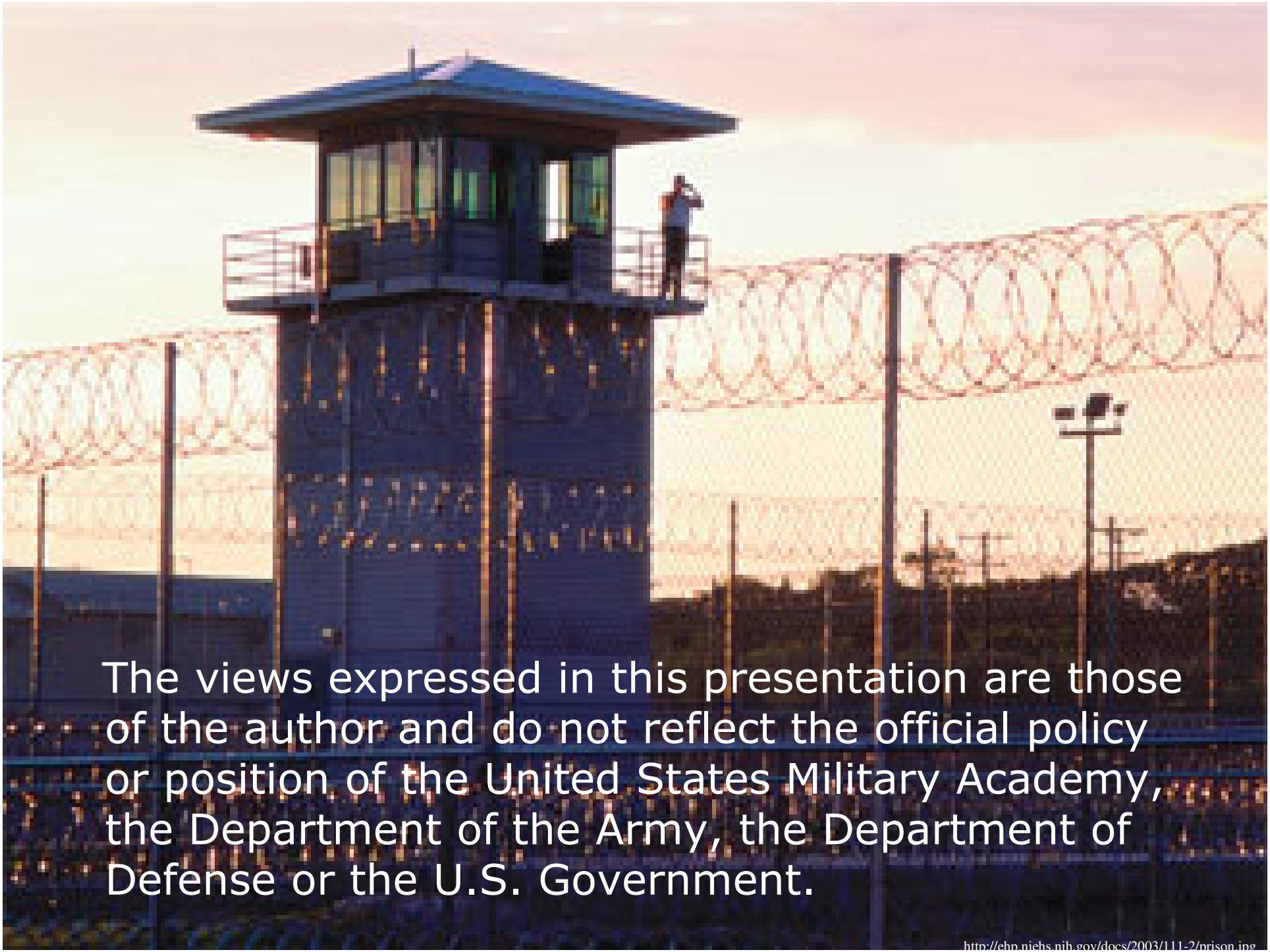
Malware Cinema

A Picture is Worth a Thousand Packets

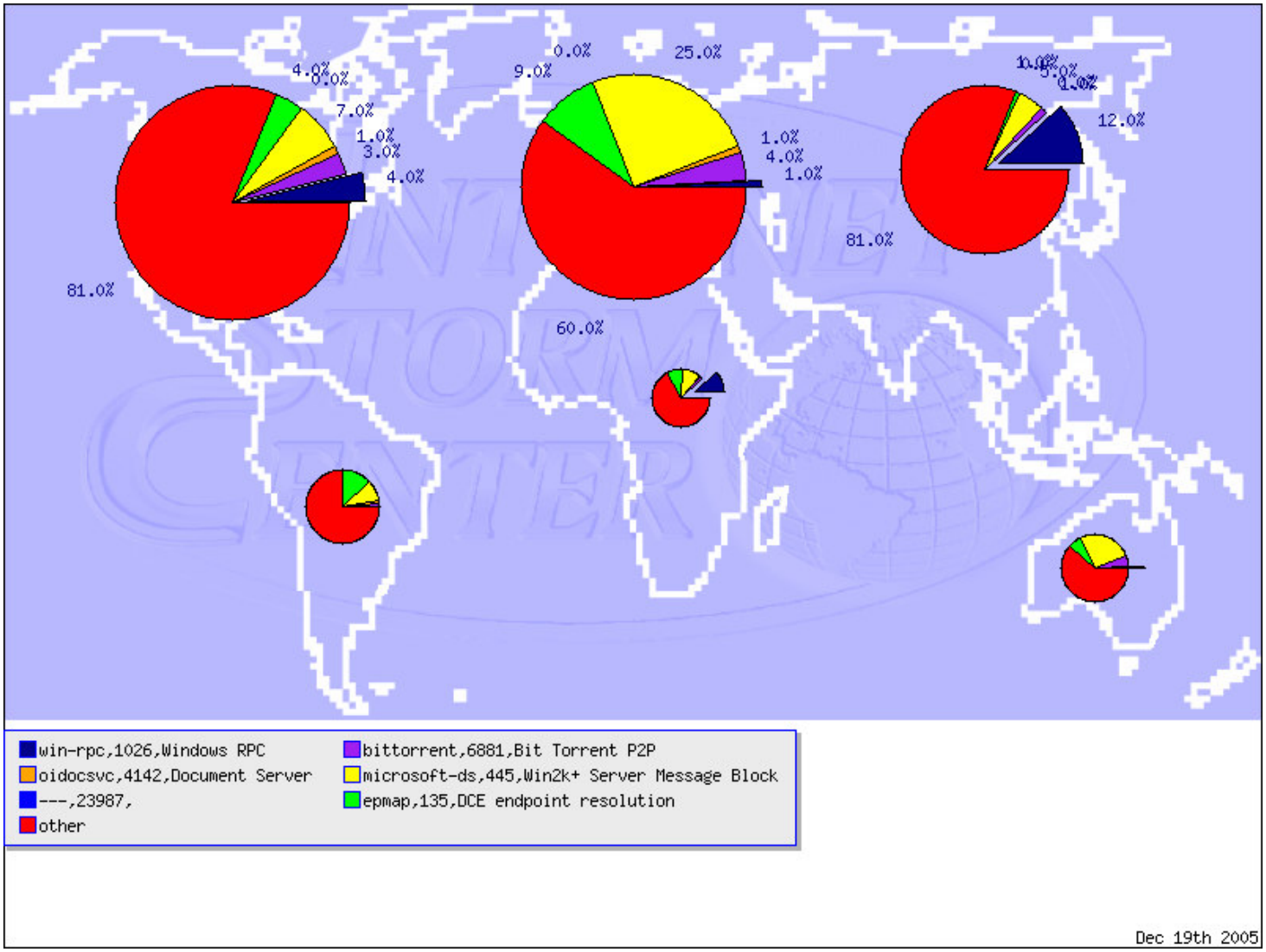
Gregory Conti
www.cc.gatech.edu/~conti
conti@cc.gatech.edu


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.
http://ehp.niehs.nih.gov/docs/2003/111-2/prison.jpg

information visualization *is the use of interactive, sensory representations, typically visual, of abstract data to reinforce cognition.*

http://en.wikipedia.org/wiki/Information_visualization

SANS Internet Storm Center

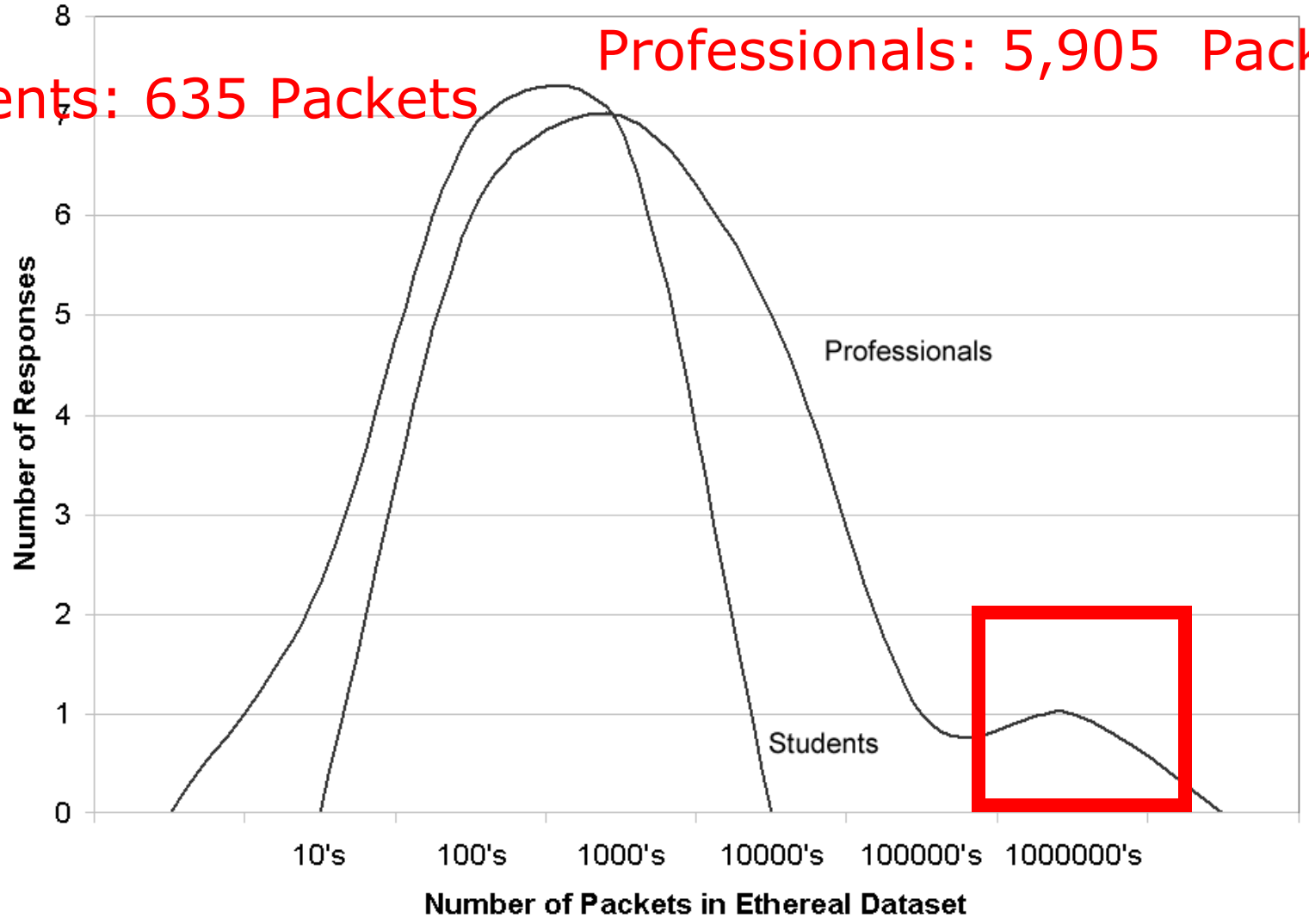


Ethereal's Tipping Point

(for the human)

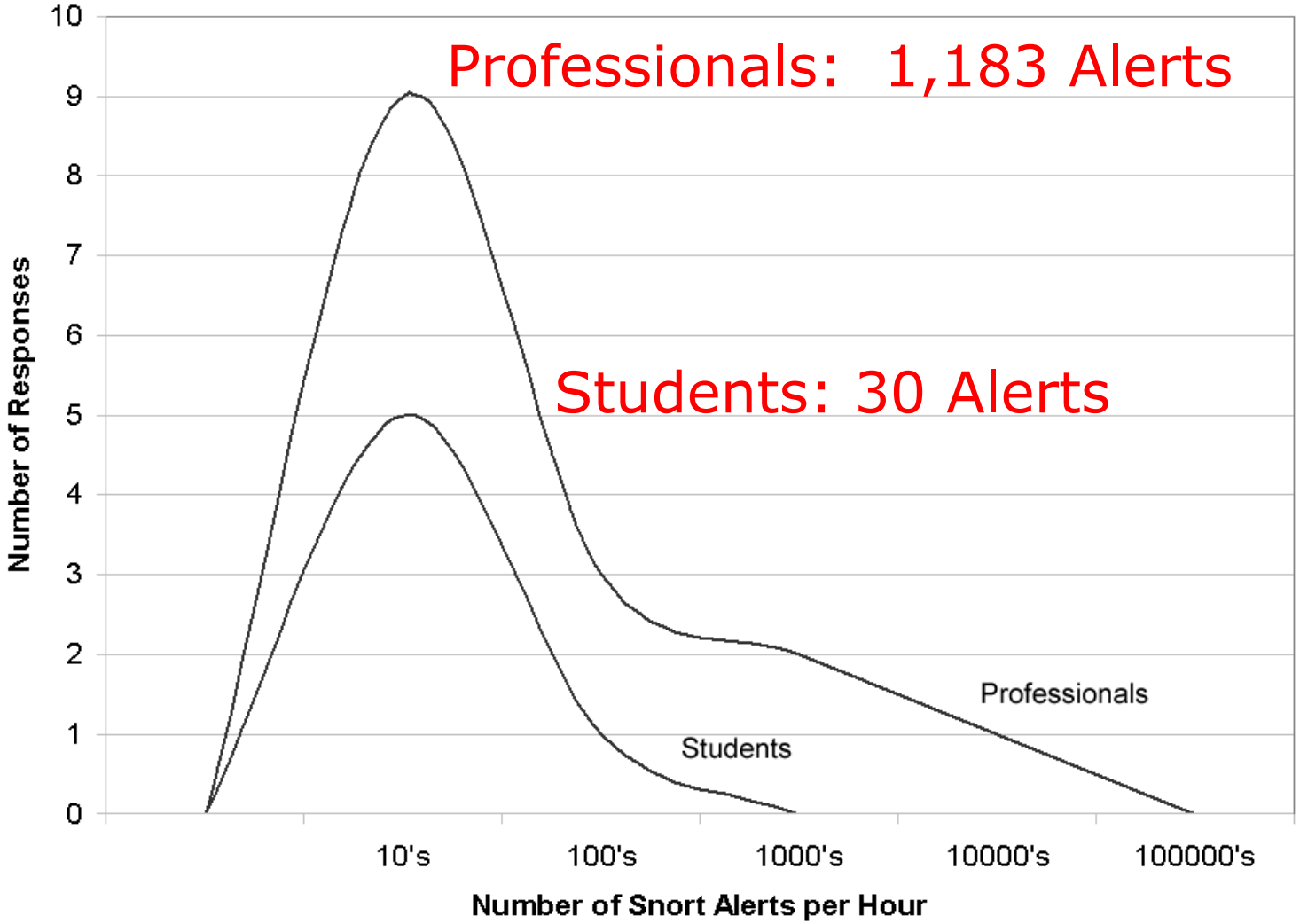
Students: 635 Packets

Professionals: 5,905 Packets



Snort's Tipping Point

(for the humans)



Potential DataStreams

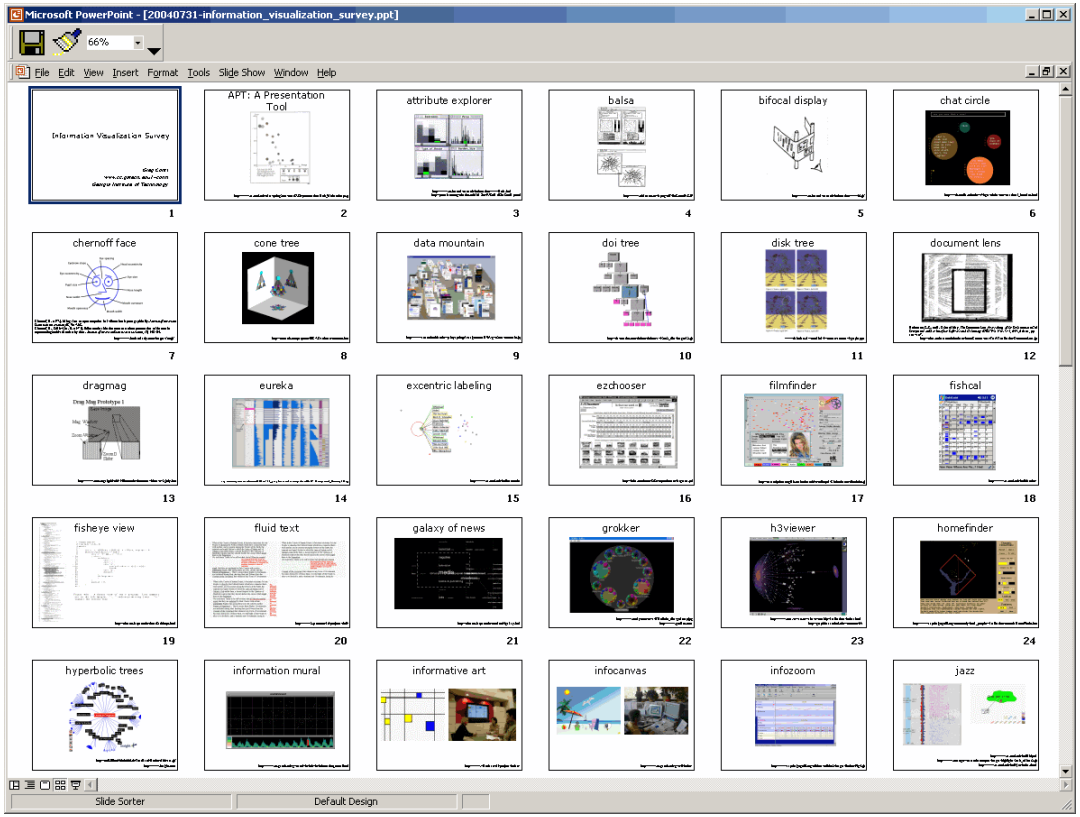
Traditional

- 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…

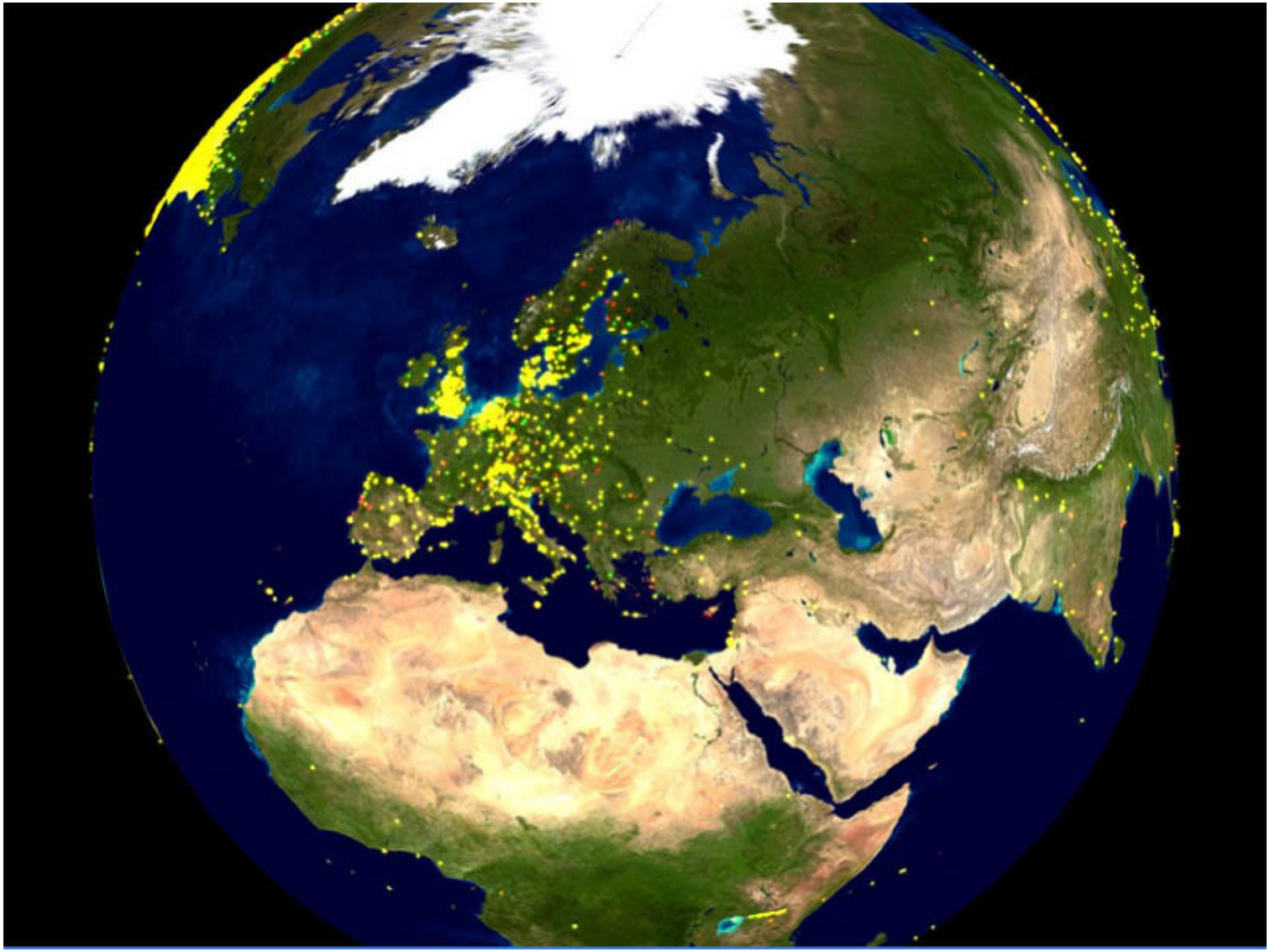


powerpoint of classic systems is here…
http://www.rumint.org/gregconti/publications/20040731-information_visualization_survey.ppt

see InfoVis proceedings for more recent work…
http://www.infovis.org/symposia.php

Rootkit Propagation

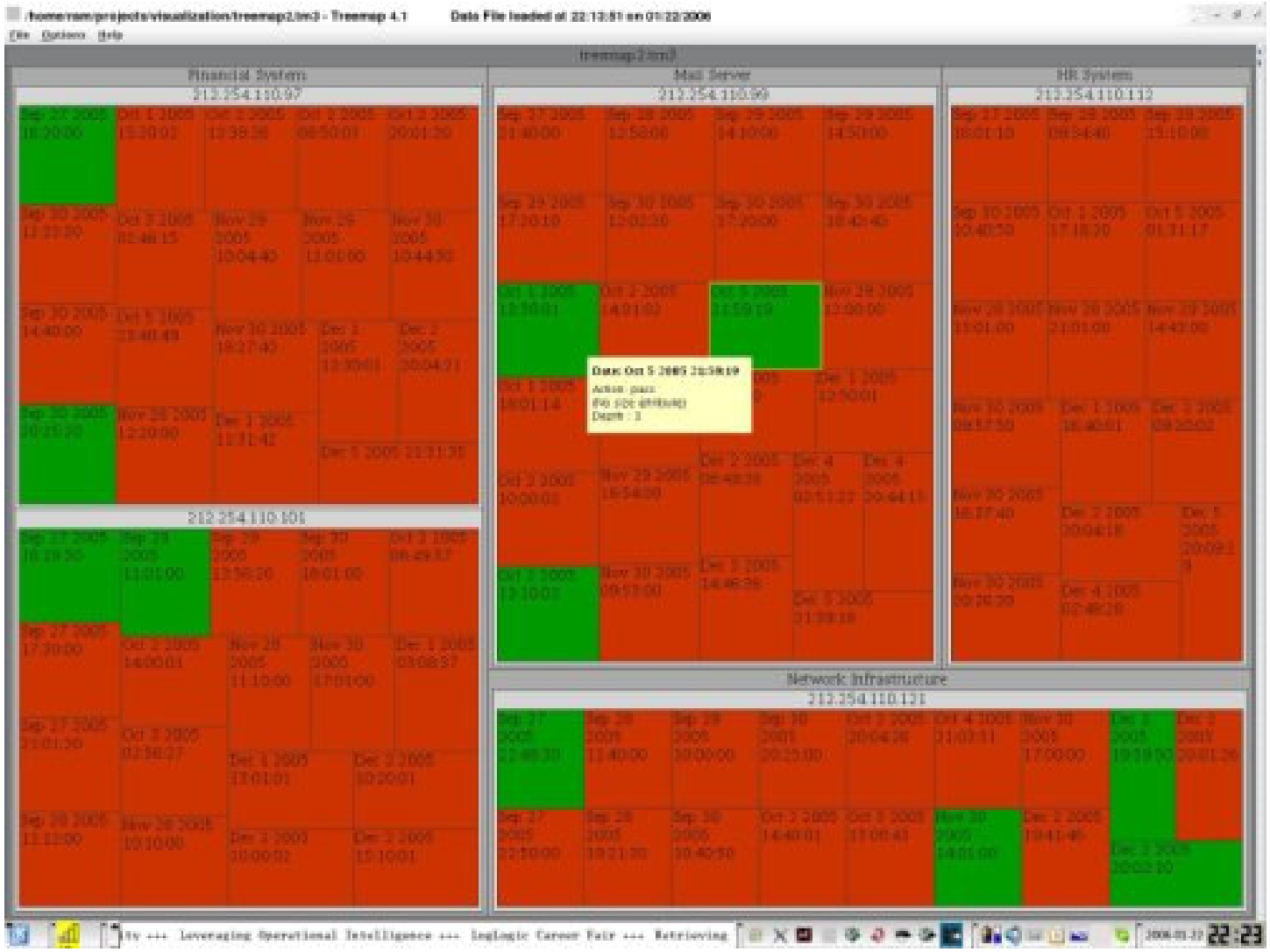
(Dan Kaminsky)



