OCEANS - Online Collaborative Explorative Analysis on Network Security

Siming Chen
Peking University
csm@pku.edu.cn

Cong Guo
Peking University
cong.guo@pku.edu.cn

Xiaoru Yuan
Peking University
xiaoru.yuan@pku.edu.cn

Fabian Merkle
Universitaet Stuttgart
merklefn@studi.informatik.uni-stuttgart.de

Hanna Schaefer
Universitaet Stuttgart
schaefha@studi.informatik.uni-stuttgart.de

Thomas Ertl
Universitaet Stuttgart
Thomas.Ertl@vis.uni-stuttgart.de

ABSTRACT

Visualization and interactive analysis can help network administrators and security analysts analyze the network flow and log data. The complexity of such an analysis requires a combination of knowledge and experience from more domain experts to solve difficult problems faster and with higher reliability. We developed an online visual analysis system called OCEANS to address this topic by allowing close collaboration among security analysts to create deeper insights in detecting network events. Loading the heterogeneous data source (netflow, IPS log and host status log), OCEANS provides a multi-level visualization showing temporal overview, IP connections and detailed connections. Participants can submit their findings through the visual interface and refer to others' existing findings. Users can gain inspiration from each other and collaborate on finding subtle events and targeting multi-phase attacks. Our case study confirms that OCEANS is intuitive to use and can improve efficiency. The crowd collaboration helps the users comprehend the situation and reduce false alarms.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces; C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*

Keywords

Network Security, Situation Awareness, Collaborative Visual Analytics

1. INTRODUCTION

VizSec '14, November 10 2014, Paris, France

ACM 978-1-4503-2826-5/14/11 ...$15.00
http://dx.doi.org/10.1145/2671491.2671493.

Network security is an important issue nowadays. Large amounts of network flows and logs describe the connection behavior and the dynamic host status. Detecting network events from the complex network data is a critical task. A network event can be regarded as an anomaly which is usually caused by an attack. Subtle event detection is challenging because of its tiny suspicious behavior among large normal connections. Event correlations for understanding of multi-phase attacks are also important and need new designs of techniques. Visual analytics provide interaction and visualization techniques that can support these tasks. Moreover, the improvement can be obtained through better support for collaboration in visual analytics.

In this paper, we introduce the OCEANS (Online Collaborative Explorative Analysis on Network Security) system, which supports fast and deep event detection by integrating visual analytics methods and collaboration features as a web application. Linked views on three different levels visualize the netflow and logs while submitting, commenting and re-analyzing the events build the collaborative features. With this collaboration possibility which makes full use of human insights, analysis becomes more efficient. Specifically, our system has the following contributions for helping security experts detecting complex events:

- **Visual event fingerprint and similarity based inspiration**: Our system describes each event as a visual fingerprint combined with three levels of information (temporal, IP connection s and detailed connection information) extracted from multiple source data. By loading others' event submission, users can understand the event pattern and target similar events, which improves efficiency.
- **Collaborative approach to tackle the event correlation for multi-phase attacks**: Experts with different specialities can put their knowledge as well as their findings of different aspects together. Visual hints and interactions help users exploring based on correlated features (e.g. suspicious IPs, neighboring time sequence). Subtle events of the multi-phase attacks are also easier to find in the multi-level visual exploration with the aid of communication.
- **Crowd-collaborative input synthesis to improve efficiency and reduce false alarms**: OCEANS integrates the crowd input from security experts and makes everyone contribute to evaluate the events. An

overview of event graph generated from all users' input makes them understand the overall connection patterns and helps detecting suspicious IPs.

2. RELATED WORKS

As security issues become ever more critical, there are many visual analytics systems aiming at analyzing network security and identifying anomaly events [17]. These systems analyze different data, including the network flows [10], status records of servers and switches [8], logs of the firewall, Intrusion Detection System (IDS) and Intrusion Protection System (IPS) [15], and the routing information [19].

There are multiple visual metaphors in security visualization. VisAlerts maps the network into a ring, showing host information, connectivity and temporal information together [9]. In another aspect, Conti et.al visualized network flow based on parallel coordinates [7]. We improved the ring layout with a three-level hierarchy and explorative filtering function. Pixel-based visualization and curve band design integrate multiple data sources into the task-specific parallel coordinates. An overall visual fingerprint could help users analyze complex patterns and subtle events.

Multi-level exploration is supported by security visual analytics systems. IDS Rainstorm visualized IDS alarms on a large network, revealing temporal patterns, IP location and severity [3]. Bunch et. al developed a 3D event detection system showing netflow's temporal information [6]. Data mining techniques are combined with visualization methods to detect events [4]. These systems did a good job in detecting different types of events, but identifying event correlations and multi-phase events is still challenging. Stoffel et. al addressed the event correlation through time-series visual analytics [18]. Our multi-level visual exploration system combines the collaboration mechanism to solve these problems with better engagement of people.

