A review of "A survey of Visualization Systems for Network Security"

Sai Rithwik M
IMT2018061
International Institute of Information Technology, Bangalore

***Abstract*—This report is a summary and a review of the survey of Visualization Systems for Network Security [1] authored by H. Shiravi, A. Shiravi and Ali A. Ghorbani published in 2011. The survey discusses various visualisation techniques under the 5 most commonly used cases in security reconnaissance. This report discusses various advantages and disadvantages of visualisation techniques in each of the 5 use-case scenarios.**

***Index Terms*—Security Visualisation, Glyphs, Node-Link diagrams, Matrices, Marks, Channels, Honeypots, Botnets, Malware, Intrusion, Recon, Treemaps, Ports, Networks**

I. Introduction

In this pandemic stricken world, all the services have moved to a virtual domain. The internet is a battlefield and computers are at continual attacks. With such a large amount of data, their protection becomes one of the key priorities. With a large number of entry points and the size of the data, it is not feasible to make sure that all the threats are assessed and entrypoints are secured. An effective security strategy uses a range of approaches to minimize vulnerabilities and target many types of cyber threats. Detection, prevention and response to security threats involve the use of security policies, software tools and IT services. This is where security visualisations play a key role in detecting and analysing bad actors or threats that have entered the system. The core aim of security visualization is to increase cognition, the processes in which information is obtained, transformed, stored, retrieved, and used.

The report starts by looking at the literature research done in this field, about the survey and then discusses various main categories needed for security visualisations that describe what and how they could be visualised. Further, the report deep dives into each category and discuss some of the most popularly used visualisations and describe the pros and cons of using them. On discussing these visualisations, a comment on the issues, concerns and guidelines charted by the authors is made and this report compares how has the security visualisations developed over the past decade by discussing the future scope.

II. Related Work

This survey details all the visualisations which have been discussed from as late as 1980s to as recent as 2011. Most of the early research has looked into the usage of Node-Link diagrams [3]–[5], [21]–[23] and has looked into modifying the channels in them to suit a specific use-case. As time progressed and the threat map started increasing more complex visualisations were required to make more inferences. VISUAL [11] and TNV [13] were one of the frontrunners in this field where they completely switched from a node-link approach to a host-based approach by using matrices and parallel plots to depict the data to the users. Some of them however lacked the scalability aspect by not being able to depict the large networks in small screen space. To solve these issues many dimensionality reduction techniques [17] were used. While some researchers got creative by automatically filtering the data and presenting only the events which trigger an alarm [19]. Most of the latest research has been shifting to contemporary and simple visualisations [21] and adding channels to them to aid novice users to understand the visualisations in an easier fashion. The survey is one of the first papers to build upon various visualisation techniques so that organisations can be well equipped to create their own tools.

III. Data and Information Classification

G. Conti in his book Security Data Visualisation [2] has discussed various use cases of security visualisations. A few of them are discussed below:

- Detection of anomalous activity
- Discovering patterns to correlate events
- Visualising Intrusion prevention, detection firewalls
- Visualising worm propagation or botnet activity
- Visualising for forensic analysis

One of the goals of conducting this survey was to develop an all-around security visualisation tool that could help the various organisations to monitor and assess attacks. The list described above has been incorporated to classify various types of security visualisations. In the survey, Shiravi et al. have classified the visualisations based on well-known Cybersecurity Management Systems principles of first implementing a monitoring tool for defence analysis and then post-attack analysis visualisations. These are the following categories the authors have divided the survey analysis into:

- **Host Server Monitoring**: This class of visualisation deals with the representation of the network ie. hosts and servers, thereby depicting the current state of the network with stats like load, unusual activities etc. This also enables the user to identify malicious servers that the host is communicating to.

