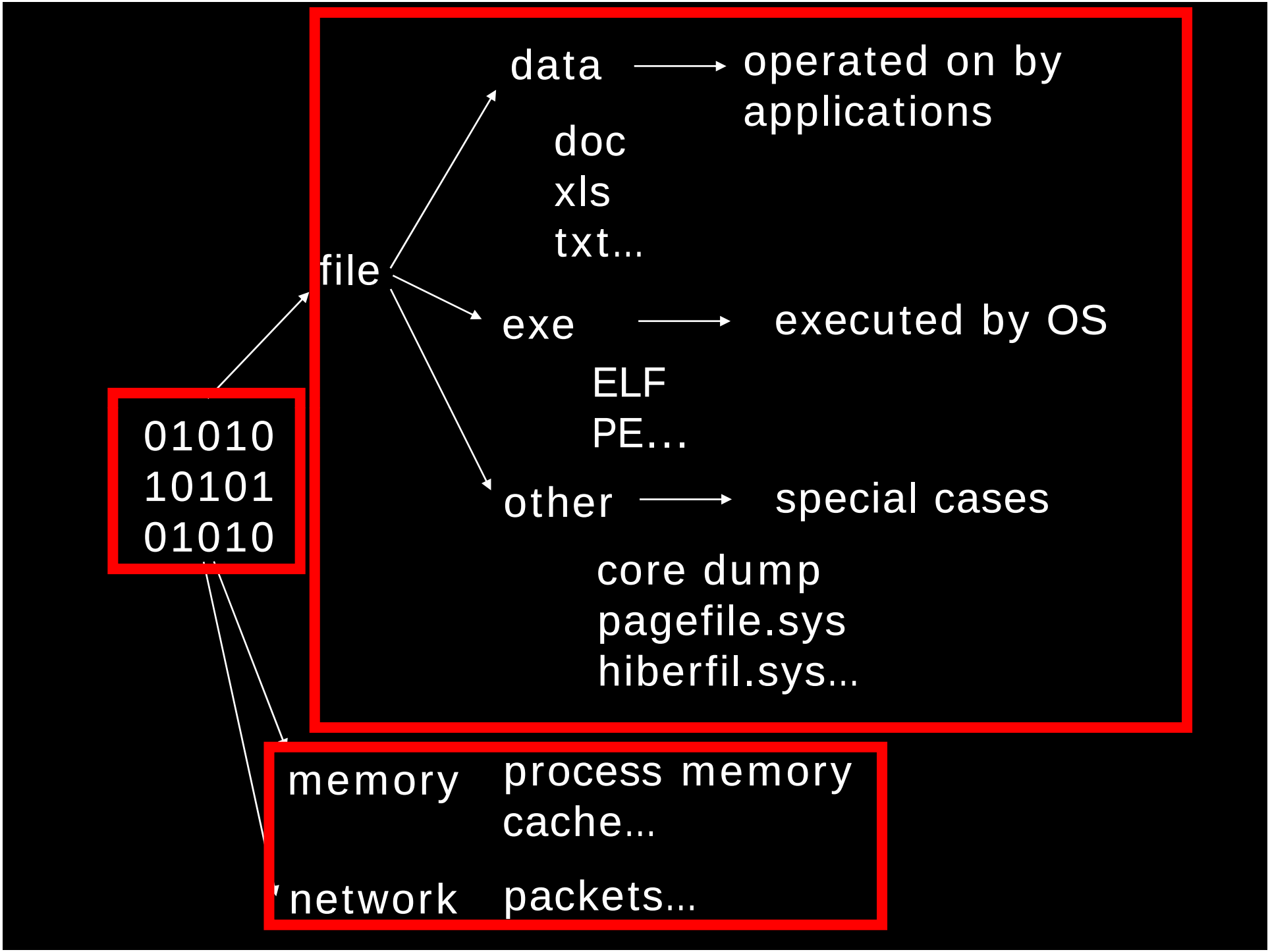
# Visual Forensic Analysis and Reverse Engineering of Binary Data

*Gregory Conti*
*Erik Dean*

*United States Military Academy*
*West Point, New York*
*gregory.conti@usma.edu*
*erik.dean@usma.edu*

# Outline

- The Problem – Tiny Windows
- Background and Motivation
- Related Work
- Moving Beyond Hex
- System Design
- Case Studies
- Demos

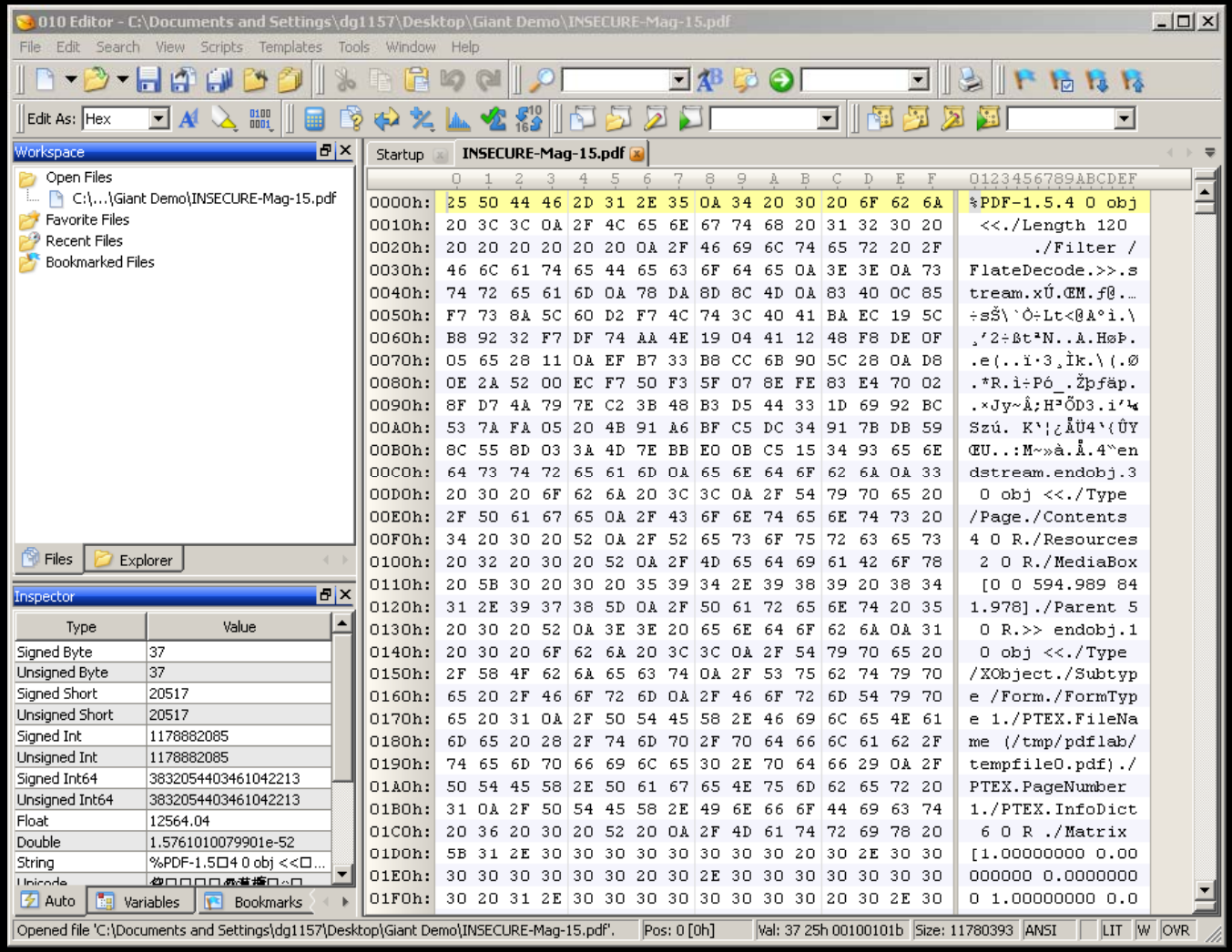

data
operated on by applications
doc
xls
txt...
file
exe
executed by OS
ELF
PE...
other
special cases
core dump
pagefile.sys
hiberfil.sys...
01010
10101
01010
memory
process memory
cache...
network
packets...


high insight
lower insight
Ida Pro
OllyDBG
BinNavi (Zynamics)
BinDiff (Zynamics)...
Filemon
Regmon...
011
hex editors
hexdump
grep & diff
strings
objdump
original application
general purpose
precise application