Collaborative analysis is useful for complex data analysis and decision making, especially with visualization and interaction [12]. Heer and Agrawala provide a design guideline of collaborative visual analytics [11]. Many Eyes [16] is a popular visualization service allowing people to upload, visualize and share their data. But it is not designed for solving specific tasks collaboratively. Sense.us [12] uses collaborative visualization systems to ask many social network users to participate in analyzing data together. Later they make use of crowdsourcing methods for collaborative data analysis and visualization interface design [20].

To our knowledge, there are only few works designed for collaborative visual analytics on network security. Traditional SNS based communication for network analysts can't directly load other findings or provide correlating interaction for deeper insight. FlowTag system [14] addresses the network security analysis by tagging the network flow of the IDS log with parallel coordinates. Analysts upload the textual analytics result to the server while others can download the description of the flow to analyze. Due to the separation of visualization and collaboration, it is hard to generate a big picture collected by all the experts. Our system generates a global overview based on everyone's input, which provides insight into the event correlation.

3. DATA DESCRIPTION AND WORKFLOW

Our system is capable of dealing with netflow data, processed Pcap data and other log data. We use the VAST 2013 Mini Challenge 3 data [2] as a benchmark for the collaboration scenario and the online deployment. The network has around 1200 user PCs and 24 servers divided into three internal parts that were monitored for two weeks (Apr.1-14, 2013). One dataset is network flow data with IPs, ports and transferred bytes, packet number and payloads. Another dataset is the health status log of internal IPs, including CPU load and memory usage, etc. For week two an IPS log is provided which records tear-down, build-up or denial activities. The size of all the processed data is around 16GB.

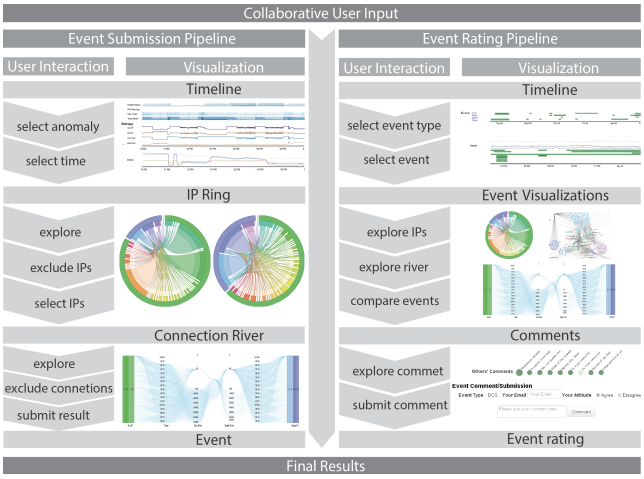

Figure 1: System workflow. Users submit events based on visual exploration as well as evaluating and further exploring others' submission.

Our system OCEANS supports fast visual analytics interaction and online collaboration. It has two integrated workflows (Figure 1). OCEANS is divided into three visualizations on different levels of detail. In the overview the users select a time range, in the ring graph they can choose a group of IPs and finally in the connection river they can examine the detailed situation. Additionally these steps are integrated into an iterative collaboration process with different domain experts using the same system. Users are able to submit events including hints and proofs from the original data. Afterwards others can reload events for verification, comment and further correlation.

4. OCEANS

OCEANS (Figure 9-left) has a timeline, a ring graph and a connection river, integrated into one collaboration platform with a submission page, a commenting panel and an event graph.

4.1 Timeline

As an overview the system provides three different types of timelines (Figure 2). The first part consists of four timelines of accumulated variables visualized as horizon graphs (Figure 2a). These variables are the summed health status and CPU load of internal IPs, the count of warning logs of denials from the IPS log and summed total bytes. Summed health status values and CPU loads provide an indication of payload-intensive attacks, such as DDoS. IPS warning logs of denials indicate the failure of connections, which is usually caused by the scanning behavior, SSH-based attacks or pass-

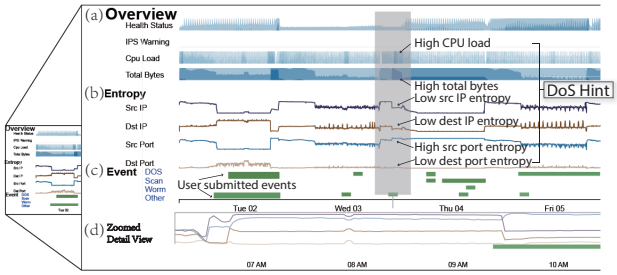

Figure 2: Timeline Overview. (a) accumulated variables timeline. (b) entropy timeline. (c) event timeline. (d) zoomed detail timeline.

word guessing. The summed total bytes indicate the overall network connection behavior. The second part shows four entropy timelines (Figure 2b). Entropy measures the distribution's degree of dispersal or concentration of features. For a given histogram $X = \{n_i, i = 1, ..., N\}$, feature i occurs n_i times in the sample. $S = \sum_{i=1}^{n}(n_i)$ is the total number of the feature observations. H(X) is defined as the following:

$$H(X) = -\sum_{i=1}^{n}(n_i/S)log_2(n_i/S) \quad (1)$$

We have four entropy timelines corresponding to source/destination IP/port (Figure 2b). Research had confirmed that attacks showing distributed or concentrated behavior in IP/port have peak or valley patterns in the entropy [13]. The third part provides a zoom function (Figure 2d). We also combine an event timeline in the overview (Figure 2c), which is discussed in Section 4.4.

