Visual Analytics Applied To Cyber Security

Vlastimil Maly[a]

[a] Department of Communication and Information Systems, University of Defence, Brno, Czech Republic,
e-mail: vlastimil.maly@unob.cz

Abstract

The field of *Visual Analytics* (VA) seeks to combine computational power, visual display, and human capabilities in order to extract knowledge from large, heterogeneous data sets. VA is in part a response to the scale and complexity of information that is now potentially available. It seeks to aid people in making sense of the world by helping them identify and exploit relevant information. Within the field of VA we are trying to exploit and enable the considerable capabilities of the human visual perceptual process. Visual analytics techniques can be extremely helpful to addressing the potential and actual consequences of *cyber attacks*. Visual analytics, when applied to anomalous network behaviours, could provide a basis for a real-time indications and warnings system usable for supposed cyber incidents.

Visual Analytics arose in the post nine-eleven eras in order to *provide tools to the analyst* to aid in the identification of threats. It was formulated primarily by experts from the field of information visualization and emphasized visual tools for information understanding and extracting patterns.

Key words: visual analytics, cyber threats, network traffic, human visual perceptual process, visual tools and technologies, patterns, decision-making process, big data, cyber-crime, graphics.

1. Introduction

Based on experiments, it was found that for the perception of the surrounding world people use from 80% their eyes. Picture information is in human brain very quickly interpreted.

Visual perception works in accordance with the rules that lead to a clear, accurate and effective representation of information. Using of this fact - various visualization methods are appropriate for data processing purposes. In the case of examination data having several parameters, it is necessary to use projections or transformation into the display geometric plane (2D).

The increase of the measured and processed data is huge in recent years. This is also important driving force for emerging of new ways of visualization and other methods for data pre-processing. *Visualization is any technique for creating images, diagrams, or animations to express information or idea*.

Big data can be a bad thing as we can't see the knowledge hidden inside. That is particularly important when we are trying to see possible relationships and linkages. Cyber security is one of the area where data visualization technology would help uncover those linkages in a way that everyone can understand, even those without a statistics degree or experience.

We need to see things as they happen – and be able to adapt to *predict future attacks*. What would help is a technology that lets us see all of the right information in a format that helps make sense of it all. Advanced analytics and data visualization make that possible.

2. Visualization – explanation and history

Visual Analytics can be seen as an integral approach combining visualization, human factors, and data analysis.[9] VA is the science of analytical reasoning supported by interactive visual interfaces. Today, data is produced at an incredible rate and the ability to collect and store the data is increasing at a faster rate than the ability to analyze it.
Over the last decades, a large number of automatic data analysis methods have been developed. VA methods allow decision makers to combine their human flexibility, creativity, and background knowledge with the enormous storage and processing capacities of today's computers to gain insight into complex problems.
Using advanced visual interfaces, humans may directly interact with the data analysis capabilities of today's computer, allowing them to make well-informed decisions in complex situations.

A few links of Visualization history:

- In the second century AD people widely start to use tables – a simple arrangement of data in rows and columns.
- Visual encoding of data appeared in the 17th century and was created for displaying two-dimensional (2D) data graph.
- Later then in the 18th and 19th century graphs were improved and most of today's graphs forms were introduced (bar chart, dependency on time, pie chart, …).
- In 1967 the new concept and connected term of "*visual language*" was developed.
- First computer graphical interface was introduced in 1984 for Apple Macintosh computers.
- Extensive research started then in the fields of computer graphics, physics, three-dimensional applications (3D), human anatomy, chemical reactions or meteorological phenomena.

Data visualization is the presentation of data in a pictorial or graphical format. [10] For centuries, people have depended on visual representations such as charts and maps to understand information more easily and quickly. As more and more data is collected and analyzed, decision makers at all levels welcome data visualization software that enables them to see analytical results presented visually, find relevance among the millions of variables or occurrences, communicate concepts and hypotheses to others, and even predict the future.

Using Visual Analytics methods and tools we can explore all relevant data quickly and easily. User can look at more options, uncover hidden opportunities, identify key relationships and make more precise decisions. Visualizations help people see things that were not obvious to them before. Even when data volumes are very large, patterns can be spotted quickly and easily. Data visualization presents the data in a way that the decision maker can easily interpret, saving time and energy. Many mathematical techniques exist for representing pattern and structure, as well as visualizing correlations, time patterns, metadata relationships, and networks of linked information.

