Network Visualization

Alex Bradley

CPSC 533C
University of British Columbia

November 30, 2009

Outline

1. Introduction
2. SeeNet: Phone (and other) networks
3. Rainstorm/Rumint: IP network security
4. OverFlow: IP network analysis/security
5. Conclusion

Outline

1. Introduction
2. SeeNet: Phone (and other) networks
3. Rainstorm/Rumint: IP network security
4. OverFlow: IP network analysis/security
5. Conclusion

Introduction

- General theme: use visualization to assimilate complexity of massive volumes of data describing network traffic
- Three papers:
  - **SeeNet: Phone (and other) networks**
    R.A. Becker, S.G. Eick, and A.R. Wilks. Visualizing Network Data. IEEE TVCG, 1995. (See also: video)
  - **Rainstorm/Rumint: IP network security**
    G. Conti, K. Abdullah, J. Grizzard, J. Stasko, J. Copeland, M. Ahamad, H. Owen and C. Lee. Countering Security Information Overload Through Alert and Packet Visualization. IEEE CG&A, 2006.
  - **OverFlow: IP network analysis/security**
    J. Glanfield, S. Brooks, T. Taylor, D. Paterson, C. Smith, C. Gates, J. McHugh. OverFlow: An Overview Visualization for Network Analysis. VizSec 2009.

Outline

1. Introduction
2. SeeNet: Phone (and other) networks
   - Overview
   - Techniques
   - Interaction
   - Critique
3. Rainstorm/Rumint: IP network security
4. OverFlow: IP network analysis/security
5. Conclusion

Visualizing Network Data

Richard A. Becker, Stephen G. Eick and Allan R. Wilks (AT&T Bell Labs)

- Goal: understand data about (telephone) network performance
- Contribution: SeeNet, a tool implementing new techniques to help network analysts cope with information overload
  - Scalability to handle larger networks and ever-increasing data volumes is important
- Three visualization techniques:
  - Link maps
  - Node maps
  - Matrix display
- Extensive support for interactive generation of visualizations
- Animation support for viewing evolution of data over time

Illustrative Example

- Tools demonstrated using AT&T long distance telephone activity on October 17, 1989 (date of Loma Prieta earthquake)
  - Magnitude 7.0 earthquake in San Francisco Bay Area

(Image credit: J.K. Nakata, U.S. Geological Survey)

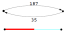
  - Coincided with 1989 World Series game, so broadcast on national TV
  - Unsurprisingly, subsequent high load on long-distance telephone network

Illustrative Example

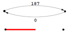
- Questions of interest to analyst in disaster scenario:
  - Where are the overloads?
  - Which links are carrying the most traffic?
  - Was there network damage?
  - Are there any pockets of underutilized network capacity?
  - Is the overload increasing or decreasing?
  - Are calls into the affected area completing or are they being blocked elsewhere in the network?

Link Maps

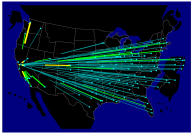
- Display data as node/link graph overlaid on map
- Link statistic value encoded through colour and line thickness
- Directed statistics can be merged into single half-line between nodes:



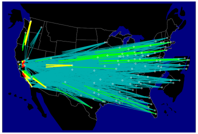
- If one value is zero, half of line may not be drawn:



Link Map (traffic to/from Oakland)



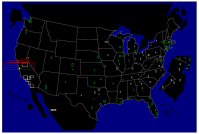
Link Map (traffic between all nodes)



Node Maps

- Link maps become cluttered if "too many", "say more than 10%" of $n^2/2$ possible links between n nodes active
- Node maps display node-oriented data through a glyph at each node
- Loses detailed information about particular links
- In next example, glyph is rectangle
  - width $\propto \sqrt{\#\ \textit{inbound calls}}$
  - height $\propto \sqrt{\#\ \textit{outbound calls}}$
  - area $\propto$ total call volume

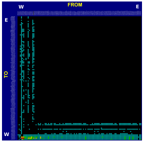
Node Map



Matrix View

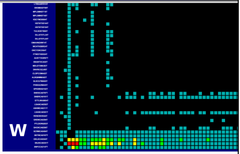
- Problems with geographical view:
  - Long lines have undue prominence
  - Clutter can obscure patterns
- Alternative: matrix view
  - Strength: solves problems above
  - Weakness: loses geographic information, poor choice of row/column order may obscure patterns