The timeline overview allows users to understand the trends and gives visual hints for event detection. For example the detail selection of Apr.3rd indicates a DoS attack, which has a large total bytes transmission, high source port entropy value and low destination IP entropy value. It means large connections targeting very few target IPs and ports (Figure 2). With the hints as starting point, users can brush the interesting time range and drill down to explore the connection behavior.

4.2 Ring Graph

The ring graph shows the connections grouped by subnets within a selected time (Figure 3). Each band of the ring indicates one subnet. The size of the band represents the amount of both inbound and outbound connections of the subnet. The three levels of the ring represent the first three levels of IP hierarchy. External IPs are visualized with colors from green to red and internal IPs from blue to purple. A connection from one band to the other shows the network connection from source IP to destination IP. The color of the connection line indicating the direction is the same as the source IP.

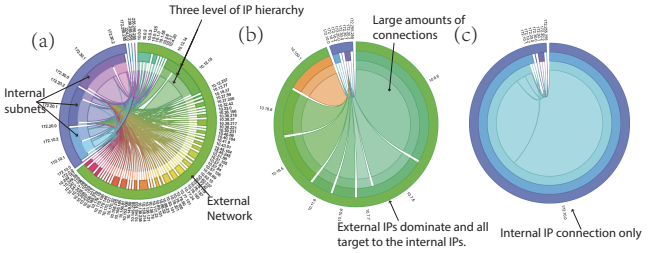

Figure 3: Ring graphs. (a) Normal connections. (b) DDoS Attack. (c) Internet broken.

The ring design gives clear visual hints for subnet connections (Figure 7). Moreover, we provide multi-level exploration by filtering out source or/and destination IP(s) with common behavior. Undo and Redo operation are supported for better exploration. Thus, users can drill down to the detail level and search for the subtle events (Figure 10e).

4.3 Connection River

The connection river shows detailed connection behavior reflected by the heterogeneous data source in a selected time range. The view has six axes. Both source and destination are displayed with IPs (Figure 4-2), connection time (Figure 4-3) and ports (Figure 4-4) as three axes. The improved design of parallel coordinates uses band and hierarchical IP blocks to convey detailed connection information. Each curve band represents a connection flow from source to destination. The vertical height of each curve band encodes the value of the selected connection variable of each connection flow (Figure 4-5). This variable can be selected from the small preview rivers below (Figure 7-1), including duration seconds, payload bytes, total bytes and packet count. Source and destination IP axes are four level hierarchical treemap-like visualizations, whose percentage is aggregated from the corresponding connection curve band height (Figure 4-2). So we can infer the different percentages of variables such as payload and packets number for each IP and subnet. For example, we can detect an file exfiltration event based on the height encoding of large bytes transmissions (Figure 7-1). Multi-level filtering operation is supported, which is similar to the ring graph.

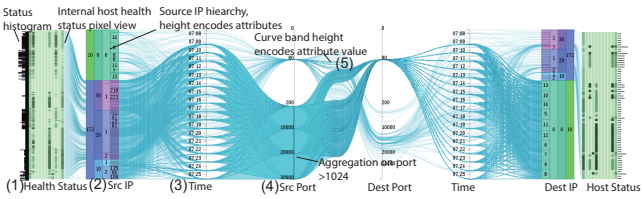

Figure 4: Connection river. It shows the detail of individual network flow, including IP, port, time and network attributes from both source and destination side, as well as health status and IPS logs.

Besides the network flow data, the connection river also conveys the information of the IPS log and health status of internal IPs. For all connections which get alerts of connection denial warnings in the IPS log, red curves are drawn in the background (Figure 7-2). These curves are blocked in the middle to signify the denial. Pixel-based visualization besides the internal IPs shows each accumulated health attribute in a shade of green (Figure 4-1). The darker the shade is, the more health of the hosts reaches a critical status. By clicking a pixel block, the health attribute values will be shown as histogram at the side.

4.4 Collaboration platform

OCEANS provides a web based collaborative interface for domain experts (Figure 5).