http://www.doxpara.com/

Firewall Data

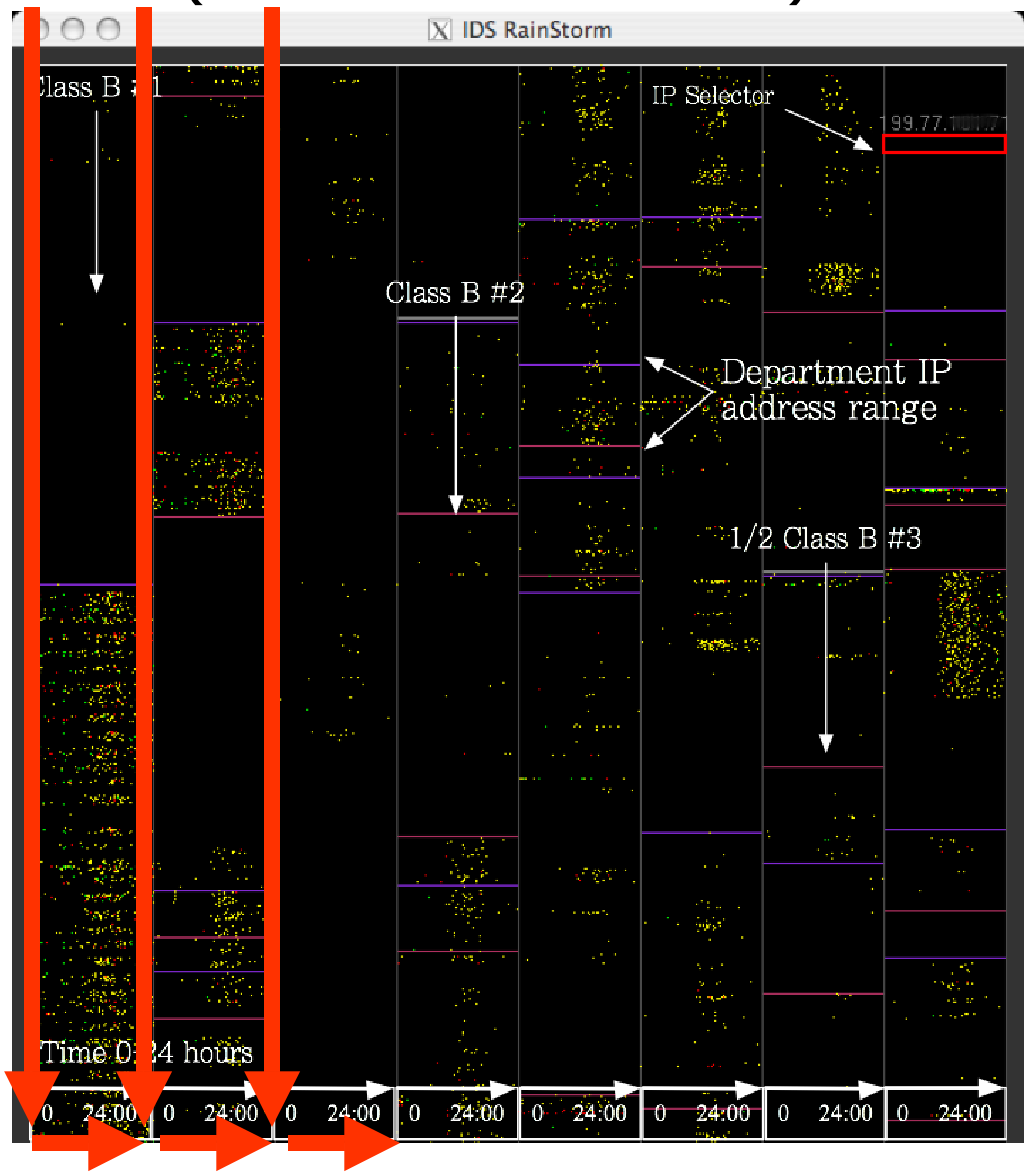
(Raffy Marty)



http://raffy.ch/blog/

IDS Alerts

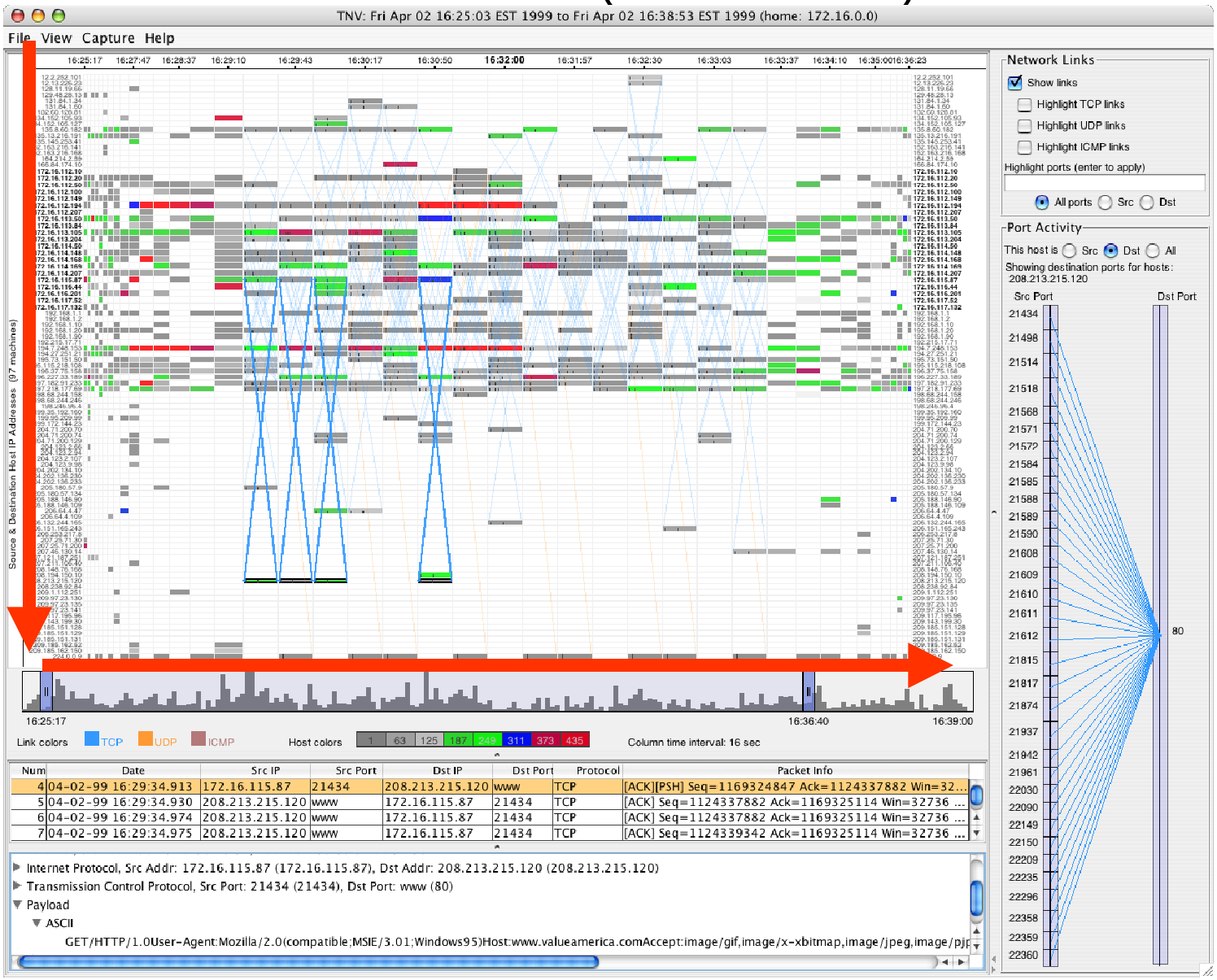
(Kulsoom Abdullah)



http://www.rumint.org/gregconti/publications/20050813_VizSec_IDS_Rainstorm.pdf

Packet Level

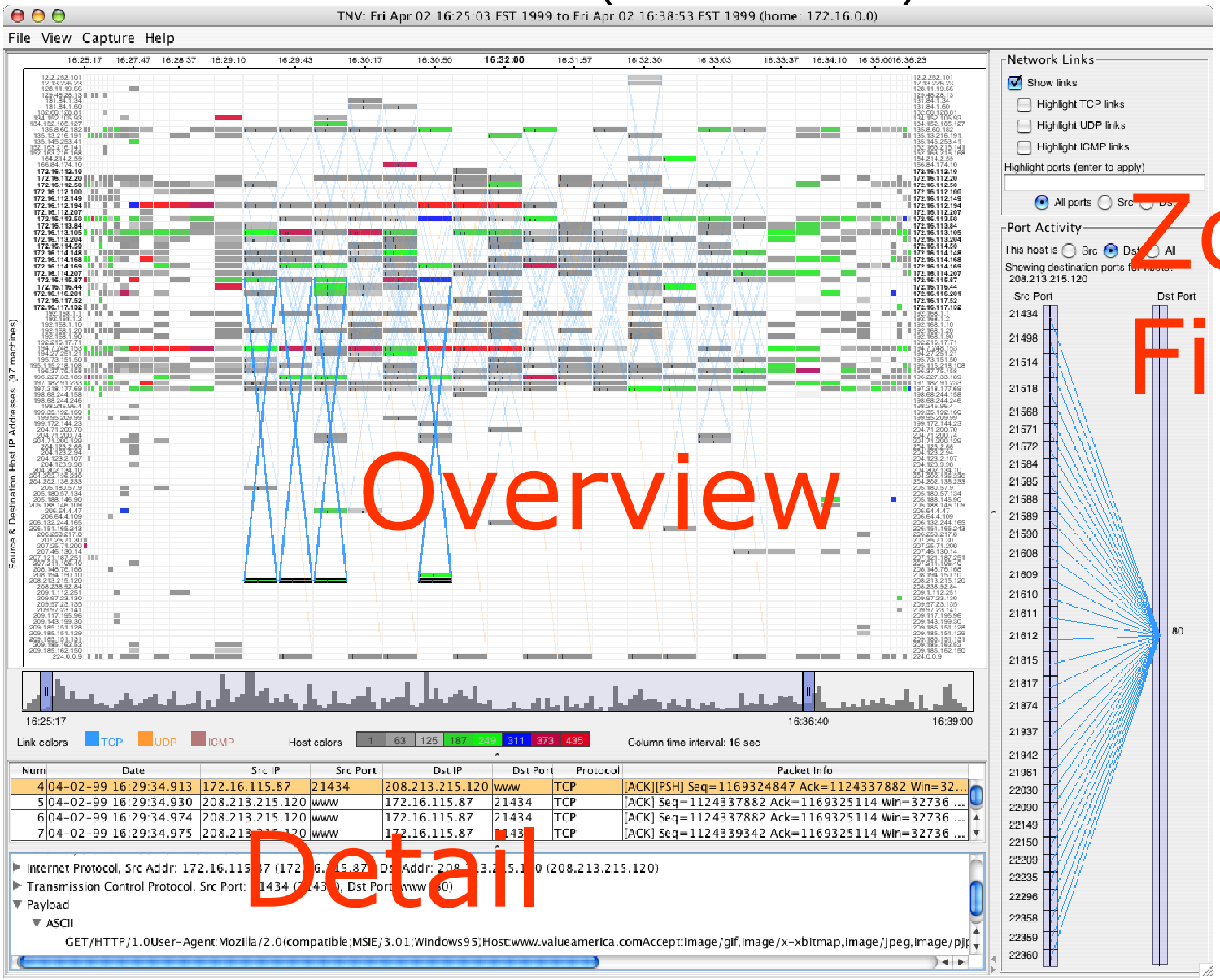
(John Goodall)



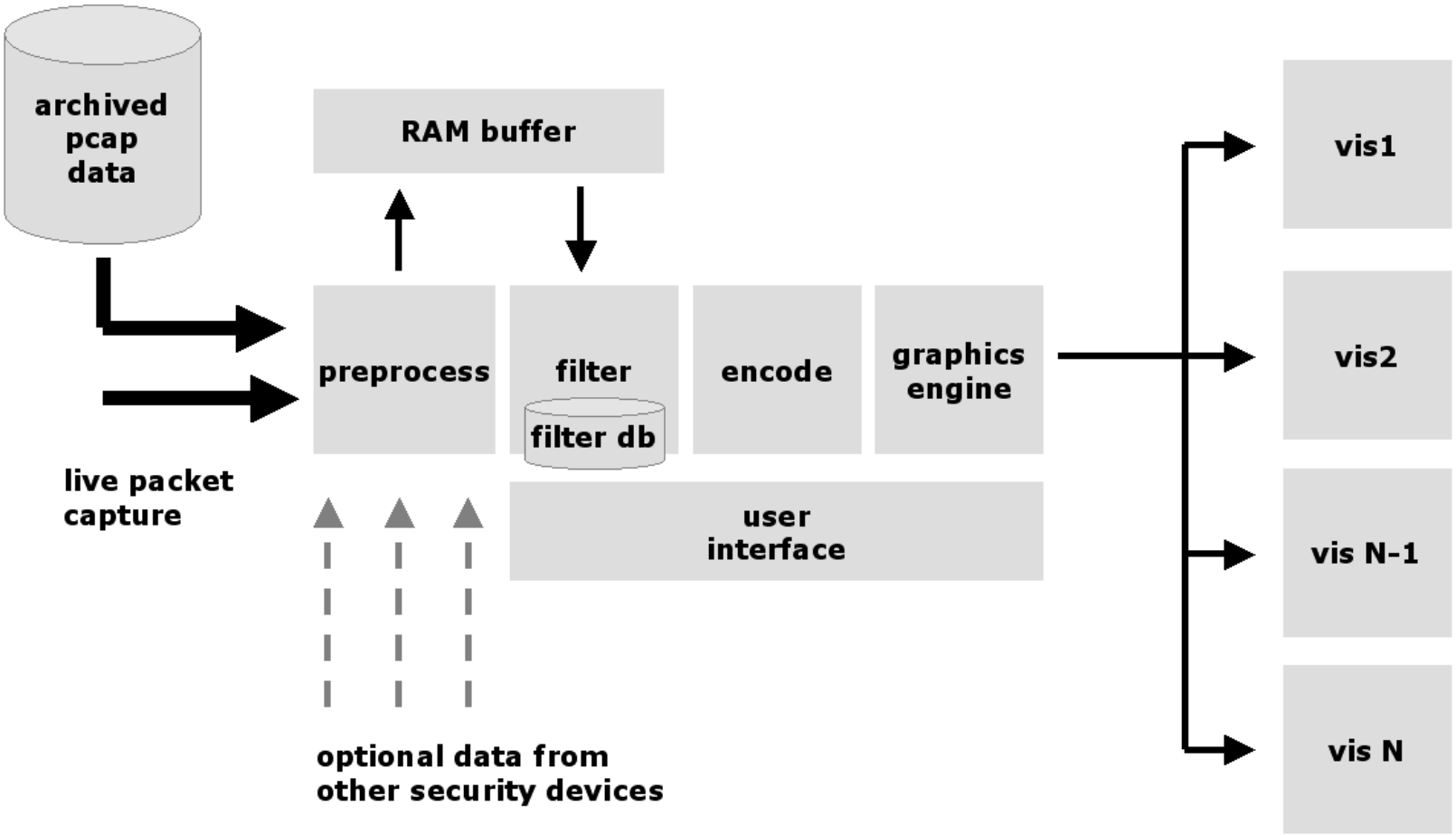
http://userpages.umbc.edu/~jgood/research/tnv/

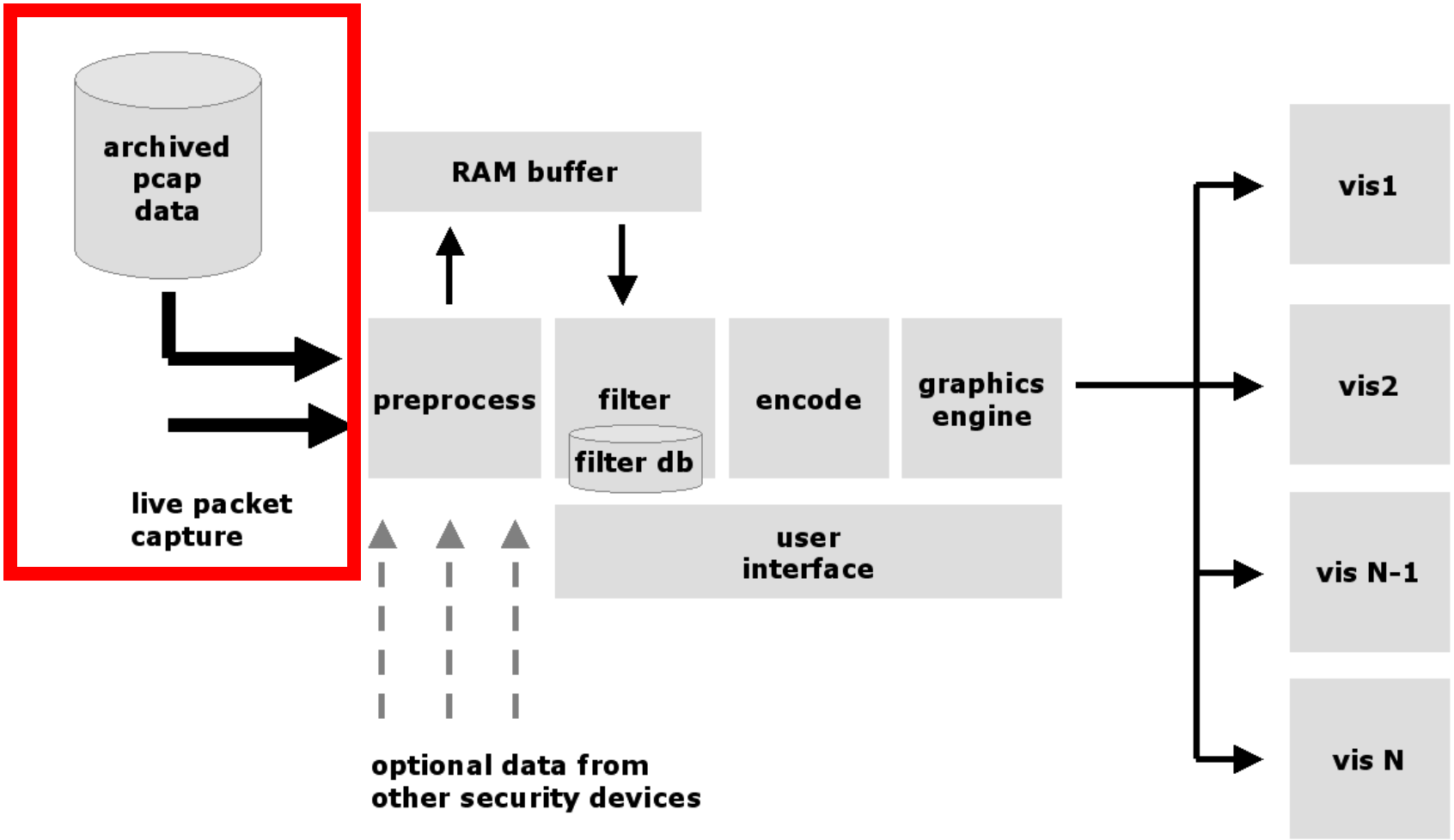
Packet Level

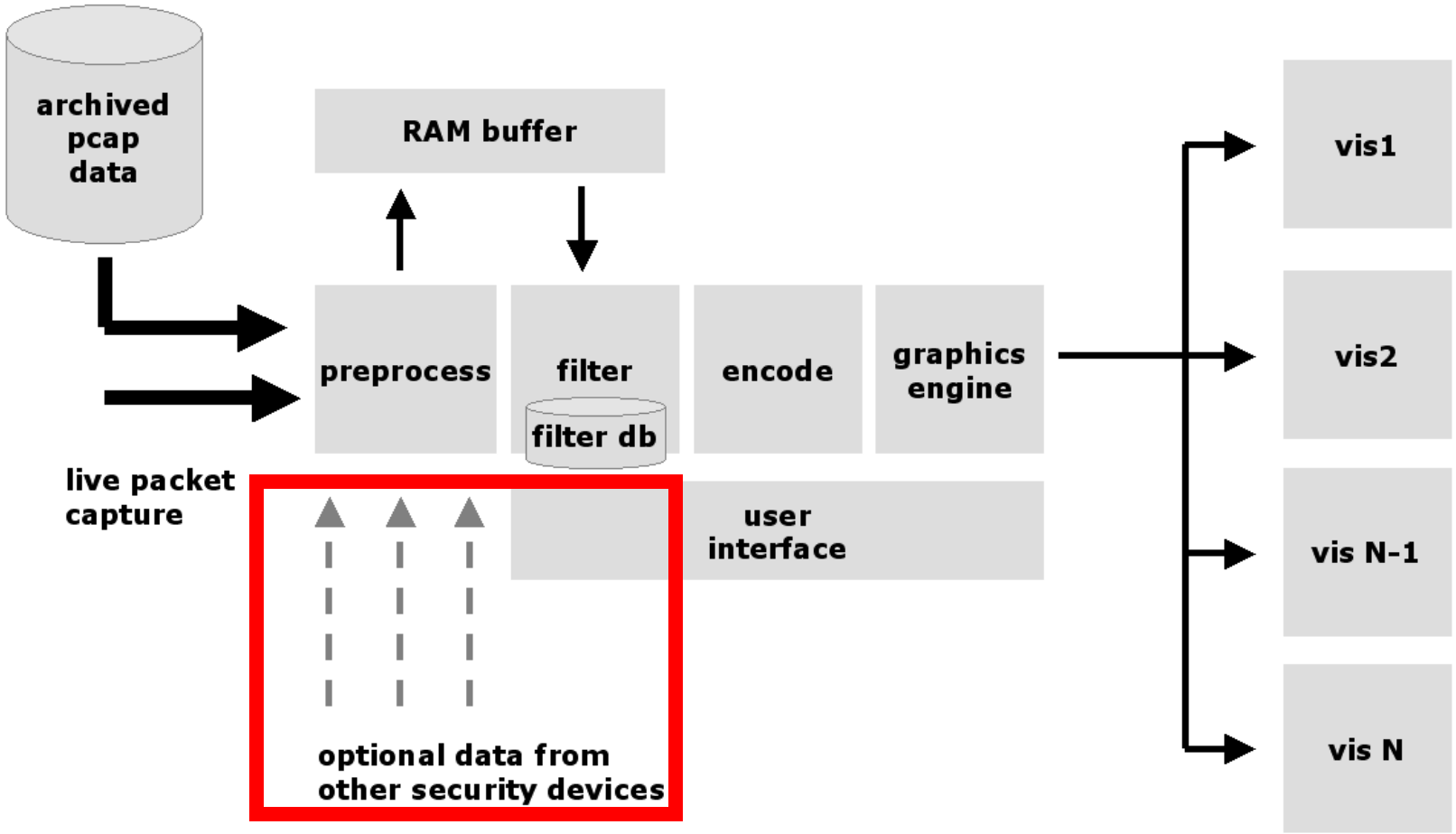
(John Goodall)

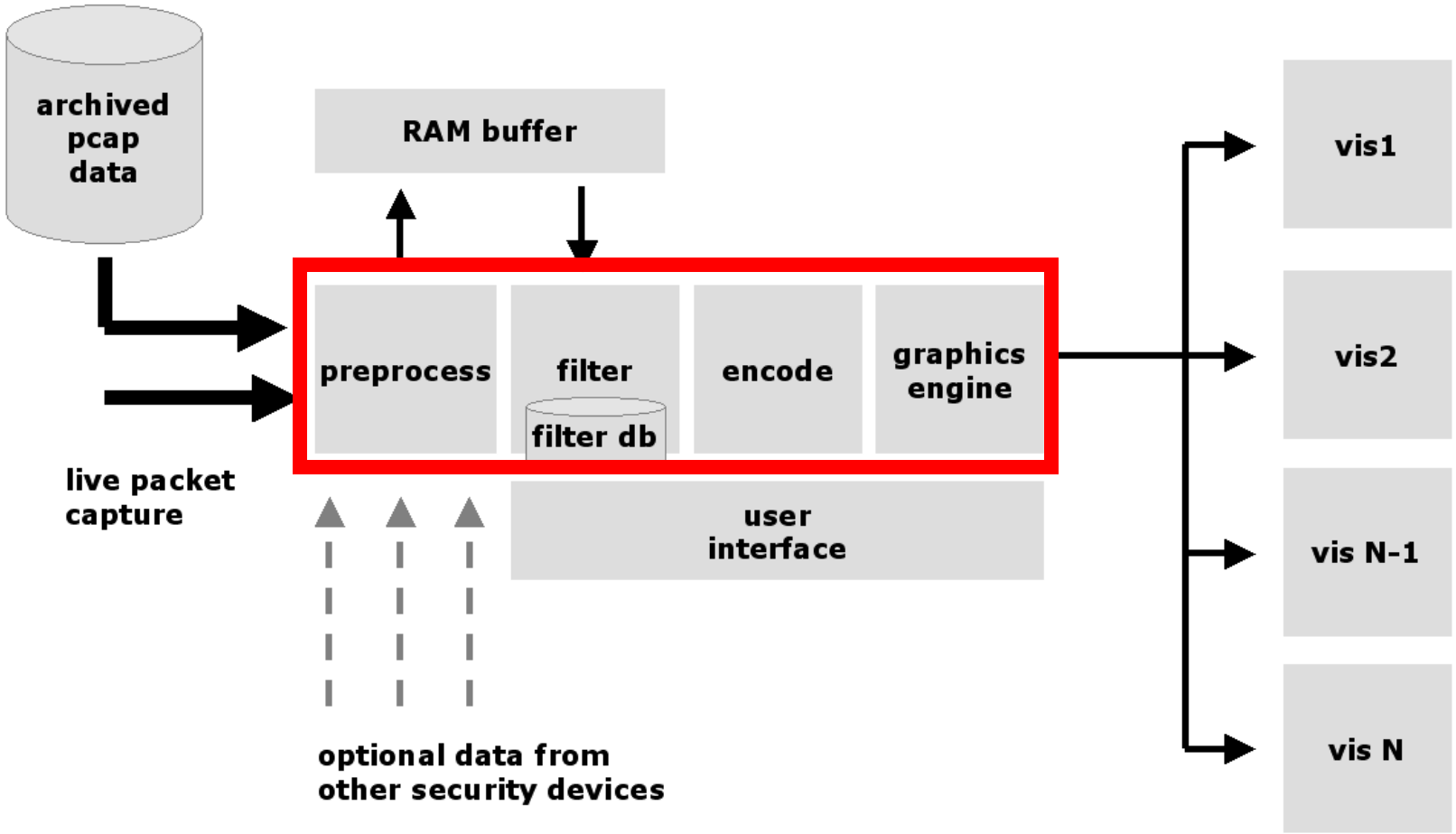


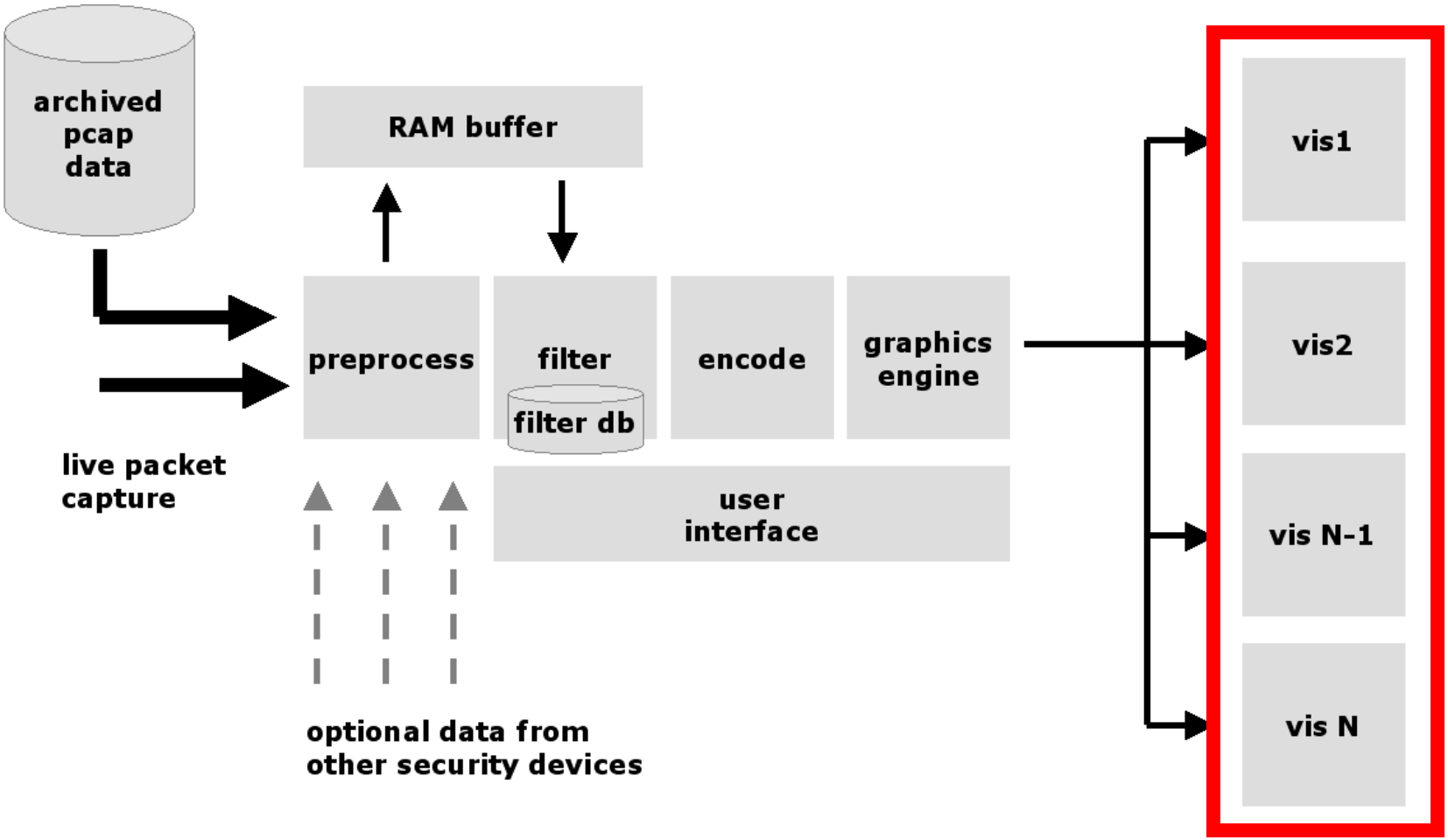
http://userpages.umbc.edu/~jgood/research/tnv/