Matrix View



("W" and "E" annotations mine)

Matrix View



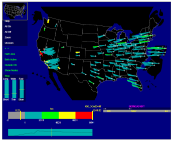
Interaction Parameters

- Interactive interface permits adjustment of many parameters with real-time graphical response:
  - Statistic displayed (e.g. absolute overload, % overload)
  - Levels (select what range of statistic is displayed)
  - Geography/topology (zoom to region, activate/deactivate nodes by location)
  - Time period displayed
  - Size of glyphs/width and length of lines displayed
  - Colour scheme (note that network data often has skewed distribution, so some care needed)
- Generation of aggregate statistics for regions/sets of nodes also important feature, but not currently supported dynamically

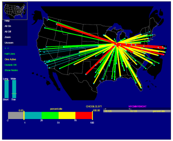
Notable Interaction Techniques

- Animation (show evolution of data over time)
- Zooming
  - Overview+detail approach: "bird's eye view" at top left
  - Which lines to display when zoomed in? (see following slides)
- Conditioning (using double-edged slider to filter out links whose statistics are not in a selected range)
- Sound (mark passage of time in animation, also some UI actions)

Interactive Adjustment: Shorten Lines

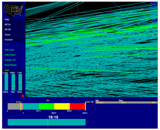


Interactive Adjustment: View Capacity To/From Chicago



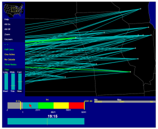
Interactive Adjustment: Zoomed View Display Choices

All lines passing through an area:



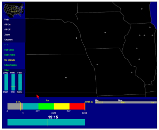
Interactive Adjustment: Zoomed View Display Choices

All lines terminating in area:



Interactive Adjustment: Zoomed View Display Choices

All lines originating and terminating in area:



Critique

- Strengths
  - Rich set of tools for interactive control of visualization
  - Tool was developed for real-world tasks in consultation with users
  - Good use of running example to show importance, usefulness of tool
- Weaknesses
  - In link map half-lines, can be difficult to find one end of a line if flow is zero
    - Would be even worse if attempted to generalize to arcs
  - Text labels in matrix view nearly illegible for large number of nodes
    - Perhaps matrix should have filtering feature?
  - No formal usability study; if user consultation led to interesting usability results or changes in interface, this isn't reported
    - Would "rules of thumb" (e.g., clutter = more than ∼10% of nodes linked) be supported by user trials?

Critique

- Strengths
  - Rich set of tools for interactive control of visualization
  - Tool was developed for real-world tasks in consultation with users
  - Good use of running example to show importance, usefulness of tool
- Weaknesses
  - In link map half-lines, can be difficult to find one end of a line if flow is zero
    - Would be even worse if attempted to generalize to arcs
  - Text labels in matrix view nearly illegible for large number of nodes
    - Perhaps matrix should have filtering feature?
  - No formal usability study; if user consultation led to interesting usability results or changes in interface, this isn't reported
    - Would "rules of thumb" (e.g., clutter = more than ∼10% of nodes linked) be supported by user trials?

Outline

1. Introduction
2. SeeNet: Phone (and other) networks
3. Rainstorm/Rumint: IP network security
   - Overview
   - Design Framework
   - Rainstorm
   - Rumint
   - Critique
4. OverFlow: IP network analysis/security
5. Conclusion

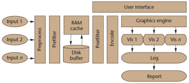
Countering Security Information Overload through Alert and Packet Visualization

G. Conti, K. Abdullah, J. Grizzard, J. Stasko, J. Copeland, M. Ahamad, H. Owen and C. Lee

- Domain: IT security
- Goal: reduce information overload on system administrators due to massive numbers of security alerts
- Rainstorm IDS: high-level view of network security alert activity
- Rumint: more detailed packet analysis
  - May still use existing tools like Ethereal (Wireshark) for really detailed analysis

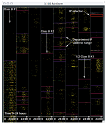
Design Framework

- Proposes framework for design of security visualization systems derived from authors' experiences developing six such systems


Input 1
Input 2
Input *n*
Preprocess
Prefilter
RAM cache
Disk buffer
Postfilter
Encode
User interface
Graphics engine
Vis 1
Vis 2
Vis *n*
Log
Report

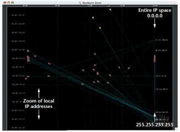
Rainstorm: Network Alert Overview

- Security alerts over 24-hour period for 163,840 IP addresses (GA Tech network). One pixel row = 20 addresses.