4.4.1 Events submission and commenting

Users submit events online and the system automatically extracts the IPs, ports and time ranges from the selection (Figure 6). Users need to tag more information such as an event type and a description and then conduct their sub-

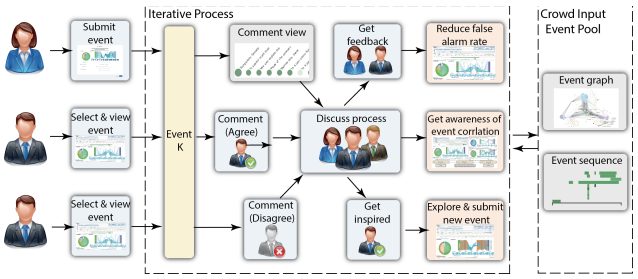

Figure 5: Crowd collaboration in OCEANS. Users submit events while others can view and comment on them. All the crowd input is synthesized into an event graph and event timeline, providing the visual and interactive hints for situation awareness.

missions. All the submitted events construct a event timeline (Figure 2c) and an event graph (Figure 7). The event timeline is positioned side by side with the other timelines, which helps users to identify event features. It also works on two levels of detail. The overview level is divided by attack type, so experts can easily focus on their main field of expertise and differentiate attacks. The height of each rectangle encodes the number of events submitted (Figure 2c). Different shadings of green indicate the certainty of events. Should more experts agree on one event, it becomes more certain and gains a darker green color.

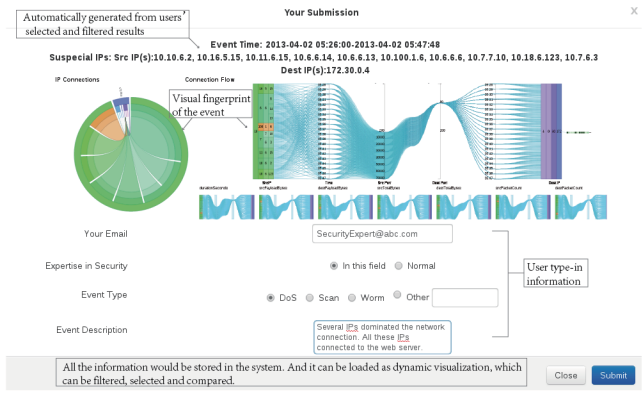

Figure 6: Event submission interface. Users submit both visual features and textual description.

Users can brush the event timeline and click to select an event. After selecting, the ring graph and detail connection river of the specified IP(s) will be loaded. Additionally the user will see the attack type, the initial description and the comments for this event in the commenting panel (Figure 9 Commenting panel). The comments (encoded as green circles) are ordered by time and provide a clear outline of how many people have been working on this event, who agrees (dark green) or disagrees (light green) and how certain the event is. After finding a new verification of the event, the user can enter a comment in the collaboration bar and tag it as agree or disagree. This comment will be shown with all other comments and the user feedback will additively change the certainty value of the event. Thus users with different background can share their findings, discussing and evaluating the events.

4.4.2 Event analysis

Our visual analytics interface works as a collaborative platform. In the event finding process, users explore the event features while viewing other events for comparison or validation. By loading other events, users can operate on the visual scene and apply further filtering and analysis. Our system provides interactive hints to correlate events for supporting users in making use of the crowd-input events. When users select one event in the timeline, other events sharing source or destination IP(s) are highlighted (Figure 9-timeline). Thus, users are able to detect the sequential relationship between different events, or identify multi-phase attacks by suspicious IPs.

The event graph (Figure 7) shows an overview of all submitted events and correlations among suspicious IPs. It is built through synthesis of the crowd input. Each user contributes to submitting and evaluating the events. Each circle represents one IP, which is extracted from the submitted events. The size of the circle is determined by its suspicion score. The score is calculated as the number of events the IP involved, adding the count of agreement on this event and subtracting the count of disagreement. IP(s) that are involved in the events with higher certainty than a threshold are highlighted with text. Internal IPs are positioned in a triangle according to different subnets. Suspicious external IP(s) will be placed next to their connecting internal IP(s). Users can evaluate the seriousness of a suspicious IP and correlated events involved with the IP. With the event graph, users gain a big picture of the network situation, and can focus on special IPs. The event graph also helps security experts viewing the clustering of several IP groups. For example, several IPs launched an DDoS attack onto one of the internal subnets (Figure 7-6).

5. IMPLEMENTATION

OCEANS's front-end is built on HTML5, using d3 [5] and jQuery. Three views load different levels of preprocessed data from the back-end. The timeline overview uses the data aggregated for every minute using a 5-minutes moving window. The ring graph constructs a hierarchy of IP layers based on the subnet. The connection river shows netflow data, IPS logs and health status logs from the server based on the selected time range and IPs. Aggregation will be used for the same IP communicating in a neighboring time range when the fetched records are too many. The users' submission and comments are stored in the center database and updated in others' working page to facilitate the collaboration. Back-end is the PHP Yii framework supported by a MySql database with the indexes based on time and IP. The explorative interaction can be nearly in real time (depending on the selected time range and IPs) with reliable network bandwidth.

6. CASES

We provided our collaboration system for several groups of security analysts and graduate students major in network security. The first case was recorded in the lab study while the second case was recorded from users' online submission. In order to record more details, the third case was based on a field study with three domain experts through our system. The findings of each case were verified by the ground truth of VAST Challenge 2013.

