Ensuring the Continuing Success of VizSec

Pin Ren
Department of EECS, Northwestern University
p-ren@cs.northwestern.edu

ABSTRACT

In this position paper, I share my research findings in the security visualization research, and my thoughts on the problems and challenges we researchers face. I also outline what we need to do to ensure the continuing success of our community. It is my belief that the future of VizSec research depends on successful addressing the following tasks:

1. Generating effective and high-level visualization design guidelines and research methodologies.
2. Conducting studies to quantify effectiveness of security visualization techniques.
3. Contributing to the two parent communities, namely computer security and visualization, by integrating visualization components into existing security tools and feeding back successful visualization designs.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/methodology; Graphical user interfaces (GUI)*

General Terms

Human Factors, Security

Keywords

Security Visualization, Visualization Evaluation, Visualization Design, Dissemination

1. INTRODUCTION

The Internet has become worldwide crucial infrastructure. However, the large amount of data transmitted over typical networked systems has made it difficult to spot activities by malicious adversaries. The massive volume of network traffic poses a task as hard as sorting out a needle in the haystack. Such challenges are well discussed and understood within our research community.

To solve network security issues, the traditional approach utilizes textual information, such as system or traffic logs. This approach requires highly experienced and well-trained administrators to do tedious monitoring and detection work. Since this approach heavily relies on human intervention, it is inherently not scalable. Even for a small network, the administrator's workload can quickly become overwhelming.

Meanwhile, numerous intrusion detection systems (IDSes) have been designed and built to address those challenges. Most of them are based on known signatures of previous attacks or traffic measurement statistics. IDSes of the first type are inflexible when dealing with challenges posed by the new and evolving threat patterns. IDSes of the latter suffer from high false positive ratio and/or the high false positive ratios because there are no universally applicable detection parameters. Even with some adaptiveness and training from historical data, the parameter setting is still far from ideal.

Recently, researchers from both the computer security and visualization communities have been looking into employing visualization and human-interaction to better address those security challenges. Our fast-growing VizSec community is the direct outcome of such efforts. We have already seen quite a few successful applications [9, 5, 6] using visualization for solving those security problems. However, many important issues remain unsolved or even unaddressed:

1. The absence of theoretical guidelines and high-level methodological principles.
2. The difficulty of quantifying and comparing the effectiveness of different visualization designs.
3. The fact that there are more solutions to toy problems than successful real-world security applications.

Based on the observations above, I would argue to rethink the whole process of designing and building security visualization tools and the way we conduct our research. I am going to elaborate on my point in the following sections.

2. DEFINE THE ARENA WHERE VISUALIZATION EXCELS

Let us first look at what visualization can do to help solve security problems. It is only possible to leverage those advantages when we are fully aware of them. Clearly, there are many potential advantages of utilizing visualization to solve computer security problems:

VizSEC'06, November 3, 2006, Alexandria, Virginia, USA.

1. Representational: convey information accurately and effectively.

2. Cognitive: ease the perceptual/cognitive burden in understanding the complex or massive dataset.

3. Analytical: facilitate thorough investigation and deep analysis with visual stimulus and human interaction.

4. Exploratory: assist dataset walkthrough and detail-on-demand tasks with intuitive interface design and interaction support.

5. Situational awareness: support effective monitoring that requires low human effort.

In general, visualization can assist domain experts to visually monitor and explore their dataset. By putting human in the loop, visualization can provide better situational awareness and augment domain experts' analytical ability. However, since this process requires human intervention, it may not be well suited for some real-time tasks that require automatic and instant reaction.

Usually visualization alone is not enough to draw conclusions on the perceived incidents. Further investigation utilizing original logs is often required to decide if it is really a malicious attack or just a system misconfiguration, loose connection or simply nothing but normal traffic.

I think it is not a promising route using security visualization alone to replace automatic IDSes and log file investigation for real-time intrusion detection. Instead, visualization should be used together with other IDSes and analytical tools for logs to improve the detection accuracy and enjoy better detection confidence. Ultimately, it will be even better to integrate visualization as a component into existing intrusion detection and analysis tools. The insight gained from visualization can guide the analysis, and the detection results and analysis outcomes can be depicted visually.