Rainstorm: Zoomed View

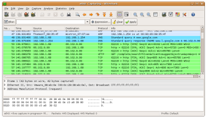
- Examine activity for a selected IP range and time span



Rumint: Detailed Packet Analysis

- Seven visualizations for detailed analysis of network packets
  - scrolling text (printable ASCII text in packets, one packet per row)
  - parallel coordinates
  - glyph-based animation
  - binary rainfall visualization
  - byte frequency display
  - detail display (hex/ASCII packet contents)
  - thumbnail toolbar
- "Personal video recorder" interface: "record" packets from a live capture for later playback

Previous packet analysis tool: Ethereal (now Wireshark)

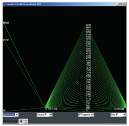

eth0: Capturing - Wireshark
File Edit View Go Capture Analyze Statistics Help
Filter:
Expression...
Clear
Apply

```
No. Time       Source          Destination     Protocol Info
47 139.931463  ThomsonT_08:35:4f Wistron_07:07:ee ARP   192.168.1.254 is at 00:90:d0:08:35:4f
48 139.931466  192.168.1.68    192.168.1.254   DNS      Standard query A www.google.com
49 139.975406  192.168.1.254   192.168.1.68    DNS      Standard query response CNAME www.l.google.com A 66.102.9.99
50 139.976811  192.168.1.68    66.102.9.99     TCP      62216 > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=2
51 140.079578  66.102.9.99     192.168.1.68    TCP      http > 62216 [SYN, ACK] Seq=0 Ack=1 Win=5720 Len=0 MSS=1430
52 140.079583  192.168.1.68    66.102.9.99     TCP      62216 > http [ACK] Seq=1 Ack=1 Win=65780 Len=0
53 140.080278  192.168.1.68    66.102.9.99     HTTP     GET /complete/search?hl=en&client=suggest&js=true&q=m&cp=1 H
54 140.086765  192.168.1.68    66.102.9.99     TCP      62216 > http [FIN, ACK] Seq=805 Ack=1 Win=65780 Len=0
55 140.086921  192.168.1.68    66.102.9.99     TCP      62218 > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=2
56 140.197484  66.102.9.99     192.168.1.68    TCP      http > 62216 [ACK] Seq=1 Ack=805 Win=7360 Len=0
57 140.197777  66.102.9.99     192.168.1.68    TCP      http > 62216 [FIN, ACK] Seq=1 Ack=806 Win=7360 Len=0
58 140.197811  192.168.1.68    66.102.9.99     TCP      62216 > http [ACK] Seq=806 Ack=2 Win=65780 Len=0

▷ Frame 1 (42 bytes on wire, 42 bytes captured)
▷ Ethernet II, Src: Vmware_38:eb:0e (00:0c:29:38:eb:0e), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
▷ Address Resolution Protocol (request)

0000  ff ff ff ff ff ff 00 0c  29 38 eb 0e 08 06 00 01   ........ )8......
0010  08 00 06 04 00 01 00 0c  29 38 eb 0e c0 a8 39 80   ........ )8....9.
0020  00 00 00 00 00 00 c0 a8  39 02                     ........ 9.

eth0: <live capture in progress> Fil...   Packets: 445 Displayed: 445 Marked: 0   Profile: Default
```

(Public domain image from Wikimedia Commons)

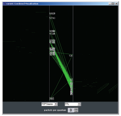
Parallel Coordinates

- View packets as multidimensional data (up to 19 dimensions)



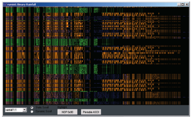
Glyph-Based Animation

- Animation of traffic based on two chosen header fields of packets
- Center pane: 2-axis parallel coords
- Left and right panes: moving glyphs showing traffic



Rainfall Visualization

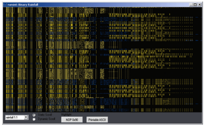
- SeeSoft-like: One packet per packet line, lines ordered by time



