VENT: Visual Exploration of Network Traffic

Final Paper

Moses Schwartz
Sandia National Laboratories and New Mexico Institute of Mining and Technology
Network Forensics, Fall 2009
mdschwa@sandia.gov

ABSTRACT

Manual analysis of packet captures is one of the most time-consuming tasks in network forensic investigations. Tools that support packet capture analysis (e.g., Wireshark) typically present the user with a table or list view of packets. VENT (Visual Analysis of Network Traffic) aims to increase human analyst efficiency by adding a visualization model to the traditional traffic capture analysis paradigm. Initial feedback on this visualization model has been positive, and suggest that this visualization model may be effective for some forensic tasks.

1. INTRODUCTION

Reading packet captures, a common component of a network forensics investigation, is typically a tedious and error-prone process. State-of-the-art analysis tools (e.g., Wireshark[1]) only serve to aid a human analyst, and can rarely be used without human knowledge and intuition to identify anomalous or malicious traffic. Term distribution visualizations of packet capture contents may serve as an effective aid for the analysis of network traffic packet captures.

1.1 Problem Statement

The current state of network forensics is that human analysts must spend a significant amount of time manually viewing and analyzing packet captures (e.g., with Wireshark). This is a very difficult and time-consuming task. The essential threat in this case is that an adversary has a highly asymmetric advantage because launching attacks may be automated, while human analysis is orders of magnitude slower and less reliable.

Malicious traffic may be thoroughly hidden or obfuscated within seemingly-legitimate network traffic. Tools such as intrusion detection systems (IDS) attempt to address this threat by identifying questionable traffic and issuing an alert. IDS do not adequately address this threat because 1) IDS can easily be bypassed by using unpublicized attacks that do not have identified signatures, and 2) in all but the most trivial cases, a human analyst is still required to analyze network traffic manually to determine the validity of an alert.

1.2 Threat Model

VENT aims to simplify the task of the human analyst. It does not attempt to introduce any fundamentally new parsing, dissection, or other analysis capabilities. Rather, it aims to provide a useful visualization that can be used to augment existing tools and simplify the analyst's task.

1.3 Solution Approach

We apply a term distribution visualization model with Focus+Context, presented in [3, 4, 5], to the problem of manual packet capture analysis. In this visualization model, we create a sequential histogram of query terms throughout a dataset and present this information as one of our set of visualization variants. The Focus+Context model consists of a brushed section of the visualization expanded into a new, full-size visualization with finer granularity[2]. Figure 1 provides a simple example of the visualization model. This example is visualizing a short packet capture of conventional web browsing, using the search specification ["protocol" is "HTTP"] and ["info" contains "GET"]. The frequency of HTTP packets is drawn on the histogram in red, and the frequency of packets containing "GET" are drawn in green.

We have demonstrated that the term distribution visualization model is effective for text search tasks [3, 4] and have explored applications for digital forensic string search [5]. This work is an exploration of the application of this visualization model to network traffic.

Throughout this paper we will refer to the research and prototype application as VENT.

2. VISUALIZATION MODEL

The Term Distribution Visualization with Focus+Context model consists of several different histogram variants, a Focus+Context mechanism, and an interaction paradigm that enables dataset exploration using the histogram as scroll bar.

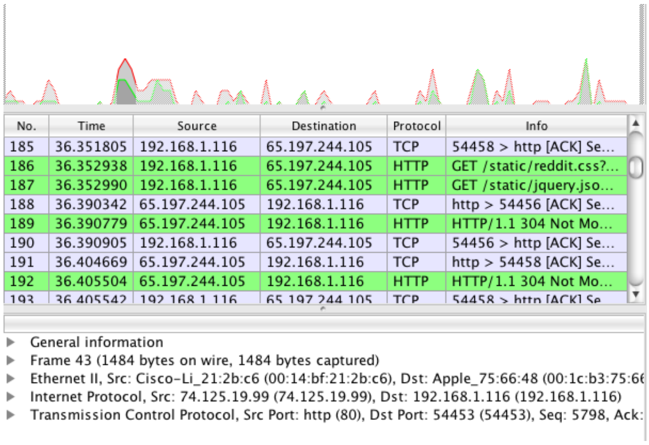
2.1 Histogram Variants

Figure 2 shows the primary histogram variants in this visualization model. From top to bottom, these visualiza-

[1] www.wireshark.org

[2] Although Focus+Context is a significant aspect of the visualization model, it has not yet been implemented in VENT

Network Forensics Fall 2009, New Mexico Tech
Copyright 2009 ACM 978-1-60558-638-0/10/03 ...$10.00.



