A Proposed Visualization for Vulnerability Scan Data

Stacey Watson
University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC 28223
swatso50@uncc.edu

Heather Richter Lipford
University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC 28223
Heather.Lipford@uncc.edu

ABSTRACT

System administrators make security decisions based on data provided by a variety of tools. Yet too often these tools do not structure the presentation of that data to support the communication and decision making needs of a variety of stakeholders within an organization. For example, consider the task of fixing system vulnerabilities based on network scans. Network vulnerability tools produce an overwhelming amount of raw data that is difficult to prioritize. The most critical vulnerabilities in the most sensitive systems must be addressed quickly, before attackers discover and exploit them. Additionally, non-security domain experts are often called upon to perform remediation and/or to make critical security decisions. As such, it is imperative that the security state of the network be communicated in such a way as to support these efforts. Unfortunately, current security tools that provide visualizations are complex and fail to provide actionable data. In this paper, we propose a new way to visualize vulnerability scan data by network zone using free and open-source tools to demonstrate how visualizations can be created to support decision making.

1. INTRODUCTION

System and network security requires that the most critical vulnerabilities be remediated before they are discovered and then exploited by an attacker. However, even a relatively small corporate network has thousands of vulnerabilities to address. The challenge, therefore, becomes identifying where to focus remediation efforts first so as to minimize the attack surface. Performing visual data analysis of vulnerability data has the potential to bring to the surface the "patterns, trends, structures, and exceptions" [7] in the hope that this would assist remediation teams to identify where to focus remediation efforts. Furthermore, such visual approaches could enable key stakeholders to quickly gain insight into the security state of a network and its systems.

However, security visualization tools are often complex and therefore are not accessible for those individuals who are not security domain experts, but who are responsible for hardening systems and networks or are called upon to make security decisions. Additionally, such data is often displayed on dashboards provided by vulnerability scanning tools. These tools are limited to pie charts, line graphs and bar charts depicting how many vulnerabilities of each severity are present, without regard to network traffic flow between systems or how far removed each system is from the external Internet. Yet, administrators must use this knowledge of their network configuration and system location within the network in order to understand the impact of vulnerabilities and determine priorities.

Thus, information displayed based solely on severity does not help decision-makers and remediation teams to quickly assess the situation and then make decisions about where to focus remediation efforts. Also, remediating a single critical vulnerability across multiple systems is not an efficient approach as frequently system-level testing must be performed to ensure that a patch or configuration change does not adversely impact the system or other systems that have dependencies upon it. It is more common for teams to remediate all vulnerabilities on particular high priority systems in a test environment before pushing the changes onto production systems. Furthermore, systems deep within a network with little to no exposure to the external Internet are at low risk of exploitation, even if they have critical vulnerabilities. Therefore, it is much more effective for teams to focus their remediation efforts on those assets that are at highest risk of exposure to the external Internet.

As others have argued, security visualizations need to instead be visual analytic systems, supporting the analysis and the "timely" and "informed" decision making of users [7]. Yet, this is not necessarily difficult to do, despite the massive volume of data generated from security tools and the complexity of modern networks. However, currently available security tools are often "inefficient and inadequate for achieving situational awareness by a human analyst" [7].

For vulnerability scan data, this means interfaces that communicate vulnerability information in such a way as to make it easier for decision-makers to focus remediation efforts on the systems with the highest risk of exposure based on not just the severity of the vulnerabilities but also where those systems are on the network.

To that end, we propose a zone-based vulnerability visualization using actual raw vulnerability data from Rapid7 Nexpose. We created this visualization using R and RStudio, both free, open-source tools with simple-to-update raw

Symposium on Usable Privacy and Security (SOUPS) 2017, July 12–14, 2017, Santa Clara, California.

CSV exports from the vulnerability tool. The paper describes our initial visualization, as well as plans for evaluating its effectiveness in helping users make critical security decisions. The visualization also demonstrates that incorporating contextual information needed to aid decision making can already be done with existing tools and techniques, and simple visualization methods.

2. BACKGROUND

Due to the volume and complexity of security data, visualizations have the potential to obscure an issue. As such, there is currently some distrust of security visualizations by security professionals [4]. Nonetheless, such approaches can be powerful aids to "transactive memory and ongoing mutual understanding" [3] and therefore researchers have suggested the employment of "more graphical metaphors" [5] in the user interfaces for security tools.