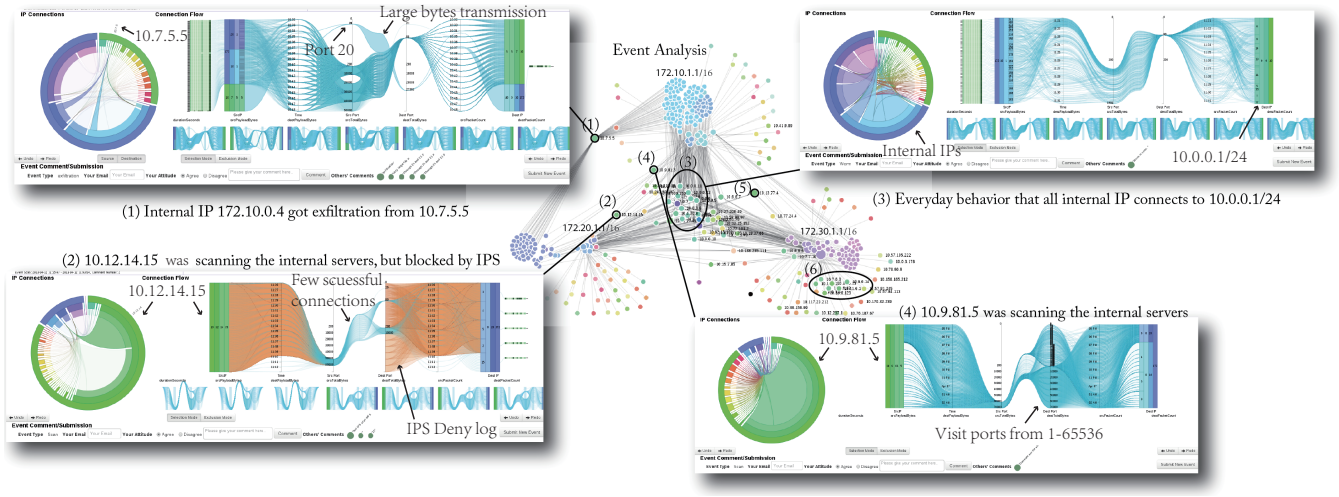

Figure 7: Overall event graph analysis. Global patterns and suspicious IP identified through the crowd input are shown. (1) Large file exfiltration through port 20. (2) Blocked network scan. (3) Clustering connections. (4) Network scan. (5) Botnet C&C identified in Figure 10-E2. (6) DDoS attack (Figure 9-E2).

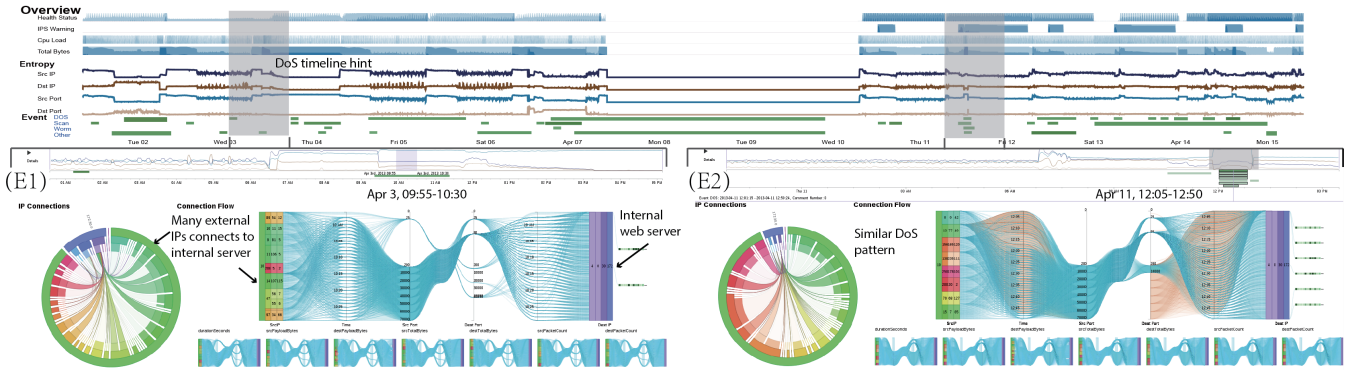

Figure 8: Similarity based inspiration. (a) Overview timeline. (b) A DDoS Event user 1A found. Connections from external IP to internal web servers dominate the network. (c) Another DDoS event found by user 1C, who was inspired by the formal visual pattern.

6.1 Case 1: DDoS Attack

Being familiar with DDoS attack, user 1A decided to search such kind of attack in the data. Firstly, he observed the overview timeline, and targeted several time ranges with high total bytes transmitted in the network. He found that the destination IP and port entropy were low (a few destination IP(s) and port(s) were targeted) while source port entropy was high (large amounts of unique source ports) between 9:30-10:30 a.m., Apr.3 (Figure 8 timeline). By selecting one of the internal subnetwork 172.30.0.0/24 to get the detail view, he found that a lot of external IPs connected to the internal web server IP 172.30.0.4 through destination port 80 (Figure 8E1). So he submitted it as a DDoS attack. Afterwards several other users saw this event and confirmed that it was a DDoS attack. For example, user 1B left a comment that *"I saw some of these IPs do bad things in other situations, eg. 10.9.81.5 had scanned the network."*.