No.	Time	Source	Destination	Protocol	Info
185	36.351805	192.168.1.116	65.197.244.105	TCP	54458 > http [ACK] Se...
186	36.352938	192.168.1.116	65.197.244.105	HTTP	GET /static/reddit.css?...
187	36.352990	192.168.1.116	65.197.244.105	HTTP	GET /static/jquery.jso...
188	36.390342	65.197.244.105	192.168.1.116	TCP	http > 54456 [ACK] Se...
189	36.390779	65.197.244.105	192.168.1.116	HTTP	HTTP/1.1 304 Not Mo...
190	36.390905	192.168.1.116	65.197.244.105	TCP	54456 > http [ACK] Se...
191	36.404669	65.197.244.105	192.168.1.116	TCP	http > 54458 [ACK] Se...
192	36.405504	65.197.244.105	192.168.1.116	HTTP	HTTP/1.1 304 Not Mo...
193	36.405542	192.168.1.116	65.197.244.105	TCP	54458 > http [ACK] Se

- General information
- Frame 43 (1484 bytes on wire, 1484 bytes captured)
- Ethernet II, Src: Cisco-Li_21:2b:c6 (00:14:bf:21:2b:c6), Dst: Apple_75:66:48 (00:1c:b3:75:66
- Internet Protocol, Src: 74.125.19.99 (74.125.19.99), Dst: 192.168.1.116 (192.168.1.116)
- Transmission Control Protocol, Src Port: http (80), Dst Port: 54453 (54453), Seq: 5798, Ack:

Figure 1: An example of our visualization model showing short packet capture of conventional web browsing, using the search specification [“protocol” is “HTTP”] and [“info” contains “GET”]. The frequency of HTTP packets is drawn on the histogram in red, and the frequency of packets containing “GET” are drawn in green.

tion variants are referred to as TileBars, Greyscale Histograms, Color Histograms, Color-Line Histograms, Line-Fill Histograms. In all of these examples, we deviate from the network forensic arena and show the visualizations as used on plain-text. In the examples, we visualize Lewis Carrol’s *Through the Looking Glass* with search terms “Alice,” “Humpty,” “Tweedledum,” and “Tweedledee.”

2.1.1 TileBars

TileBars are based on [2]. TileBars are matrices of tiles, with darkness and color blending representing the frequency of search terms.

2.1.2 Greyscale Histograms

Greyscale Histograms show term frequency as a set of overlapped greyscale histograms. The darker areas draw attention to areas with term overlap.

2.1.3 Color Histograms

Color Histograms show terms as differently colored histograms blended together to show overlap.

2.1.4 Color-Line Histograms

Color-Line Histograms simply draws each term’s histogram as a colored line.

2.1.5 Line-Fill Histograms

Line-Fill Histograms combine Color-Line Histograms with the Greyscale Histograms fill. Currently, only Line-Fill histograms are supported in VENT.

2.2 Focus+Context

Figure 3 shows the Focus+Context and interaction model. In this example, we search for the scene in which Alice tells Humpty Dumpty how old she is. We brush and drill-down twice to get to a highly granular view, then move a selection window across the lowermost visualization. The text in the box below corresponds directly to the selected part of the visualization.