Shiravi et al., in their survey of visualization systems for network security, discovered a transition to approaches that help "prioritize situations and events" from more traditional visualization approaches [7]. The work thus far has been focused on technologies that are reactive rather than proactive, such as monitoring network interactions between internet and external hosts, monitoring port activity as a means of identifying malicious programs, monitoring routing behavior as a means of detecting and correcting network disruption events, and monitoring data collected from intrusion detection systems as a means of identifying attack patterns.

Such systems require substantial "testing, tuning and evaluation" [7] before the visualizations could be deployed and used by security administrators. Furthermore, of all approaches surveyed, only two included usability studies in their evaluation process.

One such approach was Ocelot, a visualization tool meant to help security analysts make decisions about a known network intrusion event by helping the user organize the network hierarchically using node attributes such as operating system, node location, and network role [2]. The usability tests were conducted on 17 cyber-security masters students and then later on 4 cyber-defense threat analysis and penetration testing professionals. Both the less experienced students and the professionals were able to use Ocelot's features to decide how to respond to an intrusion event.

These results demonstrate that a visual approach can be effective at guiding security decisions, even for less experienced key personnel. However, little work has been done at visualizing vulnerability data so that systems and networks can be hardened, making it much less likely that an attacker will be able to breach the perimeter of a network.

It is essential to create a visualization that can be quickly learned or that draws on existing mental models for network personnel who are often called upon to perform remediation tasks, but who are not necessarily security domain experts. Network personnel understand logical and physical separation in their network and, therefore, network security zones would provide the underlying framework for this proposed vulnerability visualization.

2.1 Zone-Based Network Defense

With an eye toward protection, defense and containment of security threats, industry professionals have long advocated the use of network security zones as a means of increasing security through the strategies of defense in depth and compartmentation.

Security blogger Nige the Security Guy argues that "network security zones that separate systems based on their communication and protection needs minimize security risks while allowing information flows to continue even in the face of failures and security incidents" [8].

Security zones limit risks and exposure through role-based organizational access control into each zone, allow those monitoring the network to more quickly identify suspicious activity directed toward key assets located in critical zones and allow a zone or sub-zone to be locked down in the event of a breach or incident requiring containment. Furthermore, this approach facilitates incident response, investigation, and recovery.

Figure 1 shows the proposed zone-based architecture, in which the network is arranged into six logically and/or physically separated zones. Sub-zones could also be configured within each zone, as needed. Access into each zone is managed by a perimeter security device, such as a firewall.

The untrusted zone contains all hosts that do not belong to the organization, including the Internet Service Provider. Traffic from this zone flows into the semi-trusted (DMZ) zone, in which the systems that must be exposed externally are located. This would include the external web servers for the organization.

The trusted zone would contain all internally-exposed systems, such as employee desktops and laptops and the organization intranet. Traffic flows from it into the semi-trusted zone and out to the untrusted zone so as to allow access to all of the resources of the internet.

High risk, mission critical systems, such as database servers that hold restricted organizational data would be located in the restricted zone. Traffic from this zone flows into the trusted zone.

The management zone would hold all systems that manage the network and security for the organization. Traffic from this zone flows out into the semi-trusted, trusted and restricted zones.

Finally, the audit zone would contain all systems that collect logs and other data necessary for auditing the organization. As this data is highly sensitive, traffic only flows in from the semi-trusted, trusted and restricted zones.

We thus propose overlaying a visualization of vulnerability scan data on top of this zone architecture, providing vulnerability data in the context of the actual network layout. We will not discuss how we created our specific visualization in this paper.

3. PROPOSED VISUALIZATION

Vulnerability data, in particular, has been challenging to characterize due to the sheer volume. As a case in point, the raw Nexpose report used to generate the visualization for this paper contains 11,836 individual vulnerabilities over 35 systems and is, therefore, unwieldy and overwhelming. Nonetheless, CSV or Excel reports of such data are often distributed to teams who are responsible for the remediation

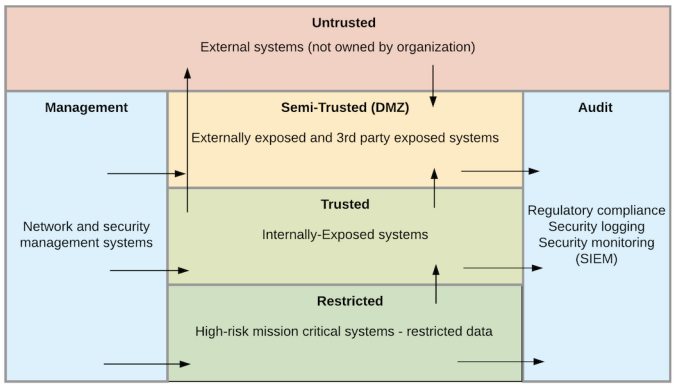

Figure 1: Zone-Based Network Architecture [8]

of vulnerabilities.