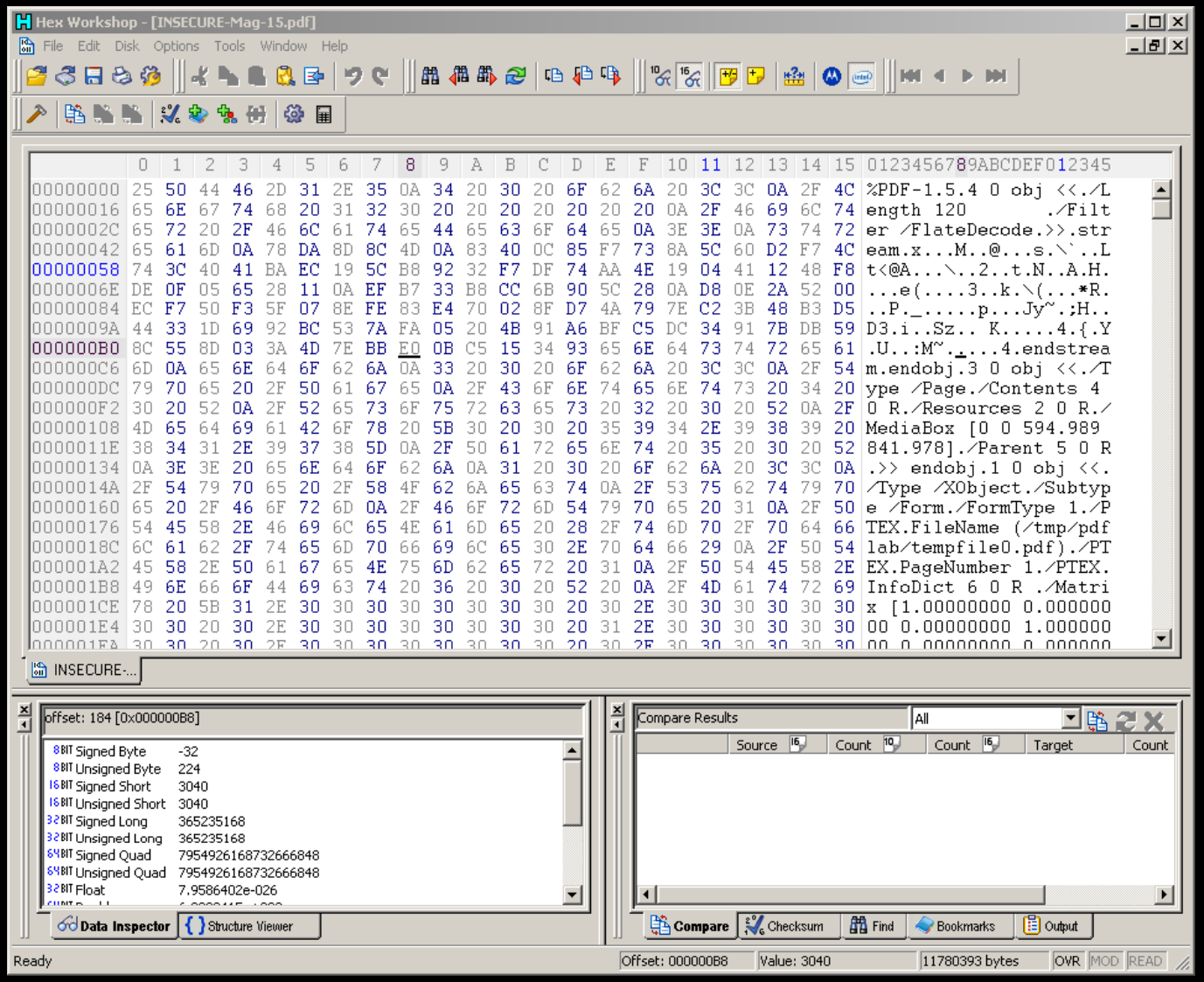
# strings /grep/diff

```
H:\Datasets>strings 20040517_homeISP.pcap | more

Strings v2.4
Copyright (C) 1999-2007 Mark Russinovich
Sysinternals - www.sysinternals.com

0hF
M@y
7bs
Z19Z
MICROSOFT NETWORKS
WINDOWS USER
Microsoft Security Bulletin MS03-043
Buffer Overrun in Messenger Service Could Allow Code Execution
(828035)
Affected Software:
Microsoft Windows NT Workstation
Microsoft Windows NT Server 4.0
Microsoft Windows 2000
...
```