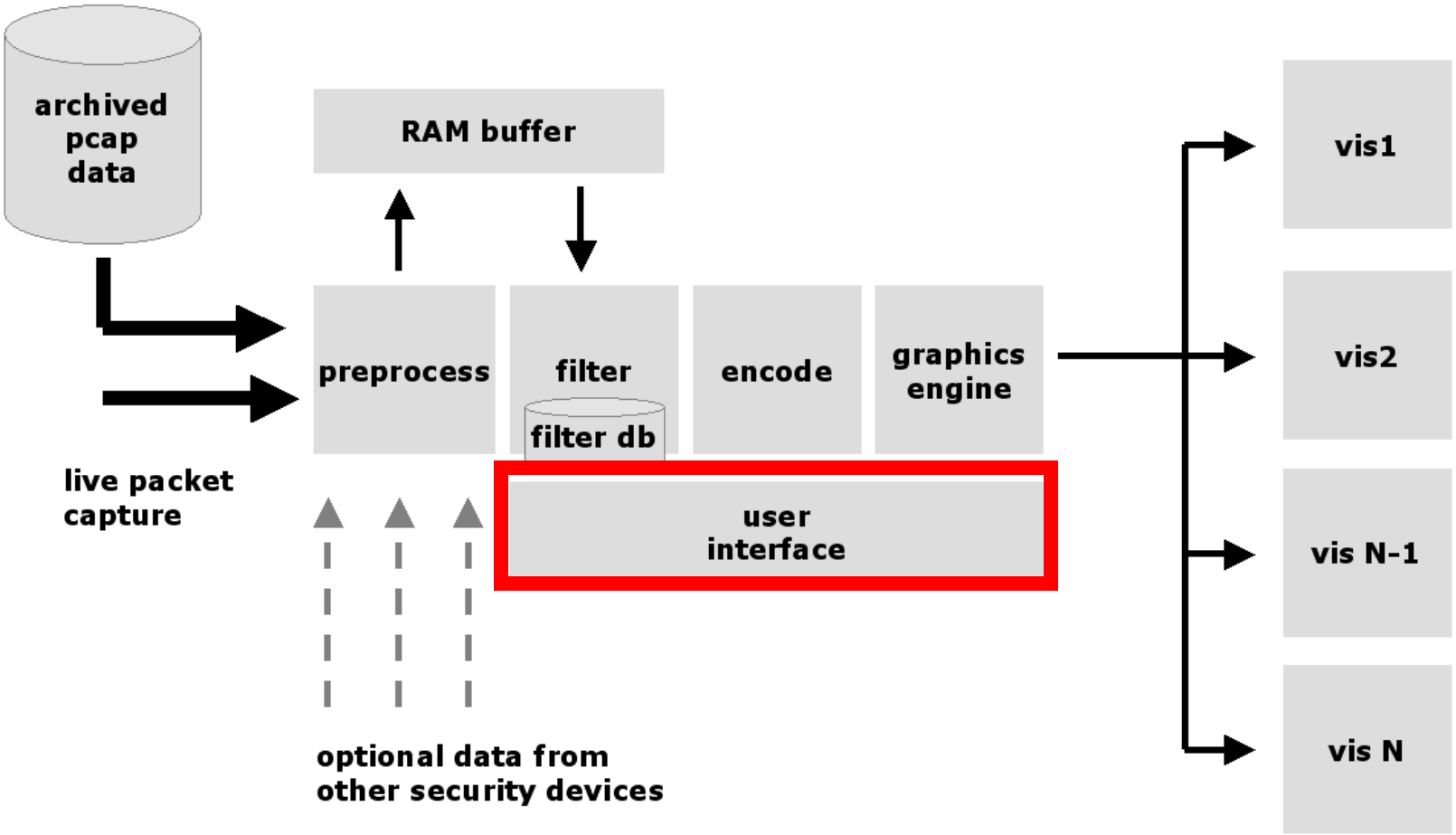

archived
pcap
data
RAM buffer
preprocess
filter
filter db
encode
graphics
engine
vis1
vis2
vis N-1
vis N
live packet
capture
user
interface
optional data from
other security devices


archived
pcap
data
live packet
capture
RAM buffer
preprocess
filter
filter db
encode
graphics
engine
user
interface
optional data from
other security devices
vis1
vis2
vis N-1
vis N


archived pcap data
RAM buffer
preprocess
filter
filter db
encode
graphics engine
vis1
vis2
vis N-1
vis N
live packet capture
user interface
optional data from other security devices


archived pcap data
RAM buffer
preprocess
filter
filter db
encode
graphics engine
live packet capture
user interface
optional data from other security devices
vis1
vis2
vis N-1
vis N


archived pcap data
RAM buffer
preprocess
filter
filter db
encode
graphics engine
vis1
vis2
vis N-1
vis N
live packet capture
user interface
optional data from other security devices


archived pcap data
RAM buffer
preprocess
filter
filter db
encode
graphics engine
vis1
vis2
vis N-1
vis N
live packet capture
user interface
optional data from other security devices

Ethereal Strengths

- Full view of all packet parameters
- Capture and display filters
- Dissect and analyze protocols

Snort Strengths

- Robust and configurable filtering
- High quality signature database
- Helps to focus human resources
- Flexibility
- Ability to access details of packets/alerts
- Open source

Ethereal Weaknesses

- Overwhelming detail / too much for human to process
- Impossible to properly visualize a large dataset without getting lost and confused
- GUI too cumbersome

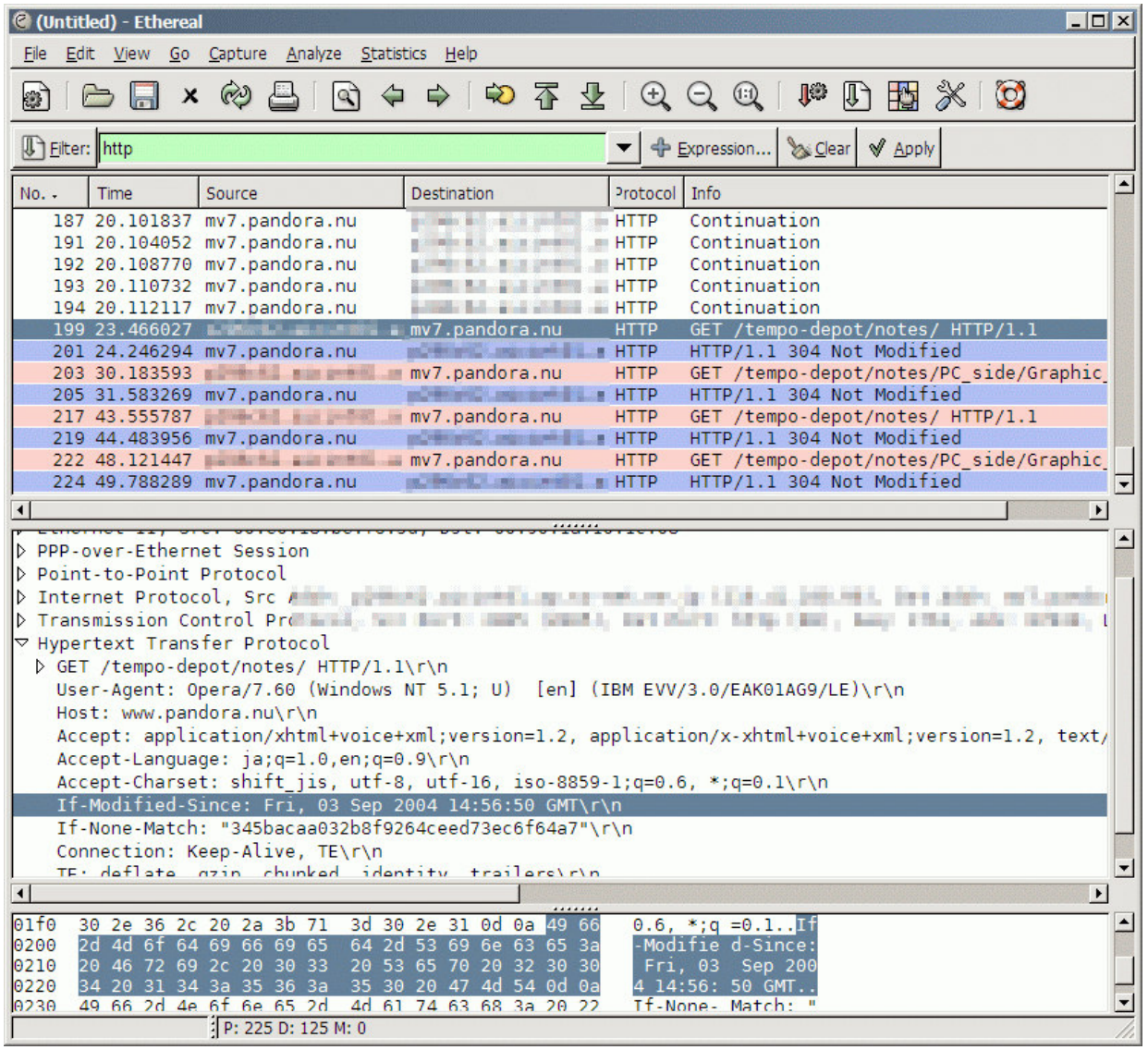
Snort Weaknesses

- Too many false positives
- Reliance on known signatures
- Time and difficulty in selecting right set of signatures for a given network.
- Front end GUIs are poor

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

More results on CD…

Ethereal



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

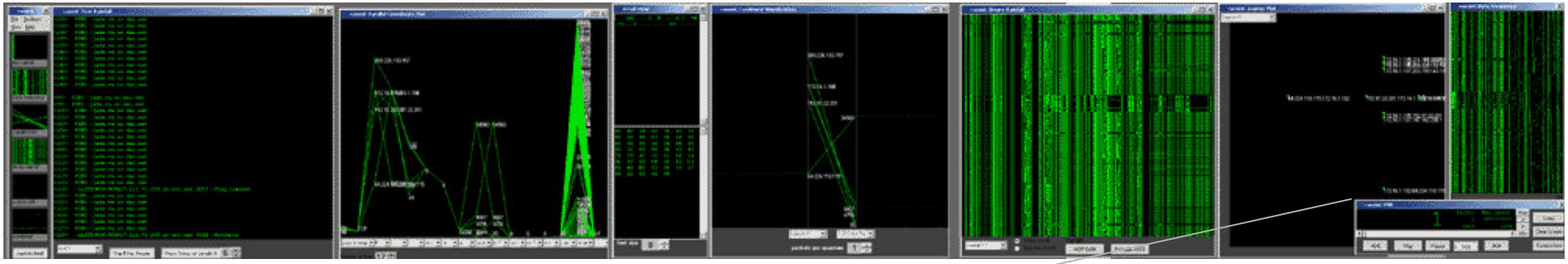
Traditional

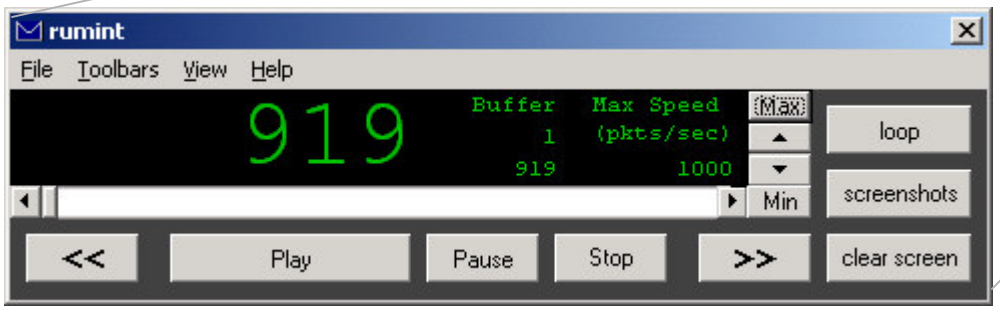
- **packet capture**
- IDS/IPS logs
- syslog
- firewall logs
- anti-virus
- net flows
- host processes
- honeynets
- network appliances

payload
byte frequency
packet length

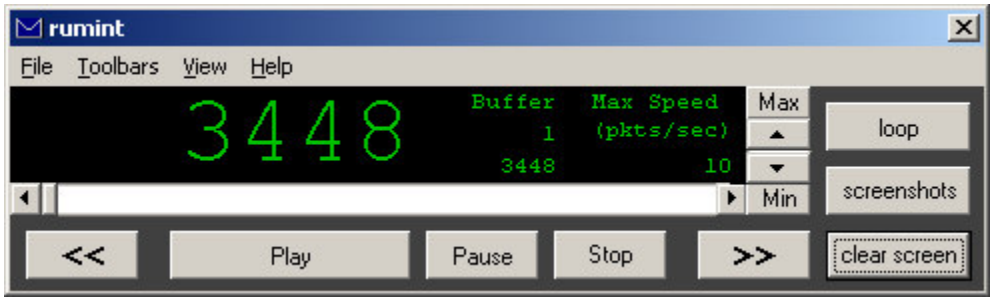
ethertype
IP version
IP header length
IP differential services
IP total length
IP identification
IP flags
IP fragment
TTL
IP transport
IP header checksum
src/dst IP
src/dst TCP&UDP port

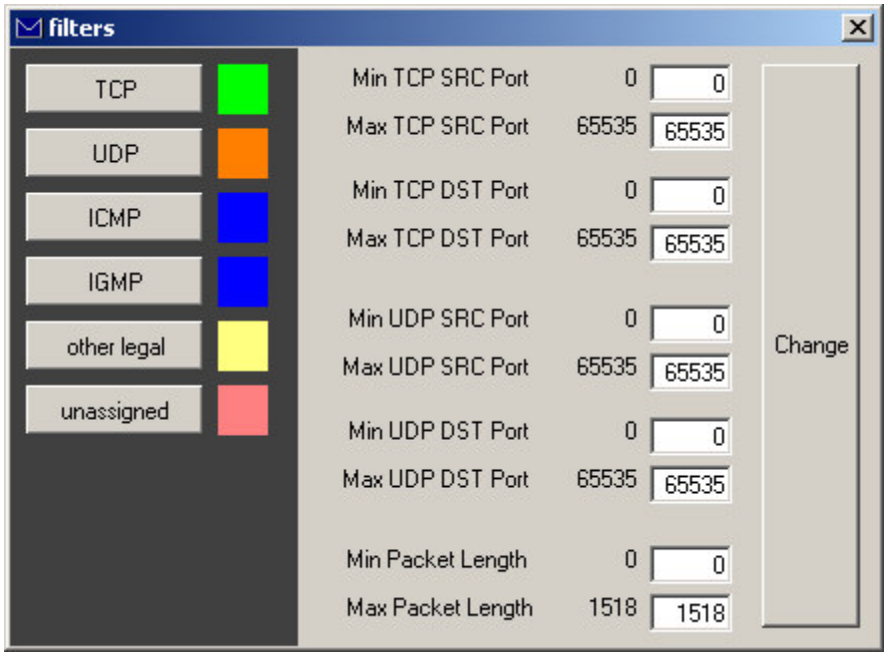
RUMINT

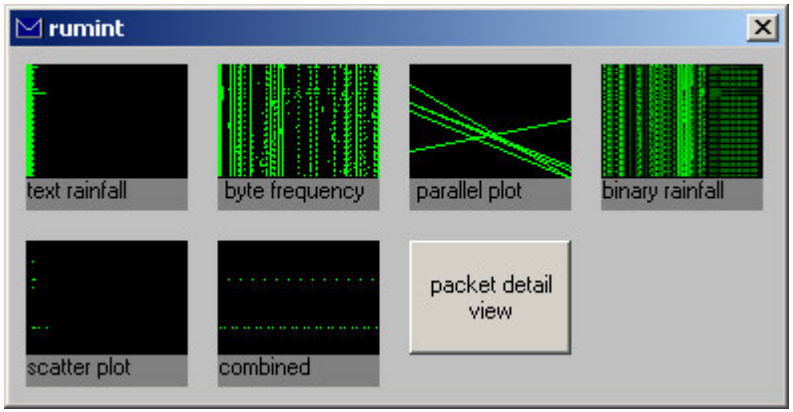



rumint
File
Toolbars
View
Help
919
Buffer
1
919
Max Speed
(pkts/sec)
1000
Max
Min
loop
screenshots
clear screen
<<
Play
Pause
Stop
>>

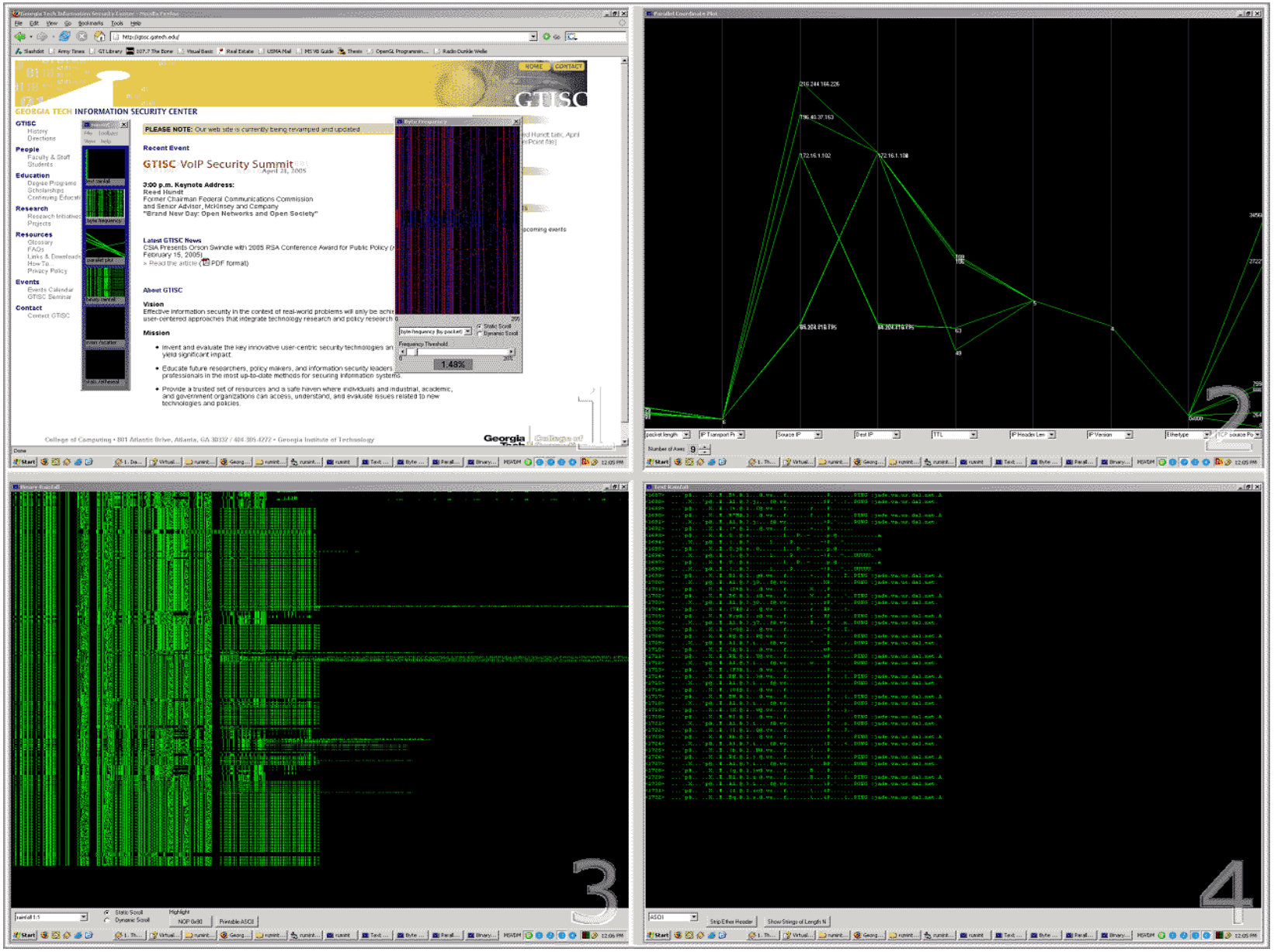
Filtering, Encoding & Interaction





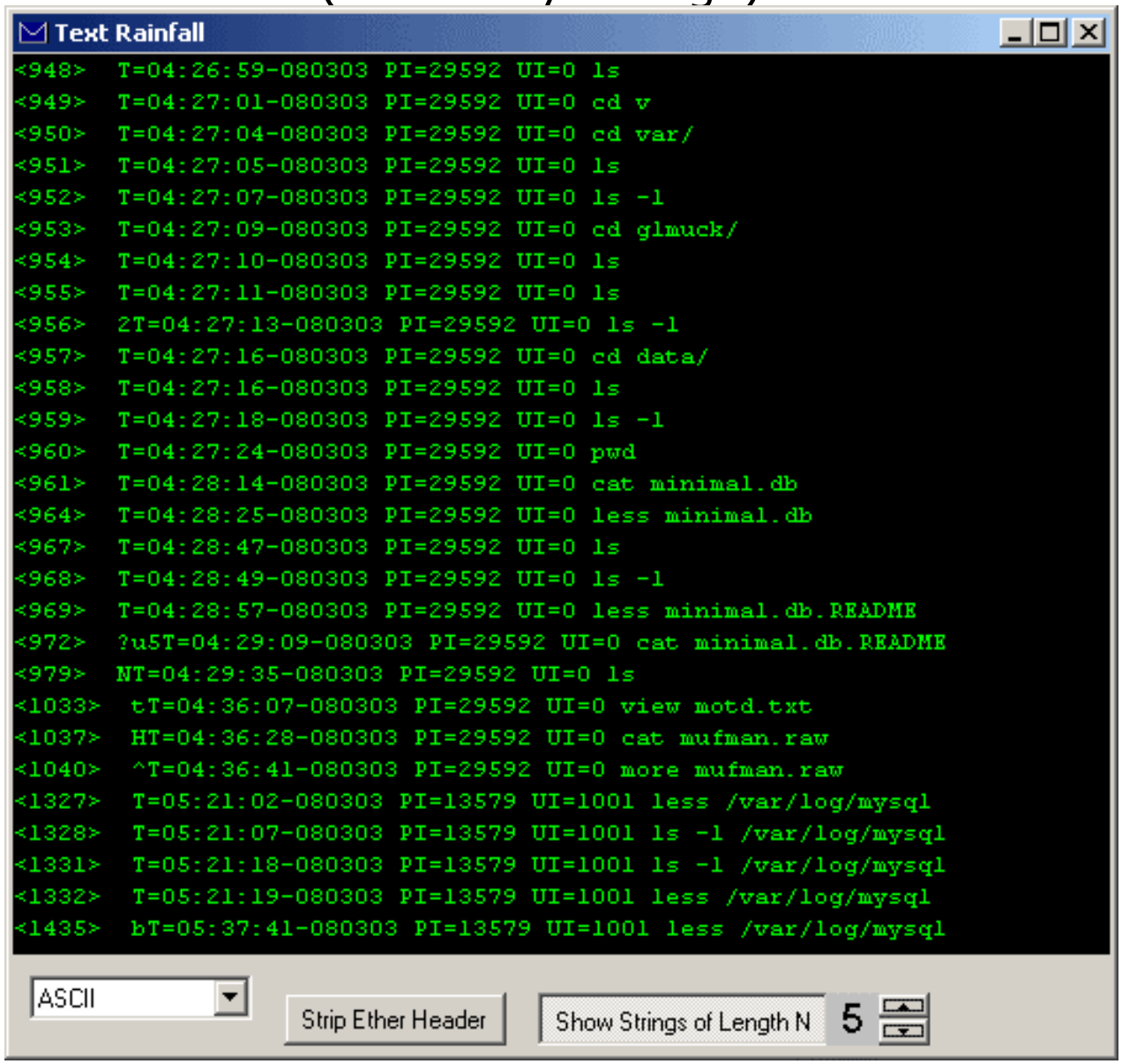


Multiple Coordinated Views…



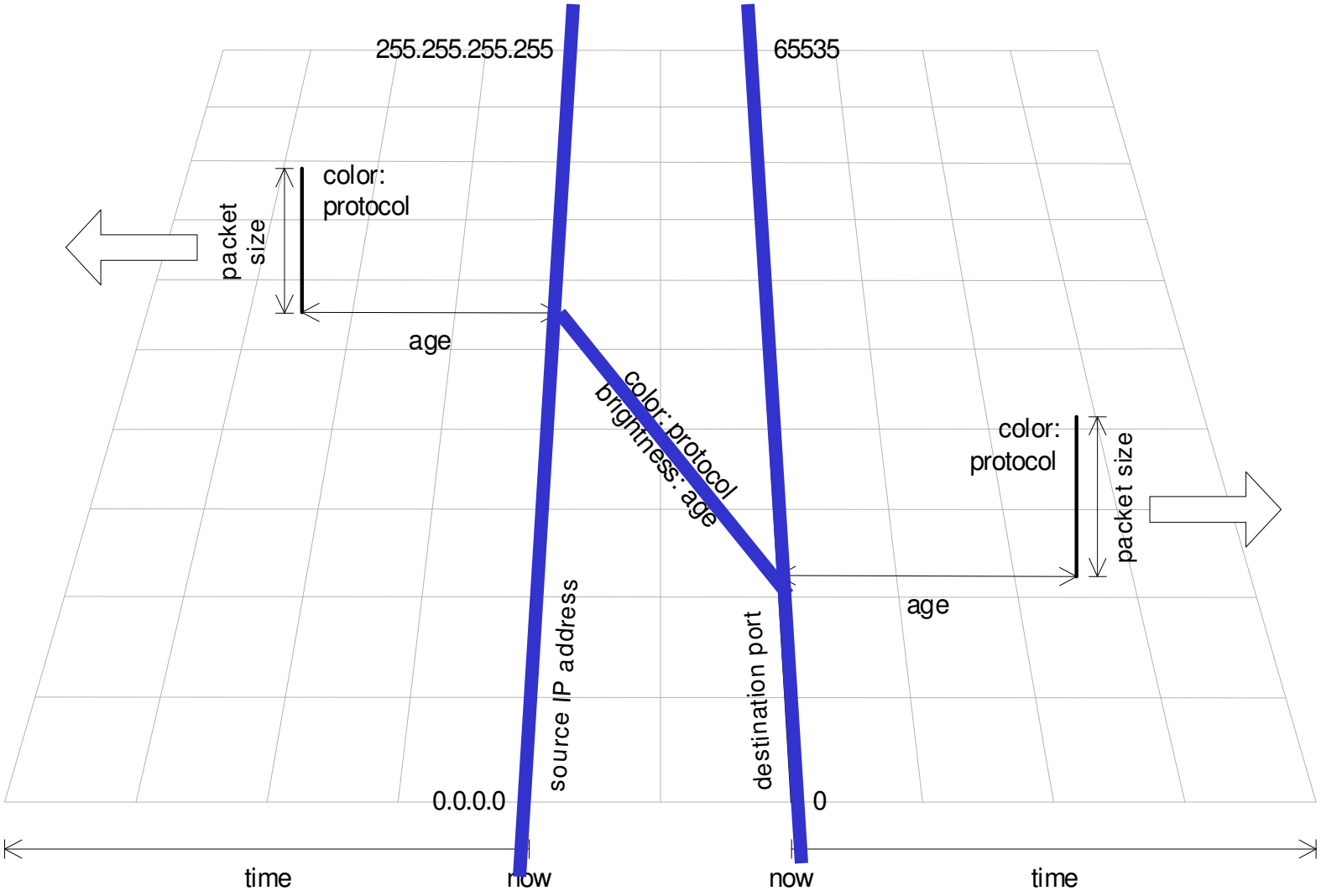
Text

(on the fly strings)