- **Internal and External Monitoring**: This is almost similar to the previous class but this solely discusses the interactions between internal hosts and external IPs.
- **Port Activity**: One of the major entrypoints to various attacks are insecure/open ports which give applications access to enter into the network. This essentially is one of the major entrypoints for various trojans and viruses to enter into the network. Hence visualisations are required to depict the port activity.
- **Attack Patterns**: An attack usually is a multistep process. It starts with recon, gaining and maintaining access and installing backdoors for access anytime in the future. Hence visualisations are required to monitor this multistep process and identify that a system is being attacked. One of the major issues with these visualisations are the number of false positives that can occur and hence weeding them out would be one of the top priorities to cover in a visualisation like this.
- **Routing Behaviours**: Understanding how the data flows through various devices in the network is one of the key factors in the identification of the problem for service unavailability in many organisations. With recent DoS attacks on various organisations and devices, identifying the clogging point for data drop is of utmost importance.

From the above list, it can be inferred that the first three classes belong to attacks and the remaining classes belong to post-attack recon and behavioural analysis.

A. Host Server Monitoring

There are two major categories in which one could divide this class of visualisation. They are

- Host-Based Approaches
- Link Based Approaches

One of the simplest visualisations that come under this category are node-link diagrams, where the hosts and servers essentially act as nodes and the respective connections to servers can be shown through edges. Erbacher et al. [3] were the first ones to work on developing a visualisation design for the same. EtherApe [4] was an implementation of the following design discussed by Erbacher et al. Figure 1 shows practically how EtherApe is able to portray the visualisations in the network through a node-link diagram in a local IIITB network for 5 minutes. This visualisation technique essentially uses circles, lines as masks and colours, thickness as channels to visualise the data. Colours essentially depict the protocols and width depict the number of requests being made to a particular server. However, EtherApe is feasible only in a small scale scenario and does not scale in large operations as there will be millions of requests and identifying that one malicious request through the visualisation in the clutter of nodes would be very hard from a visual perspective.

One of the upgrades to the Erbacher et al. research is that of Mansman et al. [5] where the authors used force-directed graph layout. A new channel of position is added to this visualisation to detect unconventional position changes

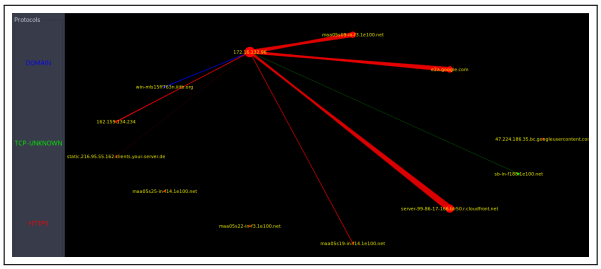

Fig. 1: EtherApe on IIITB network with host node at center and server connections spread around it

of nodes. Such changes in positions activate a flagging system to alert the user.

NVisionIP [6] takes a completely different approach compared to the one proposed by Erbacher et al. NVisionIP is a 3 layered deep visualisation. As shown in Figure 2 NVisionIP maps the entire network in a 256x256 matrix with all possible subnets on the X-axis and possible hosts on the Y-axis. In the first Galaxy view, the visualisation shows a colour which is a representation of the kind of requests a host is making to a server. In a second Sub-Multiple View, a specific subnet can be scanned and in the lowest view a particular host can be looked at and requests can be analysed from a singular host's perspective. This is a significant improvement compared to EtherApe's visualisation as the user has a larger space to analyse. The marks in this visualisation are points and channels are colours and their hues. The vast size that this visualisation covers is one of the key advantages. One place where this visualisation falls short is cognition and the ability for the users to identify potential threats due to the small area to identify a colour in the Galaxy view.

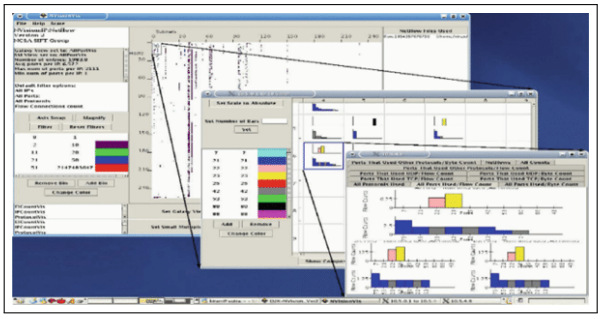

Fig. 2: Three levels in NVisionIP visualisation depicted *Image Courtesy* [6]

Some other visualisations [7] [8] include stacking the hosts on the left and connections to the right. In these visualisations channels essentially remain the same, but there is a small change in the maps they use like using splines instead of straight lines to make the visualisations aesthetically pleasing. There are some outdated techniques like Radial Network Visualisation [9] which involve pie-chart analysis in 4 concentric circles, where each circle represents the source, destination and their respective port numbers.

