysis. 2008 *IFIP International Conference on Network and Parallel Computing*, Paris, 3-5 November 2021, 147-154. https://doi.org/10.1109/NPC.2008.82

[84] Chen,T., Laih, C.S., Pouget, F. and Dacier, M. (2005) Comparative Survey of Local Honeypot Sensors to Assist Network Forensics. 1*st International Workshop on Systematic Approaches to Digital Forensic Engineering* (*SADFE*'05), Taipei, 7-9 November 2005, 120-132. https://doi.org/10.1109/SADFE.2005.6

[85] Kyriakou, A. and Sklavos, N. (2018) Container-Based Honeypot Deployment for the Analysis of Malicious Activity. 2018 *Global Information Infrastructure and Networking Symposium* (*GIIS*), Thessaloniki, 23-25 October 2018, 1-4. https://doi.org/10.1109/GIIS.2018.8635778

[86] Fraunholz, D., Zimmermann, M., Hafner, A. and Schotten, H.D. (2017) Data Mining in Long-Term Honeypot Data. 2017 *IEEE International Conference on Data Mining Workshops* (*ICDMW*), New Orleans, 18-21 November 2017, 649-656. https://doi.org/10.1109/ICDMW.2017.92

[87] Yeh, C.-H. and Yang, C.-H. (2008) Design and Implementation of Honeypot Systems Based on Open-Source Software. 2008 *IEEE International Conference on Intelligence and Security Informatics*, Taipei, 17-20 June 2008, 265-266.

[88] Fraunholz, D., Zimmermann, M., Anton, S.D., Schneider, J. and Dieter Schotten, H. (2017) Distributed and Highly-Scalable WAN Network Attack Sensing and Sophisticated Analysing Framework Based on Honeypot Technology. 2017 7*th International Conference on Cloud Computing, Data Science Engineering—Confluence*, Noida, 12-13 January 2017, 416-421. https://doi.org/10.1109/CONFLUENCE.2017.7943186

[89] Pomsathit, A. (2012) Effective of Unicast and Multicast IP Address Attack over Intrusion Detection System with Honeypot. 2012 *Spring Congress on Engineering and Technology*, Xi'an, 27-30 May 2012, 1-4. https://doi.org/10.1109/SCET.2012.6342030

[90] Lihet, M. and Dadarlat, V. (2018) Honeypot in the Cloud Five Years of Data Analysis. 2018 17*th RoEduNet Conference*: *Networking in Education and Research* (*RoEduNet*),

Cluj-Napoca, 6-8 September 2018, 1-6.
https://doi.org/10.1109/ROEDUNET.2018.8514128

[91] Koniaris, I., Papadimitriou, G., Nicopolitidis and Obaidat, M. (2014) Honeypots Deployment for the Analysis and Visualization of Malware Activity and Malicious Connections. 2014 *IEEE International Conference on Communications* (*ICC*), Sydney, 10-14 June 2014, 1819-1824. https://doi.org/10.1109/ICC.2014.6883587

[92] Wang, B., Dou, Y., Sang, Y., Zhang, Y. and Huang, J. (2020) IoTCMal: Towards A Hybrid IoT Honeypot for Capturing and Analyzing Malware. *ICC* 2020-2020 *IEEE International Conference on Communications* (*ICC*), Dublin, 7-11 June 2020, 1-7. https://doi.org/10.1109/ICC40277.2020.9149314

[93] Sethia, V. and Jeyasekar, A. (2019) Malware Capturing and Analysis Using Dionaea Honeypot. 2019 *International Carnahan Conference on Security Technology*, Chennai, 1-3 October 2019, 1-4. https://doi.org/10.1109/CCST.2019.8888409

[94] Shyla, S. and Bhatnagar, V. (2021) The Geo-Spatial Distribution of Targeted Attacks sources Using Honeypot Networks. 2021 11*th International Conference on Cloud Computing*, *Data Science Engineering* (*Confluence*), Noida, 28-29 January 2021, 600-604. https://doi.org/10.1109/Confluence51648.2021.9377145