3. Cyber Security

There are used several definitions of cyber defence nowadays. Term "cyber" is now used to refer to those parts of IT infrastructure and the threat environment that deal with countering attacks, and "cyberspace" refers to the global network of computers, networks, and people who use them. The briefest definition can be as following:

"Cyber defence is that category of products, methodologies and strategies used to counter targeted attacks."

Security Goals for Cyber are following:

- Identification of attacks or attackers before they are able to hit.
- Find previously unknown attacks.
- Find attacks that are in progress.
- Understanding the impact of any successful attack.

The ENISA Threat Landscape report analyses the "cyber enemy"; identifying and also listing the top ten (out of a total of sixteen) threats in emerging technology areas. The areas considered are Mobile Computing, Social Media/Technology, Critical Infrastructure, Trust Infrastructures, Cloud, and Big Data. Among the identified top ten threats belong [11]:

- Drive-by exploits (malicious code injects to exploit web browser vulnerabilities).
- Worms/Trojans.
- Code injection attacks.
- Exploit kits (ready to use software package to automate cybercrime).
- Botnets (hijacked computers that are remotely controlled).
- (Distributed) Denial of Service attacks (DDoS/DoS).
- Phishing (fraud mails and websites).
- Compromising confidential information (data breaches).
- Rogueware/scareware.
- Spam.

Cyber Defence needs monitoring. These technologies are expanding the depth and breadth. Detection challenges are data and variety of data is huge. We need to use data fusion to normalize field values. We need to capture the semantics of all data.

4. Security analysis of network traffic

In recent years, cyber threats have evolved beyond small-time hackers into coordinated attacks by organized cyber criminals. Because of the increase in the number and severity of these attacks, many organizations have begun using advanced analytics and data visualization technologies to find cyber-crime activity and predict future attacks. These technologies help employees who aren't data scientists or analysts to ask questions of the data to quickly and easily find patterns and spot inconsistencies.
Using the best available technology enormous amounts of network traffic data can be aggregated, manipulated, fused, visualized, processed and analysed. Analytics and data visualization gives a more complete picture of systems and networks. Using multiple types of analysis alerts and other valuable intelligence can be created for anomaly detection and predictive analytics, and to investigate slow and low network intrusion. And the analytic models get smarter over time with learning and improvement cycles.

Security analysis of network traffic

- Principles of Internet communication, TCP/IP protocol suite, important application protocols.

- Network attacks and their distribution by network layers.
- The basic elements of network security: firewall, IDS, IPS, anti-spam filter, antivirus.
- Introduction to network monitoring with an emphasis on safety.
- Basic concepts: packets, connections, flows, passive and active monitoring, measurement and data collection, analysis and visualization tools.
- Simple and advanced methods of network traffic aggregated records processing
- Quantitative values (number of bytes and packets), statistical analysis, time series prediction methods.
- Distribution of IP flows (addresses and ports) key items in temporal patterns: entropy and principal component analysis.
- Multidimensional data visualization creation.
- Research in data behaviour during time.

5. Data mining and security analysis

Data mining (DM) is an innovative way of gaining new and valuable business insights by analyzing the information held in company databases. Data mining uncovers this in-depth business intelligence by using advanced analytical and modelling techniques. With data mining, you can ask far more sophisticated questions of your data than you can with conventional querying methods. The information that data mining provides can lead to an immense improvement in the quality and dependability of business decision making.

Perfect data characteristics:

- parsed
- normalized
- understandable log records
- knowledge of log records place
- understandable the logging configuration

Cyber Security Data:

- big number of dimensions (typically between 10 - 20)
- time-series
- discrete or categorical
- abstract entities without context
- data are collected at different levels of abstraction

Known Data Mining Problems and Analytics Issues:

- Big data – problem how to distribute DM algorithms for processing.
- DM algorithms are built for numerical, not categorical data.
- DM algorithms require many parameters with some assumptions anout the data (linear separability).
- It is impossible to model all and change input.
- Anomaly detection.
- Classification and clustering.
- Association rules (high dimensionality).
- Regression.
- Rummarization.