3. LEAD TOWARDS SUCCESSFUL DESIGN

Beginning any design process, the first question we should ask is: what is the most important aspect of the problem? In the field of security visualization, what matters most is the detection/monitoring/analytical task which needs to be completed in a timely manner. Undoubtedly, the design of security visualization tools should be driven by such tasks, and most of the current security visualization designs are indeed rooted from attempting to tackle a particular real-world problem, however most of such designs are ad-hoc and not generalized.

I therefore think it is time for our community to collect and summarize efficient high-level design guidelines and methodologies, and establish the theoretical base for the design practice. Such effort will not only help us better understand why certain designs are more effective than the others on specific tasks, but also makes it easier to generalize a successful design to solve a broader range of problems and eventually helps to advance the state-of-art of our research.

When working with security datasets what we are really looking for are the following data features:

1. Statistically meaningful features, such as max/min/mean values: this can be used for supporting performing analytical tasks or tuning parameters in automatic IDSes;

2. Changes over time: in both large scale and small details, such information is the key to understanding the data dynamics, detecting anomalies, and enjoying the situational awareness.

3. Overall traffic patterns: indispensable for understanding the big picture.

4. Unusual or repeating activity patterns: essential for identifying attack signatures.

5. Correlations: extremely important to unveil hidden relationships.

6. Uncertainty: it is unavoidable and largely overlooked in noisy network data input.

To perform different types of tasks, we need to organize and present those individual data features and/or combinations of features visually in an meaningful way.

In my own experience, I find that it is helpful and effective to build the visualization system based on data feature abstractions: design and implement appropriate visualization components to emphasize data features, and then construct the task solving pipeline using suitable visualization components based on the nature of tasks.

The immediate benefit of this approach is that the visualization for data abstraction can be reused in different tasks by tweaking the workflow of visualization components, and their data interfaces. Effective working patterns can be generalized for solving a class of similar problems, and most importantly, we are able to enjoy the flexibility to try out different visualization component and choose the best one for a given task.

4. COMPARE AND EVALUATE SCIENTIFICALLY

Historically, in both the information visualization and scientific visualization research areas, visualization has long been a craft rather than science. Methods are often designed and evaluated by presenting results informally to potential users. [4], the security visualization is no exception.

However, the situation is changing quickly. The importance of quantifying effectiveness has already been recognized as one of the top research problems. [3, 2]. More and more visualization researches have put evaluation, comparison and formal user studies into the design and development process.

In the security visualization field, it is noticeably more difficult to conduct convincing evaluation and comparison than in many other visualization areas. The problem roots deeply in the volatile nature of both the dataset and tasks. Security threats are changing so quickly that visualization methods that are effective on detecting previously known threats can be soon outdated. Previous evaluation can hardly be reused due to the change of threat subject. This makes comparison between existing and new visualization designs a difficult and time-consuming job. Ideally we would like to have real-world intrusion dataset with malicious attacks and anomalies annotated by domain experts, due to the security sensitivity and rarity of annotations, such ideal testing datasets are not easy to acquire.

In my opinion, we researchers need to establish a public domain repository for evaluation datasets and tasks. This

repository should update frequently to include the newest threats data on the Internet. It is also necessary to have those datasets annotated for intrusion and anomalies by domain experts. Using standard datasets and tasks will make it easier to conduct quantitative studies on effectiveness and to compare different visualizations.

In addition to the proposed repository, a visualization contest is another good way to encourage design comparisons and evaluations. We can follow the contest model of the annual IEEE InfoVis symposium, which issues the testing data format, a sample dataset and testing tasks specification several months before the annual conference. Due to the nature of the tasks, we may require on- site demonstration working on new datasets, to test which implementation provides the most insights and best detection accuracy in the shortest possible time.