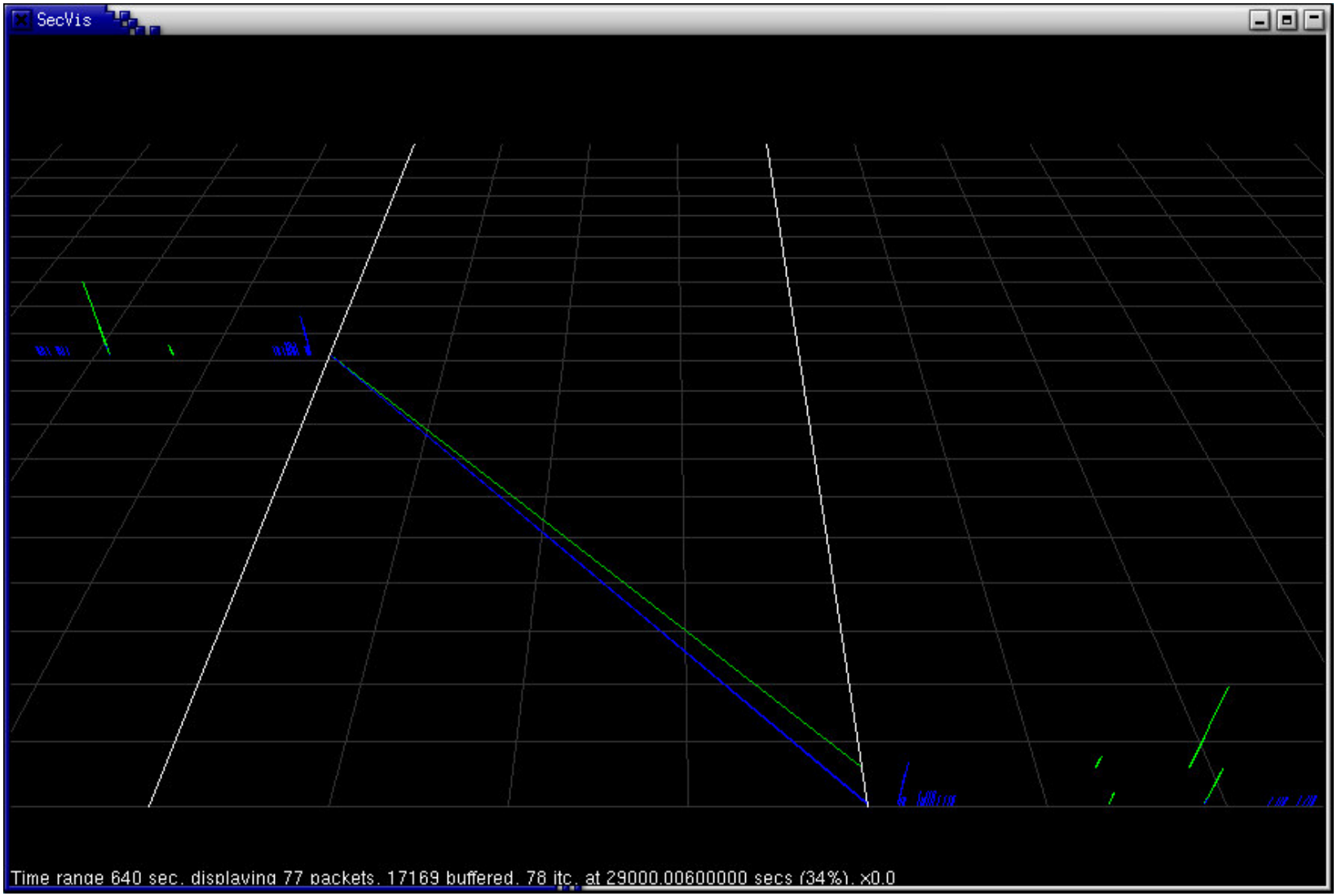


dataset: Defcon 11 CTF

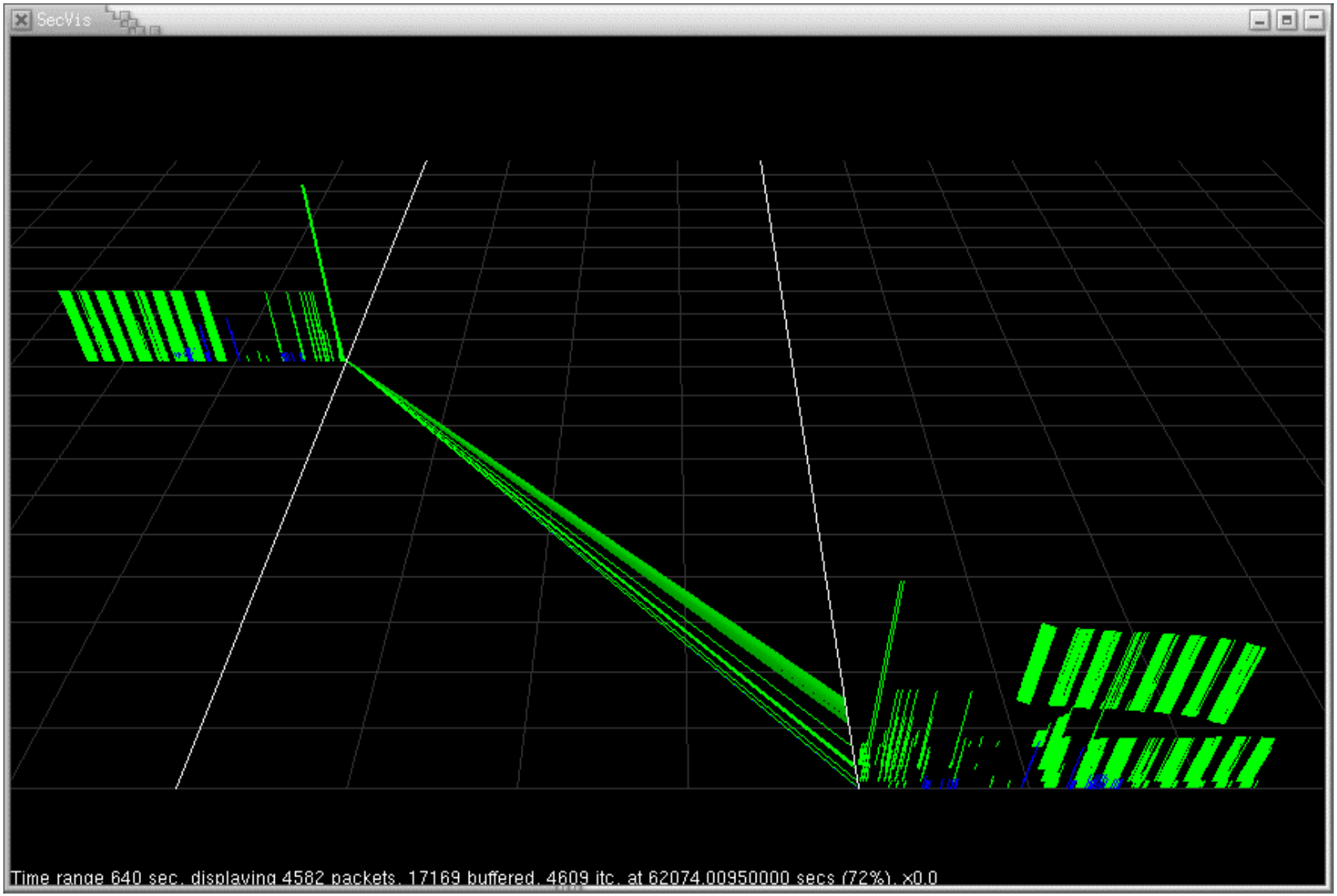
Krasser Visualization



Routine Honeynet Traffic
(baseline)



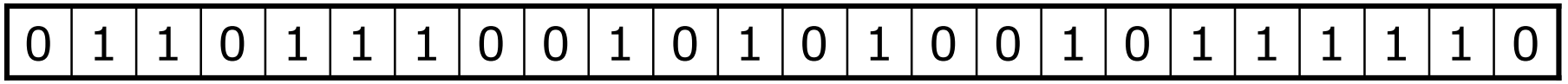
Compromised Honeypot



Binary Rainfall Visualization

(single packet)

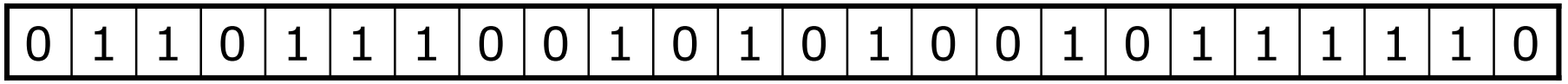
Bits on wire…



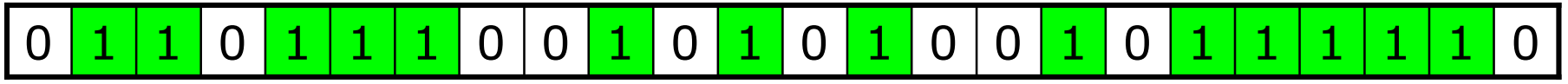
Binary Rainfall Visualization

(single packet)

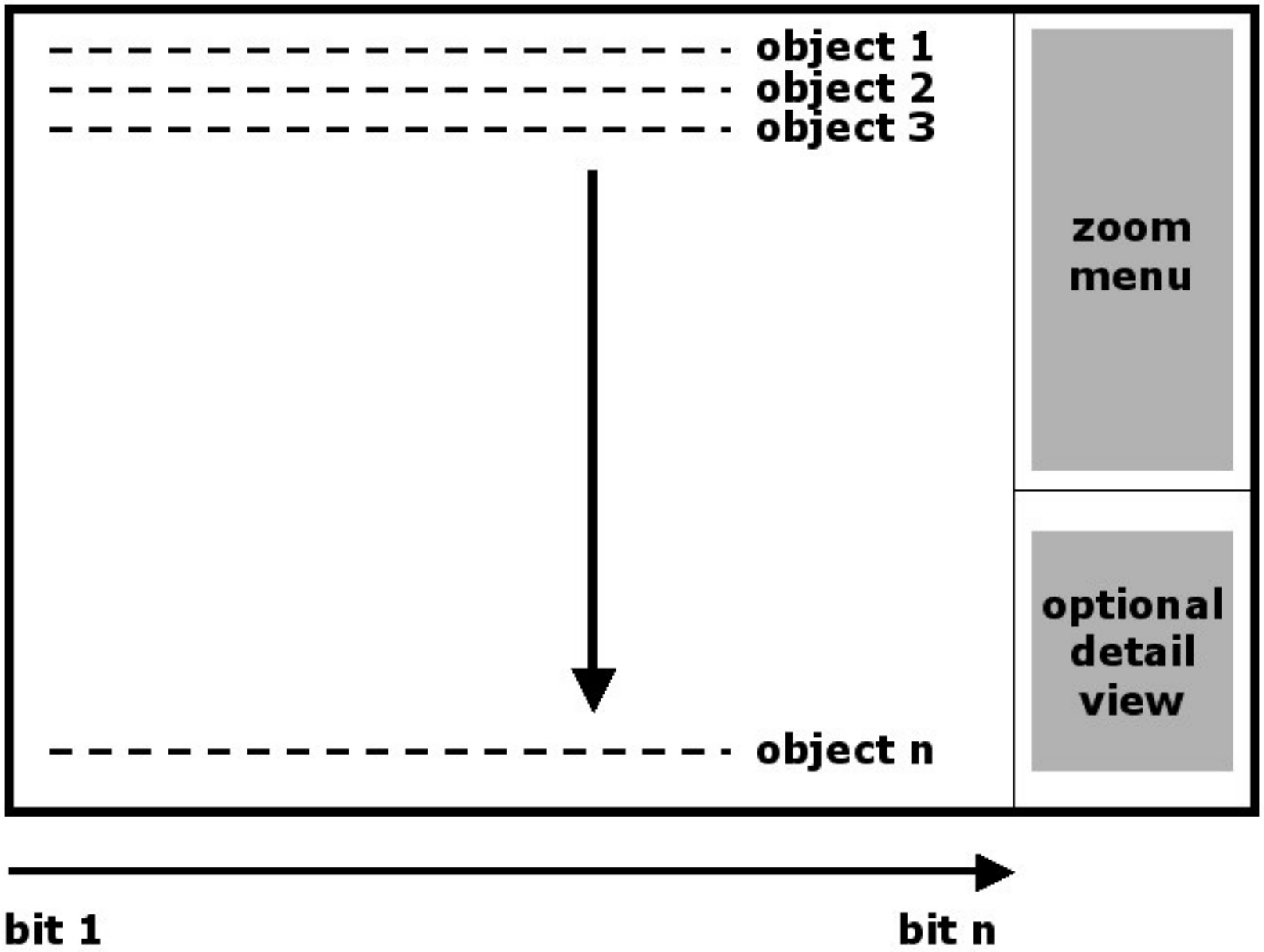
Bits on wire…



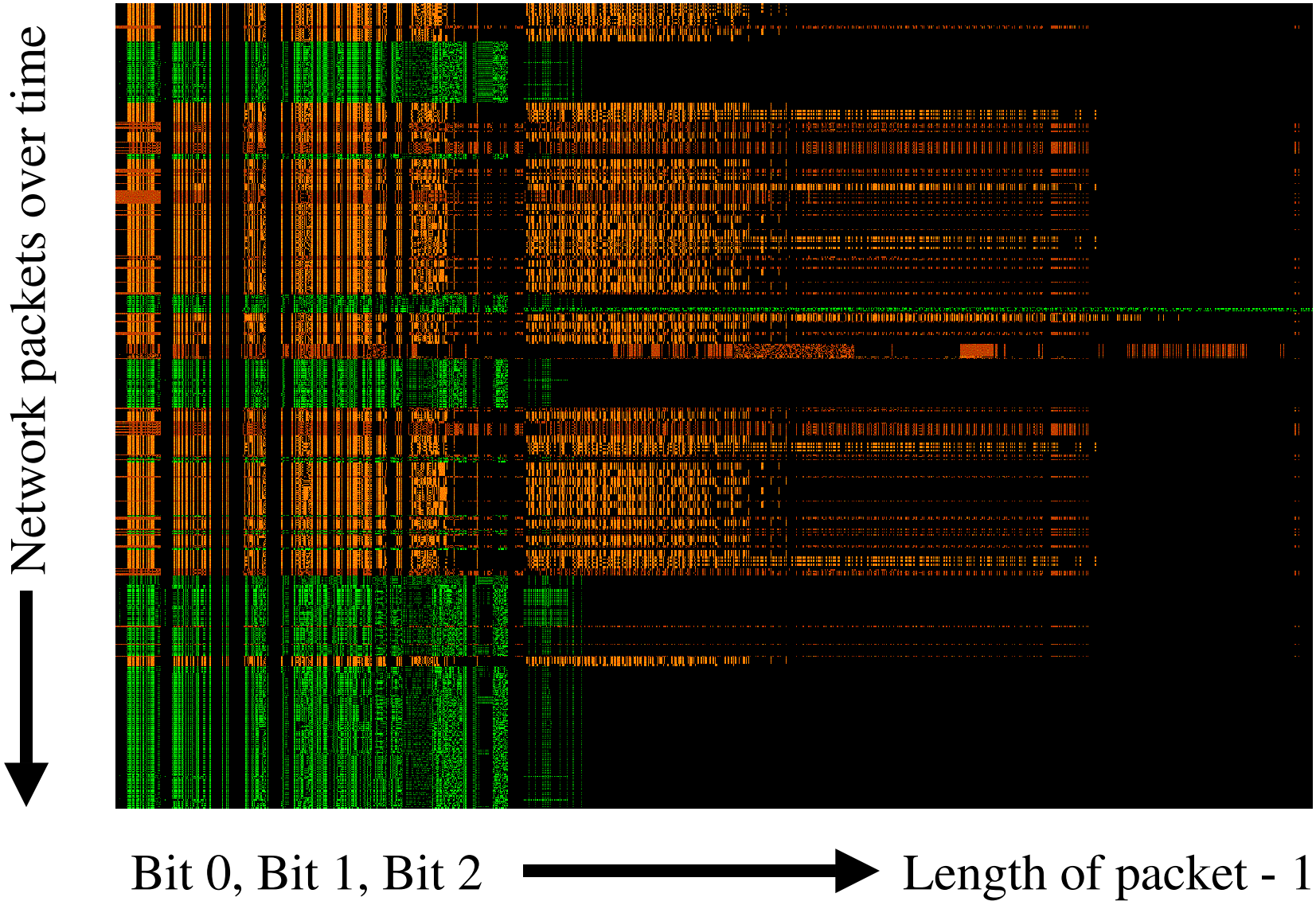
View as a 1:1 relationship (1 bit per pixel)…



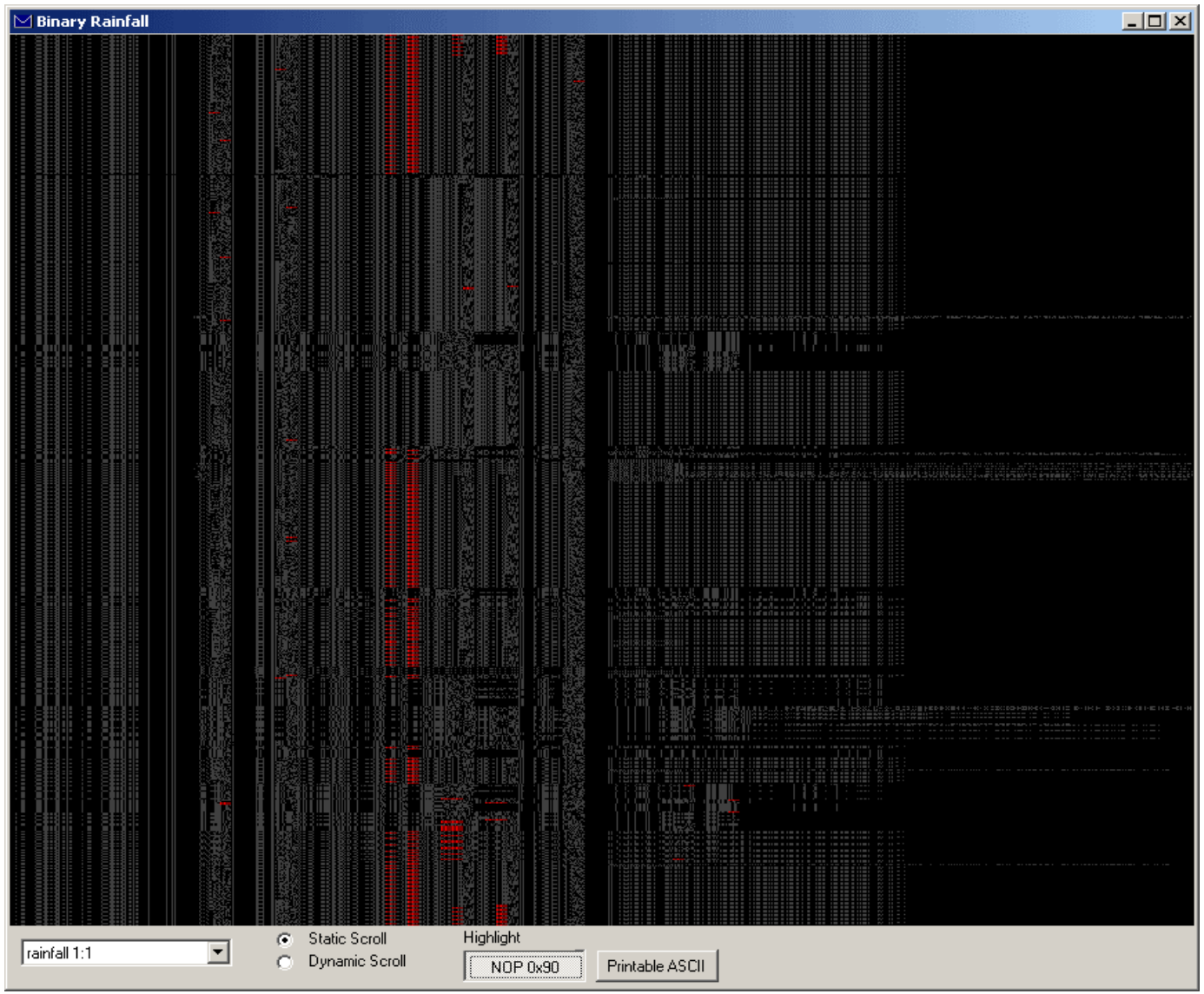
24 Pixels


object 1
object 2
object 3
zoom menu
optional detail view
object n
bit 1
bit n

Encode by Protocol



On the fly *disassembly*?