🟩 Transmission Control Protocol
🟧 User Datagram Protocol
🟦 Internet Control Message Protocol

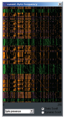
Rainfall Visualization: colour-blindness check

- Rainfall screenshot in Vischeck protanope simulation



Byte Frequency Visualization

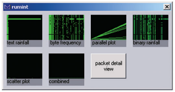
- 256 pixel columns, one packet per line
- pixel indicates byte's presence/frequency in packet



🟩 Transmission Control Protocol
🟧 User Datagram Protocol
🟦 Internet Control Message Protocol

Thumbnail Toolbar

- Overview of other visualizations; can click icons to bring one up



Critique

- Strengths
  - rich feature set developed based on some consultation with real users
  - informal user evaluations with good results
    - users liked Rumint, and particularly liked the "video recorder" idea
    - users made some good suggestions to improve RainStorm
- Weaknesses
  - "design framework" only loosely linked to rest of paper
  - weak usability claims
    - E.g.: "For [RainStorm] to be used on the network, system administrators will have to learn how to use it, how to interpret the display, and what the visual patterns mean. **People are generally good at these tasks and we are optimistic that system administrators will grasp these concepts quickly.**"
  - user evaluations not rigorous or detailed
  - as recognized by authors, Rumint lacks advanced filtering
  - some bad colour choices for the colour-blind

Critique

- Strengths
  - rich feature set developed based on some consultation with real users
  - informal user evaluations with good results
    - users liked Rumint, and particularly liked the "video recorder" idea
    - users made some good suggestions to improve RainStorm
- Weaknesses
  - "design framework" only loosely linked to rest of paper
  - weak usability claims
    - E.g.: "For [RainStorm] to be used on the network, system administrators will have to learn how to use it, how to interpret the display, and what the visual patterns mean. **People are generally good at these tasks and we are optimistic that system administrators will grasp these concepts quickly.**"
  - user evaluations not rigorous or detailed
  - as recognized by authors, Rumint lacks advanced filtering
  - some bad colour choices for the colour-blind

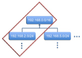
Outline

1. Introduction
2. SeeNet: Phone (and other) networks
3. Rainstorm/Rumint: IP network security
4. OverFlow: IP network analysis/security
   - Overview
   - Visualization
   - Case Study
   - Future Work
   - Critique
5. Conclusion

OverFlow: An Overview Visualization for Network Analysis

J. Glanfield, S. Brooks, T. Taylor, D. Paterson, C. Smith, C. Gates and J. McHugh

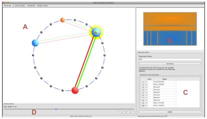
- Plugin built on prior FloVis framework[1] for network security visualizations
- Goal: complement existing FloVis visualizations by providing high-level overview of traffic between selected "organizations"
  - "Organization" = hierarchy of IP groups



  - In OverFlow, user picks one branch of the hierarchy ("active" group) for visualization

[1] T. Taylor, D. Paterson, J. Glanfield, C. Gates, S. Brooks, and J. McHugh. FloVis: Flow Visualization System. CATCH 2009.

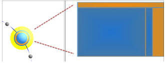
OverFlow Interface



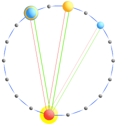
A: Circle view of communication between network "organizations". **D:** Time slider.
B: Treemap showing hierarchy of selected organization.
C: IP groups in hierarchy of selected organization.

OverFlow Interface

- Concentric circles in circle view show only the active branch of selected organization's hierarchy
- Treemap provides overview of whole hierarchy for selected organization

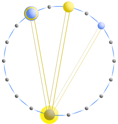


OverFlow Interface



Green and red lines show traffic from and to selected organization

OverFlow Interface: colour-blindness check

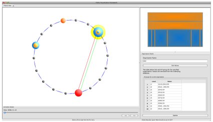


OverFlow visualization in Vischeck deuteranope simulation

Case Study

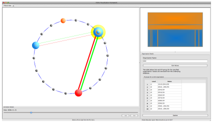
- Data collected from real network during a week in late 2008
- Network divided into four "organizations":
  - Administration
  - Security
  - Public (wireless LAN)
  - Web (external Internet)
- Visualizations provided for communications to/from Public for four days