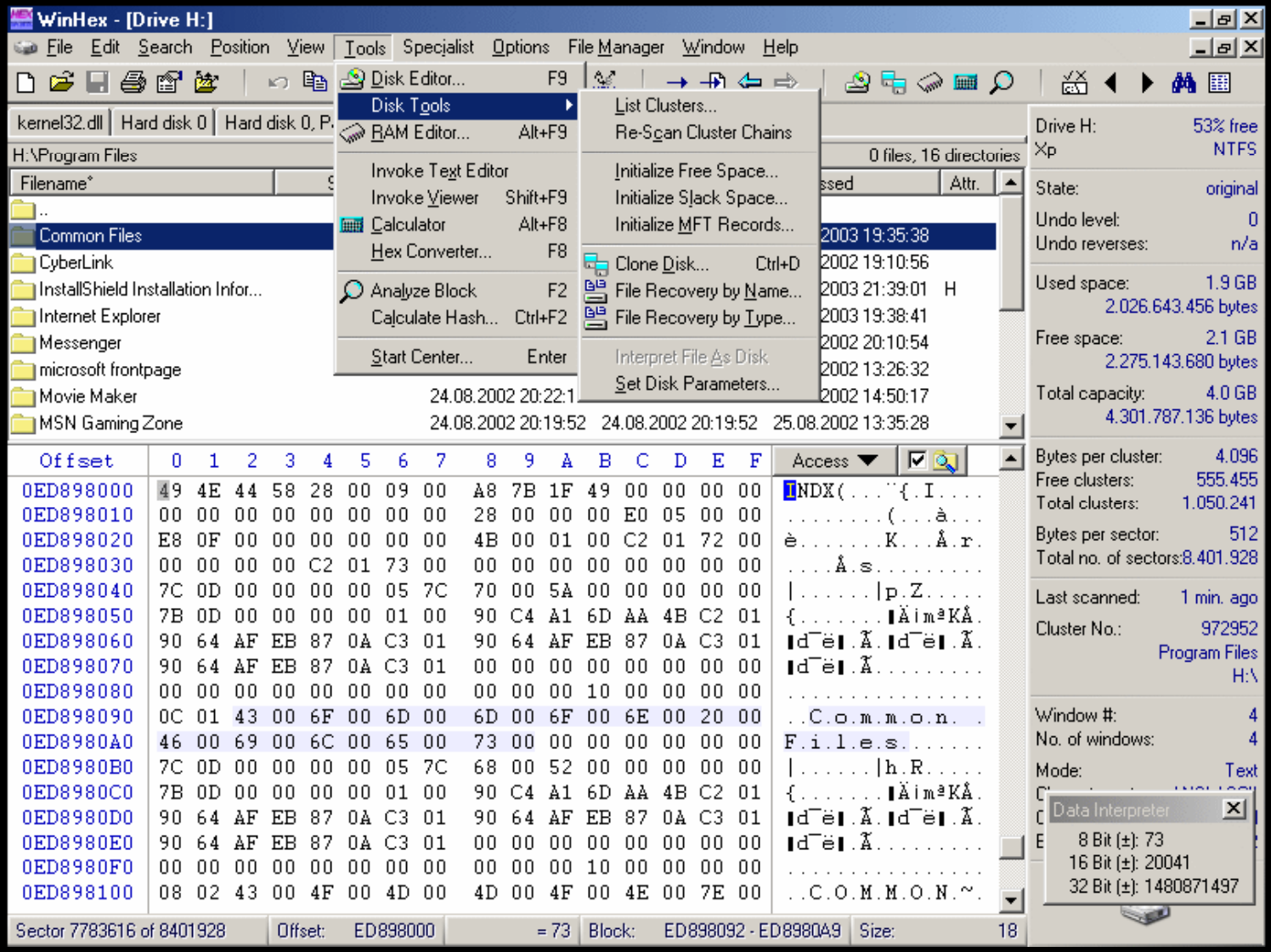
# 011 Hex Editor



# Hex Workshop



# WinHex

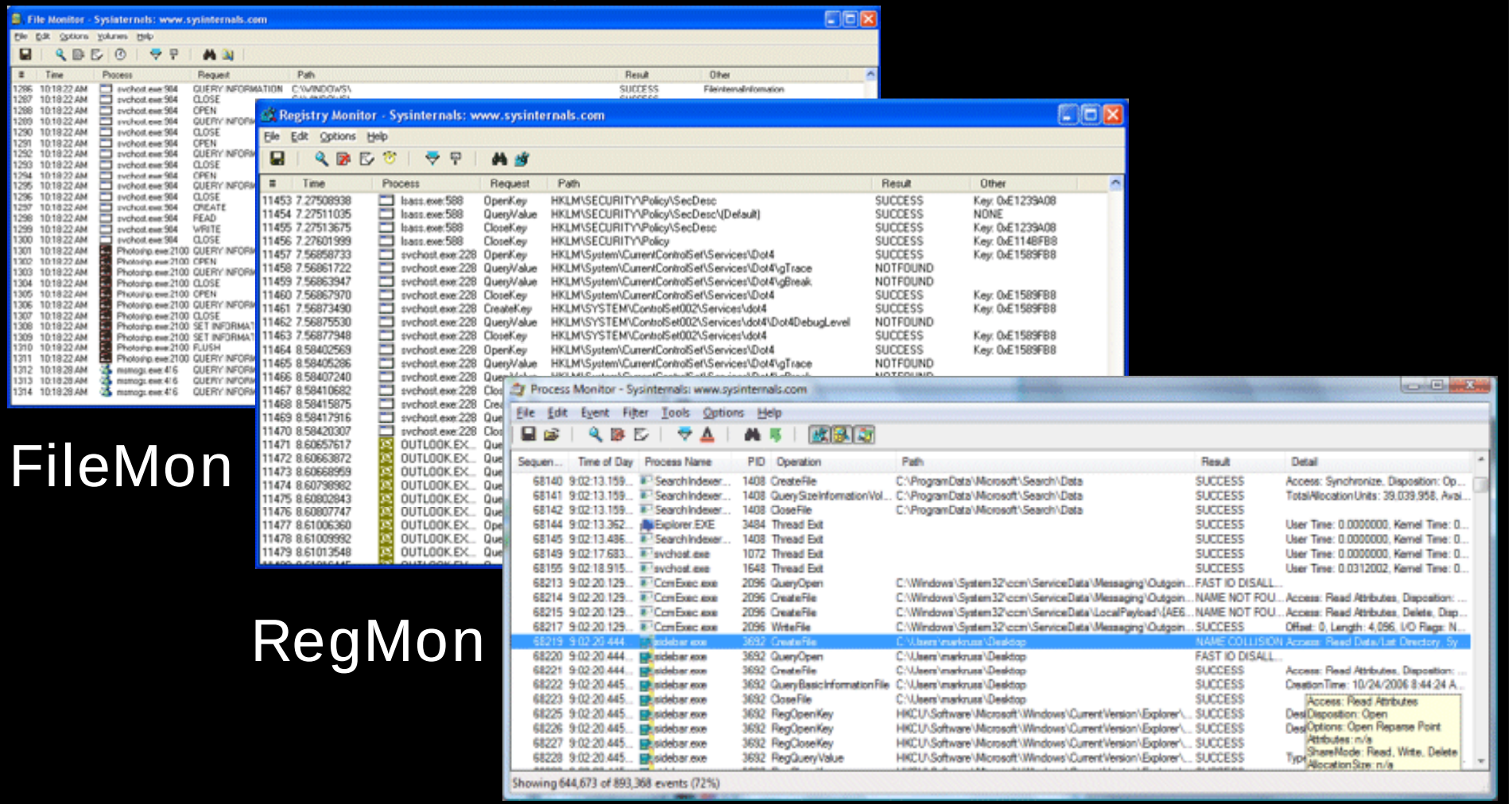



http://www.x-ways.net/pics/winhex.gif


high insight
lower insight
Ida Pro
OllyDBG
BinNavi (Zynamics)
BinDiff (Zynamics)...
Filemon
Regmon...
011
hex editors
hexdump
grep & diff
strings
objdump
original application
general purpose
precise application


high insight
lower insight
Ida Pro
OllyDBG
BinNavi (Zynamics)
BinDiff (Zynamics)...
Filemon
Regmon...
011
hex editors
hexdump
grep & diff
strings
objdump
original application
general purpose
precise application

# SysInternals




FileMon

RegMon

Process Monitor

...

http://technet.microsoft.com/en-us/sysinternals/default.aspx

# Wireshark

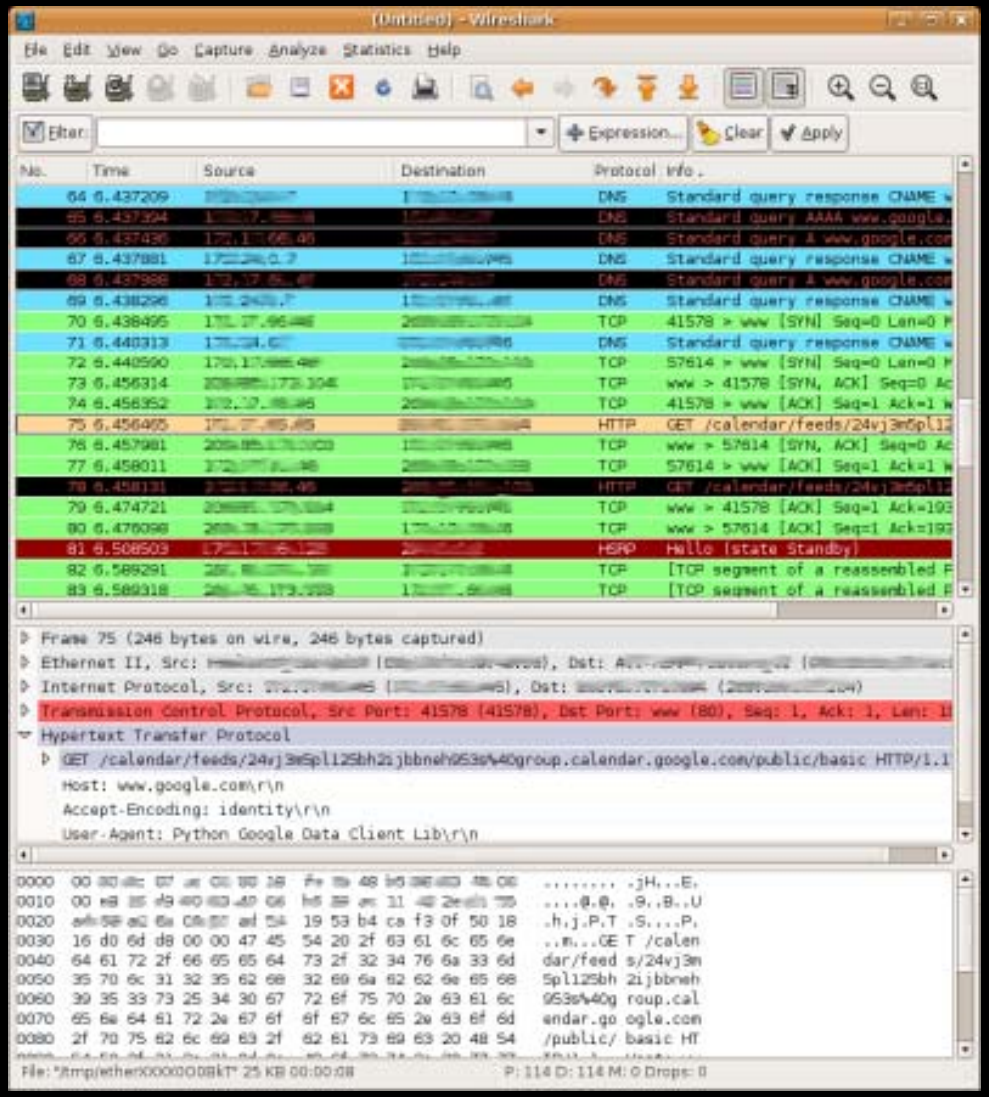
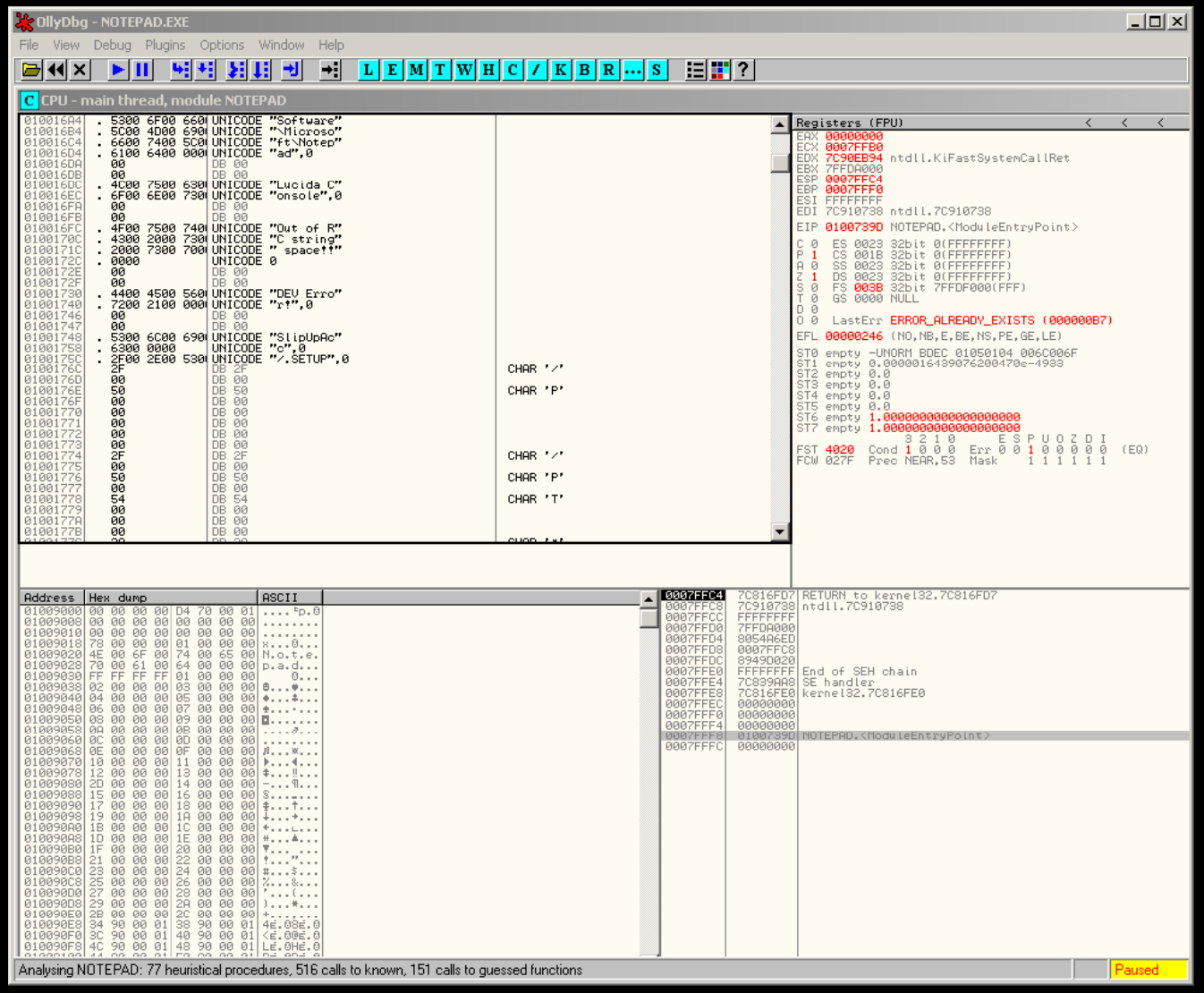



