Malware Cinema
A Picture is Worth a Thousand Packets

Gregory Conti
www.cc.gatech.edu/~conti
conti@cc.gatech.edu
The views expressed in this presentation are those of the author and do not reflect the official policy or position of the United States Military Academy, the Department of the Army, the Department of Defense or the U.S. Government.
**information visualization** is the use of interactive, sensory representations, typically visual, of abstract data to reinforce cognition.

SANS Internet Storm Center

Dec 19th 2005
Ethereal’s Tipping Point
(for the human)

Students: 635 Packets
Professionals: 5,905 Packets
Snort’s Tipping Point
(for the humans)

Professionals: 1,183 Alerts
Students: 30 Alerts
Potential DataStreams

Traditional
- packet capture
- IDS/IPS logs
- syslog
- firewall logs
- anti-virus
- net flows
- host processes
- honeynets
- network appliances

Less traditional
- p0f
- IANA data (illegal IP’s)
- DNS
- application level
- extrusion detection systems
- local semantic data (unassigned local IPs)
- inverted IDS
- geolocation (MaxMind?)
- vulnerability assessment
  - nessus, nmap ...
- system files
General InfoVis Research...

powerpoint of classic systems is here...
http://www.rumint.org/gregconti/publications/20040731-information_visualization_survey.ppt

see InfoVis proceedings for more recent work...
http://www.infovis.org/symposia.php
Rootkit Propagation
(Dan Kaminsky)

http://www.doxpara.com/
Firewall Data
(Raffy Marty)

http://raffy.ch/blog/
IDS Alerts
(Kulsoom Abdullah)

Packet Level

(John Goodall)

http://userpages.umbc.edu/~jgood/research/tnv/
Packet Level
(John Goodall)

http://userpages.umbc.edu/~jgood/research/tnv/
<table>
<thead>
<tr>
<th>Ethereal Strengths</th>
<th>Snort Strengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Full view of all packet parameters</td>
<td>• Robust and configurable filtering</td>
</tr>
<tr>
<td>• Capture and display filters</td>
<td>• High quality signature database</td>
</tr>
<tr>
<td>• Dissect and analyze protocols</td>
<td>• Helps to focus human resources</td>
</tr>
<tr>
<td></td>
<td>• Flexibility</td>
</tr>
<tr>
<td></td>
<td>• Ability to access details of packets/alerts</td>
</tr>
<tr>
<td></td>
<td>• Open source</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethereal Weaknesses</th>
<th>Snort Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Overwhelming detail / too much for human to process</td>
<td>• Too many false positives</td>
</tr>
<tr>
<td>• Impossible to properly visualize a large dataset</td>
<td>• Reliance on known signatures</td>
</tr>
<tr>
<td>without getting lost and confused</td>
<td>• Time and difficulty in selecting</td>
</tr>
<tr>
<td>• GUI too cumbersome</td>
<td>right set of signatures for a given network.</td>
</tr>
<tr>
<td></td>
<td>• Front end GUIs are poor</td>
</tr>
<tr>
<td>Ethereal Strengths</td>
<td>Snort Strengths</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td>• Full view of all packet parameters</td>
<td>• Robust and configurable filtering</td>
</tr>
<tr>
<td>• Capture and display filters</td>
<td>• High quality signature database</td>
</tr>
<tr>
<td>• Dissect and analyze protocols</td>
<td>• Helps to focus human resources</td>
</tr>
<tr>
<td></td>
<td>• Flexibility</td>
</tr>
<tr>
<td></td>
<td>• Ability to access details of packets/alerts</td>
</tr>
<tr>
<td></td>
<td>• Open source</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethereal Weaknesses</th>
<th>Snort Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Overwhelming detail / too much for human to process</td>
<td>• Too many false positives</td>
</tr>
<tr>
<td>• Impossible to properly visualize a large dataset</td>
<td>• Reliance on known signatures</td>
</tr>
<tr>
<td>without getting lost and confused</td>
<td>• Time and difficulty in selecting</td>
</tr>
<tr>
<td>• GUI too cumbersome</td>
<td>right set of signatures for a given network.</td>
</tr>
<tr>
<td></td>
<td>• Front end GUIs are poor</td>
</tr>
</tbody>
</table>

More results on CD...
Ethereal can be found at http://www.ethereal.com/

http://www.pandora.nu/tempo-depot/notes/blosxom/data/PC_side/Web_Browser/Blosxom/ethereal.png
Potential DataStreams

Traditional
- packet capture
- IDS/IPS logs
- syslog
- firewall logs
- anti-virus
- net flows
- host processes
- honeynets
- network appliances
- routing data