Many visualisations have taken some sort of inspiration from the above techniques, but one of the key lacking fea-

tures in this class of visualisations is the tradeoff between cognition and the coverage of the data. Some glyph based implementations like Tudumi [10] are being developed to look at this data from a 3D visual PoV, but this just adds additional dimensionality without aiding the cognition of the user. Since there are a vast number of hosts and servers including them all in the visualisation would impact how quickly the user is able to identify a point of interest. Trigger functions like the ones implemented in Mansman et al. [5]'s work could be used as a reference to make use of position channels to identify the vulnerable host.

B. Internal and External Monitoring

One of the most cited visualisations for this class includes VISUAL [11]. As visible in Figure 3 there is a grid where each cell represents a unique internal host. An external server is depicted as a square box and the size of each square box is dependant on the activity incoming to the external server. The masks in this visualisation are lines and squares and channels are the size and position of the squares. Filtering the data could lead to valuable insights into where most of the requests are going. Space is a huge constraint in this visualisation as unfiltered visualisation becomes cluttered.

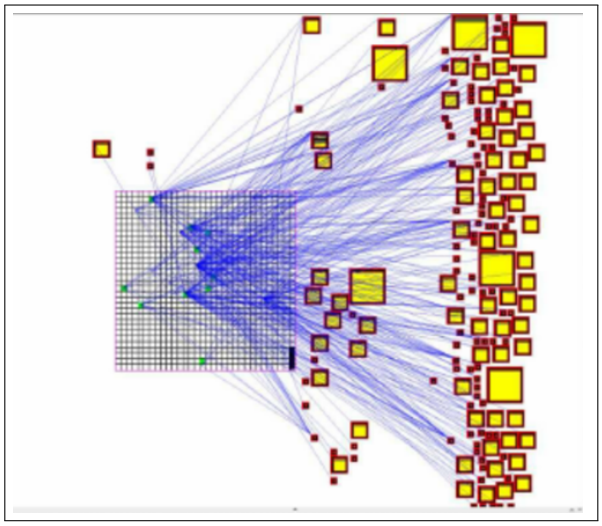

Fig. 3: VISUAL grid with hosts on the matrix and yellow boxes being the connected servers *Image Courtesy* [11]

VisFlowConnect [12] is a parallel axis based visualisation to show the user a real-time analysis of the network where ports and IP addresses are placed on the Y-Axis and the outer parallel axes depict external networks and the inner vertical axis represents an inner network. This on using techniques like shadowing on the links connecting the axes to depict time could add some context to the application. Brushing could essentially help in understanding how the requests are going from one system to another. The marks in this are lines and points and the channels are colour, hue and position. One of the biggest advantages of this is that it is light from a memory aspect and can represent a large set of information in a memory-efficient manner.

Goodall et al. have come up with a Time-Based Network Visualiser(TNV) [13] which helps to understand the high-level abstract of the data and also make sure that the nitty-gritty details are not left out. As shown in Figure 4 TNV is a bifocal visualisation that provides both contexts as well as the focal region for visualisation in the centre. The X-axis represents time and Y-axis depicts all available host IP addresses. The cells are colour-coded by making grey cells for external addresses and blue colours for internal addresses. Since this is a development on the above two visualisations, practices from the same are implemented here too.