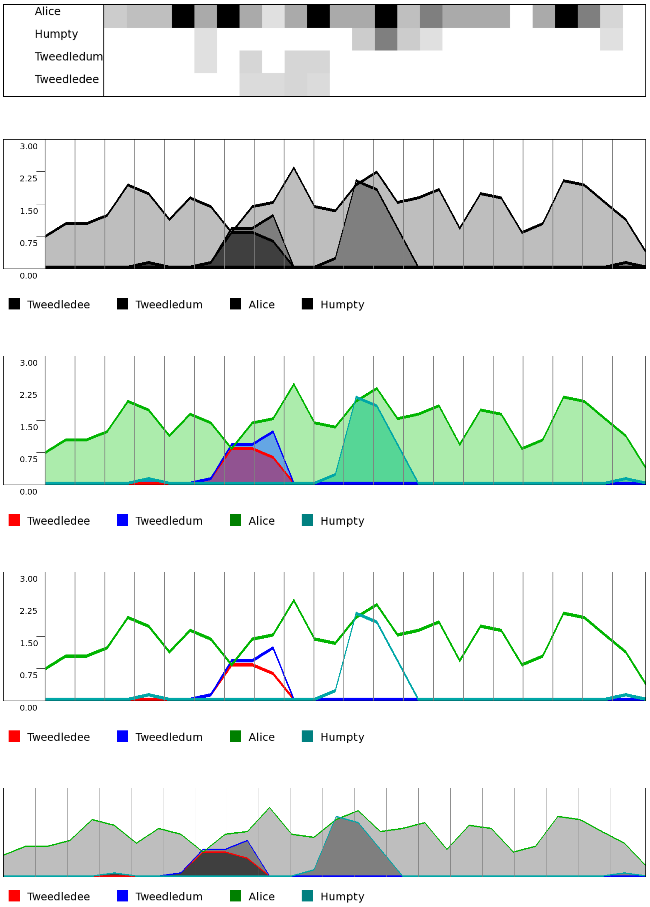


Figure 2: Examples of the primary visualization variants in the model visualizing Lewis Carrol’s *Through the Looking Glass* [1] with search terms “Alice,” “Humpty,” “Tweedledum,” and “Tweedledee.”. From top to bottom, these visualization variants are referred to as: TileBars, Greyscale Histograms, Color Histograms, Color-Line Histograms, Line-Fill Histograms. Currently, only Line-Fill histograms are supported in VENT.