Less traditional
- p0f
- IANA data (illegal IP’s)
- DNS
- application level
- extrusion detection systems
- local semantic data (unassigned local IPs)
- inverted IDS
- geolocation (MaxMind?)
- vulnerability assessment
  - nessus, nmap ...
- system files
Potential DataStreams

Traditional
• **packet capture**
• IDS/IPS logs
• syslog
• firewall logs
• anti-virus
• net flows
• host processes
• honeynets
• network appliances

payload
byte frequency
packet length

ethertype
IP version
IP header length
IP differential services
IP total length
IP identification
IP flags
IP fragment
TTL
IP transport
IP header checksum
src/dst IP
src/dst TCP&UDP port
RUMINT
Filtering, Encoding & Interaction
Multiple Coordinated Views...
Text (on the fly strings)

dataset: Defcon 11 CTF
Krasser Visualization
Routine Honeynet Traffic
(baseline)
Compromised Honeypot
Binary Rainfall Visualization
(single packet)

Bits on wire...

0 1 1 0 1 1 1 0 0 0 1 0 1 0 1 0 0 1 0 1 0 1 1 1 1 1 0
Binary Rainfall Visualization
(single packet)

Bits on wire...

```
0 1 1 0 1 1 1 0 0 0 1 0 1 0 1 0 0 1 0 1 1 1 1 1 1 1 0
```

View as a 1:1 relationship (1 bit per pixel)...

```
0 1 1 0 1 1 1 0 0 1 0 1 0 1 0 1 0 0 1 0 1 1 1 1 1 1 0
```
Encode by Protocol

Network packets over time

Bit 0, Bit 1, Bit 2  Length of packet - 1
On the fly *disassembly*?
# Binary Rainfall Visualization

(single packet)

<table>
<thead>
<tr>
<th>Bits on wire...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 1 0 1 1 1 0 0 1 0 1 0 1 0 0 1 0 1 1 1 1 1 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>View as a 1:1 relationship (1 bit per pixel)...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>0</strong> <strong>0</strong> <strong>1</strong> <strong>0</strong> <strong>1</strong> <strong>0</strong> <strong>1</strong> <strong>0</strong> <strong>0</strong> <strong>1</strong> <strong>0</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>0</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>View as a 8:1 relationship (1 byte per pixel)...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>0</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>0</strong> <strong>0</strong> <strong>1</strong> <strong>0</strong> <strong>1</strong> <strong>0</strong> <strong>1</strong> <strong>0</strong> <strong>0</strong> <strong>1</strong> <strong>0</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>1</strong> <strong>0</strong></td>
</tr>
</tbody>
</table>

3 Pixels
Byte Visualization

- object 1
- object 2
- object 3
- object n

Diagram showing byte frequency visualization with a range from 0 to 255.
Open SSH Diffie-Hellman Key Exchange
Zipped Email Attachment
Byte Presence

dictionary file via HTTP

ssh

SSL
Parallel Coordinates

- goal: plot any data fields
- dynamic columns
- change order for different insight
- intelligent lookup and translation of fields
  - e.g. IP transport protocol
Rapidly Characterize Packet Header Fields (Context)
Identify and Precisely Locate Fragmentation Anomaly
Identify and Precisely Locate x90 Anomaly
Identify and Precisely Locate Possible Random Payload Anomaly
Task Completion Time

~8 min
Task Completion Time

- Header characterization: ~16 min
- Random data anomaly: ~5 min
- Fragmentation anomaly: ~2.5 min

- Undergraduate Group
- Graduate Group
RUMINT Tipping Point

RUMINT: ~9,000 Packets
Ethereal: 635 Packets
Demo
System Requirements

- IP over Ethernet
- Tested on Windows XP
- ~256+ MB Ram
- Processor 300MHZ (minimum)
- The more screen real estate the better
- Latest winpcap
Binary Visualization
(sendmail)

original

printable ASCII in blue
Color Encode by Opcode

(ls)
## Color Encode by Disassembly?

```none
ls:     file format elf32-i386

Disassembly of section .init:

0804917c <.init>:

<table>
<thead>
<tr>
<th>Address</th>
<th>Opcode</th>
<th>Immediate</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>804917c</td>
<td>55</td>
<td></td>
<td>push %ebp</td>
</tr>
<tr>
<td>804917d</td>
<td>89 e5</td>
<td></td>
<td>mov %esp,%ebp</td>
</tr>
<tr>
<td>804917f</td>
<td>83 ec 08</td>
<td></td>
<td>sub $0x8,%esp</td>
</tr>
<tr>
<td>8049182</td>
<td>e8 8d 05 00 00</td>
<td></td>
<td>call 0x8049714</td>
</tr>
<tr>
<td>8049187</td>
<td>e8 f4 05 00 00</td>
<td></td>
<td>call 0x8049780</td>
</tr>
<tr>
<td>804918c</td>
<td>e8 9f c0 00 00</td>
<td></td>
<td>call 0x8055230</td>
</tr>
<tr>
<td>8049191</td>
<td>c9</td>
<td></td>
<td>leave</td>
</tr>
<tr>
<td>8049192</td>
<td>c3</td>
<td></td>
<td>ret</td>
</tr>
</tbody>
</table>
```
binary file analysis