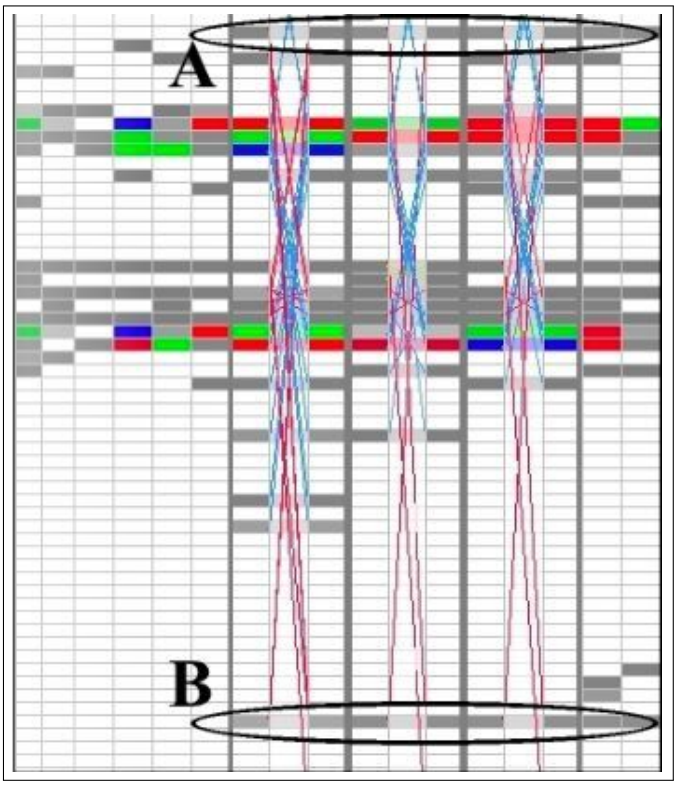

Fig. 4: TNV Visualising with color encoding and links. The above picture depicts the scenario where client B is attacking host A *Image Courtesy* [13]

We have observed that most visualisations in this category have tried to reduce the issues faced in the first category by making sure that the count of a number of hosts doesn't get compromised while visualising data. Some of the pros visible in these visualisations are how they are independent of how much time they need to be run to be memory efficient. This is a huge advantage from a security monitoring aspect as you get data from a larger sample space.

C. Port Activity

Cube root scaled histogram is a simple stacked histogram where X-Axis represents time and Y-Axis represents the requests made. Histograms can be used here as the number of majorly used ports are much lesser compared to the IP addresses. Each port is identified by a colour and whenever a query/request reaches the port it increases the stack length at a particular period of time. One of the major disadvantages of

using this is that it is hard to make inferences from a specific port PoV. This however would help in understanding when is there a possible spurt/spike of requests to multiple ports and which port is most active in this scenario. Existence Plots were proposed by Janies [14] which essentially added a new feature where each port had an activity meter and they could be analysed individually over a period of time if the user wanted to. But the cons of visualising multiple port actions at once still remain even with this added feature.

From the above visualisations, there is a clear requirement of some sort of visualisation which could be animated involving a time aspect to aid the user. McPherson et al. developed PortVis [15] as shown in Figure 5 which is a 256x256 grid of ports, where each port number is determined by the position on the matrix according to binary numbering. Colour encoding is used to depict the flow of data through a particular port. Masks in this are matrices and channels are position and colour. The main advantage of using this visualisation becomes its key disadvantage. If there is a port that has open access to external clients and it is located near other ports with high activity it essentially becomes hard to identify for the user to classify the specific port for the same.

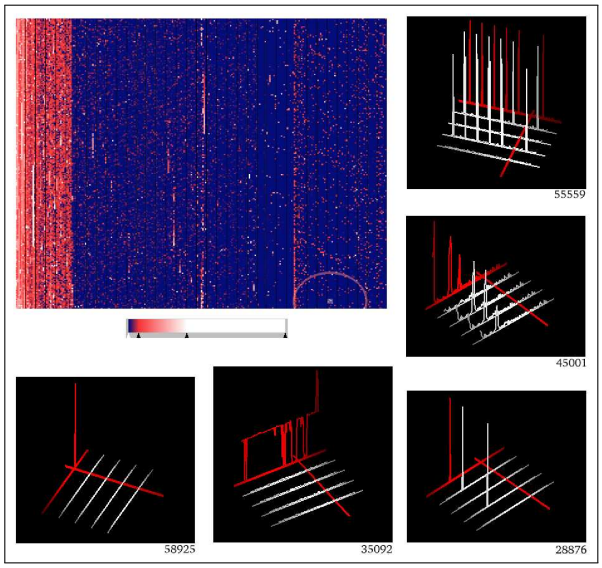

Fig. 5: PortVis represents number of sessions on the port using colors; dark blue ports have a low number of sessions, and white ports have a high number of sessions. On selecting a specific port visualisations related to that port are visible *Image Courtesy* [15]