The raw data behind our visualization was gathered from an actual scan of 35 systems at our university. As the scanned systems were all in the same subnet without any logical or physical separation between them, we manually configured security zone information for each in order to illustrate our proposed visualization.

Figure 2 shows the proposed zone-based visualization generated from this Nexpose vulnerability data with R Studio and the R igraph library.

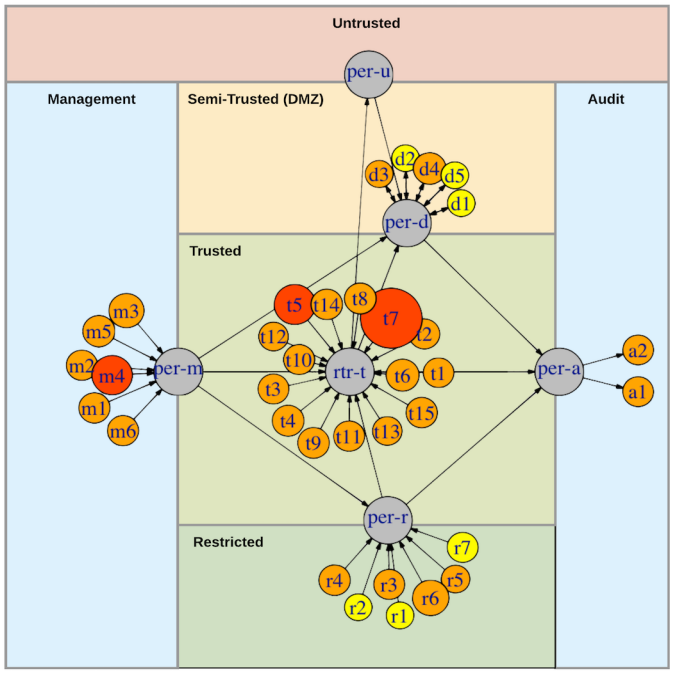

Figure 2: Zone-Based Vulnerability Visualization

3.1 Inputs

The visualization requires three CSV files as inputs:

- A file for the graph nodes, each of which correspond to a unique device on the network with the IP address for each node designated as the node id.
- A file for the edges with all connectivity to the perimeter devices represented, again designated in the file by IP address as the from/to columns.
- A file with the raw vulnerability data.

The nodes and edges files would only need to be created upon initial set up and then edited when new devices are brought online and old devices are decommissioned and/or when connectivity between devices change.

The raw vulnerability data would be exported as a CSV file from the vulnerability scanning tool and is consumed in raw format, thereby requiring no data manipulation.

3.2 Nodes

Each node in the directed graph in Figure 2 corresponds to a unique device on the network. However, as this visualization represents actual vulnerability data for live devices, a label is used on the diagram to represent each node. Furthermore, the decision to incorporate a short label to represent each device rather than a four-octet IP address avoids "occlusion and overcrowding of displays" that make it difficult for human beings to discern the patterns and trends in the data from the visualization [7].

This visualization uses a mean of all Nexpose severity scores for a given node to determine its overall severity score, though this could be easily changed to a different statistical calculation. For example, Kellet et al. propose using a combination of asset criticality as well as an organization's knowledge of adversarial interest [6] for ratings, though this would require adding additional information into the nodes CSV file and additional logic to the R Script.

Nexpose security scores are calculated based on the Common Vulnerability Scoring System, Version 2, which take into account base metrics representing the risk of a given vulnerability, including access, access complexity, authentication requirements, and data confidentiality, integrity and availability impacts [1]. All vulnerabilities are rated by the CVSS system on a scale of 0.0 to 10.0.

A moderate vulnerability, which ranges from 0.0 to 3.9 on the CVSS system, can only be exploited locally and requires authentication. These vulnerabilities would provide little to no access to restricted information to an attacker and would not provide a means by which the attacker could destroy or corrupt data or could cause system outages. The yellow nodes in Figure 2 have a moderate average severity score.

A severe vulnerability, which ranges from 4.0 to 6.9 on the CVSS system, can be exploited by a moderately experienced hacker with or without authentication. These vulnerabilities would provide partial access to restricted information to an attacker and could provide a means by which the attacker could destroy or corrupt data or cause system outages. The orange nodes in Figure 2 have a severe average severity score.

A critical vulnerability, which ranges from 7.0 to 10.0 on the CVSS system, can be easily exploited with little to no authentication. These vulnerabilities would provide full access to restricted information to an attacker and provides

a means by which the attacker can destroy or corrupt data or cause system outages. The red nodes in Figure 2 have a critical average severity score.