6. How can Visual Analytics help?

Application of Visual Analytics tools can help! We can better insight into algorithm's working. Human can be better put in the loop for

- understanding,
- validation, and
- exploration.

Using VA methods we need to enable humans to solve the problems.

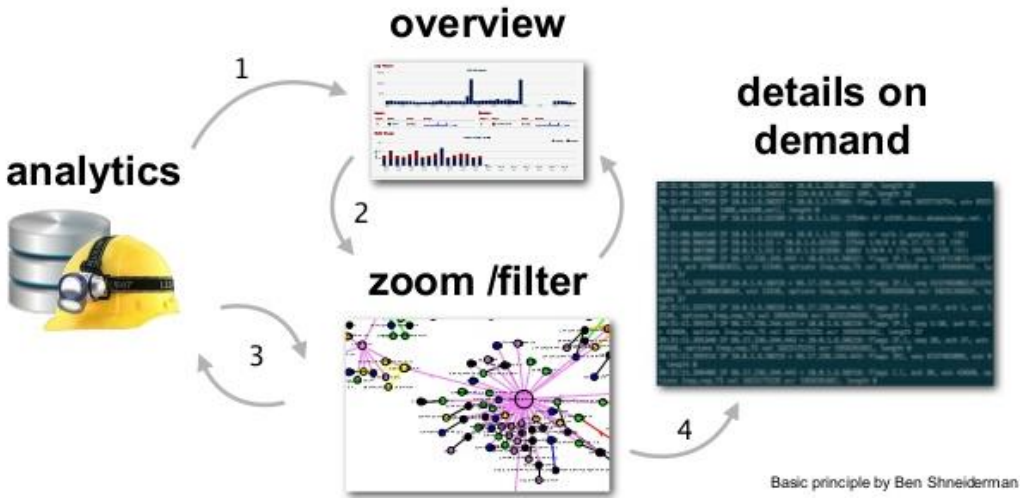

Fig. 1. Visualization process [8]

Visual analytics is the science of analytical reasoning facilitated by *interactive visual interfaces*. People use visual analytics tools and techniques to synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data; detect the expected and discover the unexpected; provide timely, defensible, and understandable assessments; and communicate assessment effectively for action. Visual analytics is a multidisciplinary field that includes the following focus areas:

- Analytical reasoning techniques that enable users to obtain deep insights that directly support assessment, planning, and decision making
- Visual representations and interaction techniques that take advantage of the human eye's broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts of information at once
- Data representations and transformations that convert all types of conflicting and dynamic data in ways that support visualization and analysis
- Techniques to support production, presentation, and dissemination of the results of an analysis to communicate information in the appropriate context to a variety of audiences.

7. Methods of improving on data analysis and visualization results

The proposed solution consists of following analytical steps.

- The analysis of time series (we can easily find trend, rate of change, discover cycles, ...) – all aspects well arranged by using of different kind of charts.
- Analysis of parts and their comparing.
- Analysis of the deviation (difference from the reference value).
- Analysis of the distribution - histograms, frequency polygons, shape, outliers, strip plots, box graph.
- Correlation Analysis - scatter matrix.
- Analysis of multidimensional data.

We need new tools beyond conventional data mining and statistical analysis. Visualization is one such tool and shown to be effective for gathering insight in big data.

The most important topics in the visualization area:

- Streaming data visualization
- Visual data mining
- Visual search and recommendation
- Big data storytelling using visualization
- Scalable parallel visualization methods
- Advanced hardware and architecture for data visualization and analysis
- HCI (Human-Computer Interaction & Knowledge Discovery) from Big Data and big data visualization
- Big data visualization applications including cyber intelligence, cyber security, business intelligence, e-commerce, scientific data analysis, education, etc.