User 1C was new to the system. At first, he played the animation in the timeline to get an overview understanding of the connection behavior. He found the same pattern as the event user 1A submitted it. He recognized the event as a DDoS attack and submitted (Figure 8E2). Based on OCEANS, users can easily understand the events and find new events based on similarity of the visual fingerprint.

6.2 Case 2: Subtle and multi-phase events

Initially user 2A had already detected a DDoS attack starting from 6:00 a.m., Apr.2 and many people commented on it (Figure 9E2). Among these, there was an interesting comment from user 2B. He mentioned the event (Figure 9E1) submitted by user 2C might have a relationship with the DDoS event. E1 is a subtle network scan in Apr.1 found by user 2C. He found out that at the beginning (Apr.1), connections were dominated by the internal connections, while there were only small amounts of external connections. He drilled down into the external connections and found an interesting pattern. At around 11:05 a.m., 10.6.6.6 was scanning the internal server of 172.30.1.0/24 (Figure 9E1).

User 2B checked the event (E1) and made an assumption that it might be a pre-scanning of the internal website before an attack (E2). Many others agreed with him and someone provided strong evidence that the 10.6.6.6 was among the DDoS attacker IPs. User 2A searched for the end period of the DDoS and found that the entropy line changed dramatically. By selecting the same victim IP at around 7:04, Apr.4, he found that 172.30.0.4 didn't respond to any connections for 8 minutes (Figure 9E3). So he correlated these events and summarized a multi-phase attack (Figure 9-bottom). When clicking event E2, users could observe the event E1 and E3

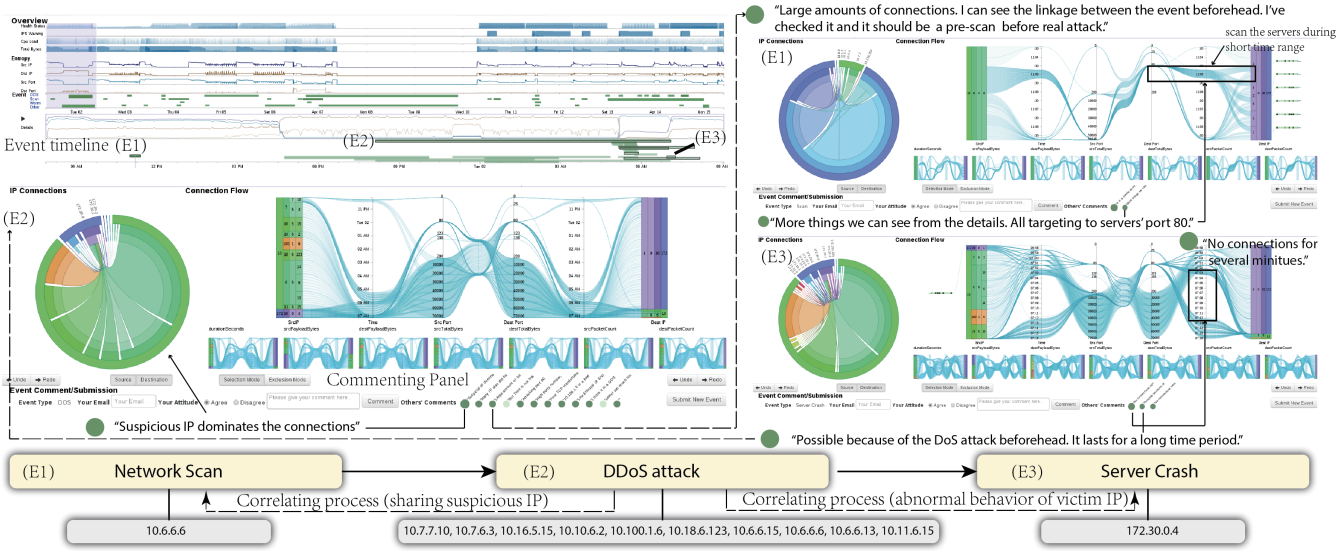

Figure 9: Event correlation and subtle events detection. (E2) DDoS attack with many comments on it. (E1) Pre-scanning (subtle) before the DDoS attack. (E3) Server crash (subtle) after the DDoS attack.

were highlighted with a black stroke in event timeline (Figure 9-Event timeline). The visual hints based on the shared IPs help users to explore the event correlation further.

6.3 Case 3: Botnet Analysis

User 3A identified a DoS-like behavior from the internal IPs in Apr.14. She found that the connection behaviors' pattern was mostly the same as the everyday behavior in the first week (Figure 7-3), but some IPs "pop up". It meant heavier connections and payloads originated from these IPs (Figure 10b). In the following two hours, these IPs kept on connecting to the 10.0.0.0/24 IP through port 80 (Figure 10c). She thought it might be a Botnet DoS attack or distributed patching progress. Due to lack of evidence, she submitted the events with the description and question.