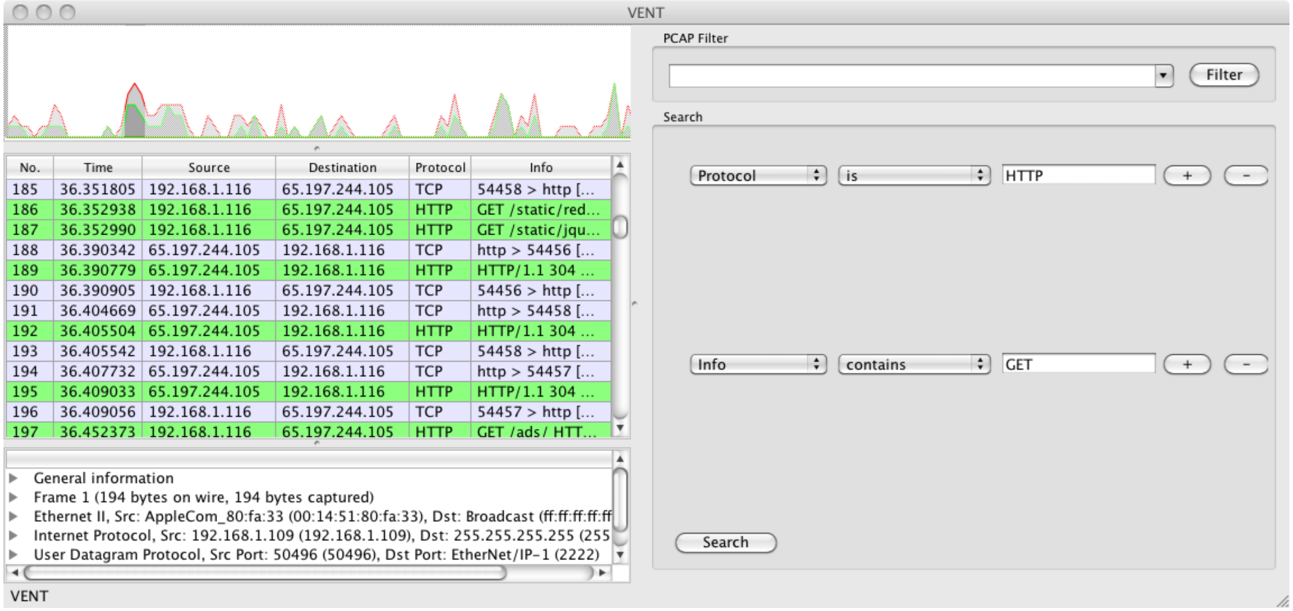
3. VENT STATUS

We have developed a prototype application to showcase VENT and solicited feedback from a professional in the field. Although there are some variations from the proposal, VENT technically fulfills its goals.

Figure 4 shows the VENT application interface.

3.1 Capabilities

VENT can display any PCAP file that Wireshark can parse. The primary view of the packet summary and details is nearly identical to Wireshark’s, although missing advanced functionality. The histogram in the upper-most pane can be used to scroll through the packet capture. Search terms are specified in the right pane. There is an arbitrary limit of 5 terms; while this may be trivially increased, usabil-



No.	Time	Source	Destination	Protocol	Info
185	36.351805	192.168.1.116	65.197.244.105	TCP	54458 > http [...
186	36.352938	192.168.1.116	65.197.244.105	HTTP	GET /static/red...
187	36.352990	192.168.1.116	65.197.244.105	HTTP	GET /static/jqu...
188	36.390342	65.197.244.105	192.168.1.116	TCP	http > 54456 [...
189	36.390779	65.197.244.105	192.168.1.116	HTTP	HTTP/1.1 304 ...
190	36.390905	192.168.1.116	65.197.244.105	TCP	54456 > http [...
191	36.404669	65.197.244.105	192.168.1.116	TCP	http > 54458 [...
192	36.405504	65.197.244.105	192.168.1.116	HTTP	HTTP/1.1 304 ...
193	36.405542	192.168.1.116	65.197.244.105	TCP	54458 > http [...
194	36.407732	65.197.244.105	192.168.1.116	TCP	http > 54457 [...
195	36.409033	65.197.244.105	192.168.1.116	HTTP	HTTP/1.1 304 ...
196	36.409056	192.168.1.116	65.197.244.105	TCP	54457 > http [...
197	36.452373	192.168.1.116	65.197.244.105	HTTP	GET /ads/ HTT...

```
▶ General information
▶ Frame 1 (194 bytes on wire, 194 bytes captured)
▶ Ethernet II, Src: AppleCom_80:fa:33 (00:14:51:80:fa:33), Dst: Broadcast (ff:ff:ff:ff:ff
▶ Internet Protocol, Src: 192.168.1.109 (192.168.1.109), Dst: 255.255.255.255 (255
▶ User Datagram Protocol, Src Port: 50496 (50496), Dst Port: EtherNet/IP-1 (2222)
```

Protocol | is | HTTP

Info | contains | GET

Figure 4: The VENT interface. The primary view of the packet summary and details is nearly identical to Wireshark's, although missing advanced functionality. The histogram in the upper-most pane can be used to scroll through the packet capture. Search terms are specified in the right pane.

ity of these visualizations typically decreases significantly with more terms.

3.2 Implementation Status

We have completed the first phase of the initial prototype development. While this prototype has limited capabilities and can only view a hard-coded PCAP file, it serves to demonstrate how the visualization model may work for packet capture analysis. Further development will add more functionality to the application.

3.3 Implementation Details

VENT is programmed entirely in Python. The user interface uses the Qt framework, accessed via the PyQt bindings. Wireshark's command line interface (tshark) is invoked on the back-end for PCAP analysis and packet dissection; its XML output is parsed and displayed in VENT.

3.4 User Evaluation

Figure 6 in the appendix shows the questionnaire used to obtain user feedback.

In the questionnaire, subjects are asked to circle one of "Strongly disagree", "Disagree", "Neutral", "Agree", "Strongly agree" as their response to the following statements:

- This application and visualization model looks useful.
- The visualization made it easier to get an overall view of the packet capture.
- I would like to see further research into these types of visualizations.