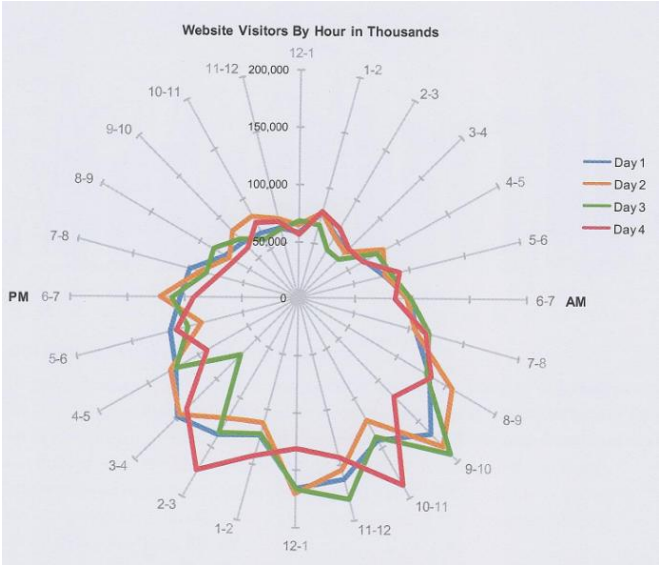

Fig. 2. Radial chart [1]

7.1 Common Techniques

There are a few basic concepts that can help us to generate the best visuals for displaying our data:

- Understand the data we are trying to visualize, including its size and cardinality (the uniqueness of data values in a column).
- Determine what we are trying to visualize and what kind of information we want to communicate.
- Know our audience and understand how it processes visual information.
- Use a visual that conveys the information in the best and simplest form for our audience.

Data visualization is an art and a science unto itself, and there are many graphical techniques that can be used to help people understand the story their data is telling.

VA relates to the areas of Information Visualization and Computer Graphics, and with respect to data analysis, it profits from methodologies developed in the fields of information retrieval, data management & knowledge representation as well as data mining.

7.2 Interactive Visualization

Interactive data visualization goes a step further – moving beyond the display of static graphics and spreadsheets to using computers and mobile devices to drill down into charts and graphs for more details, and interactively (and immediately) changing what data you see and how it is processed.

Visual analytic methods can be extended (or invented) to support distributed synchronous work such as emergency response. The challenges include:

- Developing effective interfaces to visual displays and visual analytics tools operating on multiple kinds and sizes of devices in varied circumstances (for example, mobile devices used in field operations).
- Supporting analysis of continually updating geospatially referenced information of heterogeneous form (for example, map-based field annotations, streaming video, photos and remote imagery, sensor networks).

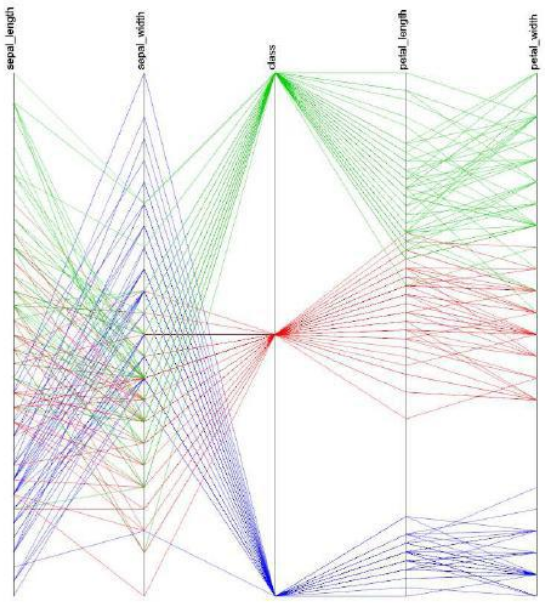

Fig. 3. Parallel coordinates chart [2]

8. Security Visualization examples

Visual analytics is in the Cyber Defence area used for:

- traffic monitoring (using specific metrics, we can better find large-scale trends),
- seasonality (exponential smoothing for data),
- high interactivity and fast data processing,
- too many nodes – optimal algorithms enable well arranged layout and clustering.