Spinning Cube of Potential Doom [16] is an update to PortVis where an animated display of network data is depicted where the data does not follow the site's usage policies, hence any data that is visible in the visualisation is practically malicious data. X and Z axis depict Destination and Source IP addresses and Y-axis represents a port number. Hence through this visualisation if one sees a transmission happening through the X and Z plane it essentially indicates that a particular port is vulnerable and needs to be immediately checked. Figure 6 shows a series of visualisations as seen in Spinning Cube of Potential Doom. The marks used in this are points and channels include position and colour.

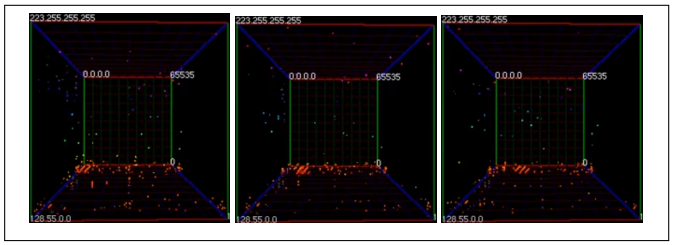

Fig. 6: Cube of Doom representation depicting data flow from source to destination on a specific port from bottom left corner to bottom right corner. *Image Courtesy* [13]

This class of visualisations keep up with the task that needs to be visualised and memory saving applications could use PortVis and devices which could afford some usage of memory could use Spinning Cube of Potential Doom. Some of the ways these visualisations could be further improved are by removing most commonly used ports like 22, 80 and 443 from the visualisation by filtering them. This could help in a better understanding of the visualisation as the user is able to infer the flow of data through some uncommonly active ports.

D. Attack Patterns

Since there is a large amount of data to cater to and identifying the false positives become very important. One of the earliest researches in this field was by Girardin [17] in which they designed a visualisation system that incorporated dimensionality reducing techniques like Self-Organising Maps(SOM). As depicted in Figure 7 this SOM uses various channels like colours, size and position to show how different an attribute on the map is from the others. This clearly helps to weed out dissimilar events as similar entities look like other similar entities. One of the major disadvantages of using this is that without supervised training this system is far from actually predicting the attack patterns. Some of the latest additions to classifying the data according to similar or dissimilar activity are by using many clustering and dimensionality reduction algorithms. One of the latest developments in this field is using t-Distributed Stochastic Neighbor Embedding (t-SNE) [18]. This essentially is a scatter plot and reduces the dimensions and clusters all the similar events together with a similar colour. The colour contrast used helps to clearly identify an anomaly.

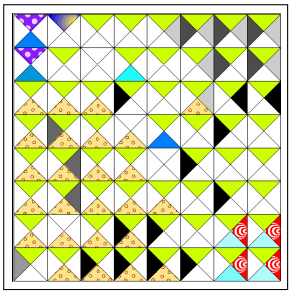

Fig. 7: SOM dimensionality reduction to depict patterns *Image Courtesy* [17]

As discussed earlier, identifying false positives is of utmost importance and SnortView [19] helps in identifying the same. From Figure 8 we can see that, SnortView has 3 major panels address, alert and destination. The X-Axis depicts time and Y-Axis depicts the addresses. An intrusion into the system is depicted by using colour channels. The major disadvantage that this brings is from the scalability aspect as not many IP addresses can be mapped to the Y-Axis. Some approaches improved upon this to use the density of logs on the Y-Axis instead of IP addresses and trying to identify intrusions from the same.