image: http://code.google.com/support/bin/answer.py?answer=71567

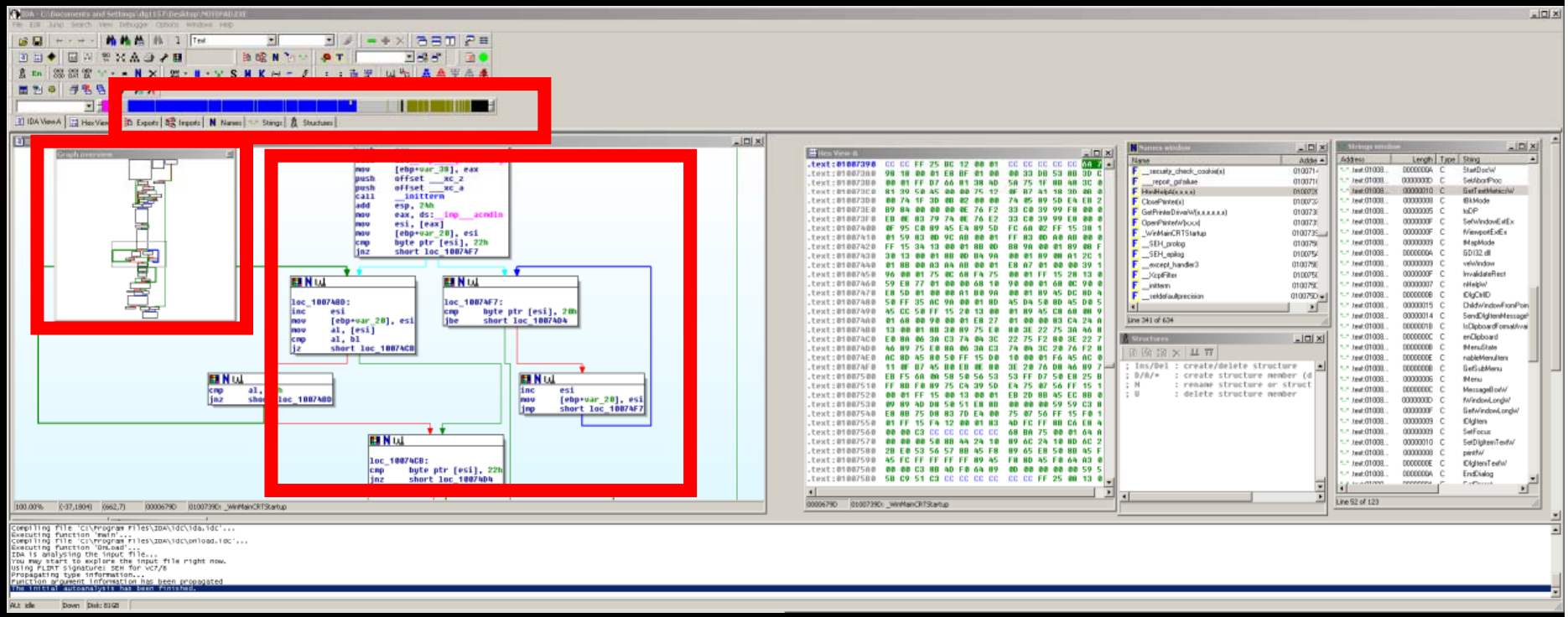
# OllyDbg




http://www.ollydbg.de/

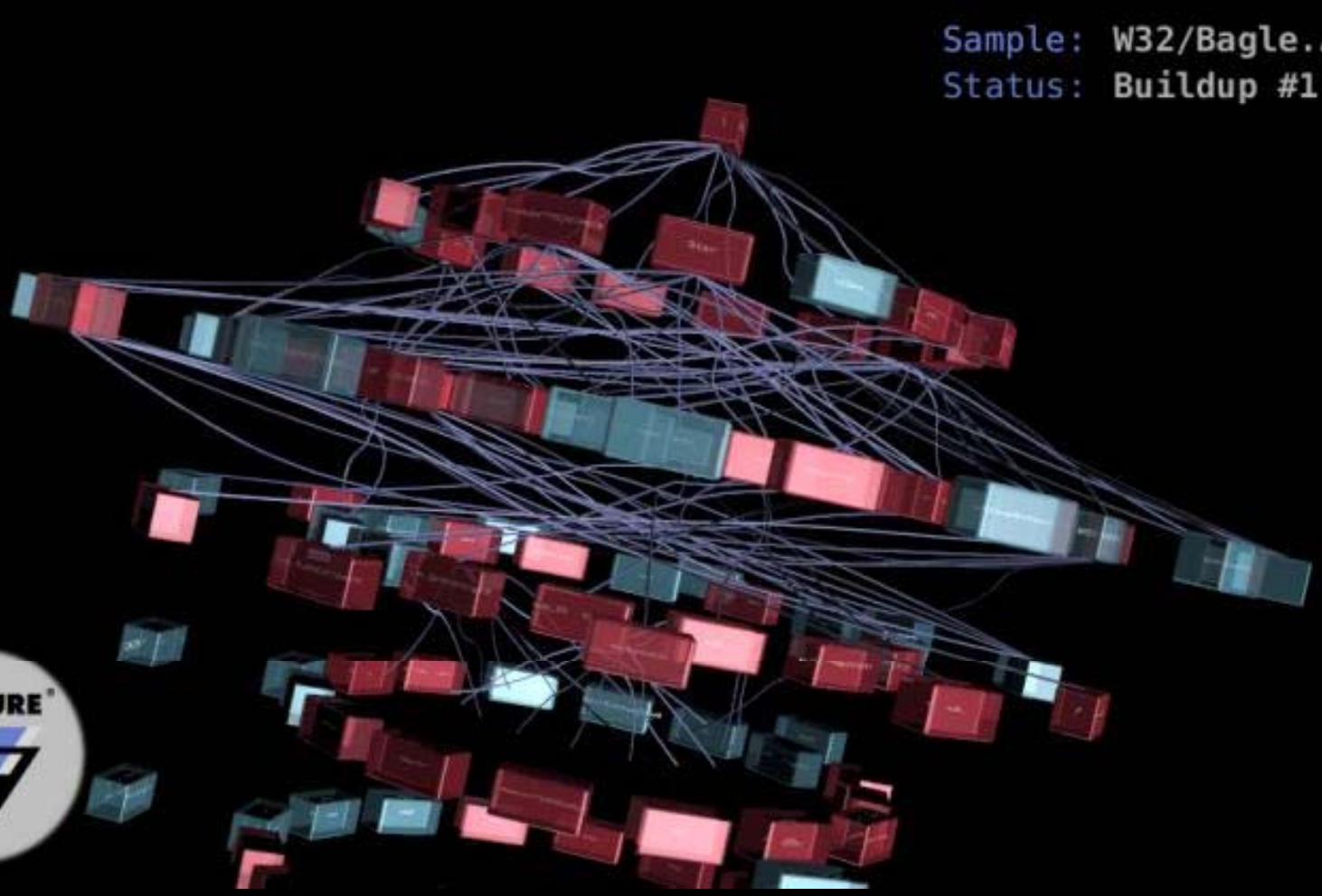
# IDA Pro

v5.1



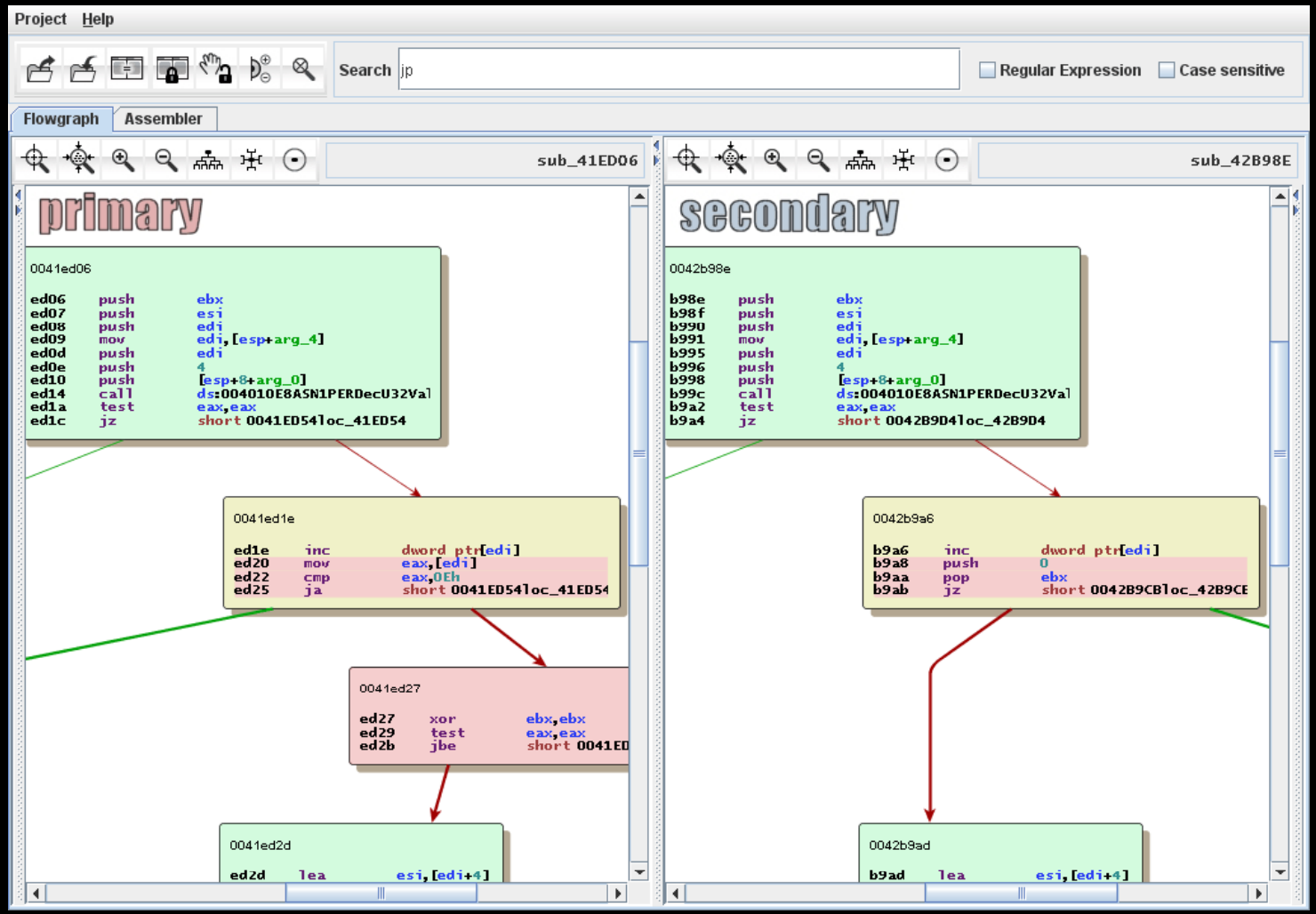
http://www.hex-rays.com/idapro/

# F-Secure Malware



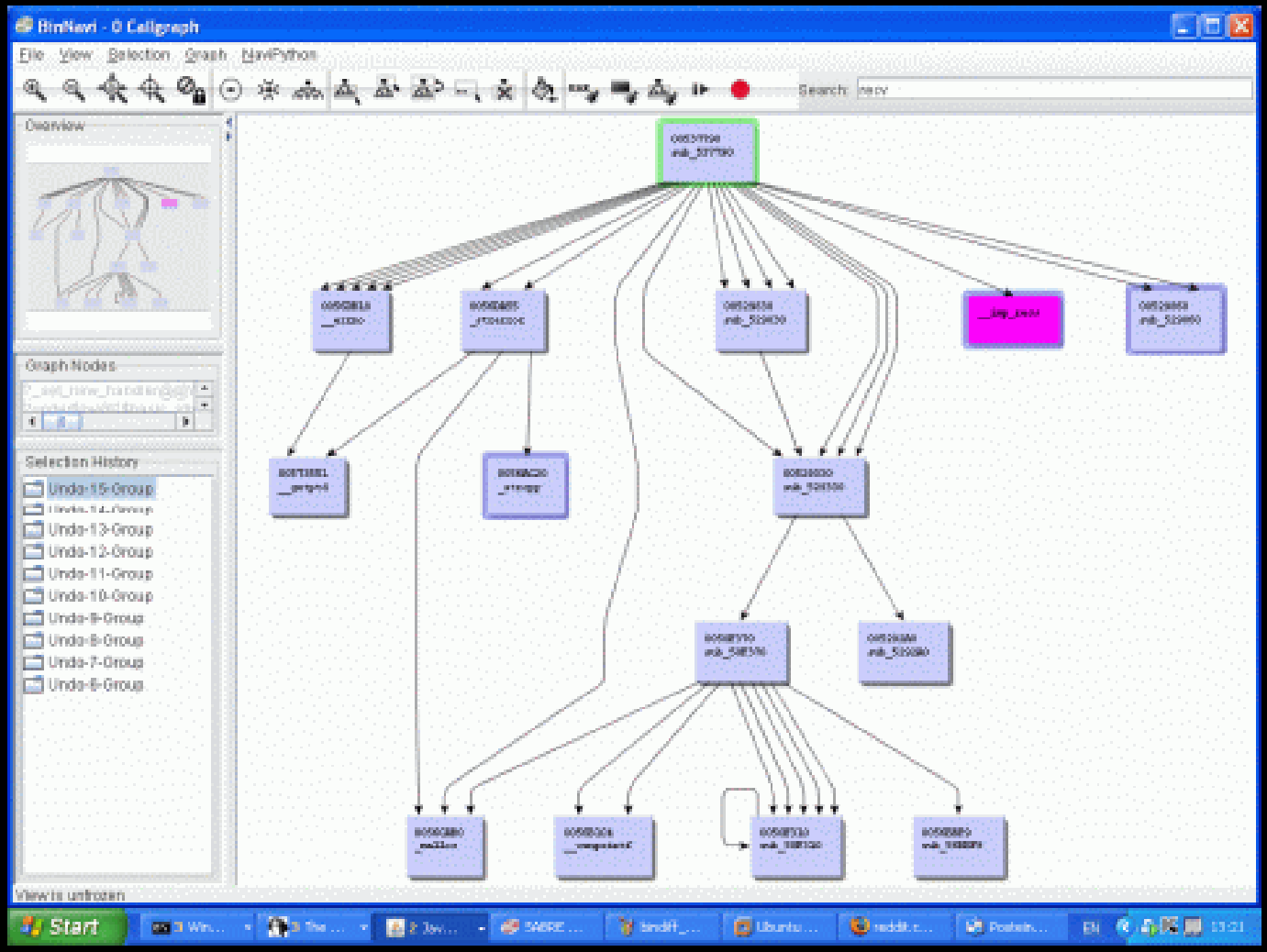

http://www.f-secure.com/weblog/archives/00000662.html

# Zynamics BinDiff

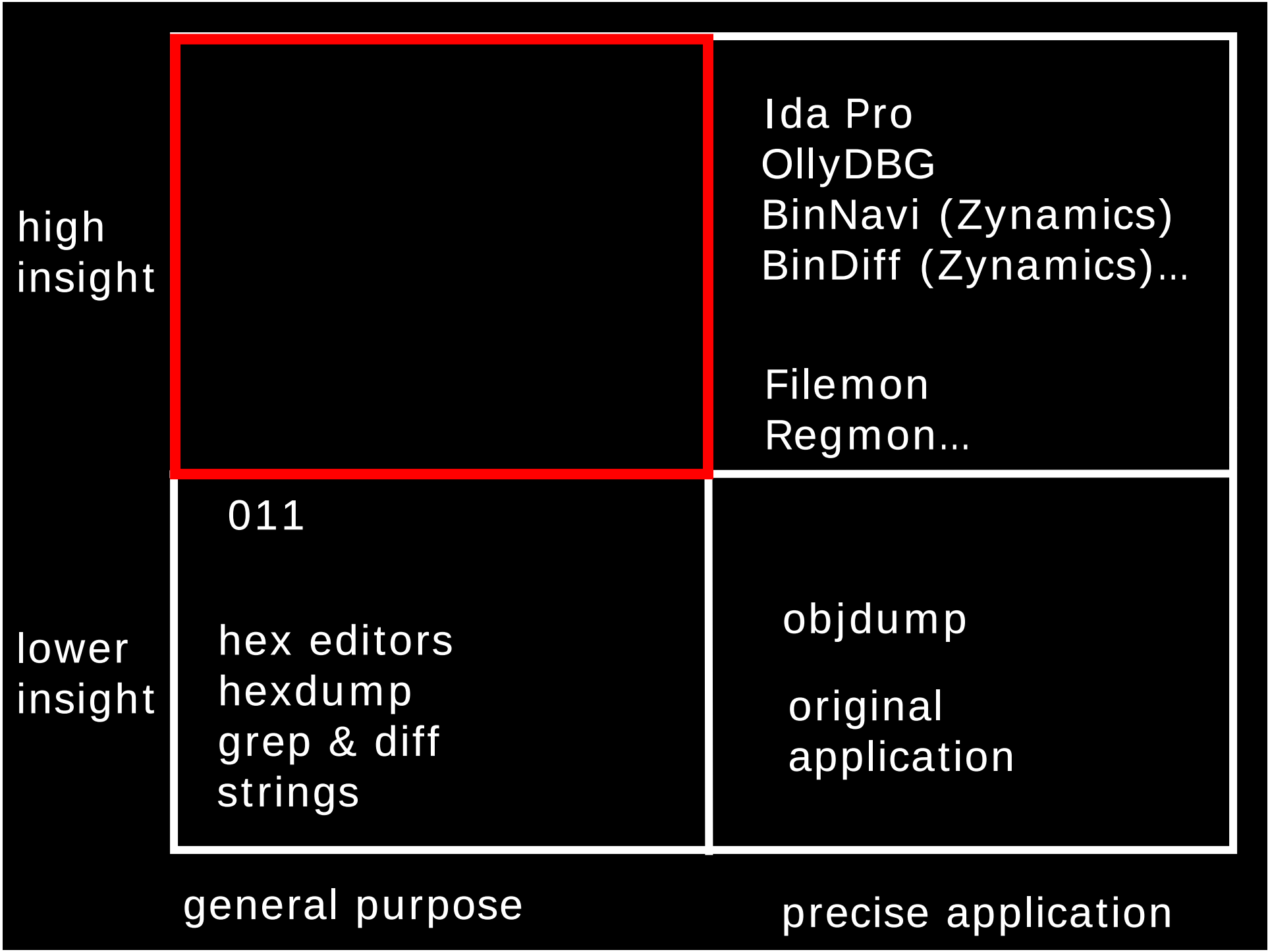



http://www.zynamics.com/content/_images/bindiff_scr2.gif

# Zynamics BinNavi




http://www.zynamics.com/index.php?page=binnavi


high insight
lower insight
Ida Pro
OllyDBG
BinNavi (Zynamics)
BinDiff (Zynamics)...
Filemon
Regmon...
011
hex editors
hexdump
grep & diff
strings
objdump
original application
general purpose
precise application

# Framework

- File Independent Level
  - Entropy
  - Byte Frequency
  - N-Gram Analysis
  - Strings
  - Hex / Decimal / ASCII
  - Bit Plot  (2D/3D)
  - File Statistics
- File Specific Level
  - Complete or Partial Knowledge of File Structure
  - For Example, Metadata