Subjects are also asked to write a free-form response to the following questions:

- Do you think a visualization like this would be useful for real forensic investigations? Why or why not?
- Do you have any suggestions for improvements?
- Do you have any other comments?

3.5 User Feedback

We solicited feedback about VENT from one information security professional familiar with network forensics. The subject's responses to the multiple choice section were:

- This application and visualization model looks useful: **AGREE**
- The visualization made it easier to get an overall view of the packet capture: **AGREE**
- I would like to see further research into these types of visualizations: **STRONGLY AGREE**

To the other questions, the subject's responses were:

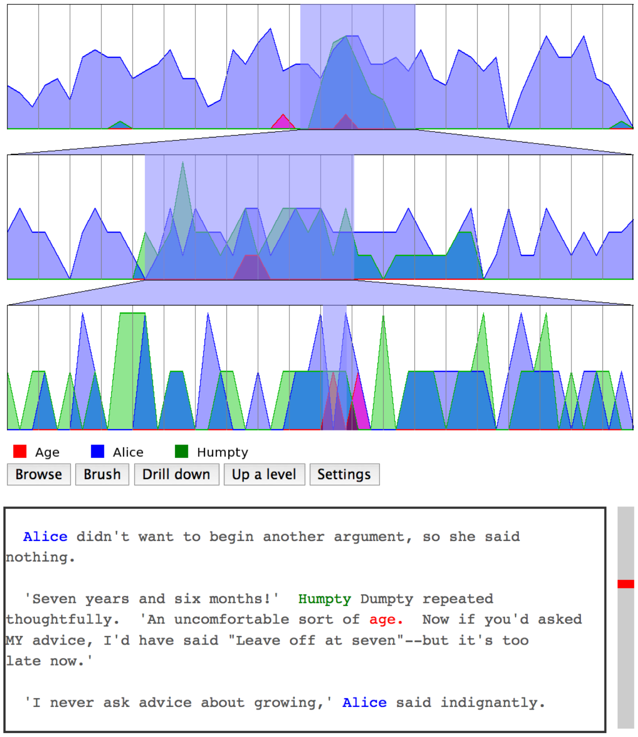
- Do you think a visualization like this would be useful for real forensic investigations? Why or why not?

"Yes, one can get an instant idea of the most frequently occuring IP addresses, and/or protocols etc. in the traffic being investigated. This helps in identifying specific areas of interest in the investigation process faster."

- Do you have any suggestions for improvements?

"1) A drill-down feature can be useful to focus on specific packets of interest, 2) If a user finds a packet of specific interest, a feature that allows to build the session that packet is associated to can be useful, 3) Using the tool on near real time/live traffic"

The subject did not have any other comments.



Alice didn't want to begin another argument, so she said nothing.

'Seven years and six months!' Humpty Dumpty repeated thoughtfully. 'An uncomfortable sort of age. Now if you'd asked MY advice, I'd have said "Leave off at seven"--but it's too late now.'

'I never ask advice about growing,' Alice said indignantly.

Figure 3: The visualization model showing the use of Focus+Context to identify a particular section of *Through the Looking Glass*: Alice telling Humpty Dumpty how old she is. We brush and drill-down twice to get to a highly granular view, then move a selection window across the lowermost visualization. The text in the box below corresponds directly to the selected part of the visualization.