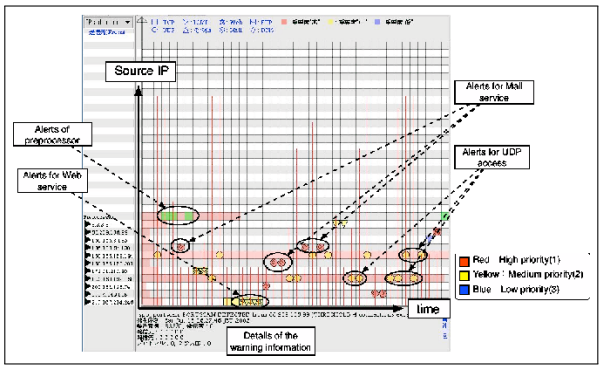

Fig. 8: SnortView matrix visualisation to depict anomalies in the network *Image Courtesy* [19]

In large networks, there is a huge probability of tons of alerts being generated and as discussed earlier some of them could also be false negatives. In 2005 Abdullah et al. [20] came up with a novel approach to address the issue of multiple IP addresses present. They used parallel axes to fit in the large range of IP addresses. This is similar to the TNV [13] visualisation depicted earlier, but with better usage of colours and position to depict the severity of data as it comes from an external IP to an internal IP. One of the major drawbacks is that this is suited only for Class B IP addresses which are currently outdated. Other techniques like binary rainfall which includes data from every single packet transfer can also be used to visualise intrusion detection. But one of the major drawbacks is the amount of time required to continuously monitor the rainfall on the screen in real-time which is not ideal from a threat analysis perspective. The marks used in these visualisations are points and channels are colours, position and lines.

To identify the true negatives with a high probability Livnat et al. came up with VizAlert [21] which correlates heterogeneous events. As depicted in Figure 9 VizAlert is a circular ring with logs surrounded around the ring. On identifying a trigger, a line goes from the outer ring to the map in the interior. Line widths are used to reinforce the number of alerts in the visualisation and the radius of nodes within the interior of the ring represents a number of unique alerts being received. This is a really good visualisation to make multiple inferences. Some of the major inferences could be how many attacks a particular host was receiving. If the radius of the node is larger, it would mean it has some serious vulnerability required to be patched. One of the major disadvantages of this visualisation is that it cannot be scaled for a very large network.

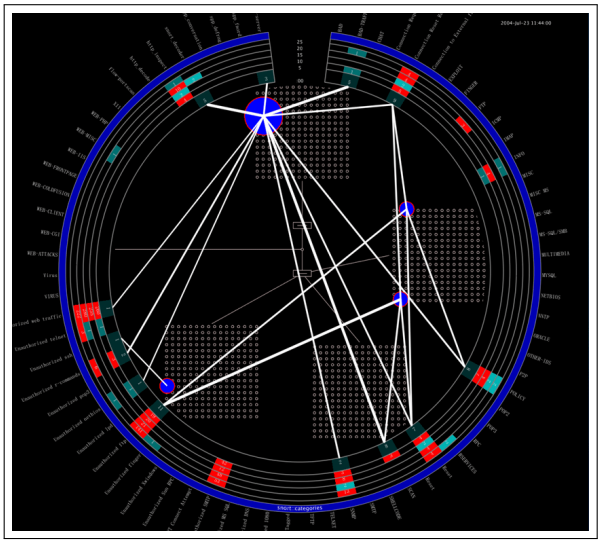

Fig. 9: VisAlert shows the devices under attack from external services. The blue node is a representation of host under attack from the red nodes around the circle *Image Courtesy* [21]

There are a few more techniques that were similar to the above ones. Treemaps are extensively used to identify botnet spreads in Operating Systems. Some of them had used 3D visualisation spaces to layer intrusions based on the importance and attack spread. Some visualisations resorted to simple techniques like circular node-link diagrams by placing hosts and alert categories and used techniques like edge bundling for improved aesthetics and better visualisation. While some of the visualisations developed on their reinforcement and cognition by using different channels like Kernel Density Estimation over Grid Thematic Mapping and usage of splines over lines.

Most of the visualisations in this class have been vastly researched over the years and hence so much insight is provided through such minimalistic visualisations. The visualisations in this class can be merged with visualisations from other classes for suppose port detection to identify how the bad actors are able to gain access.

E. Routing Behaviour

In this set of visualisations, the main aim is to identify and correct router misconfigurations due to frequent updates in Border Gateway Protocol(BGP). Describing a network of devices is usually more feasible through a node-link diagram and in this class of visualisations the same has been used. Important routing changes are visualised in an animated fashion in LinkRank Visualisation [22] and the technique proposed by Wong et al. [23]. LinkRank visualisation essentially depicts the network of routers and depicts the path with variable width. Thicker the width indicates more the number of hijacks. Depending on this a rank is also allotted to a node indicating the priority to traverse through that particular point.