# Syntax Highlighting for Hex Dumps

(Kaminsky)

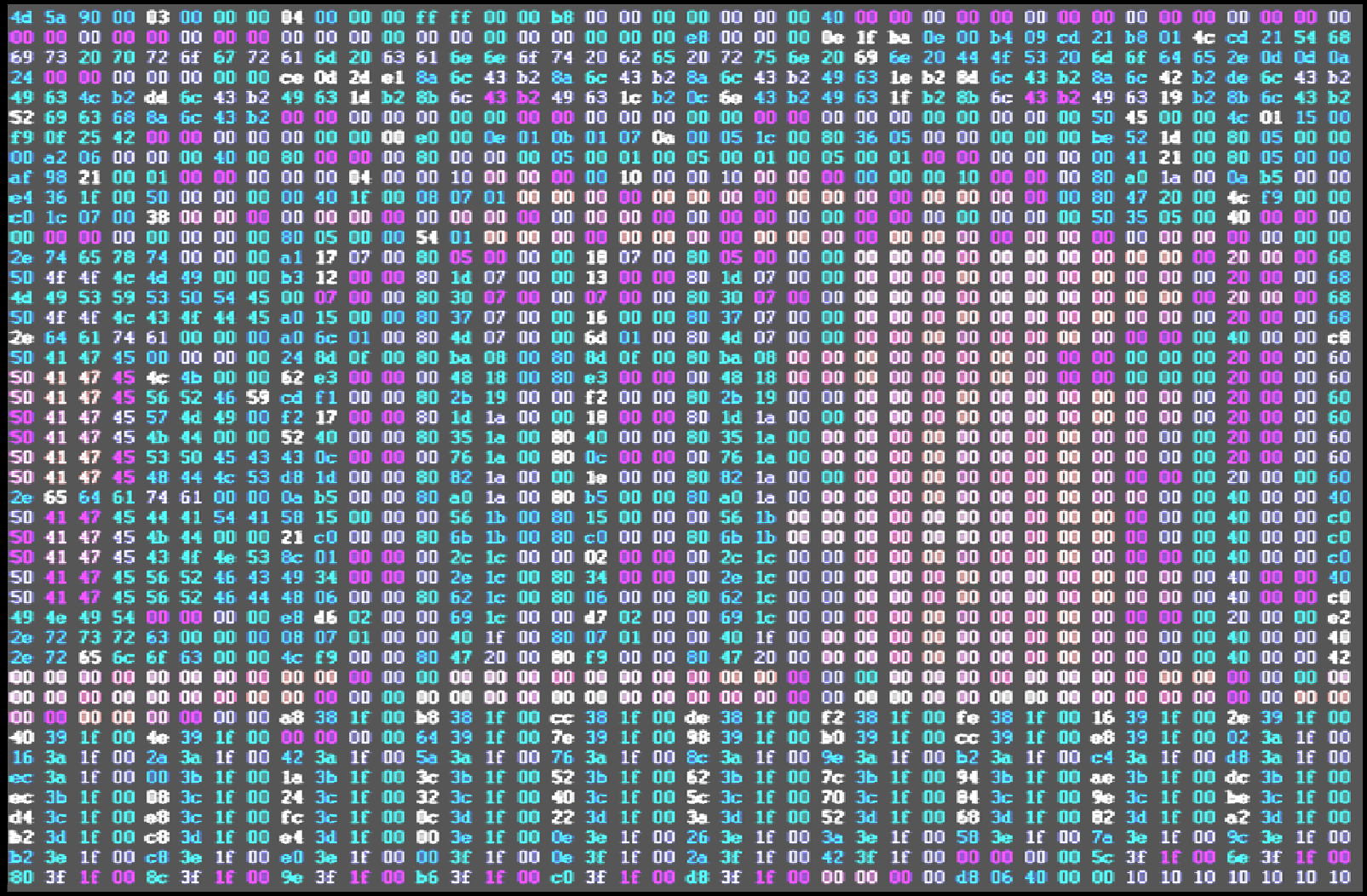
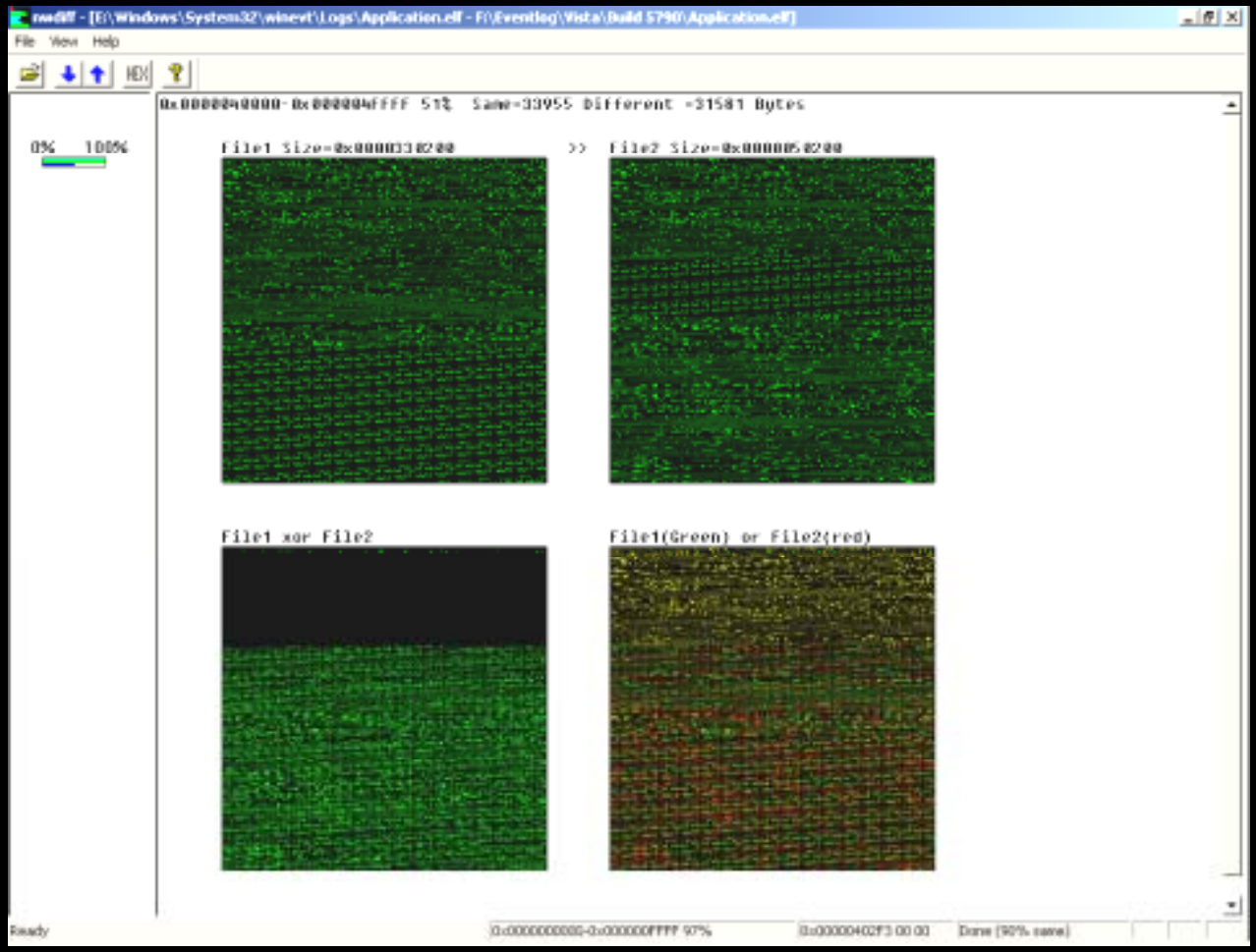