3.6 Variation From Proposal

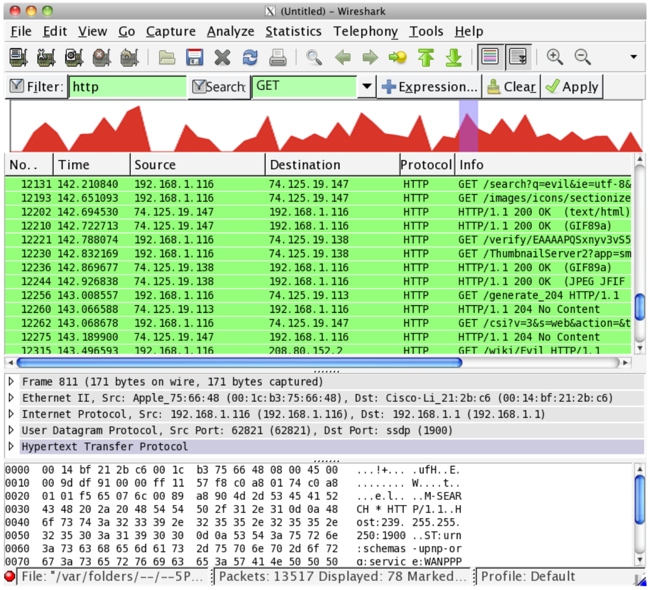
The project proposal listed two final deliverables:

1. a prototype implementation of a term distribution visualization for network traffic, and
2. a report detailing the challenges faced, solutions, and a brief user evaluation suitable for publication in an appropriate venue.

The proposal also stated that the effectiveness of the visualization model will be tested with a small-scale, subjective user study. This study was to include students in this network forensics class, as well as input from professionals in the field. This aspect has been scaled back due to the longer-than-expected time to implement the tool. Feedback from one professional was obtained, and a form to record other subject's feedback has been created.

For the prototype implementation, the proposal specified that: *The prototype implementation will support at least one protocol at each effective layer of the OSI model (i.e., Ethernet->IP/TCP->HTTP), though it may support more if such capabilities can be borrowed from another library. It will support some simple filters. The interface will include an interactive visualization, though Focus+Context may not be supported.* The final deliverable supports far more protocols, thanks to the use of Wireshark. The interactive visualization, though not fully polished, is implemented and functional. Focus+Context was not implemented.

Figure 5 shows the proposal's mockup of the VENT interface. Comparison to the implemented tool, shown in Figure 4, demonstrates that the final product is very similar in concept and execution to the mockup.



Filter: http
Search: GET

```
No. .  Time        Source          Destination     Protocol Info
 12131 142.210840  192.168.1.116   74.125.19.147   HTTP     GET /search?q=evil&ie=utf-8&
 12193 142.651093  192.168.1.116   74.125.19.147   HTTP     GET /images/icons/sectionize
 12202 142.694530  74.125.19.147   192.168.1.116   HTTP     HTTP/1.1 200 OK  (text/html)
 12210 142.722713  74.125.19.147   192.168.1.116   HTTP     HTTP/1.1 200 OK  (GIF89a)
 12221 142.788074  192.168.1.116   74.125.19.138   HTTP     GET /verify/EAAAAPQSxnyv3vS5
 12230 142.832169  192.168.1.116   74.125.19.138   HTTP     GET /ThumbnailServer2?app=sm
 12236 142.869677  74.125.19.138   192.168.1.116   HTTP     HTTP/1.1 200 OK  (GIF89a)
 12244 142.926838  74.125.19.138   192.168.1.116   HTTP     HTTP/1.1 200 OK  (JPEG JFIF
 12256 143.008557  192.168.1.116   74.125.19.113   HTTP     GET /generate_204 HTTP/1.1
 12260 143.066588  74.125.19.113   192.168.1.116   HTTP     HTTP/1.1 204 No Content
 12262 143.068678  192.168.1.116   74.125.19.147   HTTP     GET /csi?v=3&s=web&action=&t
 12275 143.189900  74.125.19.147   192.168.1.116   HTTP     HTTP/1.1 204 No Content
 12315 143.496593  192.168.1.116   208.80.152.2    HTTP     GET /wiki/Evil HTTP/1.1
```

```
▷ Frame 811 (171 bytes on wire, 171 bytes captured)
▷ Ethernet II, Src: Apple_75:66:48 (00:1c:b3:75:66:48), Dst: Cisco-Li_21:2b:c6 (00:14:bf:21:2b:c6)
▷ Internet Protocol, Src: 192.168.1.116 (192.168.1.116), Dst: 192.168.1.1 (192.168.1.1)
▷ User Datagram Protocol, Src Port: 62821 (62821), Dst Port: ssdp (1900)
▷ Hypertext Transfer Protocol
```

```
0000  00 14 bf 21 2b c6 00 1c  b3 75 66 48 08 00 45 00   ...!+... .ufH..E.
0010  00 9d df 91 00 00 ff 11  57 f8 c0 a8 01 74 c0 a8   ........ W....t..
0020  01 01 f5 65 07 6c 00 89  a8 90 4d 2d 53 45 41 52   ...e.l.. ..M-SEAR
0030  43 48 20 2a 20 48 54 54  50 2f 31 2e 31 0d 0a 48   CH * HTT P/1.1..H
0040  6f 73 74 3a 32 33 39 2e  32 35 35 2e 32 35 35 2e   ost:239. 255.255.
0050  32 35 30 3a 31 39 30 30  0d 0a 53 54 3a 75 72 6e   250:1900 ..ST:urn
0060  3a 73 63 68 65 6d 61 73  2d 75 70 6e 70 2d 6f 72   :schemas -upnp-or
0070  67 3a 73 65 72 76 69 63  65 3a 57 41 4e 50 50 50   g:servic e:WANPPP
```