IV. Conclusion, Remarks and Future Scope

The main aim of any visualisation is to give the user a high-level understanding of what is being analysed so that the user can make informed decisions. One of the major observations in the analysis was the fact that while the user had a high-level understanding of what was happening, most of the visualisations failed to cater to the part where nitty-gritty events were left out. These small events might be a major cause a system could be attacked. Thereby the visualisations must be smart enough to weed out the unnecessary data and automatically identify and evaluate certain events and present them to the user. To include these features like Zooming, Distortions or Filtering on queries could be used. These techniques also ensure that the data used can be scalable enough.

From a visualisation perspective, more specifically a dimensional analysis perspective, most of the 3D systems were very hard for the user to understand compared to a 2D visualisation. One of the key reasons owing to this is the fact that users tend to lose focus on what they were looking for by zooming and moving around the 3D visualisations. A lot of research is being done in this field and there have also been guidelines for developers to propagate the right ways to incorporate 3D visualisations as mentioned by Shneiderman [24]. Since the data is related to security, it might contain a lot of sensitive details about organisations or individuals. Solutions must be developed that maintains the privacy of the users and also conveys the necessary information through visualisations.

This survey has been very influential in developing various visualisation tools which are widely used in the industry. Big Data analysis is one of the most widely spoken topics in the current days and the insights provided by this survey has been a benchmark for other security analysis and has helped to improve anomaly detection using Big Data Processing techniques [25]. Tools like BotViz [26] were developed which is used widely to visualise botnets. This also helped researchers look into multivariate time series data and analyse them [27]. Of late researchers have been looking into similarity matrices to classify malware and intrusions [28] into the system. Many more applications like honeypot analysis and predictive visualisations are being looked at given that computers are getting stronger in the current timeframe.

References

[1] Shiravi, Hadi & Shiravi, Ali & Ghorbani, Ali. (2011). "A Survey of Visualization Systems for Network Security" *Visualization and Computer Graphics, IEEE Transactions on. 18. 1 - 1. 10.1109/TVCG.2011.144.*

[2] G. Conti (2007), "Security Data Visualization." *No Starch Press.*

[3] Erbacher, Robert & Walker, Kenneth & Frincke, Deborah. (2002). Intrusion and Misuse Detection in Large-Scale Systems. *Computer Graphics and Applications, IEEE. 22. 38-47. 10.1109/38.974517.*

[4] EtherApe: A graphical network monitor

[5] F. Mansman, L. Meier, and D.A. Keim, "Visualization of Host Behavior for Network Security," *Proc. Workshop Visualization for Computer Security (VizSEC '07), pp. 187-202, 2008.*

[6] K. Lakkaraju, W. Yurcik, and A. Lee, "NVisionIP: Netflow Visualizations of System State for Security Situational Awareness," *Proc. ACM Workshop Visualization and Data Mining for Computer Security, vol. 29, pp. 65-72, 2004.*

[7] G. Fink, P. Muessig, and C. North, "Visual Correlation of Host Processes and Network Traffic," *Proc. IEEE Workshop Visualization for Computer Security (VizSEC 05), pp. 11-19, 2005.*

[8] G. Fink, V. Duggirala, R. Correa, and C. North, "Bridging the Host-Network Divide: Survey, Taxonomy, and Solution," *Proc. 20th USENIX Conf. Large Installation System Administration, pp. 247-262, 2006.*

[9] D. Keim, F. Mansmann, J. Schneidewind, and T. Schreck, "Monitoring Network Traffic with Radial Traffic Analyzer," *Proc. IEEE Symp. Visual Analytics Science and Technology, pp. 123-128, 2006.*

[10] Takada, Tetsuji & Koike, Hideki. (2002). Tudumi: information visualization system for monitoring and auditing computer logs. *570- 576. 10.1109/IV.2002.1028831.*

[11] R. Ball, G.A. Fink, and C. North, "Home-Centric Visualization of Network Traffic for Security Administration" *Proc. ACM Workshop Visualization and Data Mining for Computer Security, pp. 55-64, 2004.*