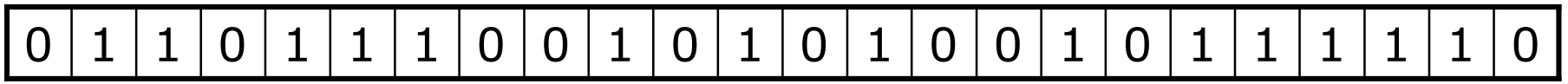


dataset: Honeynet Project Scan of the Month 21

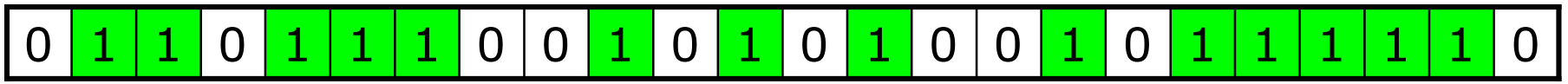
Binary Rainfall Visualization

(single packet)

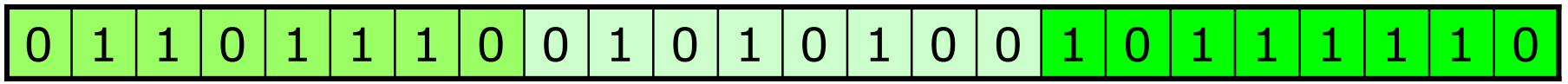
Bits on wire…



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

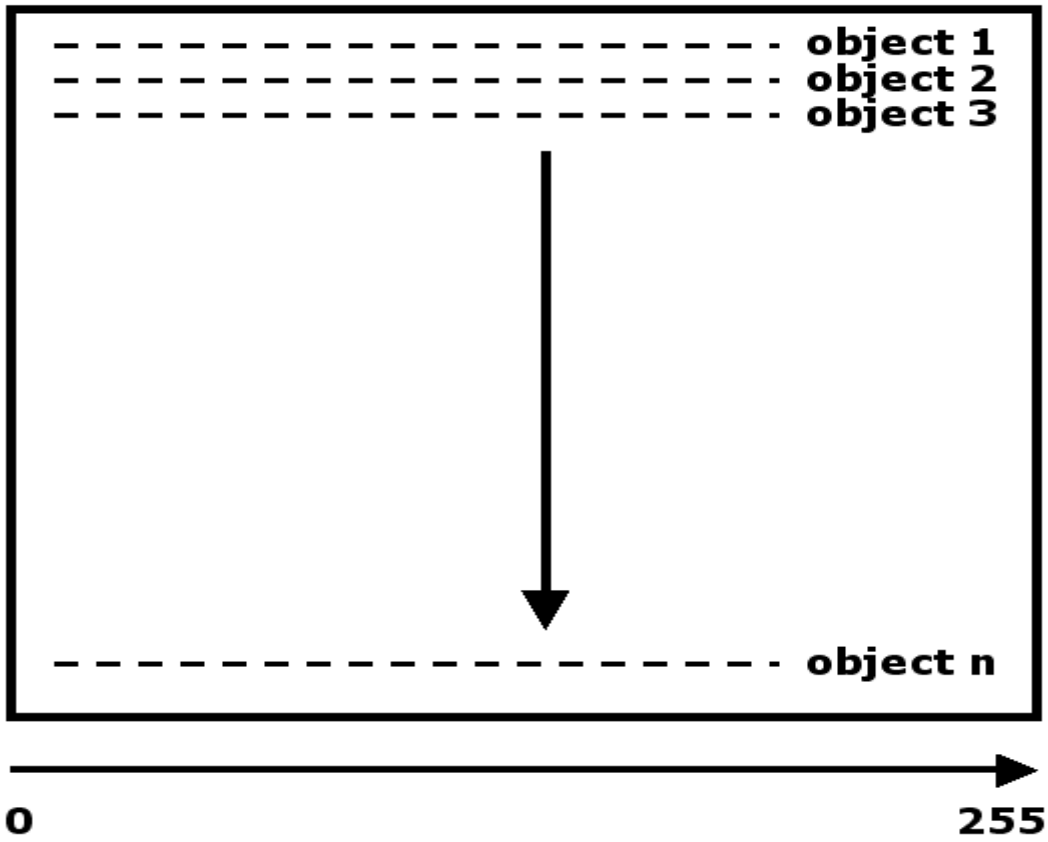


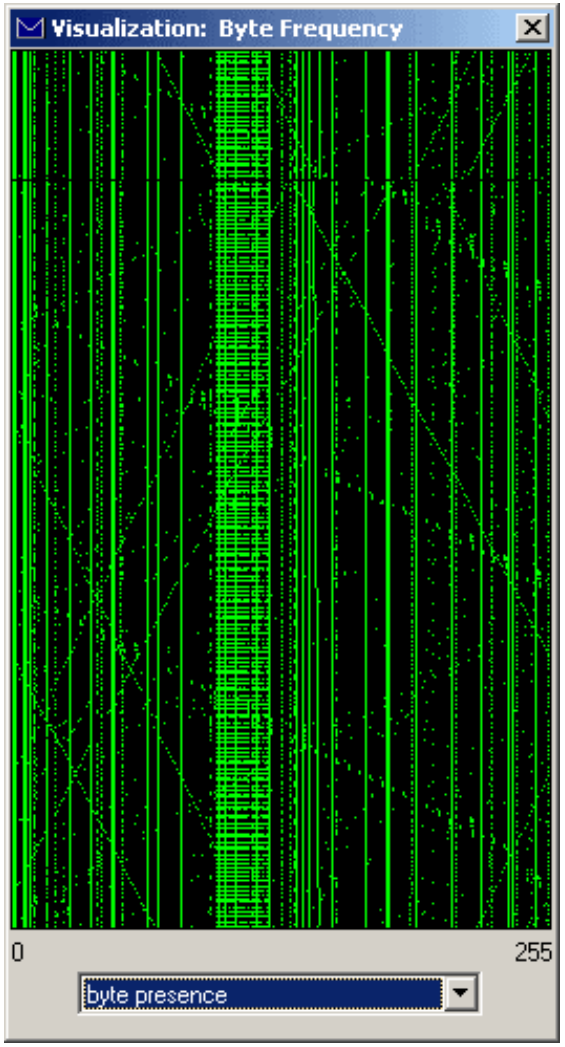
View as a 8:1 relationship (1 byte per pixel)…

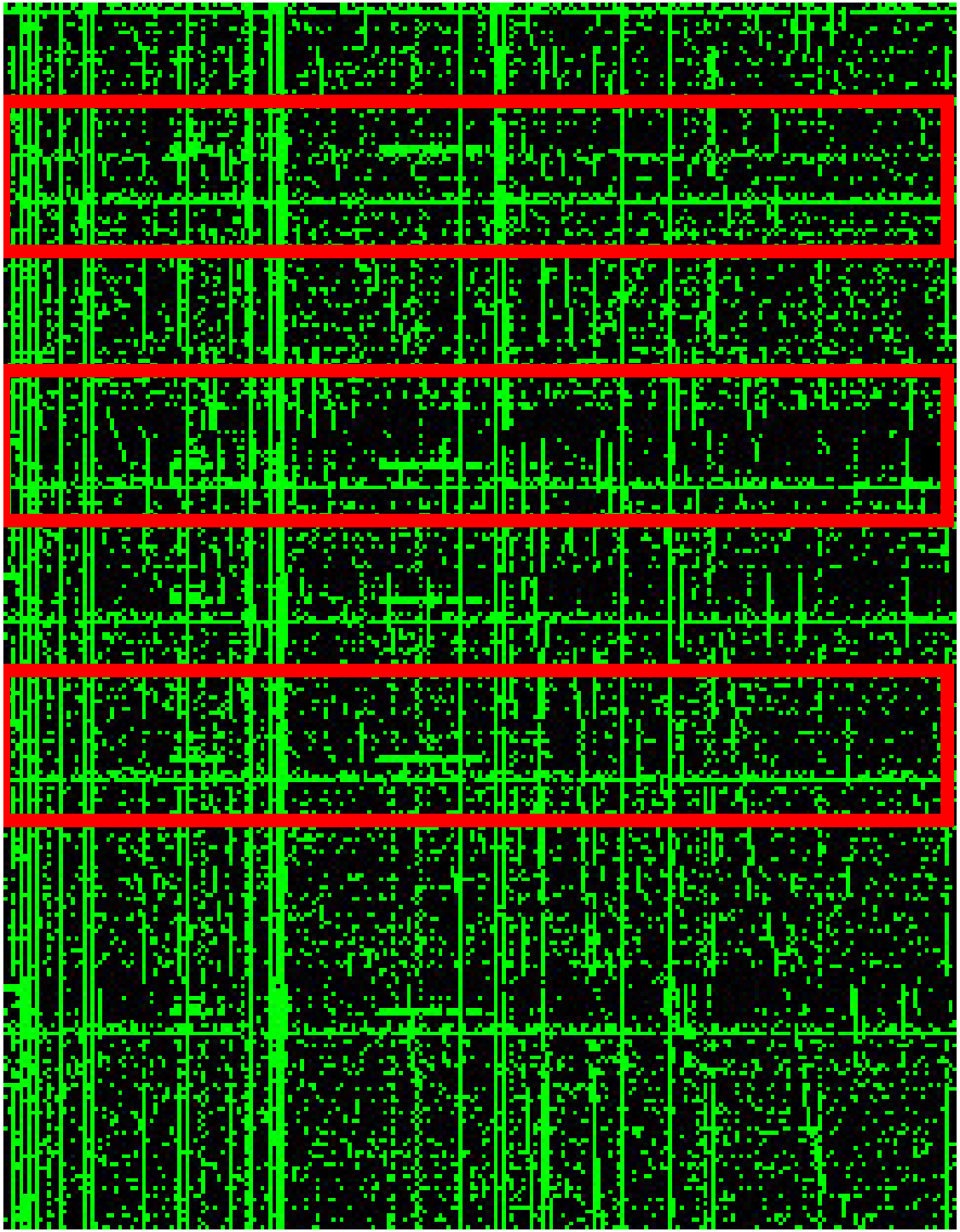


3 Pixels

Byte Visualization

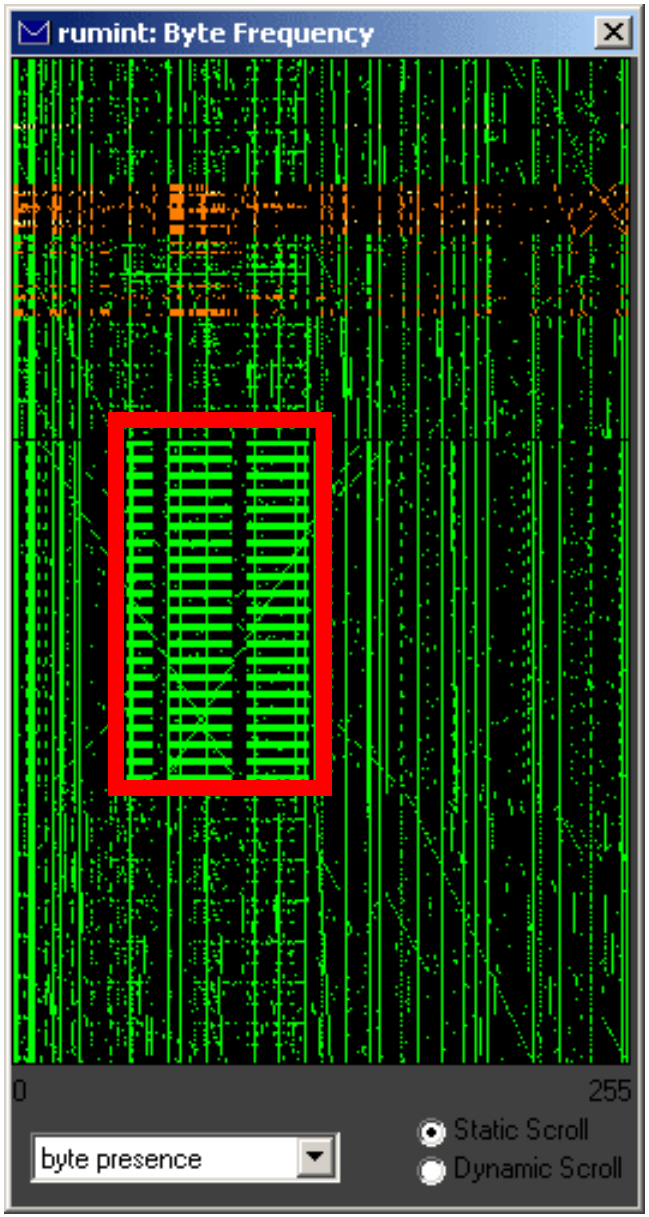




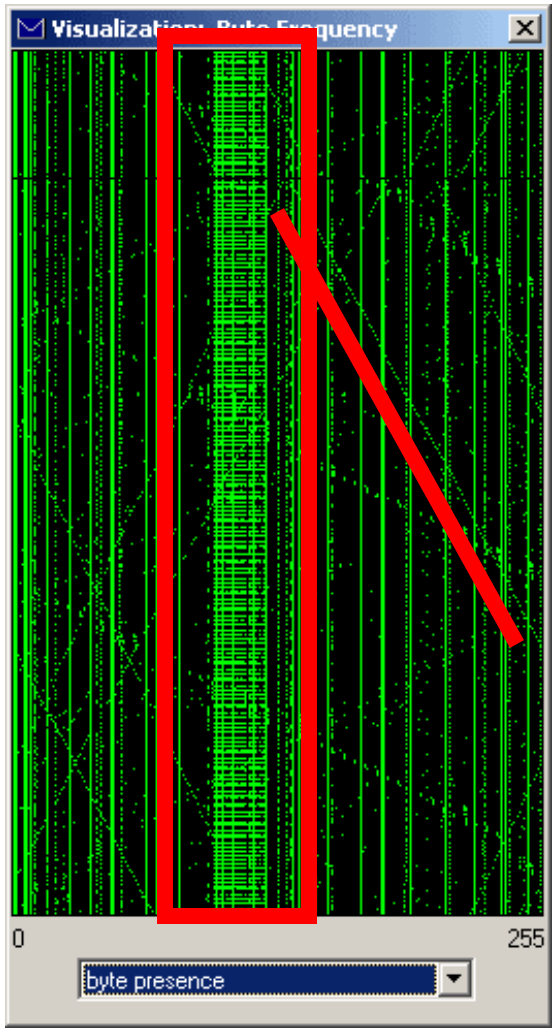


Open SSH Diffie-Hellman Key Exchange

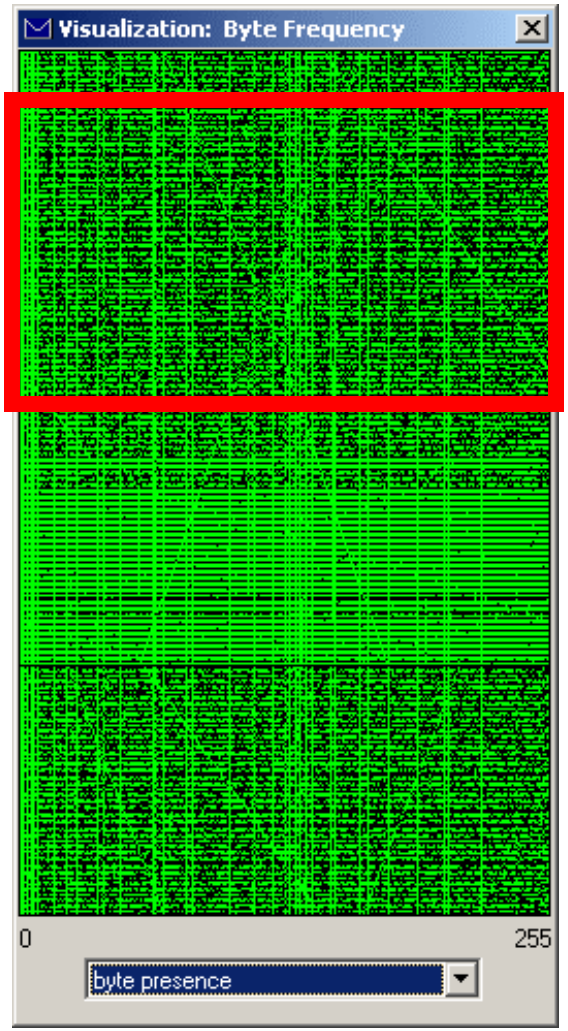
Zipped Email Attachment



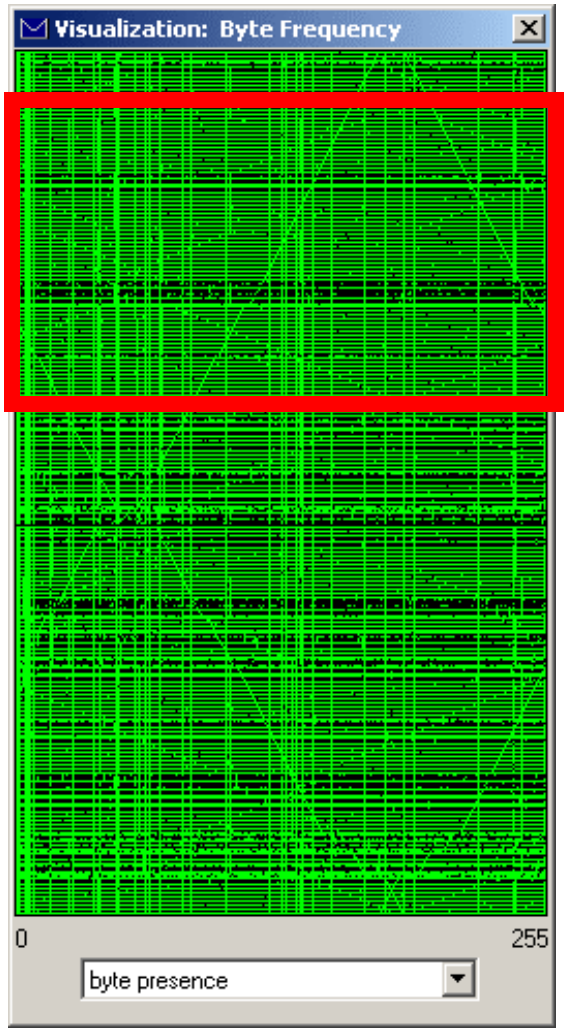
Byte Presence



dictionary file via HTTP

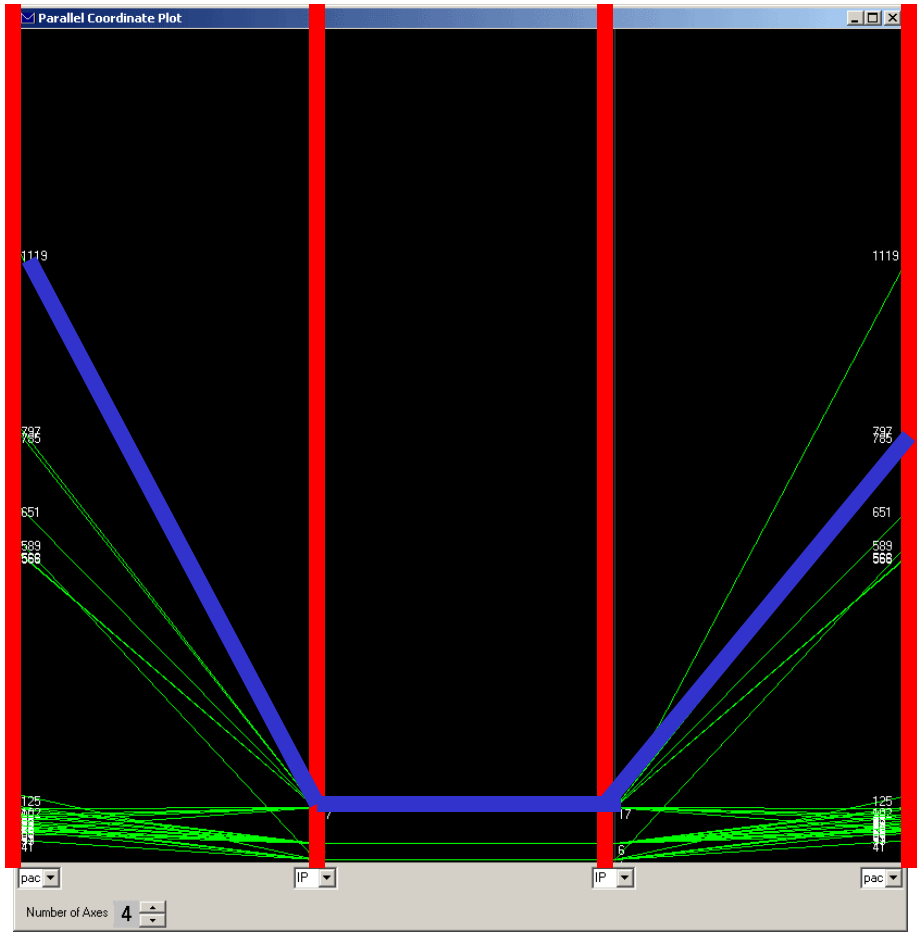


ssh



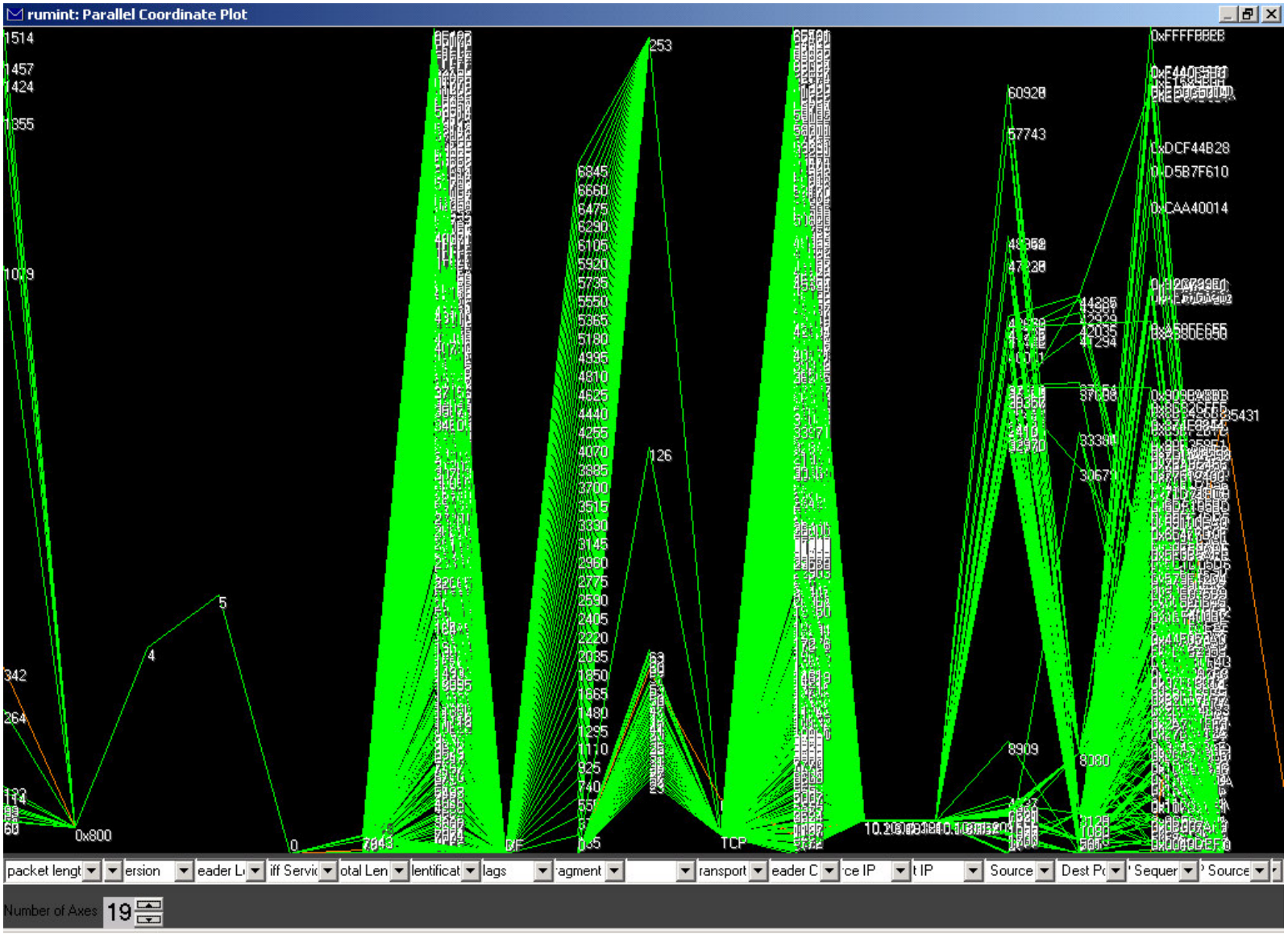
SSL

Parallel Coordinates



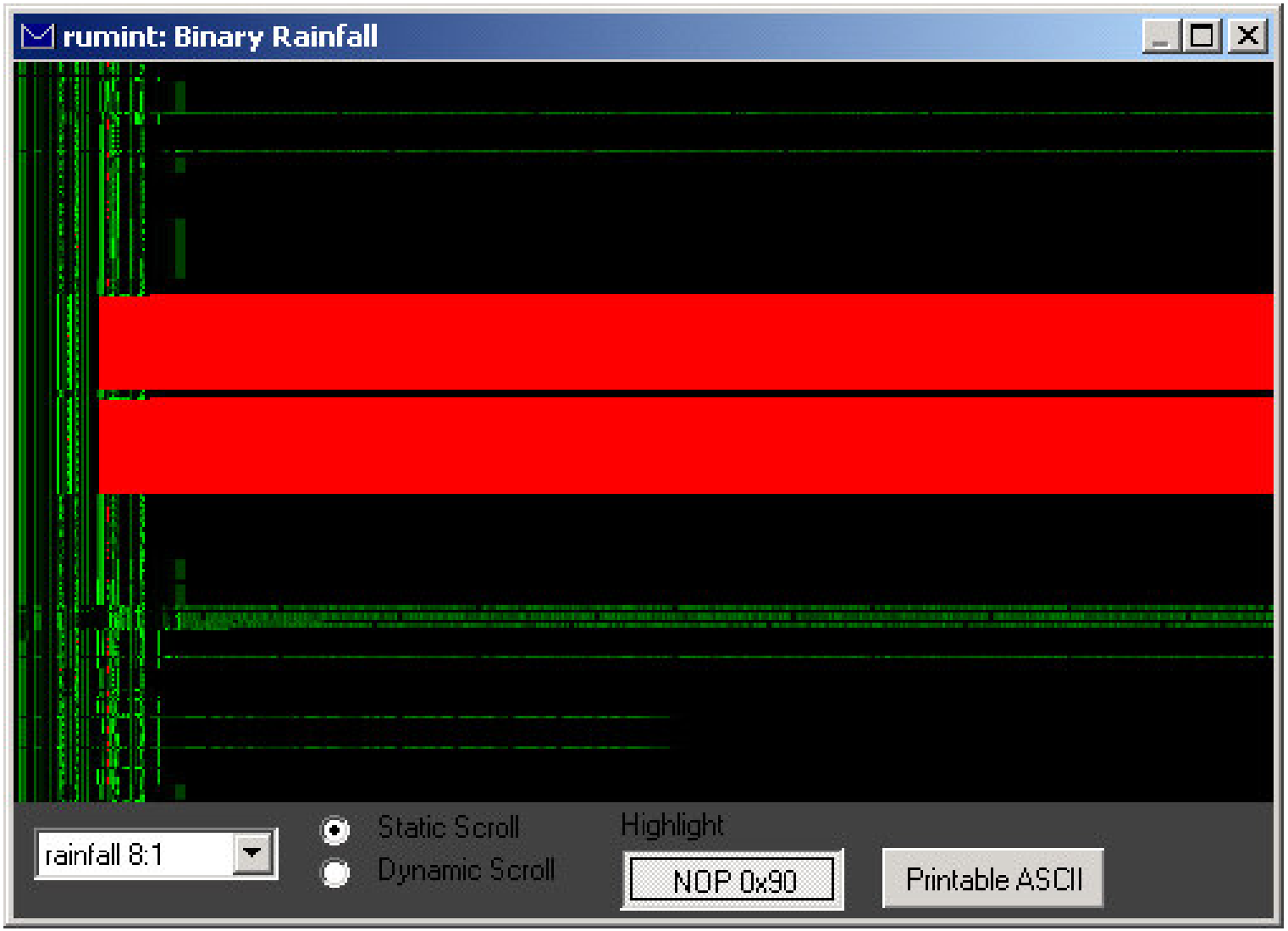
- goal: plot any data fields
- dynamic columns
- change order for different insight
- intelligent lookup and translation of fields
  - e.g. IP transport protocol

Rapidly Characterize Packet Header Fields (Context)