- Overview – Situational Awareness
  - goal: insight into million node network, real-time immediate exploration,
  - data: infrastructure – traffic flows, host data, application data, contextual data,
  - analytics: data reduction, automatic restriction,
  - VA methods scales to lots of data,
  - data aggregation – loss of information,
  - highlight anomalies.
- Analytics and filter – exploration and discovery
  - goal: find previously unknown and changing attacks, build patters and models,
  - understanding applications and infrastructure to later spot changes and anomalies,
  - data reduction (high volume, variety, velocity).
- Details on Demand – forensic investigation
  - goal: reduction of main time to repair (MTTR) for incidents, validation of attack, find origin, assessment of impact
  - analytics challenges: fast data queries, data understanding, finding meaningful patterns

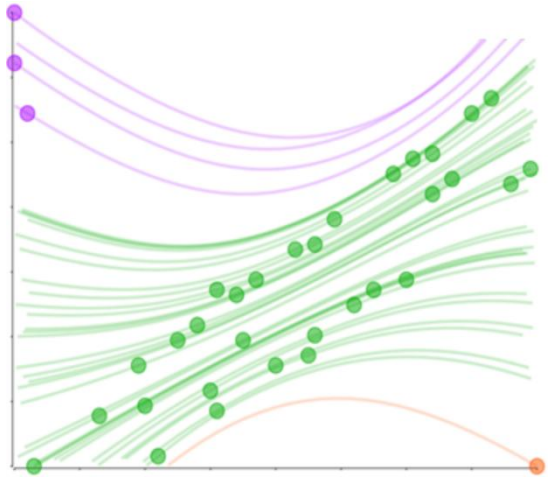

Fig. 4. Cluster by flow lines [5]

9. Conclusion

Data visualization is the presentation of data in a pictorial or graphical format. For centuries, people have depended on visual representations such as charts and maps to understand information more easily and quickly. Many organizations have begun using advanced analytics and data visualization technologies to find cyber-crime activity and predict future attacks.

This article was created to address and explain some of the methods and visual technologies currently utilized in the field of Cyber Defence. Monitoring of network traffic brings the huge volumes of data. Patterns detection within these data allows effectively performing of visual analytics methods and approach mentioned in this article.

Acknowledgements

The work presented in this paper has been supported by the Ministry of Defence of the Czech Republic (Research Plan No. PRO K209).

References

[1] FEW, Stephen C.; *Now You See It – Simple Visualization Techniques of Quantitative Analysis*. Oakland : Analytics Press, 2009. 329 s. ISBN 0-9706019-8-0.
[2] NOVÁKOVÁ, Lenka; *Visualization data for Data Mining*. Praha, 2009. 99 s. Dizertační práce, ČVUT, Fakulta elektrotechnická
[3] THOMAS, J., COOK, K.: *Illuminating the Path: Research and Development Agenda for Visual Analytics*. IEEE-Press (2005)
[4] KIELMAN, Joseph. *Visual Analytics: Dealing with Cyber Storms*. IST-116 Symposium on Visual Analytics, Shrivenham (UK), 2013
[5] ROSENBLUM, Larry. *From Visual Analytics to "Big Data": Progress and Directions*. IST-116 Symposium on Visual Analytics, Shrivenham (UK), 2013
[6] TUKEY, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics, 33*(1), 1-67
[7] WARE, C. (2004). *Information visualization: Perception for design* (2nd ed). San Francisco, CA: Morgan Kaufmann Publishers.
[8] HEER, J., & SHNEIDERMAN, B. (2012). Interactive dynamics for Visual Analytics. *Communications of the ACM, 55(4)*, 45-54.
[9] http://www.visual-analytics.eu/faq
[10] http://www.sas.com/data-visualization/overview
[11] http://www.enisa.europa.eu

Assoc. Prof. Ing. Vlastimil MALY, Ph.D.

Vlastimil MALY was born in 1964. He received his M.Sc. from Military Academy in Brno (CZE) in 1987. He received his Ph.D. from Military Academy in Brno in 1994. He became Associate Professor at University of Defence, Brno in 2007. His research interests include Software Methodology, Information Systems Development and Programming, Knowledge Management, Service Oriented Architecture and C3I SW development in connection with military (NATO) Network Enabled Capability concept. He was appointed in 2007 as the Head of Communication and Information Systems Dept. at University of Defence (UoD), Brno (CZE). He is from 2012 in new position as Vice-Dean for Public Relations and Development, Faculty of Military Technology, UoD. He is also Czech panel member (representative) in NATO STO (recently RTO) - Information Systems Technology (IST) Panel.