Besides the data and task repositories, successful security visualization designs can also form a repository and it should be made public and open source online. This idea is inspired by the Information visualization Cyberstructure[1] from Indiana University. The major benefit of such repository is that users and researchers can have access to existing security visualization designs, and can build on top of previous implementation to solve their problem faster. This effort requires continuing and long-term commitment, and its success calls for the involvement of the whole VizSec community.

5. DEPLOY, INTEGRATE AND DISSEMINATE

Our VizSec community is an emerging interdisciplinary research community. It sits right in the middle of many more traditional research areas, such as computer security, visualization (especially information visualization) and human-computer interaction. The future of this interdisciplinary research effort not only depends on its own development but also on the advancement of its parent research domains. More importantly, the outcomes of our investigations in VizSec have the potential to significantly impact the development of those parent areas.

As I discussed, standalone visualization tools may not be quite helpful in solving real-world problems, thus they have not been able to make big impact. However, such tools are more effective and decisive when used together with existing IDSes and security analysis tools. In my opinion, it is time to integrate visualization as a component into those existing tools. Only when visualization is integrated, deployed, and widely used by domain experts on a daily basis, can it achieve greater impact and better solve real-word problems.

The abstract, dynamic nature of security datasets and tasks provides an excellent opportunity for security visualization to exert greater impact in the visualization community as well. More and more visualization sub-domains are interested in datasets with similar characteristics as seen in security, such as stock market data visualization and road traffic visualization. They may benefit from the development of security visualization with similar format of data and tasks requirement. Meanwhile, our community can also borrow successful existing techniques from those visualization domains. For example, the visualization portion of IDGraphs [7] was originally designed for financial market visualization, and the *Flying Term* visual metaphor designed for DNS query data visualization [8] is now appreciated in the visual analysis of news articles.

6. CONCLUSION

As our research endeavor in security visualization dives deeper into real world challenges and problems, the future of the VizSec community and our research depends on the way we choose to address those challenges and tasks. In this position paper I have explained my positional statements regarding those important issues, and would welcome any feedback and comment.

7. ACKNOWLEDGEMENTS

I would like to thank my advisor Bruce Gooch for his many helpful comments. I would also like to thank Holger Winnemöller for his proofreading and editing help.

8. REFERENCES

[1] J. Baumgartner and K. Börner. Interactive Poster: Towards an XML Toolkit for a Software Repository Supporting Information Visualization Education and Research. In *Poster Compendium of the 2003 IEEE Symposium on Information Visualization*, 2003. http://iv.slis.indiana.edu/.

[2] C. Chen. Top 10 Unsolved Information Visualization Problems. *IEEE Computer Graphics and Application*, 25(4):12–16, 2005.

[3] C. Johnson. Top Scientific Visualization Research Problems. *IEEE Computer Graphics and Application*, 24(4):13–17, 2004.

[4] R. Kosara, C. G. Healey, V. Interrante, D. H. Laidlaw, and C. Ware. User Studies: Why, How, and When? *IEEE Computer Graphics and Application*, 23(4):20–25, 2003.

[5] K. Lakkaraju, W. Yurcik, and A. J. Lee. NVisionIP: NetFlow Visualizations of System State for Security Situational Awareness. In *Proc. of VizSEC/DMSEC*, 2004.

[6] J. McPherson, K.-L. Ma, P. Krystosk, T. Bartoletti, and M. Christensen. PortVis: A Tool for Port-Based Detection of Security Events. In *Proc. of VizSEC/DMSEC*, 2004.

[7] P. Ren, Y. Gao, Z. Li, Y. Chen, and B. Watson. IDGraphs: Intrusion Detection and Analysis Using Stream Compositing. *IEEE Computer Graphics and Application*, 26(2):28–39, 2006.

[8] P. Ren, J. Kristoff, and B. Gooch. Visualizing DNS Traffic. In *Proc. of VizSEC*, 2006.

[9] X. Yin, W. Yurcik, M. Treaster, Y. Li, and K. Lakkaraju. VisFlowConnect: NetFlow Visualizations of Link Relationships for Security Situational Awareness. In *Proc. of VizSEC/DMSEC*, 2004.