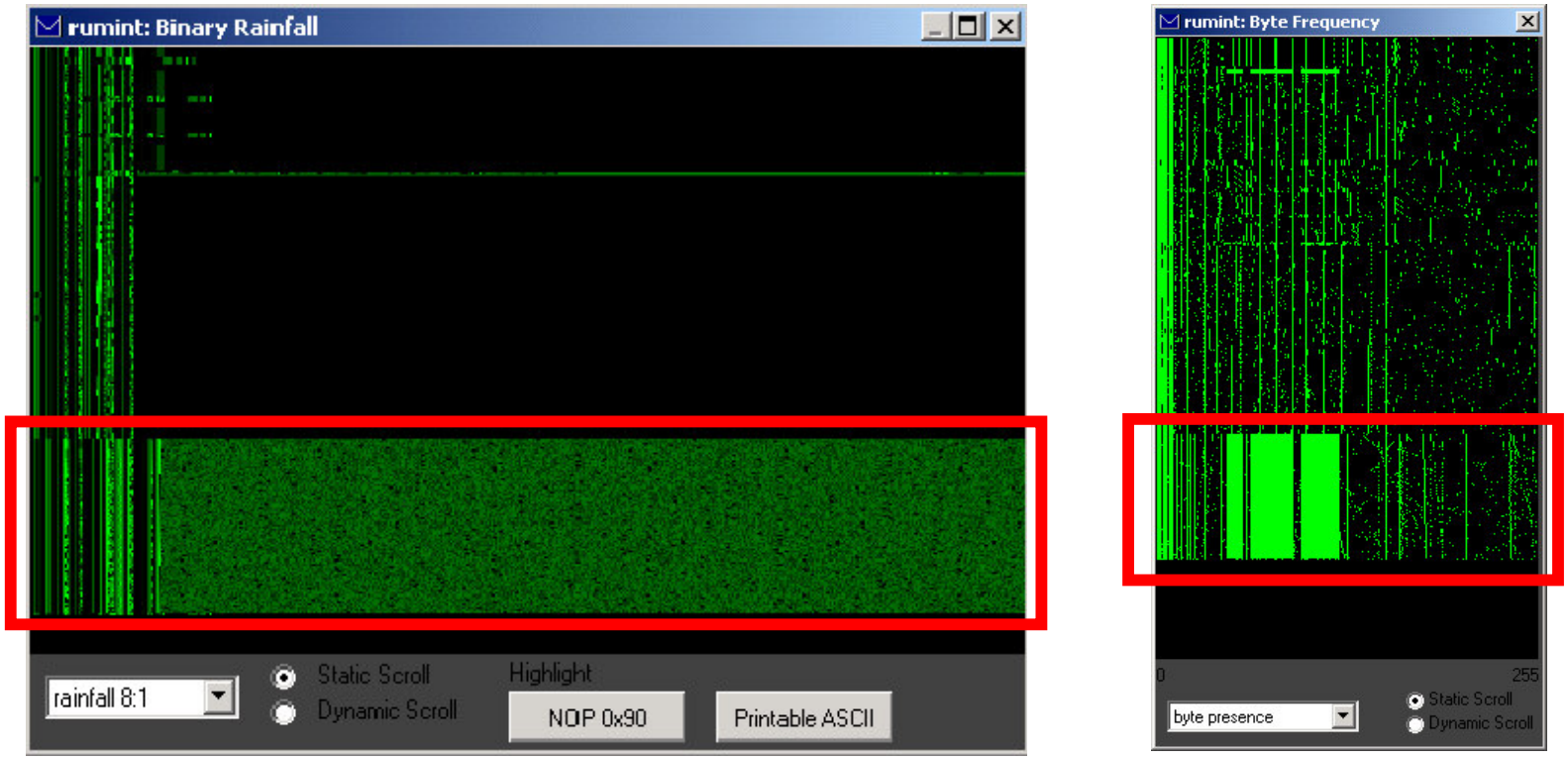


Identify and Precisely Locate Fragmentation Anomaly

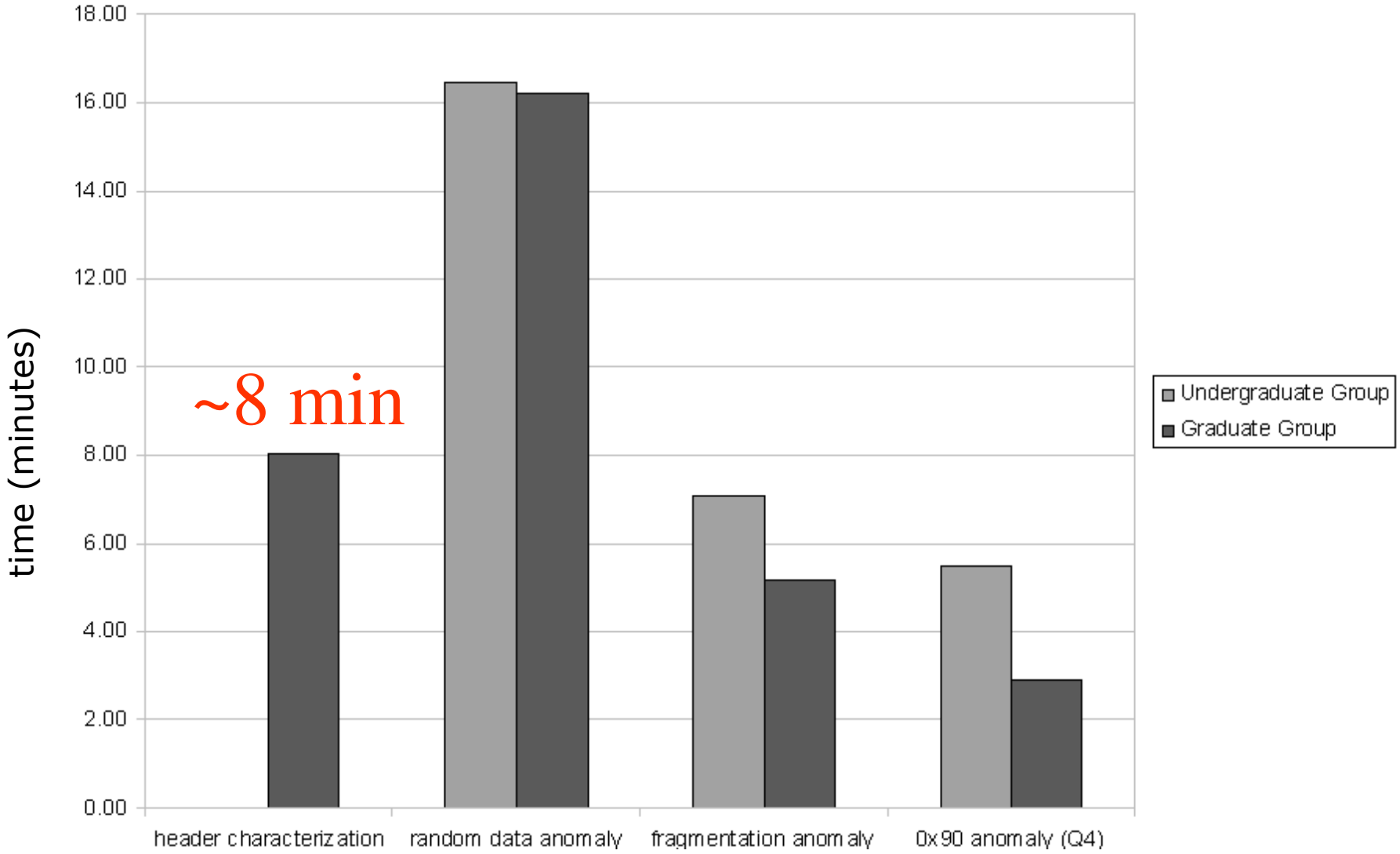
Identify and Precisely Locate x90 Anomaly



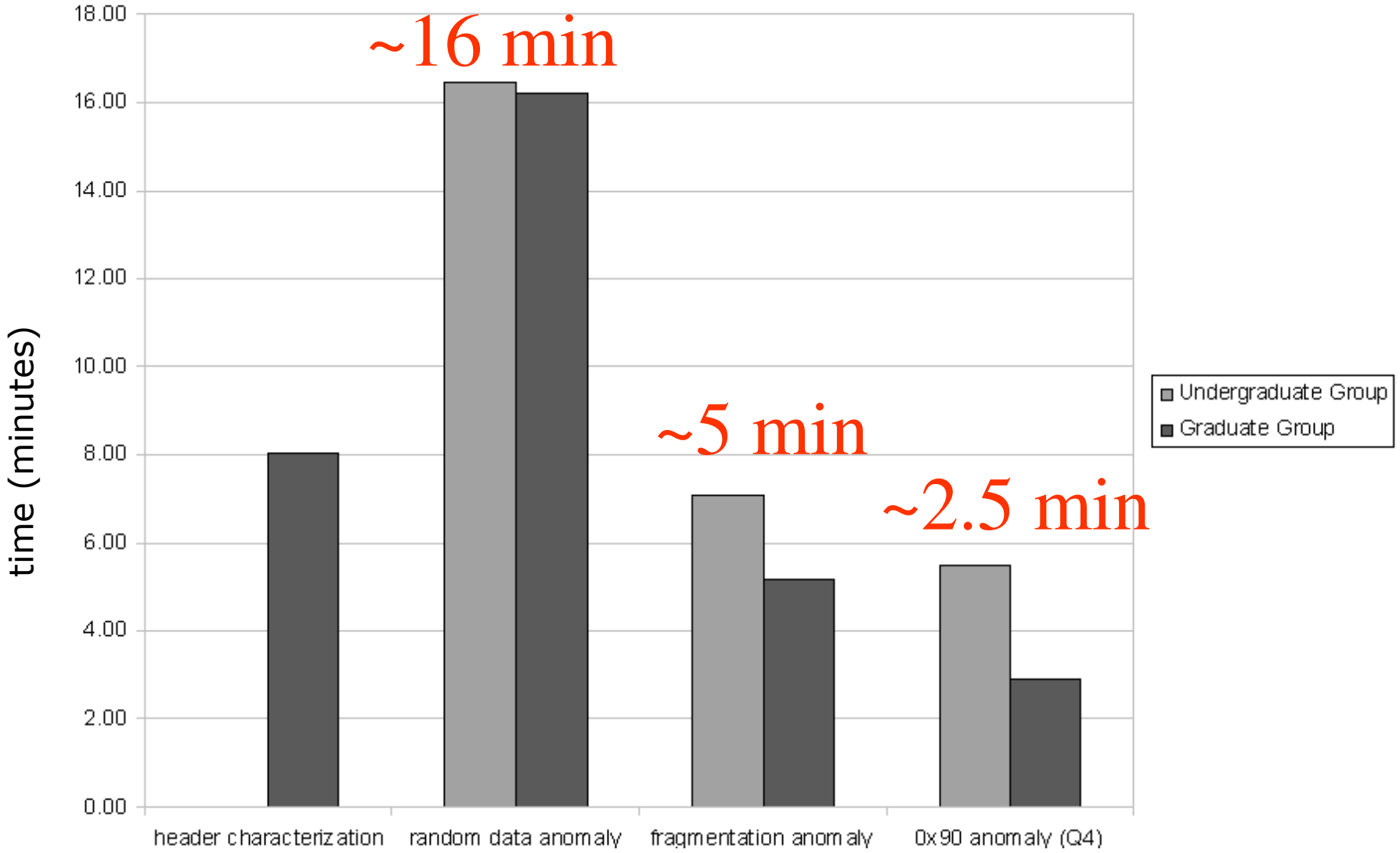
Identify and Precisely Locate Possible Random Payload Anomaly



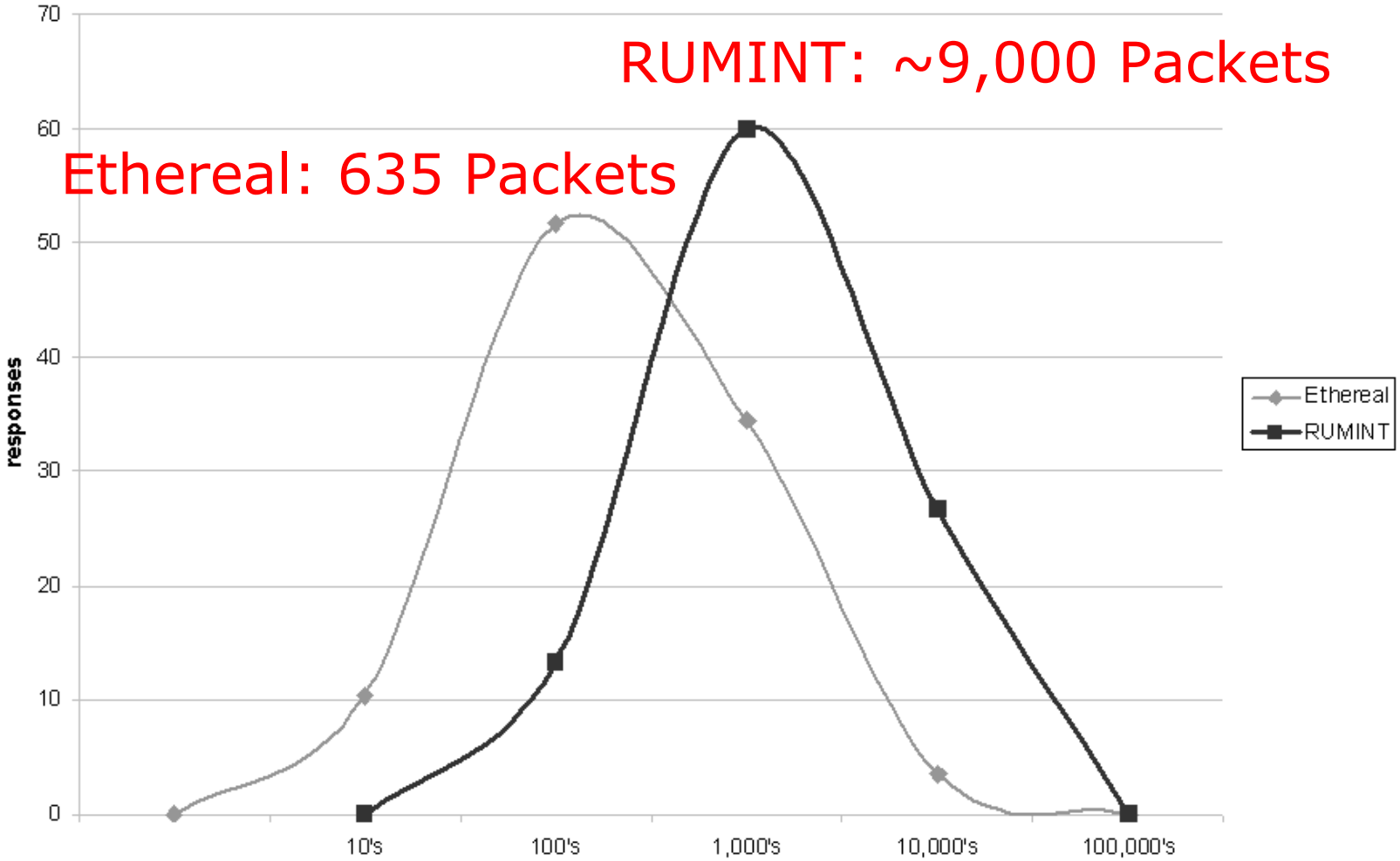
Task Completion Time



Task Completion Time



RUMINT Tipping Point



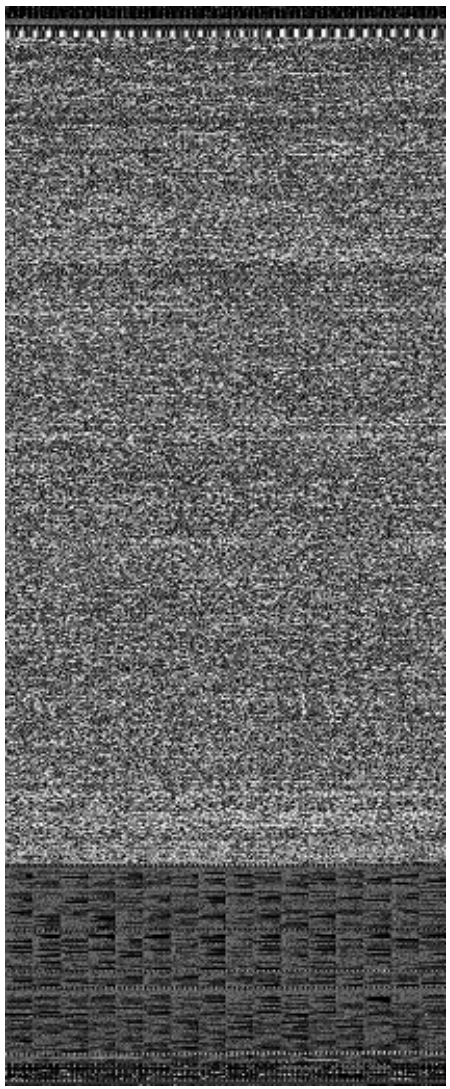
Demo

System Requirements

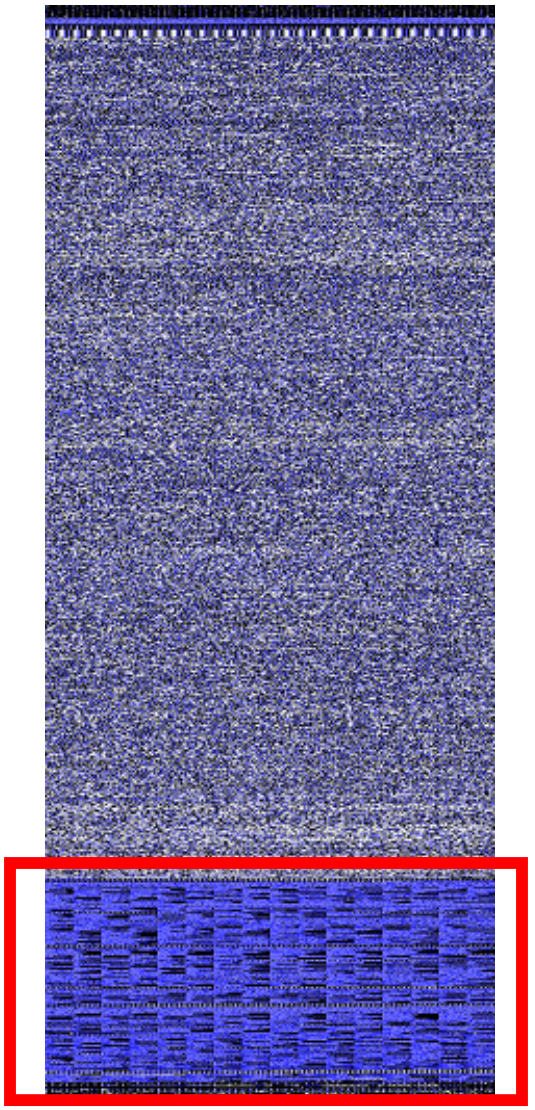
- IP over Ethernet
- Tested on Windows XP
- ~256+ MB Ram
- Processor 300MHZ (minimum)
- The more screen real estate the better
- Latest winpcap

Binary Visualization

(sendmail)



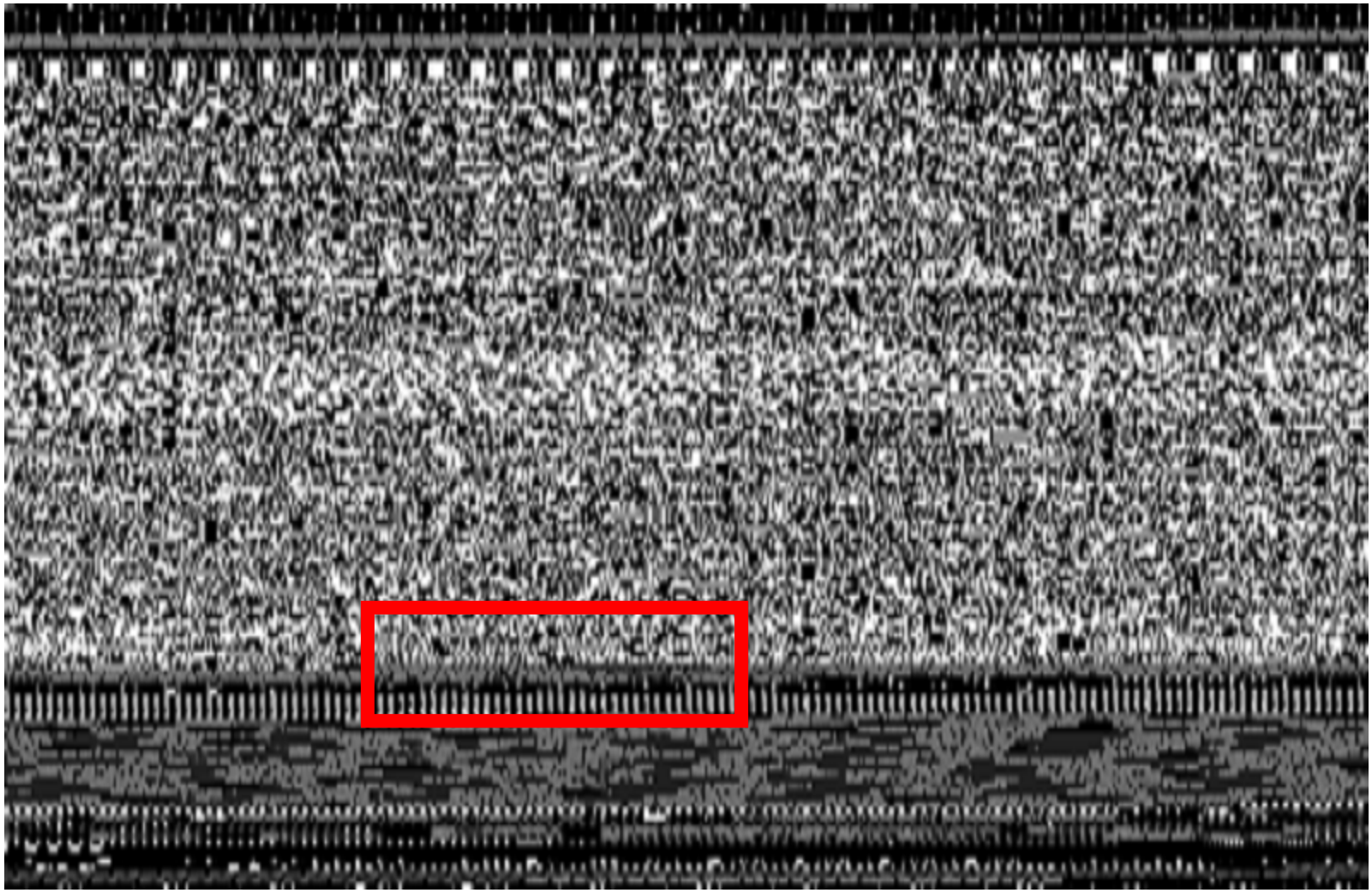
original

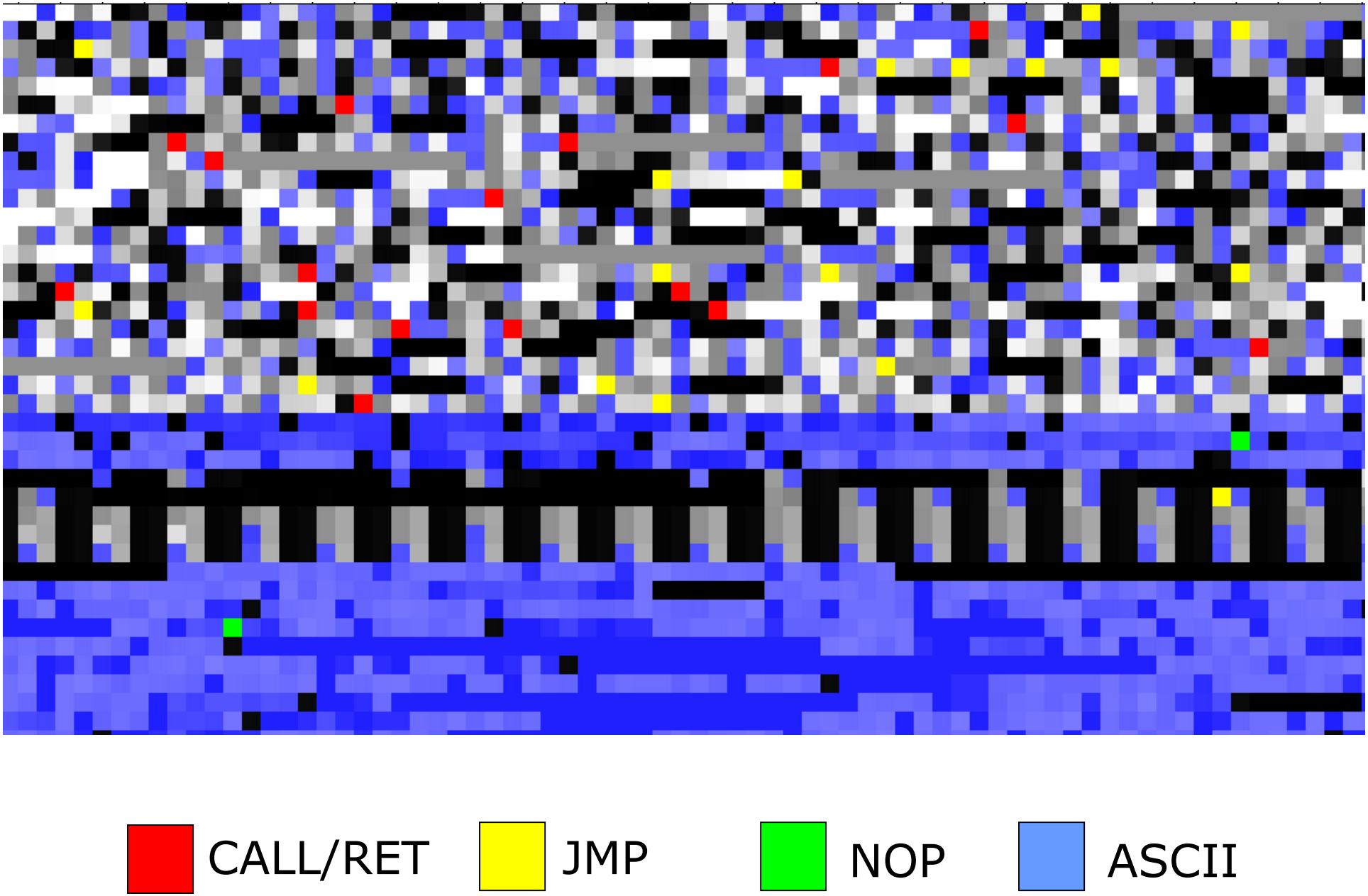


printable ASCII in blue

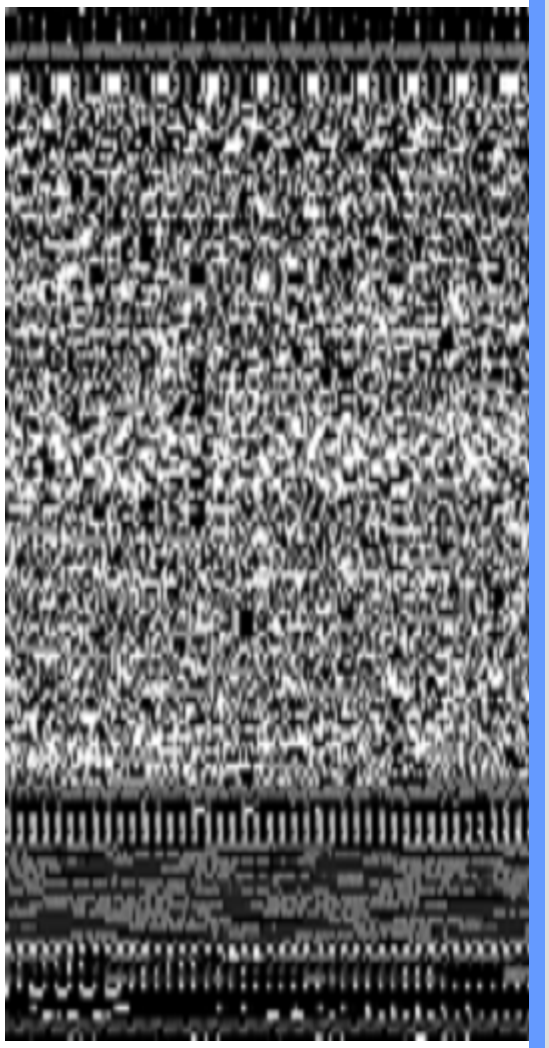
Color Encode by Opcode

(ls)



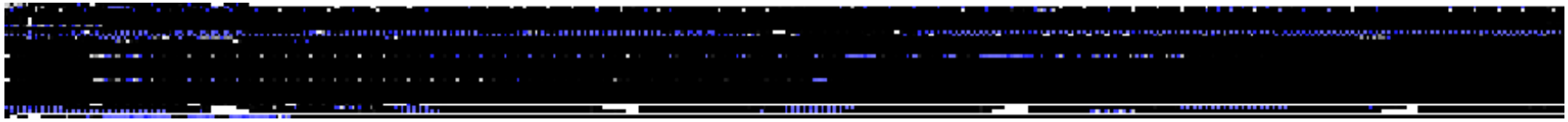

CALL/RET
JMP
NOP
ASCII

Color Encode by Disassembly?