[12] X. Yin, W. Yurcik, M. Treaster, Y. Li, and K. Lakkaraju, "VisFlow-Connect: Netflow Visualizations of Link Relationships for Security Situational Awareness," *Proc. ACM Workshop Visualization and Data Mining for Computer Security, pp. 26-34, 2004.*

[13] J. Goodall, W. Lutters, P. Rheingans, and A. Komlodi, "Preserving the Big Picture: Visual Network Traffic Analysis with TNV," *Proc. IEEE Workshop Visualization for Computer Security (VizSEC '05), pp. 47-54, 2005.*

[14] J. Janies, "Existence Plots: A Low-Resolution Time Series for Port Behavior Analysis," *Proc. Fifth Int'l Workshop Visualization for Computer Security (VizSec '08), pp. 161-168, 2008.*

[15] J. McPherson, K. Ma, P. Krystosk, T. Bartoletti, and M.Christensen, "PortVis: A Tool for Port-Based Detection of Security Events," Proc. the ACM Workshop Visualization and Data Mining for Computer Security, pp. 73-81, 2004.

[16] S. Lau, "The Spinning Cube of Potential Doom," *Comm. the ACM, vol. 47, no. 6, pp. 25-26, 2004.*

[17] L. Girardin, "An Eye on Network Intruder-Administrator Shootouts," Proc. *First Conf. Workshop Intrusion Detection and Network Monitoring, vol. 1, pp. 3-13, 1999.*

[18] L.J.P. van der Maaten. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research 15(Oct):3221-3245, 2014.*

[19] H. Koike and K. Ohno, "SnortView: Visualization System of Snort Logs," *Proc. ACM Workshop Visualization and Data Mining for Computer Security, vol. 29, pp. 143-147, 2004.*

[20] K. Abdullah, C. Lee, G. Conti, J. Copeland, and J. Stasko, "Ids Rainstorm: Visualizing ids Alarms," *Proc. IEEE Workshop Visualization for Computer Security (VizSEC '05), pp. 1-10, 2005.*

[21] Livnat, Yarden & Agutter, James & Moon, S. & Erbacher, Robert & Foresti, S.. (2005). A visualization paradigm for network intrusion detection. *2005. 92 - 99. 10.1109/IAW.2005.1495939.*

[22] M. Lad, D. Massey, and L. Zhang, "Visualizing Internet Routing Changes," *IEEE Trans. Visualization and Computer Graphics, vol. 12, no. 6, pp. 1450-1460, Nov./Dec. 2006.*

[23] T. Wong, V. Jacobson, and C. Alaettinoglu, "Internet Routing Anomaly Detection and Visualization," *Proc. Int'l Conf. Dependable Systems and Networks (DSN '05), pp. 172-181, 2005.*

[24] B. Shneiderman, "Why Not Make Interfaces Better than 3D Reality?," *IEEE Computer Graphics and Applications, vol. 23, no. 6, pp. 12-15, Nov./Dec. 2003.*

[25] Ariyaluran Habeeb, Riyaz Ahamed & Nasaruddin, Fariza & Gani, Abdullah & Hashem, Ibrahim & Ahmed, Ejaz & Imran, Muhammad. (2018). *Real-time big data processing for anomaly detection: A Survey. International Journal of Information Management. 45.10.1016/j.ijinfomgt.2018.08.006.*

[26] I. Sharafaldin, A. Gharib, A. H. Lashkari and A. A. Ghorbani, "BotViz: A memory forensic-based botnet detection and visualization approach," 2017 *International Carnahan Conference on Security Technology (ICCST), 2017, pp. 1-8, doi: 10.1109/CCST.2017.8167804.*

[27] A. Thomson, M. Graham and J. Kennedy, "Pianola - Visualization of Multivariate Time-Series Security Event Data," 2013 *17th International Conference on Information Visualisation, 2013, pp. 123-131, doi: 10.1109/IV.2013.15.*

[28] S. Venkatraman and M. Alazab, "Classification of Malware Using Visualisation of Similarity Matrices," 2017 *Cybersecurity and Cyberforensics Conference (CCC), 2017, pp. 3-8, doi: 10.1109/CCC.2017.11.*