User 3B who checked the event thought there should be "Controllers" if it was a botnet. So he filtered the connection from internal IPs to external IPs and explored the time before the event happened. He found out that eight IPs of 172.20.1.0/24, 172.30.1.0/24, 172.10.2.0/24 were connecting to 10.0.3.77 through port 22 every 10 minutes (Figure 10d). He submitted the new event as Botnet C&C (Command-and-Control) and commented on the original event about what he found. These two events could be evidence for each other and confirmed that these eight internal IPs were controlled as parts of the botnet.

However, the source of botnet was not detected yet. Based on the observation of the two events, user 3C traced to the starting time of the suspicious SSH behavior (Figure 10a-E3). Firstly, he excluded the normal dominating connections which happened everyday. Secondly, by filtering the connections from internal to external IPs (Figure 10e), he found that there were quite short connections from port 80 to external IP 10.4.20.8 before the SSH periodical situation started (Figure 10f). Collecting all the event pieces, analysts could get a bigger picture of the botnet behavior. Firstly, several internal IPs connected to the malicious IP 10.4.20.8 through http and got infected. After being infected, it opened the backdoor and kept SSH connections to the controller 10.0.3.77 and listened to its command. In the following days, the machines started a DDoS attack to the victims of external IP 10.0.0.0/24 (Figure 10g).

Multi-level exploration with filtering is useful for detecting the subtle events. The collaboration platform provides new perspectives for detecting the complex attacks.

7. USER FEEDBACK

We evaluated our system from three parts: VAST Challenge 2013 submission, the lab study and online-deployment.

7.1 VAST Challenge 2013 MC3

We used an initial version of OCEANS without collaboration feature to detect events in the VAST Challenge 2013 Mini Challenge 3 [2]. Our system successfully identified most of the "loud events", such as DDoS, network scan and was recognized as the "Outstanding Situation Awareness" award among 11 submissions [1]. The result confirmed our visual analytics techniques could help domain experts. However, one limitation that our submitted system could't easily find subtle events or multi-phase attacks was addressed by the reviewers. So in this work, we arm our system with more explorative interactions and collaborative features. This results in a better utilization of the human knowledge.

7.2 Lab study

We recruited 16 people (13 male, 3 female) with computer network knowledge to participate in the scenario. After providing a tutorial for 15 minutes, we divided them into two groups randomly to finish the following tasks. Group A used the visual analytics function without collaborative features while Group B could explore the system with full functionality. For each group, we got two people answering questions about the system.

- Task 1 (Guided Exploration): Write down your steps for finding the network breakdown in Apr.14 or 15.
- Task 2 (Detect Suspicious IPs): Find a DoS attack in Apr.2 and describe time and suspicious IPs.
- Task 3 (Event Identification): Find at least one event and write down the exploring steps.

Figure 10: Botnet analysis. (a) Overview timeline. (b,c)-E1, firstly detected events, which turned out to be a botnet DDoS attack. (d)-E2, secondly detected events. By applying filters, user found the suspicious outbounce SSH connections. (e,f)-E3, botnet infection as subtle event. (g) Summarized attack pattern.

In our observation, group A asked more questions in the test. For the time usage, group A used 38.23 minutes while group B used 22.64 minutes on average. We also evaluated all the participants' task accuracy (Figure 11). It proved that our system can be correctly used and help users finding events with visual analytics. With collaborative features, the accuracy is higher and the speed is improved. Afterwards, we provided all of them the questionnaire and got the feedback (Figure 12). Most feedback was positive while some complained about the learning curve and comfortableness. More discussion is in Section 8.

7.3 Crowd-collaboration online deployment

We also conducted one month of live deployment. We sought for people from network security communities to use the system. For easier self-learning goal, we provided basic operation illustration in our system. We also provided a ranking system for their submissions to attract more people participate in it. We received 343 page views from 134 unique visitors over the course of the deployment. They submitted 80 events and gave 239 comments in all. People who participated were including but not limited to security engineers from industry, faculties and graduate students.

Besides the findings illustrated in Section 6, we got much valuable feedback. For example, an engineer from a security company told us: *"The system is useful to view from different aspects, like time, IPs and ports."*. For the collaboration feature, one user said: *"Viewing others' submission made me*

understand the features of network scan and DDoS. I also submitted three events.". A security analyst wrote: *"I viewed the comments on my submitted events, and some mentioned what I didn't think about.".* A faculty member of the security technology department wrote: *"It's good and it is very suitable for me to use the tool as a teaching aid in the network security course. Everyone can participate in detecting events!".* However, some drawbacks were mentioned too. One PhD student in the field said: *"I like the fancy visualization, but it took time to learn to do all the operations.".*

Figure 11: Task accuracy percentage summary.