Node size in Figure 2 is representative of the overall vulnerability exposure for each node and is calculated based on the number of severe and critical vulnerabilities discovered for a given node. In other words, nodes with a larger number of severe and critical vulnerabilities will be larger than nodes with a smaller number.

The grey nodes in Figure 2 represent perimeter network and security devices such as firewalls, routers and so forth. The grey nodes placed on the lines between network zones represent the perimeter device that creates logical or physical separation between the network zones.

3.3 Edges

Each edge in the directed graph in Figure 2 corresponds to the flow of traffic in the network, based on the firewall rules implemented on perimeter devices. Notice that in this case, traffic between the grey perimeter devices flows according to the recommended configuration in figure 1, which was proposed by the Adaptive Zone Defense blog article [8].

4. USE CASES AND STRATEGIES

This zone-based vulnerability visualization in figure 2 could be used as a part of a regular review of network architecture and flow to ensure that devices are in the correct security zone and that firewall and routing rules are configured properly to maintain proper logical and physical separation between zones.

Additionally, this visualization provides a high-level overview of the security state of the network that could potentially be used when making key decisions about where additional security devices or sub zones might be necessary or when additional network or security personnel might be required for remediation efforts.

Furthermore, remediation strategies could be developed based on the visualization in the hope that it would help prioritize the remediation teams' efforts in the remediation of vulnerabilities. For example, the strategy might be to re-mediate all nodes from largest to smallest with a critical (red) severity score and then all nodes with severe (orange) severity scores in each zone in the following order:

- Semi-trusted (DMZ)
- Trusted
- Restricted
- Management
- Audit

Once this has been accomplished, the remediation team can go back to address the moderate vulnerabilities by zone, in the above order.

Alternatively, the strategy could be to address all nodes from largest to smallest with a critical (red) severity score and then all nodes with severe (orange) and, finally, all moderate (yellow) nodes in each security zone in the above order. Either way, the visualization supports discussion of which nodes are most critical to fix and why.

5. BENEFITS

The zone-based vulnerability visualization in Figure 2 has a number of potential key benefits, the first of which is that it is intended to provide a high-level overview of the security state of the network. It is hoped that this overview would clearly depict the data for all stakeholders, including key personnel with limited domain knowledge in security and those in management who need a holistic sense of security before making critical decisions. As found by Werlinger et al., security professionals regularly communicate with stakeholders with little security training, who have varying perceptions of risks and who do not consider security to be their primary priority [9].

Once a remediation strategy has been adopted, it is hoped that the visualization would provide actionable data for prioritizing remediation efforts for network and security personnel. The remediation team, using the node color, node size and zone location of each node could quickly identify which nodes should be addressed first.

Additionally, it is hoped that inclusion of network architecture and traffic flow in the visualization would help network architects identify nodes in the wrong zone or insecure traffic patterns.

The visualization could be readily adapted to other types of network structures according to the organization's planned defense in depth and compartmentation strategies.

Finally, after the nodes and edges documents are configured, the visualization is extremely easy to generate with R and RStudio, both open-source tools. A CSV export of vulnerabilities is all that is required to re-run the visualization with fresh data and can even be configured to update an R-powered website dashboard for key personnel.

6. LIMITATIONS

The proposed visualization in Figure 2 has several limitations, the first of which is that it requires an accurate knowledge of network architecture and traffic flow. As such, initial set up of the nodes and edges CSV files could take some time and research. Inaccurate network architecture data could lead to a false sense of security or to remediation efforts being directed away from the highest priority devices.

Other limitations arise whenever attempting to capture vast data in a simplified, visual representation. When Fink et al. interviewed eight cyber-security analysts, concerns about visualizations obscuring an issue were cited as one of the reasons analysts distrust such systems [4]. The choice of how to calculate the overall security risk, for example, could misrepresent risk by choosing a measure that reduces the number of critical (red) nodes. The prevalence of false positives in automated security systems also could misrepresent the risk, though it is arguably better to over-represent risk than to under-represent it.

Furthermore, there is the potential side effect that the visualization might mislead tool users should they confuse node size with node color or misunderstand that edge direction represents firewall rules.

Finally, scalability could be an issue with a one to one mapping between devices and nodes on the connected graph for very large networks. Networks with hundreds or thousands

of devices would result in overcrowding and occlusion and therefore would have to be depicted differently. Some careful thought needs to go into this issue, but some ideas for addressing it include representing multiple devices in each node or presenting an overview of the entire network and then separate visualizations for each network zone.