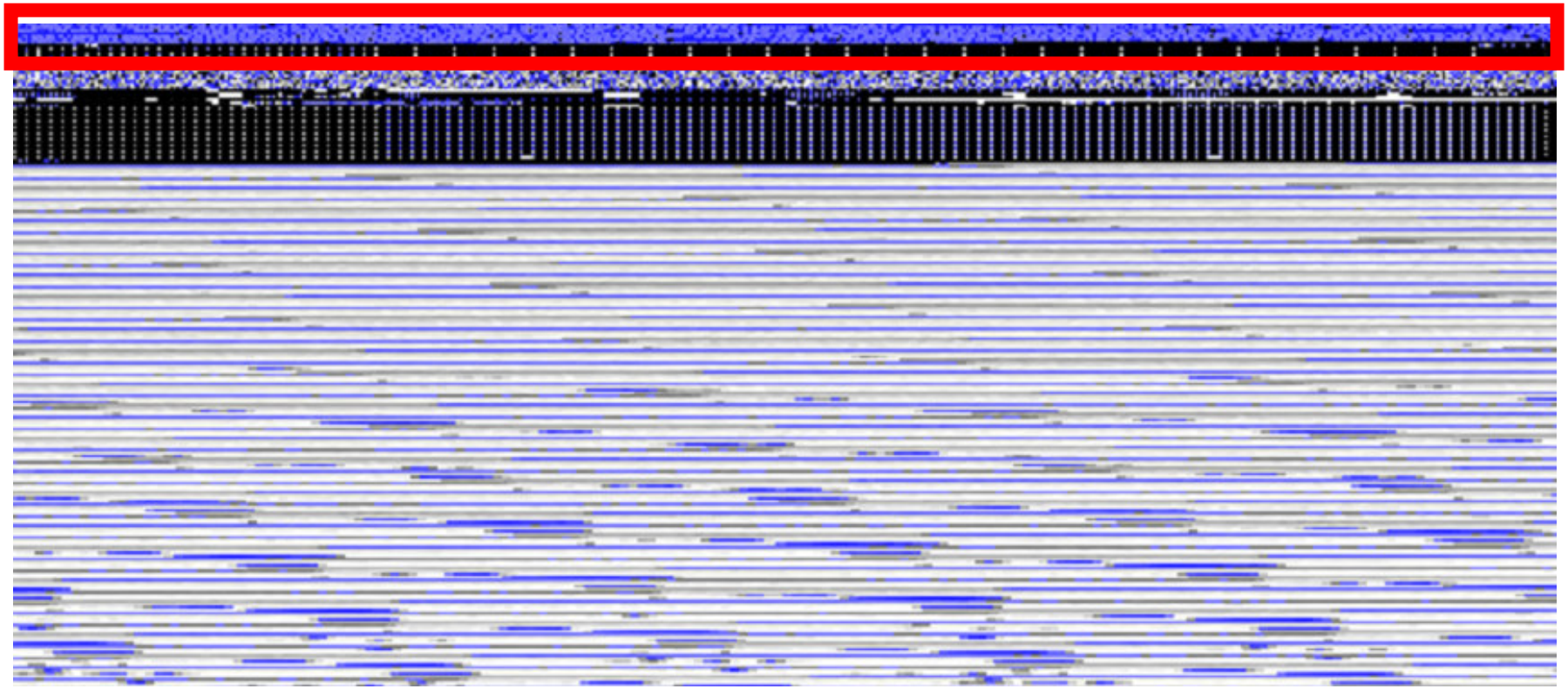
```
ls:     file format elf32-i386


Disassembly of section .init:

0804917c <.init>:
 804917c:       55                      push   %ebp
 804917d:       89 e5                   mov    %esp,%ebp
 804917f:       83 ec 08                sub    $0x8,%esp
 8049182:       e8 8d 05 00 00          call   0x8049714
 8049187:       e8 f4 05 00 00          call   0x8049780
 804918c:       e8 9f c0 00 00          call   0x8055230
 8049191:       c9                      leave
 8049192:       c3                      ret
```

empty word document



text

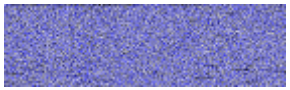
full word document (truncated)



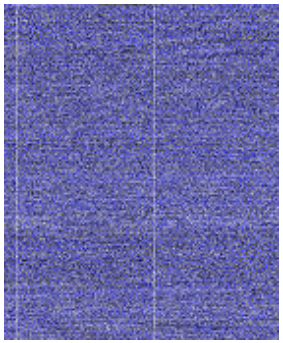
binary file analysis

original image

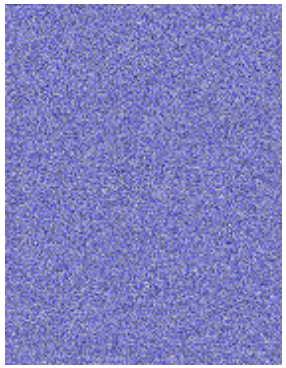
jpg



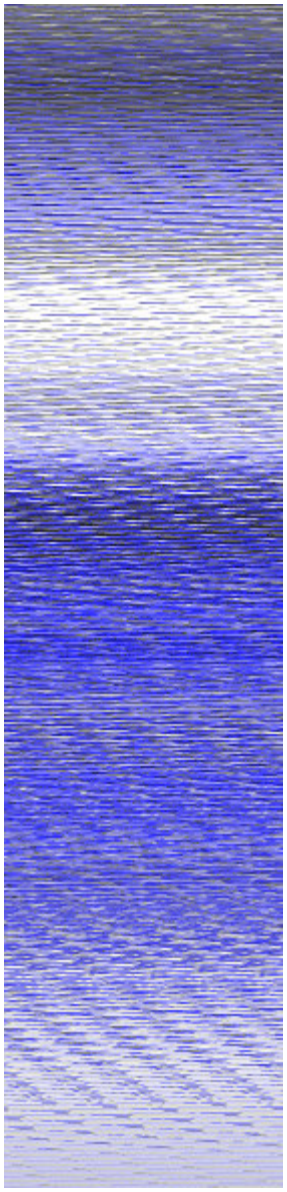
gif



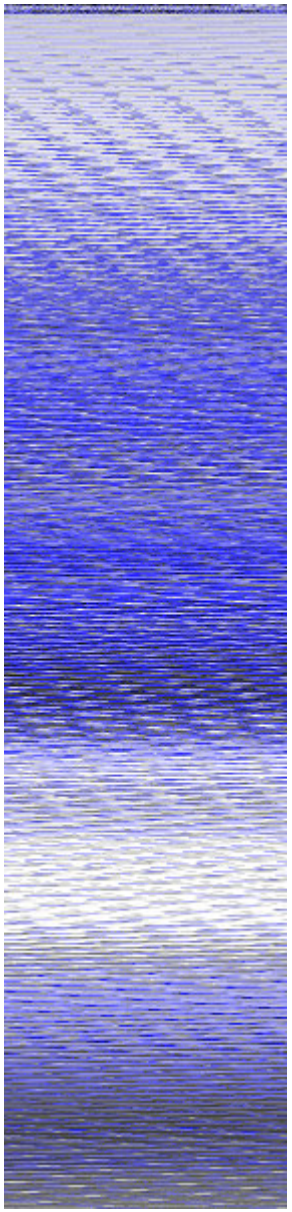
png



bmp

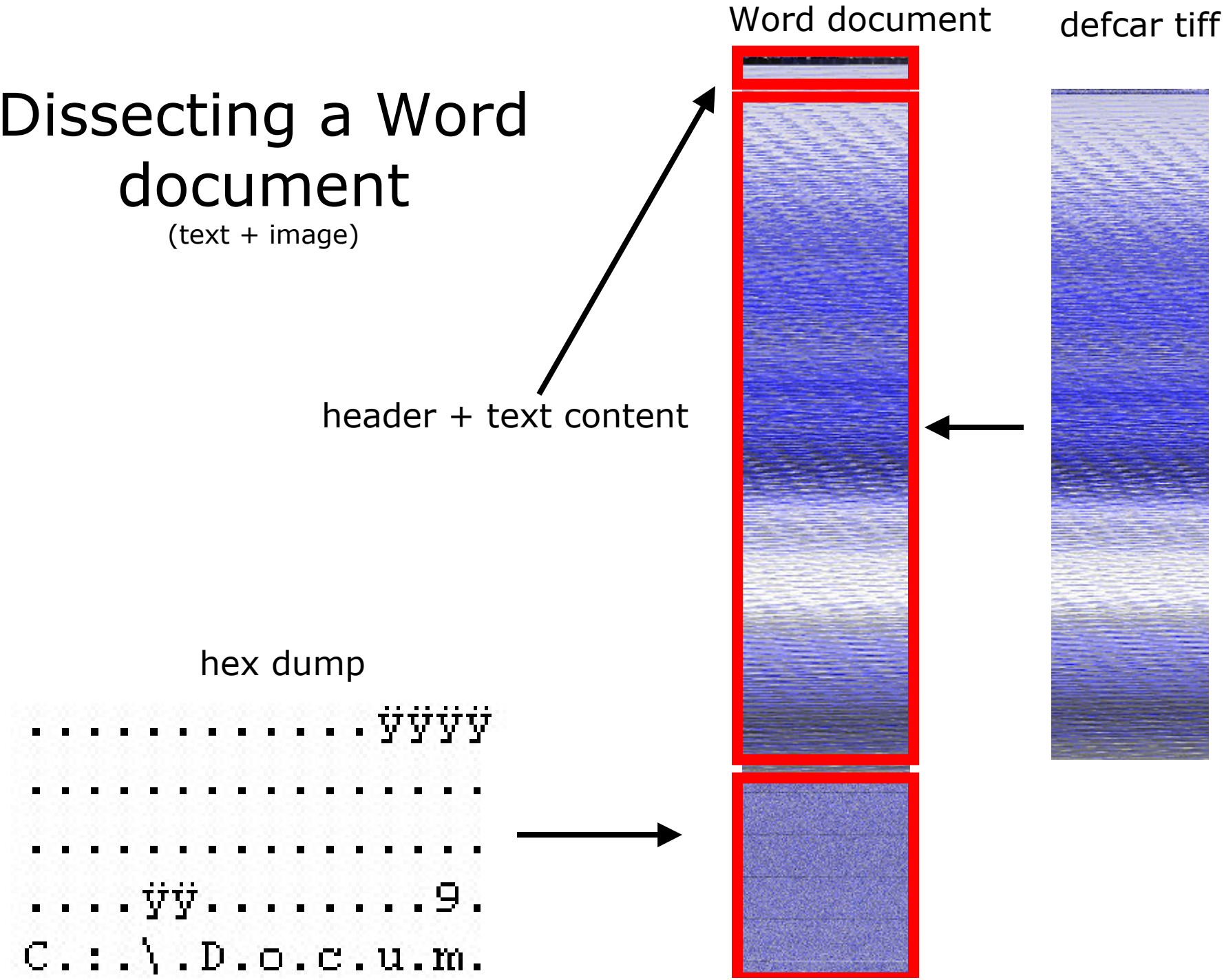


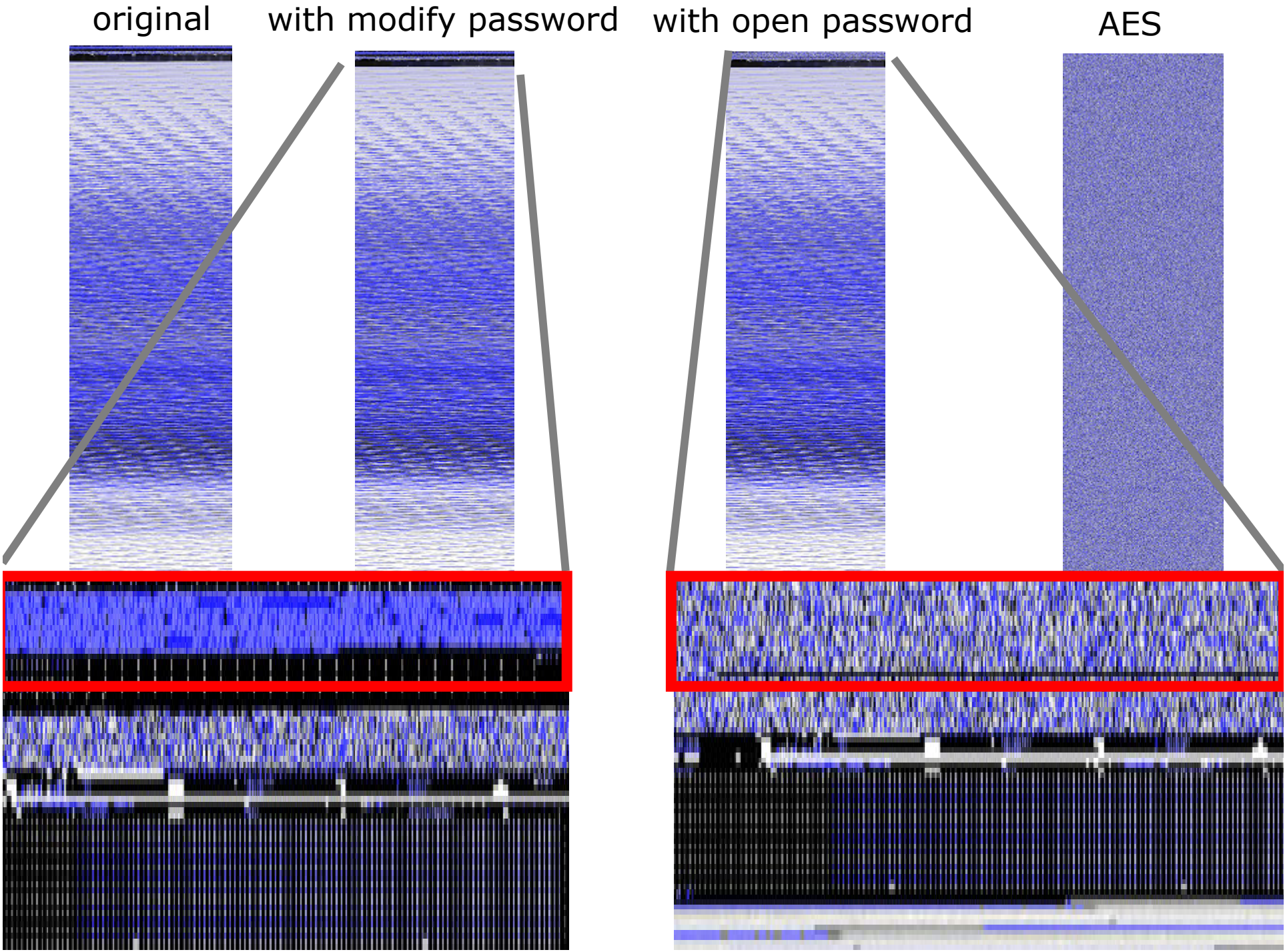
tiff



Dissecting a Word document

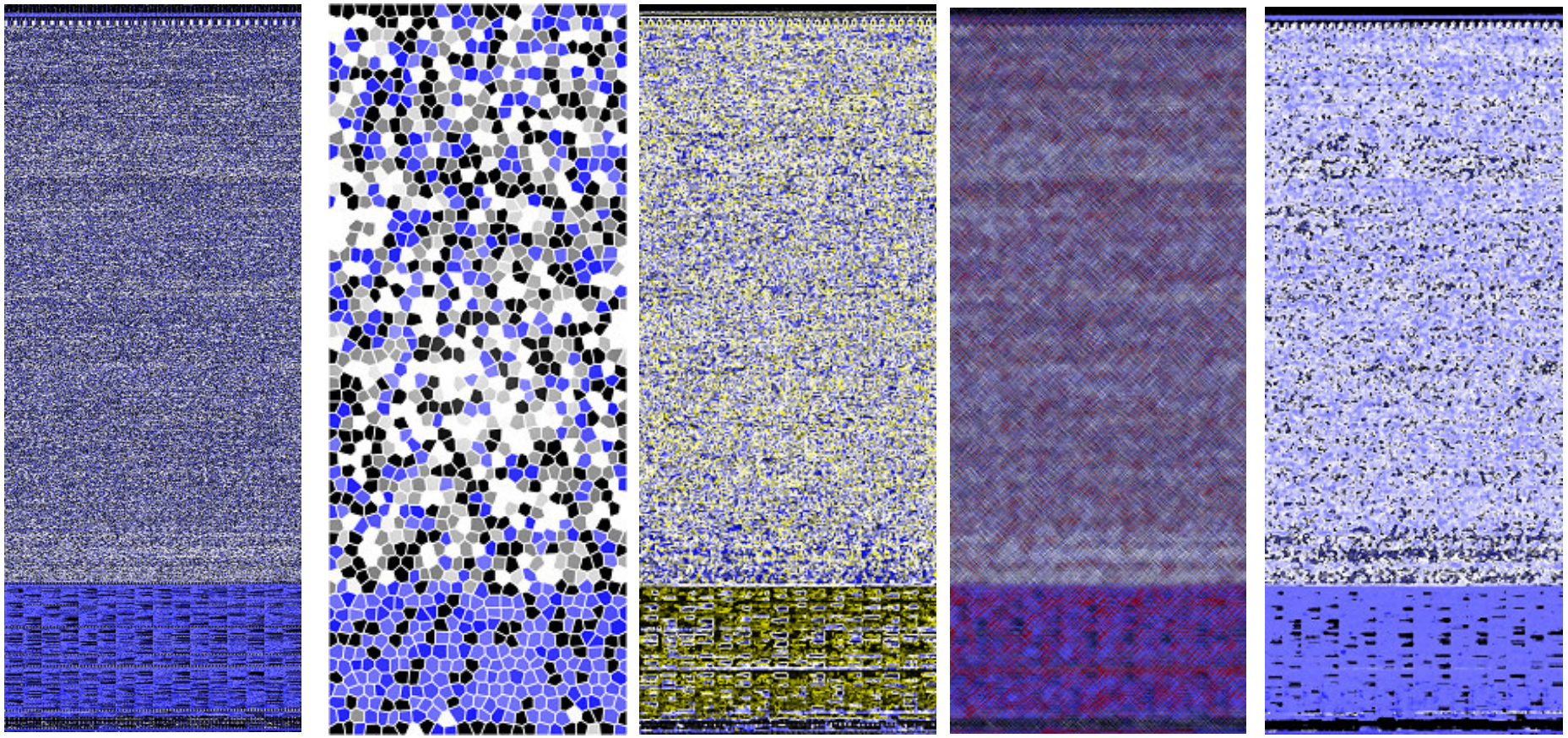
(text + image)




original
with modify password
with open password
AES

Apply Image Filtering Algorithms

(sendmail)



original stained glass glowing edges colored pencil dry brush

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…

Attacking Information Visualization System Usability
Overloading and Deceiving the Human

Gregory Conti
College of Computing
Georgia Institute of Technology

Mustaque Ahamad
College of Computing
Georgia Institute of Technology

John Stasko
College of Computing
Georgia Institute of Technology

ABSTRACT
Information visualization is an effective way to easily comprehend large amounts of data. For such systems to be truly effective, the information visualization designer must be aware of the ways in which their system may be manipulated and protect their users from attack. In addition, users should be aware of potential attacks in order to minimize or negate their effect. These attacks target the information visualization system as well as the perceptual, cognitive and motor capabilities of human end users. To identify and help counter these attacks we present a framework for information visualization system security analysis, a taxonomy of visualization attacks and technology independent principles for countering malicious visualizations. These themes are illustrated with case studies and working examples from the network security visualization domain, but are widely applicable to virtually any information visualization system.

CR Categories: H.5.2 [**Information Systems**]: Information Interfaces and Presentation - User Interfaces; C.2.3 [**Computer-Communication Networks**]: Network Operations: Network monitoring; C.2.0 [**Computer-Communication Networks**]: General - Security and Protection

Keywords: malicious visualizations, usability attacks, denial of information, secure visualization, information visualization

1 INTRODUCTION

All but the most trivial visualization systems must be designed with security in mind. Information visualization systems are constantly exposed to attack, either from malicious entities attempting to overwhelm, mislead or distract the human viewer or from non-malicious entities that accomplish the same result by accident. Some might believe that today's systems are not potential targets for attack. Clearly there are many domains where security is of minimal importance, but increasingly information visualization systems are being used to support critical decision making. For example, intelligence analysis, law enforcement, network security and business decision-support systems exist in an adversarial environment where it is likely that malicious entities are actively attempting to manipulate human end users. We believe that there is a clear threat today and there will be a growing problem into the foreseeable future. For information visualization systems to maintain relevance security must be considered. Information visualization systems inherently have the human tightly coupled in the system loop. In most cases, the human is the decision maker who will act upon (or not act upon) the information presented and, as a result, the human is a high-payoff and likely target. Any point in the information visualization system may be attacked, from data collection to processing to final visualization, in order to impact human interpretation. A "minor" compromise of a single bit may have significant impact on the human (consider a change in the foreground color of a scatter plot to the background color). Major compromises may have far greater impact. Our primary goal is to identify these threats and vulnerabilities, as well as develop principles to counter or mitigate these attacks. By identifying the threats and weaknesses of their system, designers can make appropriate decisions to mitigate these vulnerabilities.

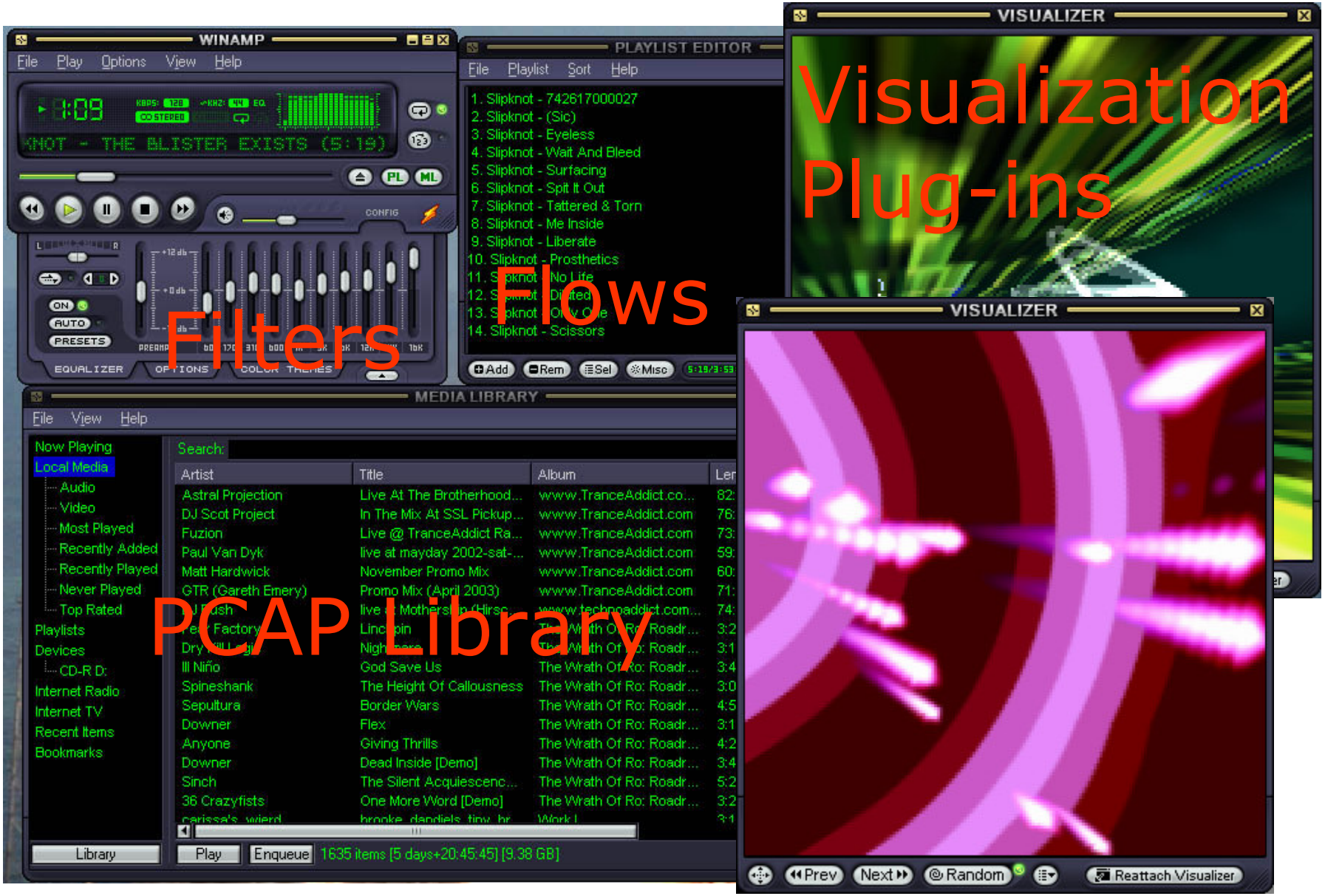
To see a sample attack in action, consider a visual intrusion detection system designed to supplement classical anomaly-based and signature-based intrusion detection systems. Such systems are typically co-located with a firewall at the border between the internal institutional network and the public Internet. This vantage point allows the system to observe and collect selected data from network traffic at entry and egress from the internal network. Our example system collects header data from network traffic and visualizes it in real-time. In particular, it captures the source and destination addresses of communicating network nodes, network protocols in use, source and destination ports (used for process to process communication across an Internet Protocol (IP) network, e.g. port 80 for a web server) as well as calculates a timestamp for each record. An adversary may easily inject arbitrary data into the visualization system, intermingled with legitimate users' traffic, due to weaknesses in current networking protocols. In our example, the adversary knows the system operator on the night shift is red-green colorblind. They also know that the default settings on the visualization system map the very common (99+% of traffic) Transmission Control Protocol (TCP) to green, the User Datagram Protocol (UDP) to blue and the Internet Control Management (ICMP) protocol to red. In addition, the attacker knows that the target node has serious ICMP and UDP vulnerabilities. The attacker waits until late in the operator's shift and launches an ICMP based attack. The already tired operator does not notice the red packet amidst the much greater noise of green packets. In this case, the attacker took advantage of the visualization system's color mapping to target a specific user, but many other techniques could have been used. We will describe and illustrate these attacks in later sections.