Case Study



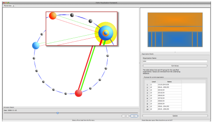
Day 1: Public communicates with Web

Case Study



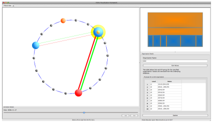
Day 2: Public communicates with Administration and Web

Case Study



Day 3: Public communicates with Security (bad!), Admin, and Web

Case Study



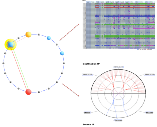
Day 4: Public communicates with Administration and Web

Case Study

- Further investigation showed that Security should not, in fact, have been communicating with Public
- Weakness (my comment): Visualization only shows communications to/from selected organization; no apparent option to show all communications
  - E.g., for day 1: "…there was communication between all of the different subnets except for (1) the Security subnet only communicated with the Web subnet, and neither of the other two subnets, and (2) the Administration and PUblic [*sic*] subnets."
  - But screenshot only shows Public and Web communicating

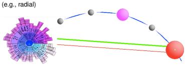
Future Work

- Launch other FloVis plugins to see details



Future Work

- Launch other FloVis plugins to see details
- Use visualizations other than concentric circles for organizations (e.g., radial)

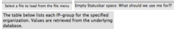


Critique

- Strengths
  - Interesting concept that appears potentially useful
  - Developed as a rapid response to suggestions from potential expert users
- Weaknesses
  - User interface design needs improvement
    - Text labels would make organizations in circle view easier to identify
    - Why display hierarchy as two-column table? Tree seems more natural
    - Only shows communications to/from selected organization
    - UI junk: redundant descriptions, leftover scaffolding
    - Red/green line choice suboptimal for colour-blind users
  - Case study provides only basic "it worked" result; no lessons drawn about UI or desired features to add/remove

Critique

- Strengths
  - Interesting concept that appears potentially useful
  - Developed as a rapid response to suggestions from potential expert users
- Weaknesses
  - User interface design needs improvement
    - Text labels would make organizations in circle view easier to identify
    - Why display hierarchy as two-column table? Tree seems more natural
    - Only shows communications to/from selected organization
    - UI junk: redundant descriptions, leftover scaffolding



Select a file to load from the file menu

Empty Statusbar space: What should we use me for??

The table below lists each IP-group for the specified organization. Values are retrieved from the underlying database.

    - Red/green line choice suboptimal for colour-blind users
  - Case study provides only basic "it worked" result; no lessons drawn about UI or desired features to add/remove

Outline

1. Introduction
2. SeeNet: Phone (and other) networks
3. Rainstorm/Rumint: IP network security
4. OverFlow: IP network analysis/security
5. Conclusion

Conclusion

- Common themes
  - Vast amount of data about network traffic ⟹ need to mitigate operator information overload
  - Often supporting exploratory analysis
    - Is anything interesting/bad going on here?
  - Provide high-level overview and detail views
    - Sometimes, wide range of visualizations provided (SeeNet, Rumint)
  - Give operator a large, detailed picture of network status; anomalies can be detected as changes in the picture
  - No formal usability testing
- Contrasts
  - SeeNet focused on network capacity/load; Rainstorm/Rumint and OverFlow on content of transmissions (finding anomalies, potential security problems)
  - SeeNet and Rainstorm/Rumint have many techniques for both high-level and detail analysis packed into one paper; OverFlow paper focuses on one high-level technique

Conclusion

- Common themes
  - Vast amount of data about network traffic ⟹ need to mitigate operator information overload
  - Often supporting exploratory analysis
    - Is anything interesting/bad going on here?
  - Provide high-level overview and detail views
    - Sometimes, wide range of visualizations provided (SeeNet, Rumint)
  - Give operator a large, detailed picture of network status; anomalies can be detected as changes in the picture
  - No formal usability testing
- Contrasts
  - SeeNet focused on network capacity/load; Rainstorm/Rumint and OverFlow on content of transmissions (finding anomalies, potential security problems)
  - SeeNet and Rainstorm/Rumint have many techniques for both high-level and detail analysis packed into one paper; OverFlow paper focuses on one high-level technique

Questions?