File: "/var/folders/--/--5P... | Packets: 13517 Displayed: 78 Marked... | Profile: Default

Figure 5: Initial mockup of the visualization interface.

4. ANALYSIS

4.1 Forensic Impact

The ultimate goal for VENT is to create a highly effective visualization to simplify manual packet capture analysis. If prototypes show that the visualization techniques are valuable and useful to analysts, we will attempt to develop a fully-featured application or integrate the visualizations into an existing tool (such as Wireshark). Figure 4 shows the prototype implementation of VENT.

If effective, the VENT model may have a significant impact on forensic analysis of traffic captures. It will reduce the time required for manual analysis of traffic captures, and may make it easier to identify trends or anomalies that might have been missed otherwise.

Existing tools have relatively few effective visualization capabilities for low-level packet or protocol analysis. VENT represents a departure from the current interaction paradigm by providing useful visualizations. As VENT matures, more visualization techniques may be integrated: multiple views (multiple visualizations as well as text-based) with seamless brushing and filtering may be a very powerful platform for network forensic tasks.

4.2 Usability

User feedback—both as recorded in previous sections and verbally delivered to the researchers—suggests that there may be significant interest in a more fully-featured implementation of VENT. Even with the minimal functionality available in the prototype, the subject was able to operate the interface with little training and reported that it would likely be useful in real situations.

5. FUTURE WORK

One of the primary results from the research and development of VENT was simply the identification of avenues for future work.

5.1 Focus+Context

Implementing Focus+Context is a top priority for future work.

5.2 Visualizing Dataset Reduction

Applying PCAP filters to a dataset—essentially dataset reduction—is a natural (and simple) extension to VENT's current capabilities. However, *visualizing* that reduction to maintain context is an interesting problem. A variation on the existing Focus+Context model may be useful to create such a visualization.

5.3 Multiple Views

Multiple visualizations of a packet capture tightly coupled with shared brushing and filtering is an ultimate goal for VENT. The Model/View architecture makes implementing such a capability feasible.

5.4 Modular Visualizations

In conjunction with Multiple Views, the ability to rapidly develop new visualizations and load them as plugins or modules is a long-term goal.

5.5 Improved Wireshark Interface

The most significant issue making advanced features in VENT difficult to implement is the interface to Wireshark. Direct access to the functions and data structures in libwireshark will be eventually needed. This is a technically challenging task simple because of the size and complexity of the Wireshark codebase; however, completing such a project would have enormous benefits for the network analysis community.

6. CONCLUSIONS

Although still an early prototype, VENT may ultimately have a significant impact on forensic analysis of network traffic captures. While it does not introduce fundamentally new analysis capabilities, it does augment some existing capabilities by delivering more information to the analyst. This enables a “bird's-eye” view of the dataset, while still allowing low-level packet analysis. Initial user feedback is promising, though it highlights the need for continued development.

7. REFERENCES

[1] L. Carroll. *Through the Looking Glass.* Project Gutenberg, 1991.