original image
Dissecting a Word document
(text + image)

header + text content

hex dump

..............\yyyy
..............
..............
..............
..............

................

..............
..............
..............
..............
..............

................

C:\\\.D.o.c.u.m.
Apply Image Filtering Algorithms
(sendmail)
Uses...

- visual diff
  - revision changes
  - datafile changes
  - malware (including malmedia)
- context
- encryption
- binary navigation
- dissecting file formats
- unfamiliar binaries
- visual strings / pattern matching
  - checksums
- integrate with hex editors / disassemblers
- cool t-shirts :)

???
Attacking the Analyst...

G. Conti, M. Ahamad and J. Stasko; "Attacking Information Visualization System Usability: Overloading and Deceiving the Human;" Symposium on Usable Privacy and Security (SOUPS); July 2005. On the CD...

G. Conti and M. Ahamad; "A Taxonomy and Framework for Countering Denial of Information Attacks;" IEEE Security and Privacy. (accepted, to be published) Website...
Attack Demo
Future Vision

Visualization Plug-ins

Filters Flows

PCAP Library
Directions for the Future...

We are only scratching the surface of the possibilities

- attack specific community needs
- plug-ins (vis, filters, processing)
- launch network packets?
- protocol specific visualizations
  - including application layer (e.g. VoIP, HTTP)
- Open GL
- graph visualization+
- screensaver/wallpaper snapshot?
- work out GUI issues
- database of filters / smart books
- stress testing
- evaluate effectiveness
- human-machine bridging technologies
entropy (bits)*

For more information...


S. Krasser, G. Conti, J. Grizzard, J. Gribschaw and H. Owen; "Real-Time and Forensic Network Data Analysis Using Animated and Coordinated Visualization;" IEEE Information Assurance Workshop (IAW); June 2005.

G. Conti; "Countering Denial of Information Attacks with Information Visualization;" Interz0ne 4; March 2005.


G. Conti; "Network Attack Visualization;" DEFCON 12; August 2004.

G. Conti; "Network Security Data Visualization;" Interz0ne3; April 2004.

www.cc.gatech.edu/~conti
www.rumint.org
On the CD...

- Talk slides
- Code
  - rumint
- Papers
  - SOUPS Malicious Visualization paper
  - Hacker conventions article
  - Ethereal / Snort Survey

See also: www.cc.gatech.edu/~conti and www.rumint.org
Acknowledgements

404.se2600, Kulsoom Abdullah, Sandip Agarwala, Mustaque Ahamad, Bill Cheswick, Chad, Clint, Tom Cross, David Dagon, DEFCON, Ron Dodge, EliO, Emma, Mr. Fuzzy, Jeff Gribschaw, Julian Grizzard, GTISC, Hacker Japan, Mike Hamelin, Hendrick, Honeynet Project, Interz0ne, Jinsuk Jun, Keshoto, Oleg Kolesnikov, Sven Krasser, Chris Lee, Wenke Lee, John Levine, Michael Lynn, David Maynor, Neel Mehta, Jeff Moss, NETI@home, Henry Owen, Dan Ragsdale, Rockit, Byung-Uk Roho, Charles Robert Simpson, Ashish Soni, SOUPS, Jason Spence, John Stasko, StricK, Susan, USMA ITOC, IEEE IAW, VizSEC 2004, Grant Wagner and the Yak.
Greg Conti
conti@cc.gatech.edu

www.cc.gatech.edu/~conti
www.rumint.org
Labeling Attack

CDX 2003 Dataset
X = Time
Y = Destination IP
Z = Destination Port
Precision Attack

http://www.nersc.gov/nusers/security/Cube.jpg

http://developers.slashdot.org/article.pl?sid=04/06/01/1747223&mode=thread&tid=126&tid=172
The Conscience of a Hacker... by The Mentor ... 1/8/86

Another one got caught today. It's all over the papers. "Teenager Arrested in Computer Crime Scandal", "Hacker arrested after Bank Tampering"...

Damn kids. They're all alike...