To help combat usability attacks against visualization systems this work includes several novel contributions: a framework for information visualization system security analysis, a taxonomy of malicious attacks as well as technology independent principles for designing information visualization systems that will resist attack. We illustrate and validate these contributions with results from the design, implementation and real-world use of a visual network intrusion detection system [1].

Symposium On Usable Privacy and Security '05, July 6-8, 2005, Pittsburgh, PA, USA.

Attack Demo

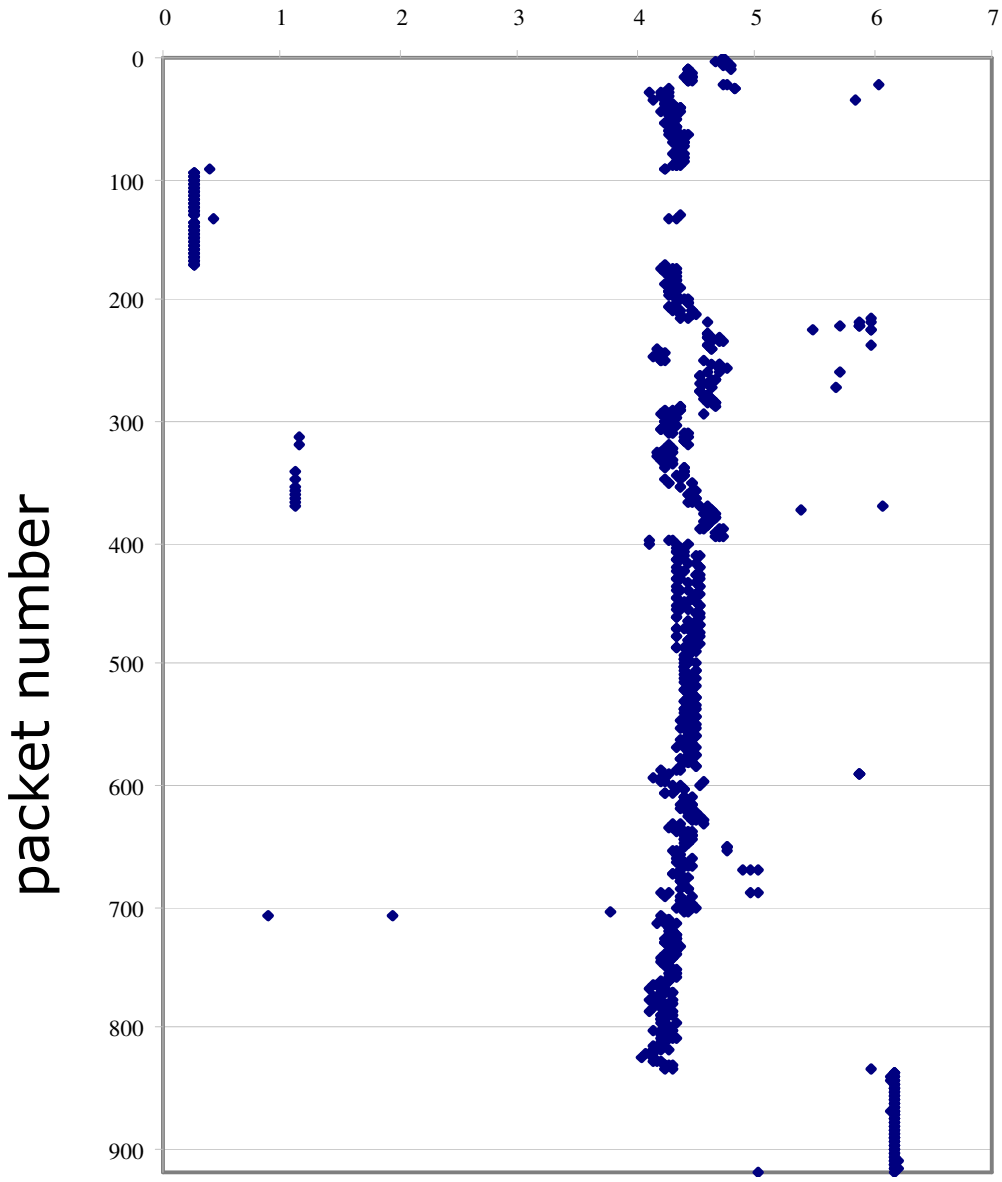
Future Vision

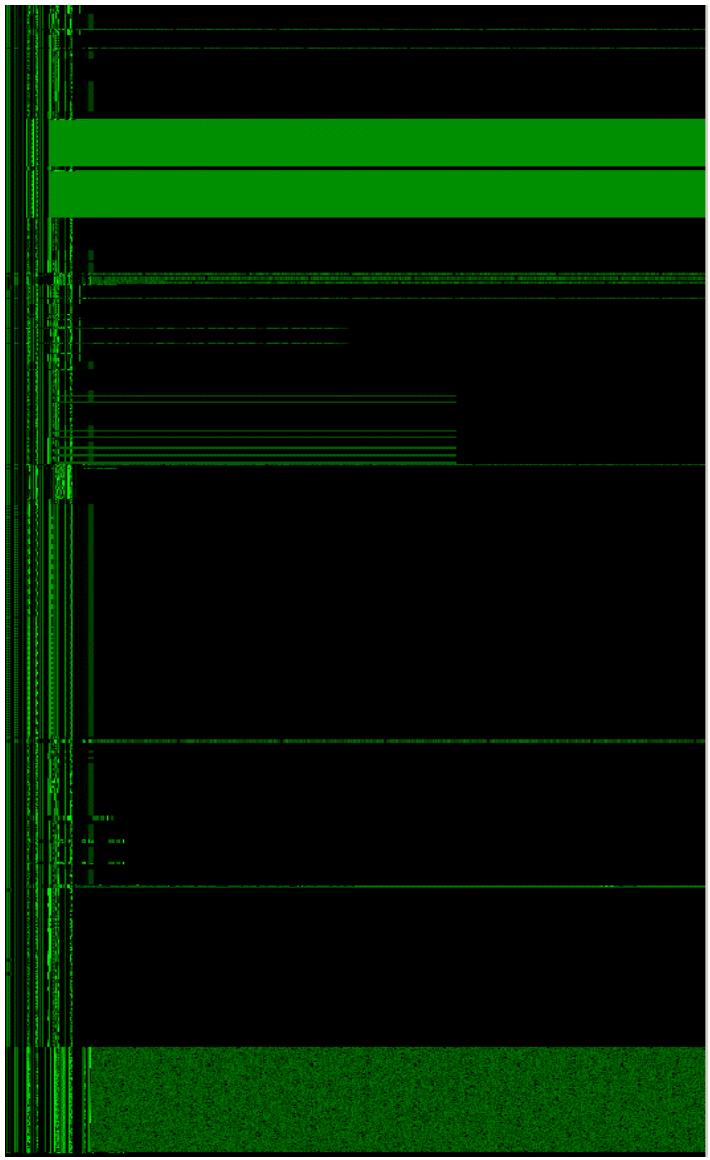


Directions for the Future…

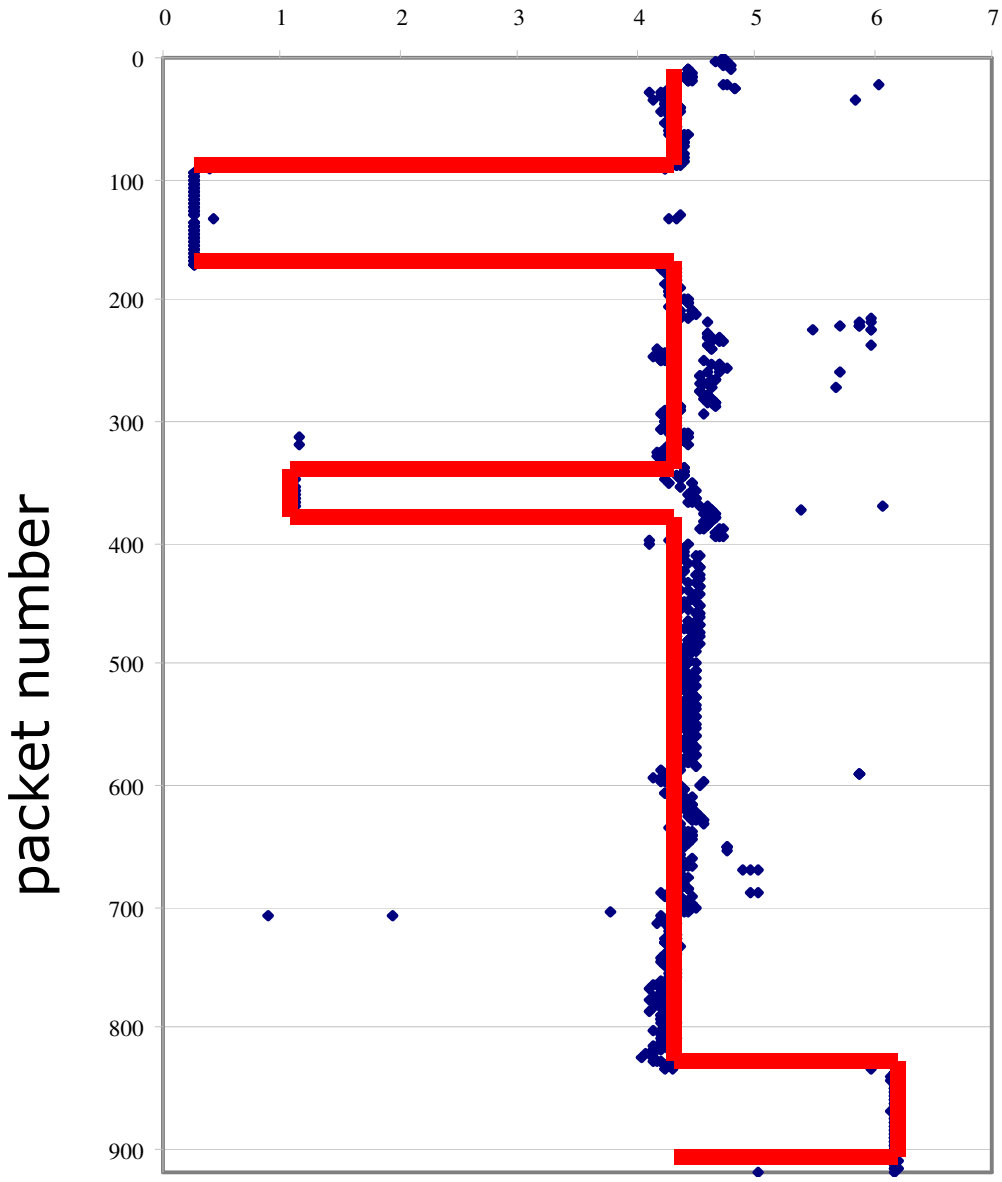
We are only scratching the surface of the possibilities

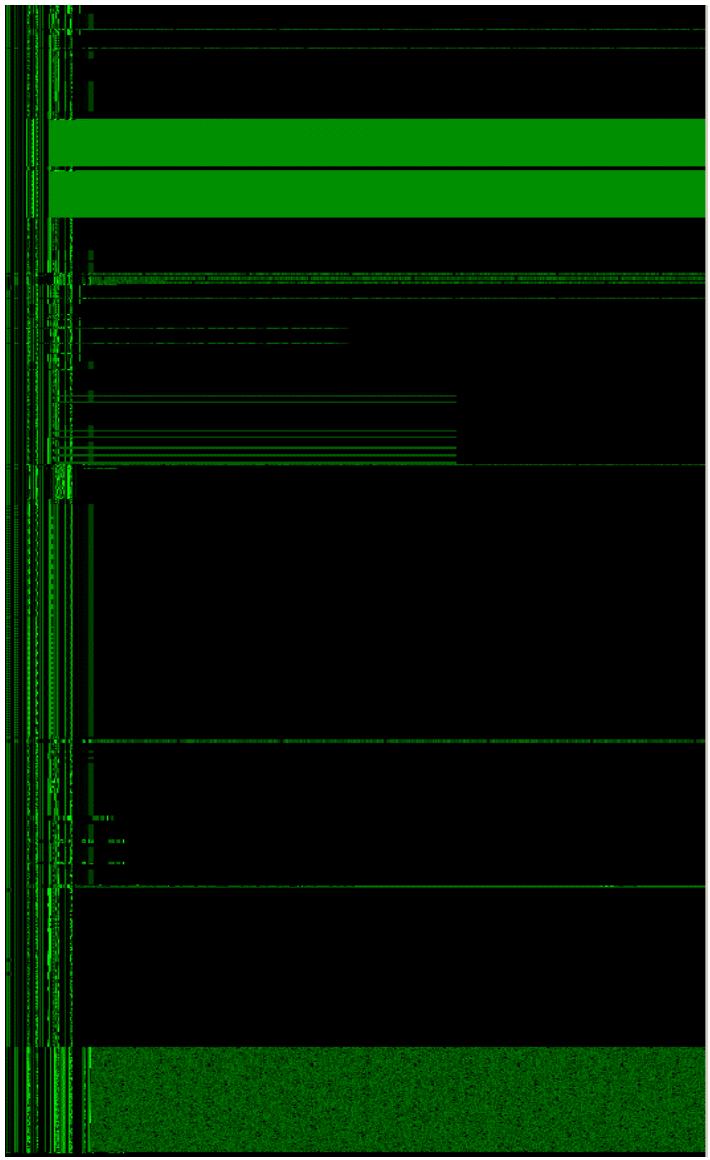
- attack specific community needs
- plug-ins (vis, filters, processing)
- launch network packets?
- protocol specific visualizations
  - including application layer (e.g. VoIP, HTTP)
- Open GL
- graph visualization+
- screensaver/wallpaper snapshot?
- work out GUI issues
- database of filters / smart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





* Hamming, Richard W. *Coding and Information Theory*, 1980, pp. 104-108





* Hamming, Richard W. *Coding and Information Theory*, 1980, pp. 104-108

For more information…

G. Conti, K. Abdullah, J. Grizzard, J. Stasko, J. Copeland, M. Ahamad, H. Owen and C. Lee;"Countering Security Analyst and Network Administrator Overload Through Alert and Packet Visualization;" IEEE Computer Graphics and Applications (CG&A), March 2006.

G. Conti, J. Grizzard, M. Ahamad and H. Owen; "Visual Exploration of Malicious Network Objects Using Semantic Zoom, Interactive Encoding and Dynamic Queries;" IEEE Symposium on Information Visualization's Workshop on Visualization for Computer Security (VizSEC); October 2005.

G. Conti; "Beyond Ethereal: Crafting A Tivo for Security Datastreams;" Black Hat USA; July 2005.

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.

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 and K. Abdullah; " Passive Visual Fingerprinting of Network Attack Tools;" ACM Conference on Computer and Communications Security's Workshop on Visualization and Data Mining for Computer Security (VizSEC); October 2004.

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

CACM

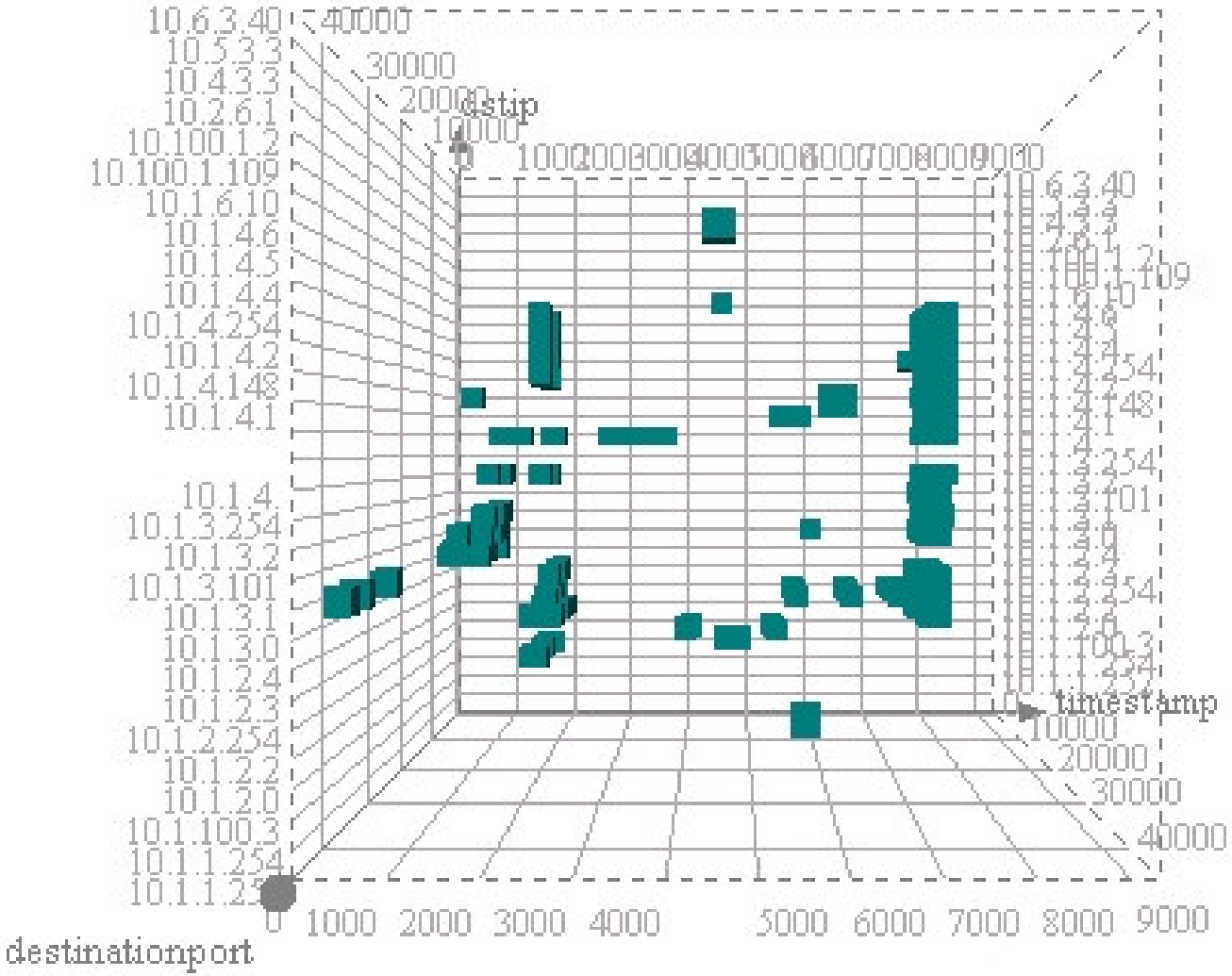
Acknowledgements

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, Kenshoto, 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

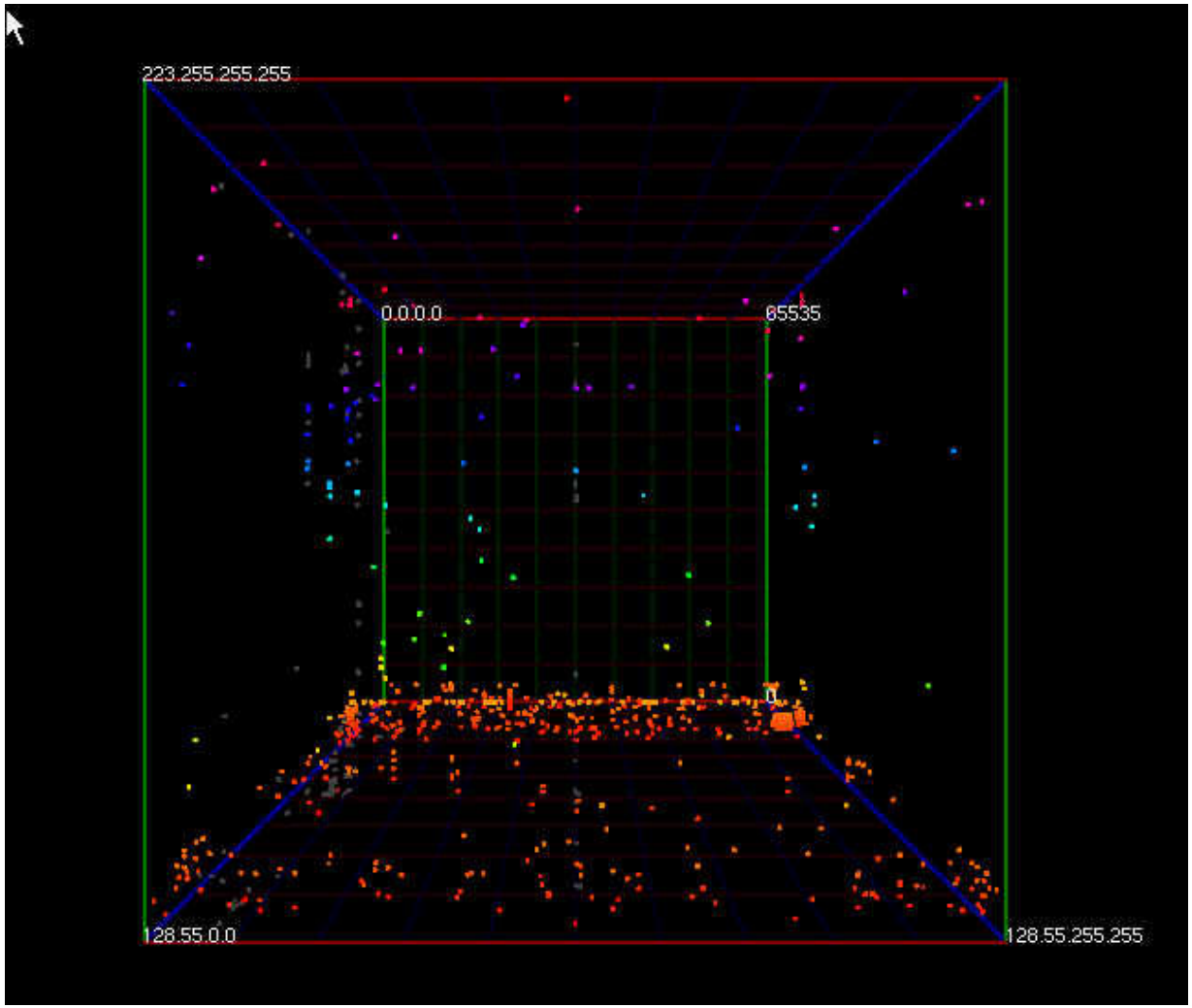
www.cc.gatech.edu/~conti
www.rumint.org

Labeling Attack



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

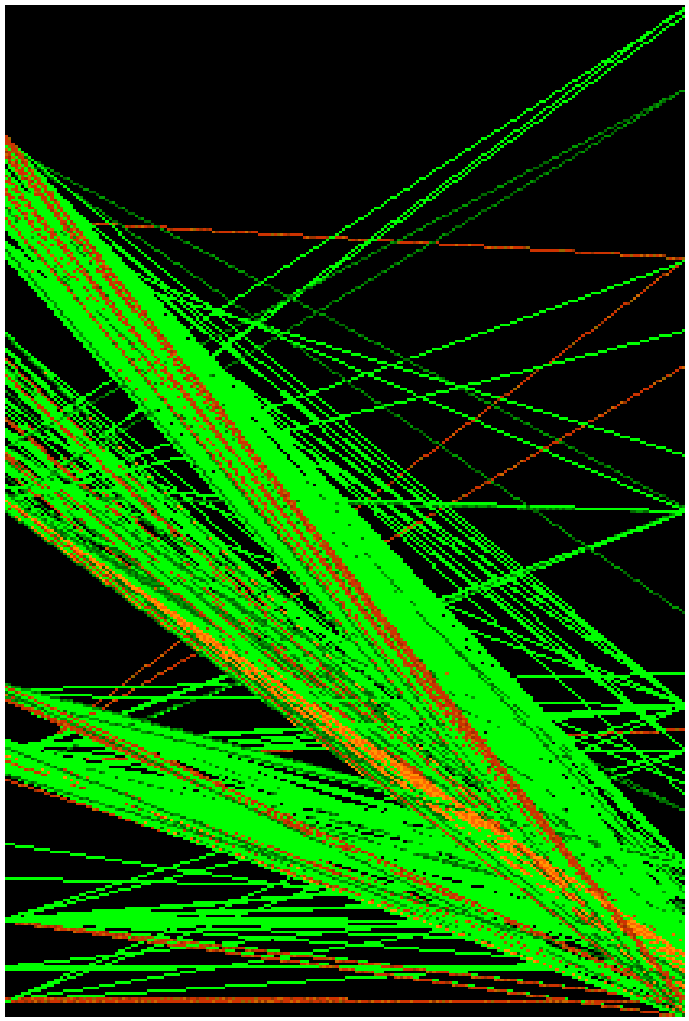
Precision Attack



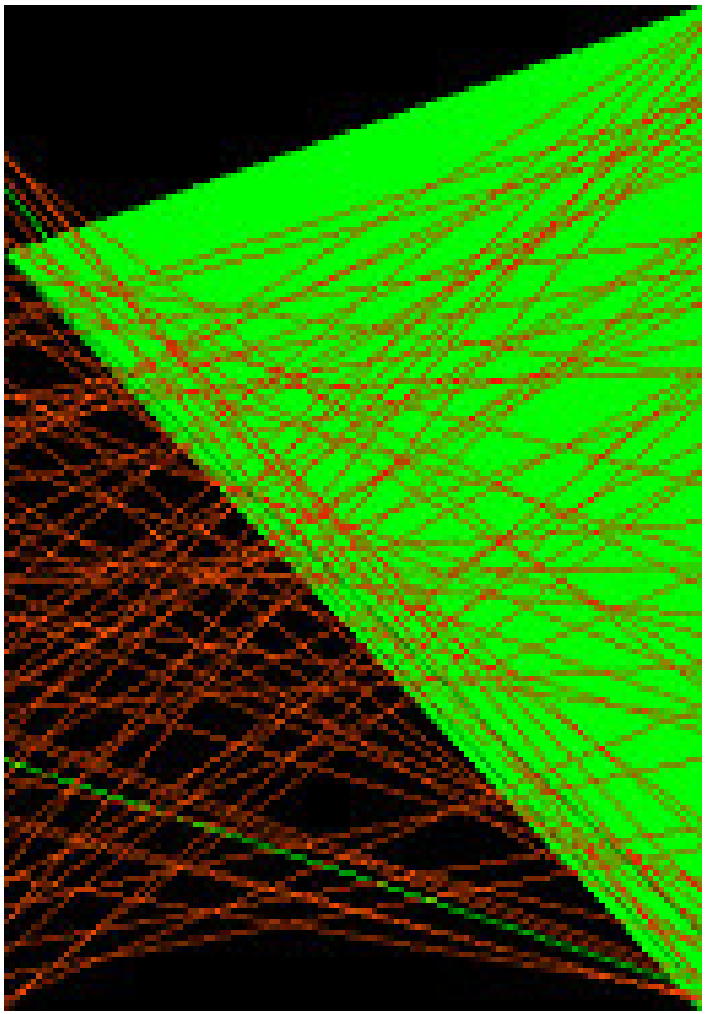
http://www.nersc.gov/nusers/security/Cube.jpg

http://developers.slashdot.org/article.pl?sid=04/06/01/1747223&mode=thread&tid=126&tid=172

Occlusion



Jamming



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…