7. FUTURE WORK

The next steps for this work are to assess whether or not the visualization fosters analysis and communication with actual users to help them prioritize remediation efforts. Also, it would be useful to compare this approach with what remediation teams are currently using to assess whether this visualization is more effective and efficient at remediating the most critical assets.

Additionally, the visualization itself could be extended and scaled to handle larger networks and more sophisticated measures of criticality. For example, we would like to investigate incorporating some of the model proposed by Kellett et al. where asset criticality is determined based on adversarial interest [6]. This would require more research when setting up the nodes CSV file, but may be well worth the extra front-end set up. The additional logic for processing asset criticality could be added to the R Script.

More generally, we would like to examine more deeply the needs for supporting security-related decision making through dashboards and other methods of information presentation to security administrators and stakeholders. Our visualization here demonstrates that for some kinds of tasks, there already exist straightforward ways to incorporate critical environmental information with raw security data to help inform security decisions. We would like to examine more broadly what kinds of information stakeholders need, how dynamic interaction with that data can help decision making, the trade-offs between high level visualizations such as our example here which are more abstract, and low level visualizations providing greater details of the raw data, and ways to move between those representations.

Due to the risks, vulnerability scan data is not typically shared outside of an organization. However, we believe organizations could be recruited to participate in the study with the understanding that our team would not have access to the vulnerability data. Detailed configuration instructions and the R script could be provided to organization personnel and we could provide some tutoring using sample data so they could use the visualization on their own data. Usability could be measured and compared with their existing vulnerability management strategies through detailed survey instruments and interviews.

8. CONCLUSIONS

It is hoped that the proposed zone-based vulnerability visualization presented in this paper would help address the overwhelming volume of data produced by modern network scanning tools such as Rapid7 Nexpose. It uses free, open-source tools to generate information about the security state of each device based on the CVSS-based severity score provided by Nexpose and the overall exposure based on how many vulnerabilities were found. Furthermore, it also provides information about which devices are in which security zone and the traffic flow between devices and zones.

This visualization may help address a current problem in vulnerability management, namely, that visualizations in vulnerability scanning tools are limited to pie charts, line graphs and bar charts depicting how many vulnerabilities of each severity are present without regard to network traffic flow between systems. This results in data that is not actionable as it does not provide the situational awareness necessary for informed analysis and decision-making.

Our goal is to help teams more effectively focus on those assets that are most exposed to the external Internet systems, allowing decision makers to identify where time and personnel should be focused to address the highest risk zone and devices or where changes would proactively reduce the potential attack surface of the network. We believe that this specific example of combining existing tool output with environmental information demonstrates the importance of designing interfaces to support the kinds of real world decisions security administrators need to make.

9. REFERENCES

[1] *PCI, CVSS, & Risk Scoring Frequently Asked Questions*, 2017 (accessed May 21, 2017). https://help.rapid7.com/nexpose/en-us/Files/Risk_scoring_FAQ.html.

[2] D. L. Arendt, R. Burtner, D. M. Best, N. D. Bos, J. R. Gersh, C. D. Piatko, and C. L. Paul. Ocelot: user-centered design of a decision support visualization for network quarantine. In *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pages 1–8, Oct 2015.

[3] D. Botta, K. Muldner, K. Hawkey, and K. Beznosov. Toward understanding distributed cognition in it security management: the role of cues and norms. *Cognition, Technology & Work*, 13(2):121–134, 2011.

[4] G. A. Fink, C. L. North, A. Endert, and S. Rose. Visualizing cyber security: Usable workspaces. In *2009 6th International Workshop on Visualization for Cyber Security*, pages 45–56, Oct 2009.

[5] E. Haber and E. Kandogan. Security administrators: A breed apart. *SOUPS USM*, pages 3–6, 2007.

[6] M. Kellett. Ranking assets based on criticality and adversarial interest. 2016.

[7] H. Shiravi, A. Shiravi, and A. A. Ghorbani. A survey of visualization systems for network security. *IEEE Transactions on Visualization and Computer Graphics*, 18(8):1313–1329, Aug 2012.

[8] N. the Security Guy. *Adaptive Zone Defense*, 2013 (accessed May 21, 2017). https://nigesecurityguy.wordpress.com/tag/network-security-zones/.

[9] R. Werlinger, K. Hawkey, D. Botta, and K. Beznosov. Security practitioners in context: Their activities and interactions with other stakeholders within organizations. *International Journal of Human-Computer Studies*, 67(7):584 – 606, 2009.