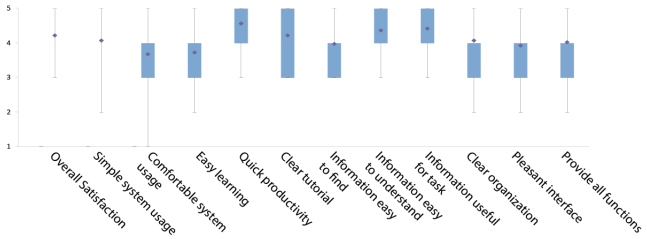

Figure 12: Feedback of the questionnaire.

8. DISCUSSION AND CONCLUSION

Our studies demonstrate that collaborative visual analytics can help domain experts to identify and verify events more efficiently. In the crowd collaboration system, people of different speciality and background knowledge can contribute their ideas and solve a broader range of problems. Compared to the traditional single expert analysis system, the online collaborative visual analytics has advantages: better engagement of people, more discussion of specific events and deeper understanding of correlated features.

We provide a new perspective to detect the complex and subtle events. Everyone's input can be shared and overall analysis would bring a big-picture of the suspicious IPs and correlation of events. After evaluating all the events submitted to the system, we conclude that submitted events with higher certainty turned out correct.

There are some limitations to OCEANS. The complexity of the dataset and the close interrelation of the systems features lead to a slow initial learning curve. We give a tutorial on the website and the possibility to leave questions or feedback inside the tool. We will make the system easier to learn and use in the future. However, our results show that after users got familiar with the tool, he/she could use it efficiently and gain the insights. Another issue is the threshold setting for the synthesis of crowd input: currently the setting is based on our trial to gain highlighting for suspicious IP(s). Statistical models for the suspicious score distribution should be applied for setting the threshold.

9. ACKNOWLEDGEMENTS

The authors wish to thank the reviewers for their comments. The authors also wish to thank IEEE VAST Challenge (2013) for the data and review comments. This work is supported by NSFC No.61170204.

10. REFERENCES

[1] Peking university submission for vast challenge 2013, mc3, http://hcil2.cs.umd.edu/newvarepository/vast

[2] Vast challenge 2013: http://vacommunity.org/vast challenge 2013: Mini-challenge 3.

[3] K. Abdullah, C. Lee, G. Conti, J. A. Copeland, and J. Stasko. Ids rainstorm: Visualizing ids alarms. In *Proc. of VizSec*, 2005.

[4] A. Boschetti, L. Salgarelli, C. Muelder, and K.-L. Ma. Tvi: a visual querying system for network monitoring and anomaly detection. In *Proc. of VizSec*, 2011.

[5] M. Bostock, V. Ogievetsky, and J. Heer. D3 data-driven documents. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2301–2309, 2011.

[6] L. Bunch, J. M. Bradshaw, and M. Vignati. The netflow observatory: An interactive 3-d event visualization. In *Proc. of VizSec*, 2013.

[7] G. J. Conti and K. Abdullah. Passive visual fingerprinting of network attack tools. In *Proc. of ACM CCS*, pages 45–54, 2004.

[8] G. A. Fink, P. Muessig, and C. North. Visual correlation of host processes and network traffic. In *Proc. of VizSec*, pages 11–19, 2005.

[9] S. Foresti, J. Agutter, Y. Livnat, S. Moon, and R. F. Erbacher. Visual correlation of network alerts. *IEEE Comput. Graph. Appl.*, 26:48–59, 2006.

[10] L. Harrison, X. Hu, X. Ying, A. Lu, W. Wang, and X. Wu. Interactive detection of network anomalies via coordinated multiple views. In *Proc. of VizSec*, 2010.

[11] J. Heer and M. Agrawala. Design considerations for collaborative visual analytics. *Information Visualization*, 7(1):49–62, 2008.

[12] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun. ACM*, 52(1):87–97, 2009.

[13] A. Lakhina, M. Crovella, and C. Diot. Mining anomalies using traffic feature distributions. *SIGCOMM Comput. Commun. Rev.*, 35(4), 2005.

[14] C. P. Lee and J. A. Copeland. Flowtag: A collaborative attack-analysis, reporting, and sharing tool for security researchers. In *Proc. of VizSec*, 2006.

[15] A. Luse, K. P. Scheibe, and A. M. Townsend. A component-based framework for visualization of intrusion detection events. *Information Security Journal: A Global Perspective*, 17:95–107, 2008.

[16] ManyEyes. http://www-958.ibm.com/software/analytics/manyeyes/.

[17] A. Shiravi, H. Shiravi, and A. A. Ghorbani. A survey of visualization systems for network security. *IEEE Trans. Vis. Comput. Graph.*, 18(8):1313–1329, 2012.

[18] F. Stoffel, F. Fischer, and D. A. Keim. Finding anomalies in time-series using visual correlation for interactive root cause analysis. In *Proc. of VizSec*, pages 65–72, 2013.

[19] S. T. Teoh, K. liu Ma, and S. F. Wu. A visual exploration process for the analysis of internet routing data. In *Proc. of Vis*, pages 523–530, 2003.

[20] W. Willett, J. Heer, and M. Agrawala. Strategies for crowdsourcing social data analysis. In *Proc. of SIGCHI*, pages 227–236, 2012.