[2] M. A. Hearst. Tilebars: visualization of term distribution information in full text information access. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 59–66, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

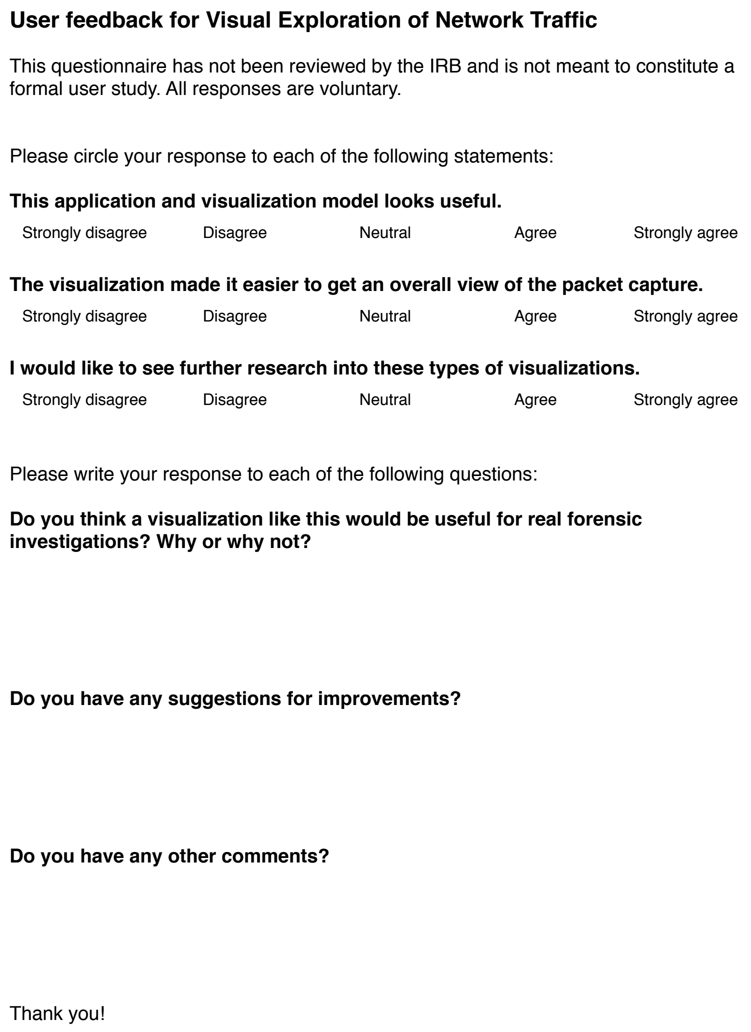
[3] M. Schwartz. Term distribution visualizations: Usability evaluation and application to digital forensic string search. Master's thesis, New Mexico Institute of Mining and Technology, 2008.

[4] M. Schwartz, C. Hash, and L. M. Liebrock. Term distribution visualizations with focus+context. In *SAC '09: Proceedings of the 2009 ACM Symposium on Applied Computing*, pages 1792–1799, New York, NY, USA, 2009. ACM.

[5] M. Schwartz and L. M. Liebrock. A term distribution visualization approach to digital forensic string search. In *Proceedings of VizSEC 2008: Visualization for Computer Security, 5th International Workshop*, Lecture Notes in Computer Science, pages 36–43. Springer Berlin / Heidelberg, 2008.

APPENDIX

Questionnaire



User feedback for Visual Exploration of Network Traffic

This questionnaire has not been reviewed by the IRB and is not meant to constitute a formal user study. All responses are voluntary.

Please circle your response to each of the following statements:

This application and visualization model looks useful.

Strongly disagree Disagree Neutral Agree Strongly agree

The visualization made it easier to get an overall view of the packet capture.

Strongly disagree Disagree Neutral Agree Strongly agree

I would like to see further research into these types of visualizations.

Strongly disagree Disagree Neutral Agree Strongly agree

Please write your response to each of the following questions:

Do you think a visualization like this would be useful for real forensic investigations? Why or why not?

Do you have any suggestions for improvements?

Do you have any other comments?

Thank you!

Figure 6: Questionnaire used to record user feedback