image: Dan Kaminsky, CCC2006

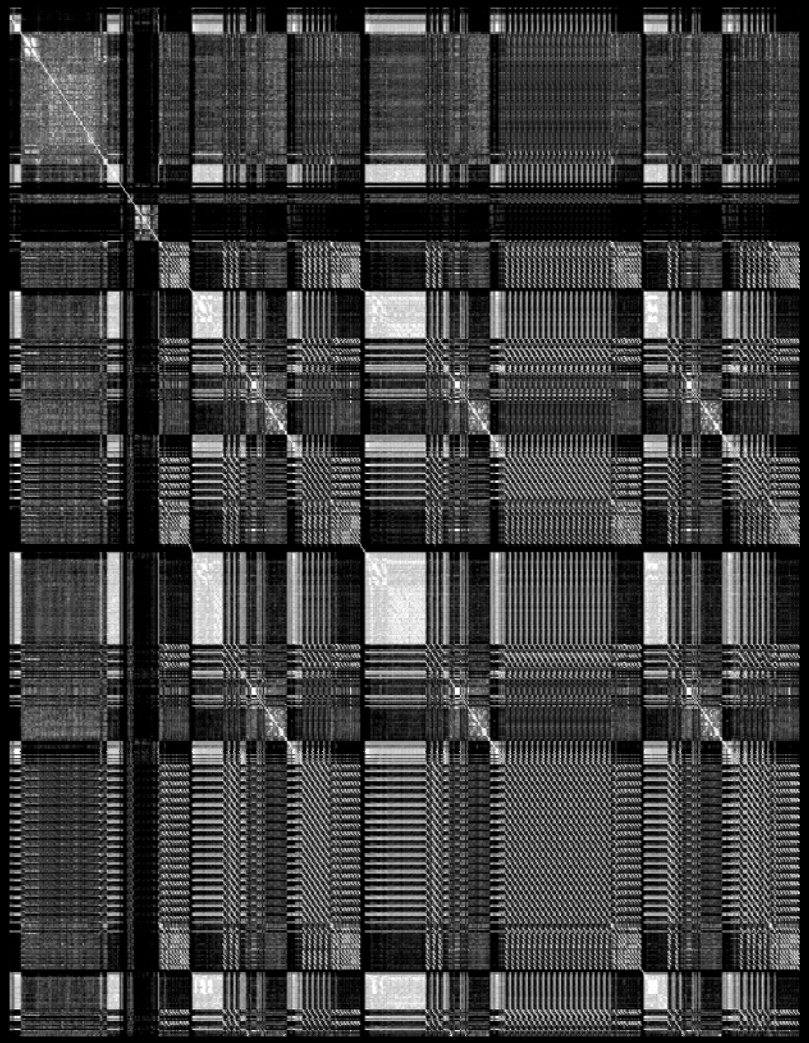
# nwdiff



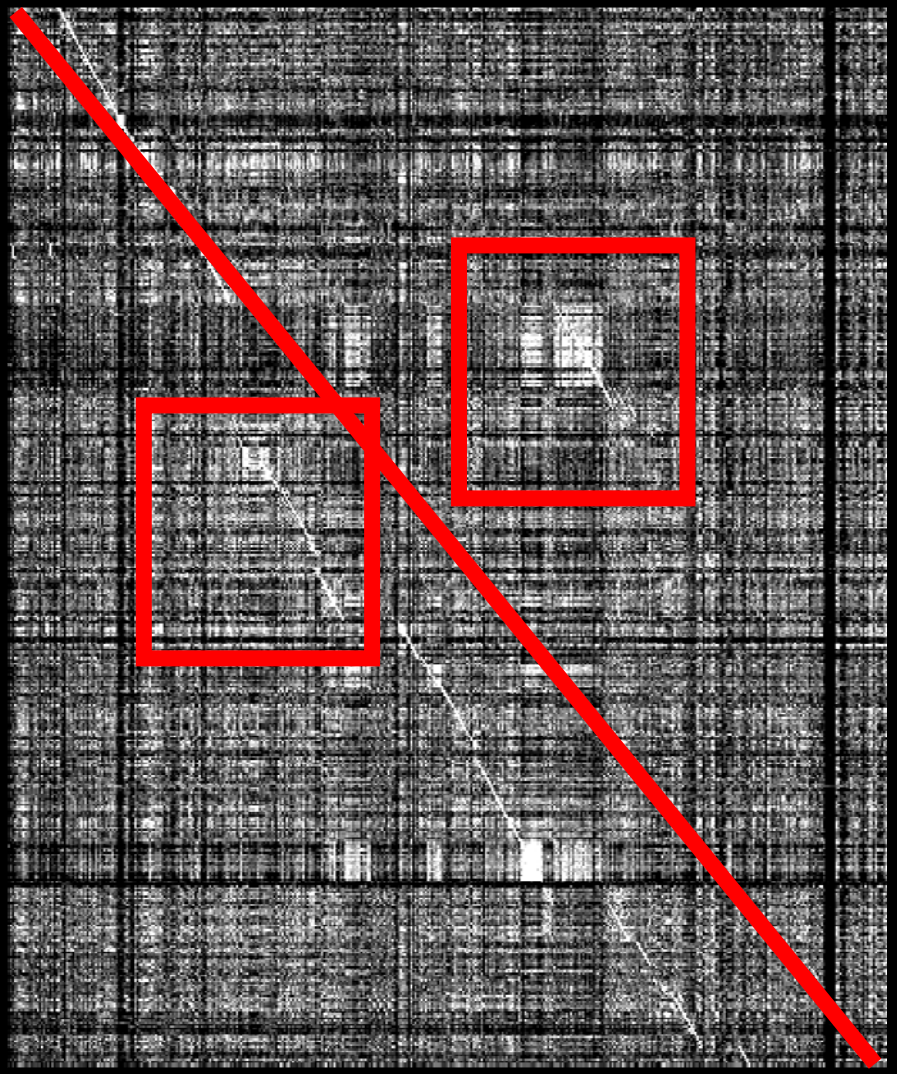

http://computer.forensikblog.de/en/2006/02/compare_binary_files_with_nwdiff.html
http://www.geocities.jp/belden_dr/ToolNwdiff_Eng.html

# Dot Plots & Visual BinDiff

(Kaminsky)

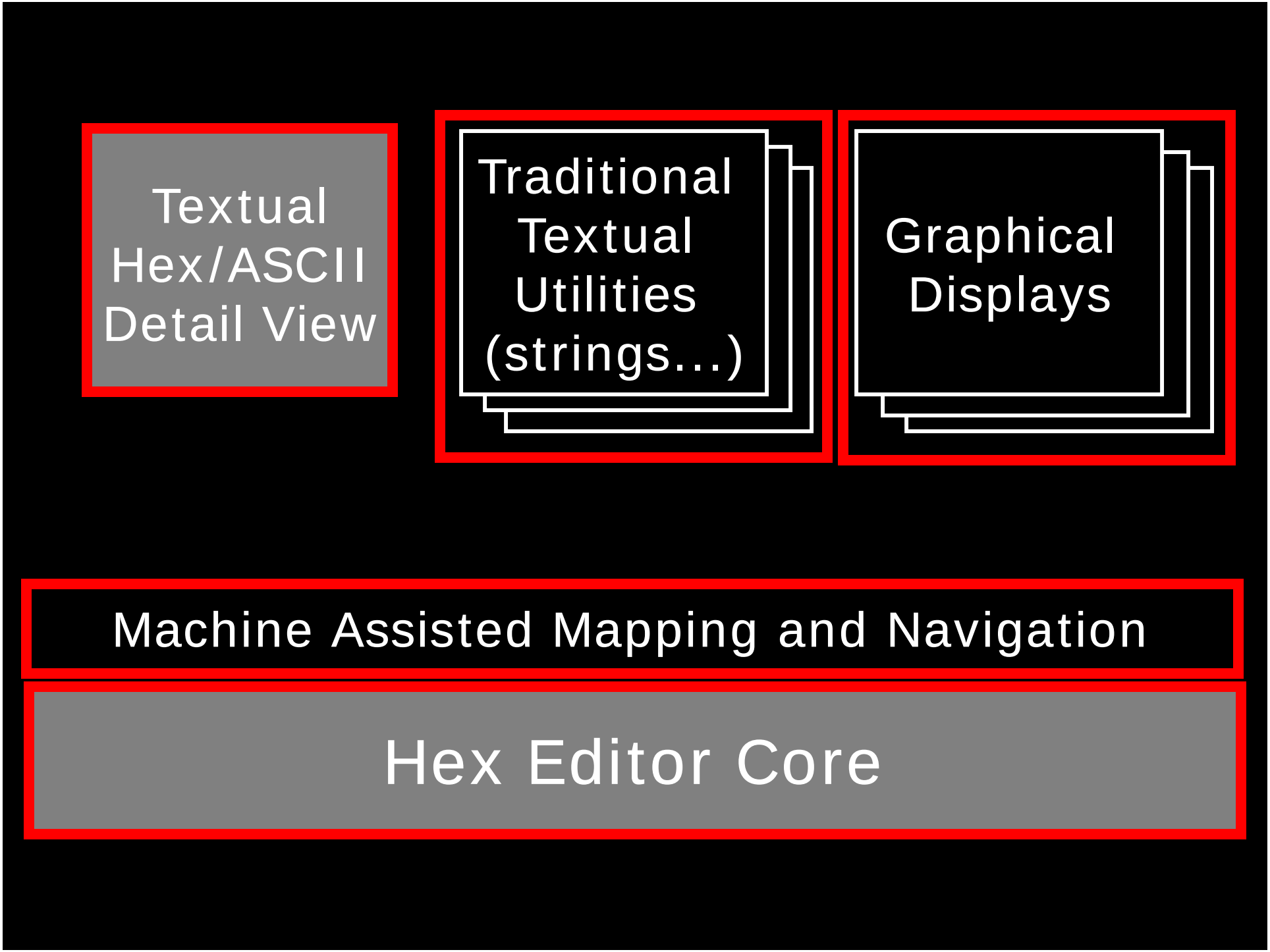


Self-Similarity in
a single file. (.NET Assembly)



Diffing Two Files

images: Dan Kaminsky, CCC2006


Textual
Hex/ASCII
Detail View
Traditional
Textual
Utilities
(strings...)
Graphical
Displays
Machine Assisted Mapping and Navigation
Hex Editor Core

# Towards a Visual Hex Editor

- Identify Unknown Binaries
- Malware Analysis
- Analyze Unknown/Undocumented File Format
- Locate Embedded Objects
  - Encoding / Encryption
- Audit Files for Vulnerabilities
- Compare files (Diffing)
- Cracking
- Cryptanalysis
- Perform Forensic Analysis
- File System Analysis
- Reporting
- File Fuzzing

# Goals

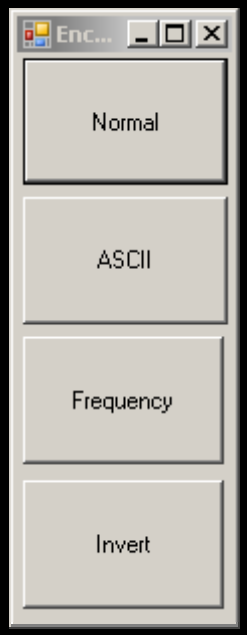
- Handle Large Files
- Many Insightful Windows
- Big Picture Context
- Improved Navigation
- Data Files / Executable Files
- Hex Editor best practices is the foundation
- Support Art & Science