[95] Dowling, S., Schukat, M. and Melvin, H. (2017) Using Analysis of Temporal Variances within a Honeypot Dataset to Better Predict Attack Type Probability. 2017 12*th International Conference for Internet Technology and Secured Transactions* (*ICITST*), Cambridge, 11-14 December 2017, 349-354.
https://doi.org/10.23919/ICITST.2017.8356416

[96] Lakh, Y. and Shymkiv, R. (2019) Using Honeypot Programs for Providing Defense of Banking Network Infrastructure. 2019 *IEEE International Scientific-Practical Conference Problems of Infocommunications*, *Science and Technology* (*PIC S T*), Kyiv, 8-11 October 2019, 527-532.
https://doi.org/10.1109/PICST47496.2019.9061550

[97] Moore, C. and Al-Nemrat, A. (2015) An Analysis of Honeypot Programs and the Attack Data Collected. In: Jahankhani, H., Carlile, A., Akhgar, B., Taal, A., Hessami, A.G. and Hosseinian-Far, A., Eds., *Global Security*, *Safety and Sustainability*: *Tomorrow's Challenges of Cyber Security*, Springer, Berlin, 228-238.
https://doi.org/10.1007/978-3-319-23276-8_20

[98] Sánchez, R., Herrero, Á. and Corchado, E. (2015) Clustering and Neural Visualization for Flow-Based Intrusion Detection. *Computational Intelligence in Security for Information Systems Conference*, Burgos, 15-17 June 2015, 333-345.
https://doi.org/10.1007/978-3-319-19713-5_29

[99] Cao, J., Li, W., Li, J. and Li, B. (2018) DiPot: A Distributed Industrial Honeypot System. In: Qiu, M.K., Ed., *Smart Computing and Communication*, Springer, Berlin, 300-309. https://doi.org/10.1007/978-3-319-73830-7_30

[100] Abbas, C.J.B., Villalba, L.J.G. and López, V.L. (2007) Implementation and Attacks Analysis of a Honeypot. In: Gervasi, O. and Gavrilova, M.L., Eds., *Computational Science and Its Applications*, Springer, Berlin, 489-502.
https://doi.org/10.1007/978-3-540-74477-1_46

[101] Nicomette, V., Kaâniche, M., Alata, E. and Herrb, M. (2011) Set-Up and Deployment of a High-Interaction Honeypot: Experiment and Lessons Learned. *Journal in Computer Virology*, **7**, 143-157. https://doi.org/10.1007/s11416-010-0144-2

[102] Agrawal, N. and Tapaswi, S. (2017) The Performance Analysis of Honeypot Based Intrusion Detection System for Wireless Network. *International Journal of Wireless Information Networks*, **24**, 14-26. https://doi.org/10.1007/s10776-016-0330-3

[103] Zurutuza, U., Ezpeleta, E., Herrero, Á. and Corchado, E. (2011) Visualization of Misuse-Based Intrusion Detection: Application to Honeynet Data. In: Corchado, E., *et al.*, Eds., *Soft Computing Models in Industrial and Environmental Applications*, 6*th International Conference SOCO* 2011, Springer, Berlin, 561-570. https://doi.org/10.1007/978-3-642-19644-7_59

[104] Dumas, M., Robert, J.-M. and Mcguffin, M.J. (2012) Alertwheel: Radial Bipartite Graph Visualization Applied to Intrusion Detection System Alerts. *IEEE Network*, **26**, 12-18. https://doi.org/10.1109/MNET.2012.6375888

[105] Leaden, G., Zimmermann, M., DeCusatis, C. and Labouseur, A.G. (2017) An API Honeypot for DDoS and XSS Analysis. 2017 *IEEE MIT Undergraduate Research Technology Conference* (*URTC*), Cambridge, 3-5 November 2017, 1-4. https://doi.org/10.1109/URTC.2017.8284180

[106] Aupetit, M., Zhauniarovich, Y., Vasiliadis, G., Dacier, M. and Boshmaf, Y. (2016) Visualization of Actionable Knowledge to Mitigate DRDoS Attacks. 2016 *IEEE Symposium on Visualization for Cyber Security* (*VizSec*), Baltimore, 24 October 2016, 1-8. https://doi.org/10.1109/VIZSEC.2016.7739577