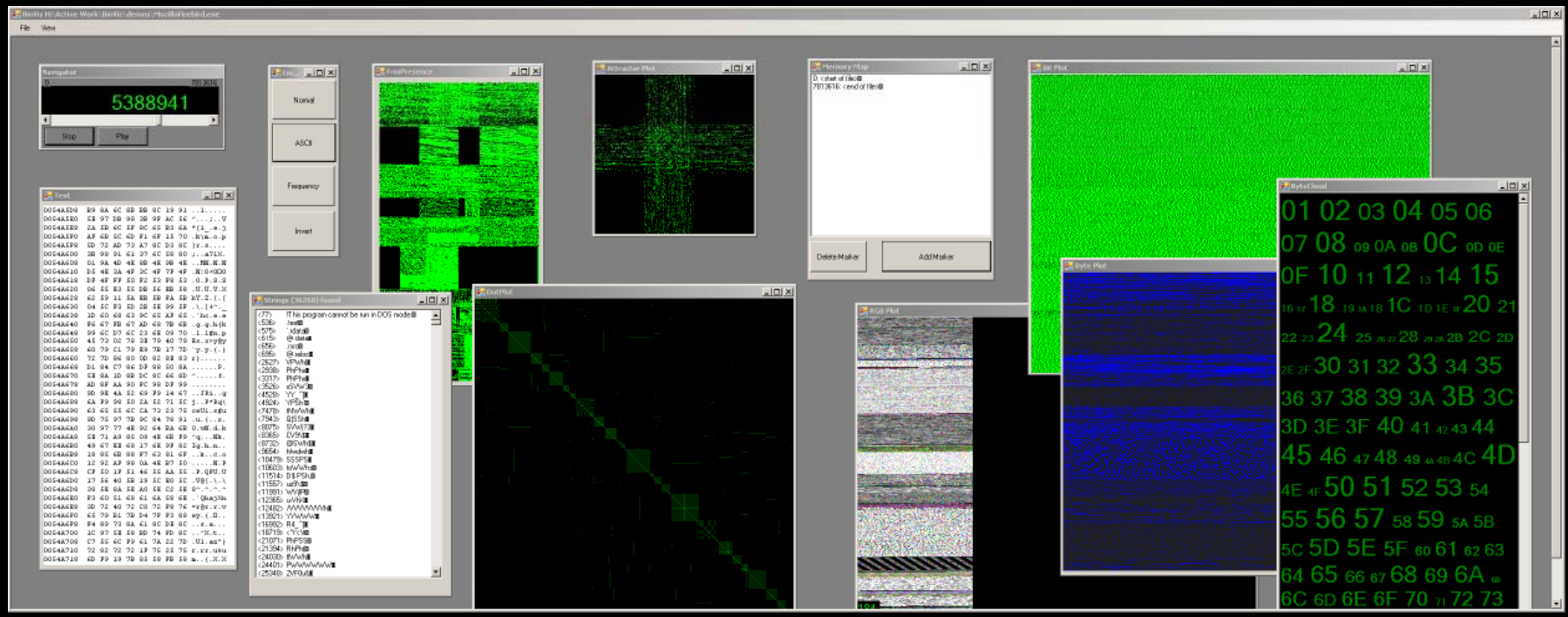
# Design

- Robust extensible framework
- Open source
- Context Independent File Analysis
- Semantic File Analysis
- Useful
- Multiple coordinated views
- Combine Functionality of current tools and extend with visuals

# Filtering + Encoding

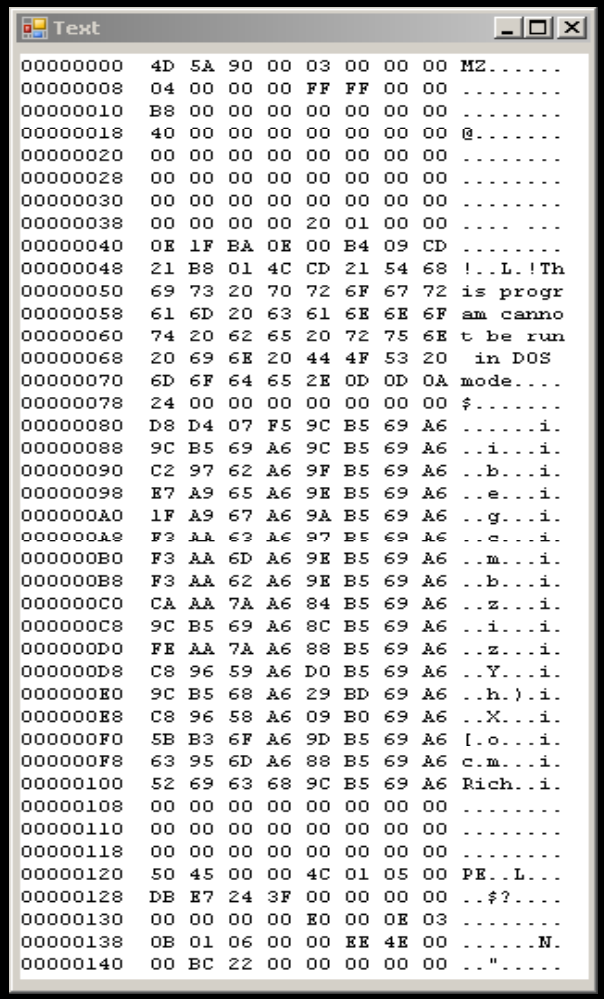
- Identifying something
  - REGEX → algorithmic
- Using this knowledge to..
  - highlight
  - fade
  - remove
- Interactive or automated






- **Textual**: Text/ASCII, Strings, ByteCloud
- **Graphical**: Bitplot, BytePlot, RGBPlot, BytePresence, ByteFrequency, Digram, Dotplot
- **Interaction**: VCR, Memory Map, Color Coding

# Traditional Views



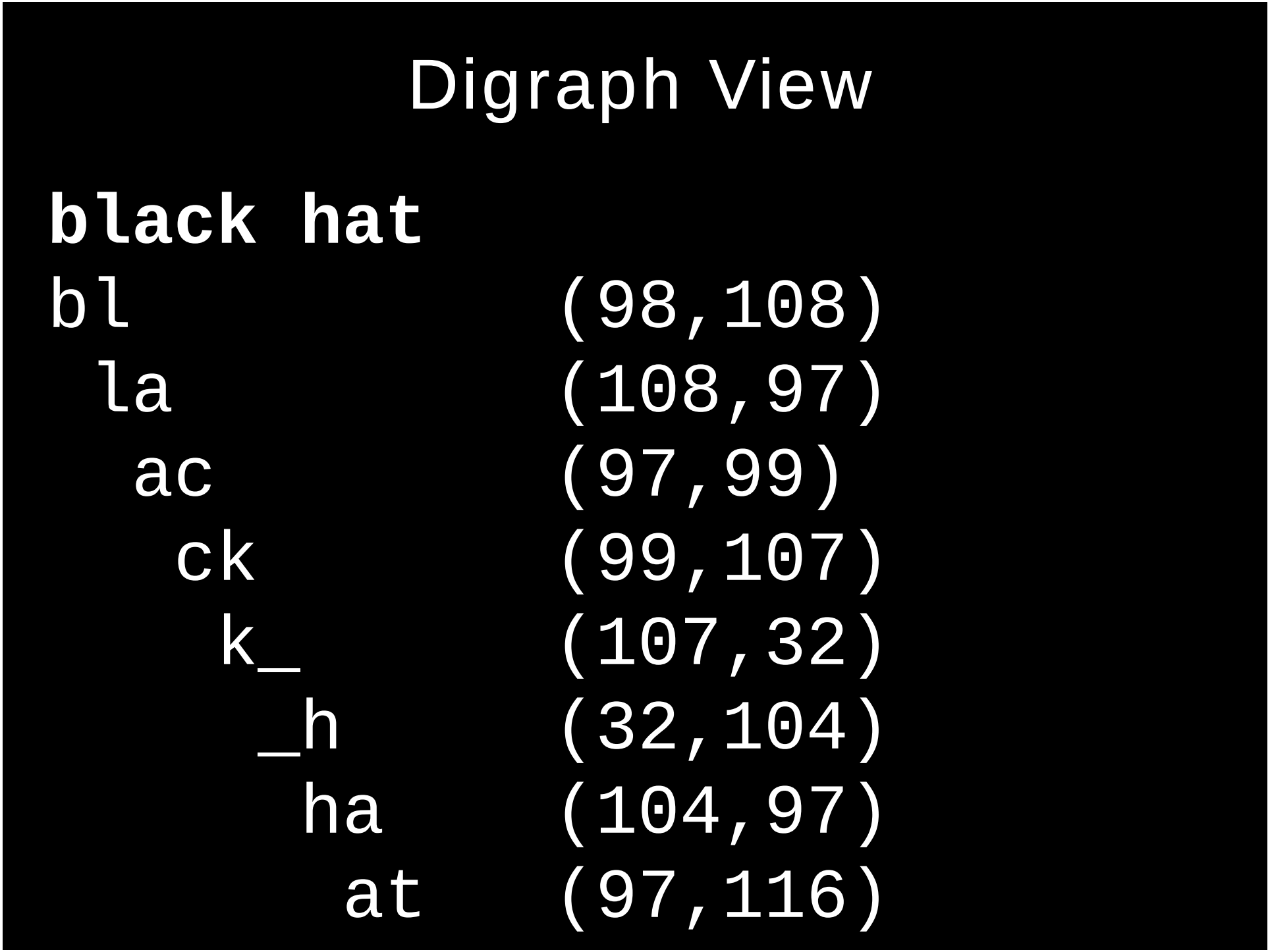

Hex / ASCII View


Strings

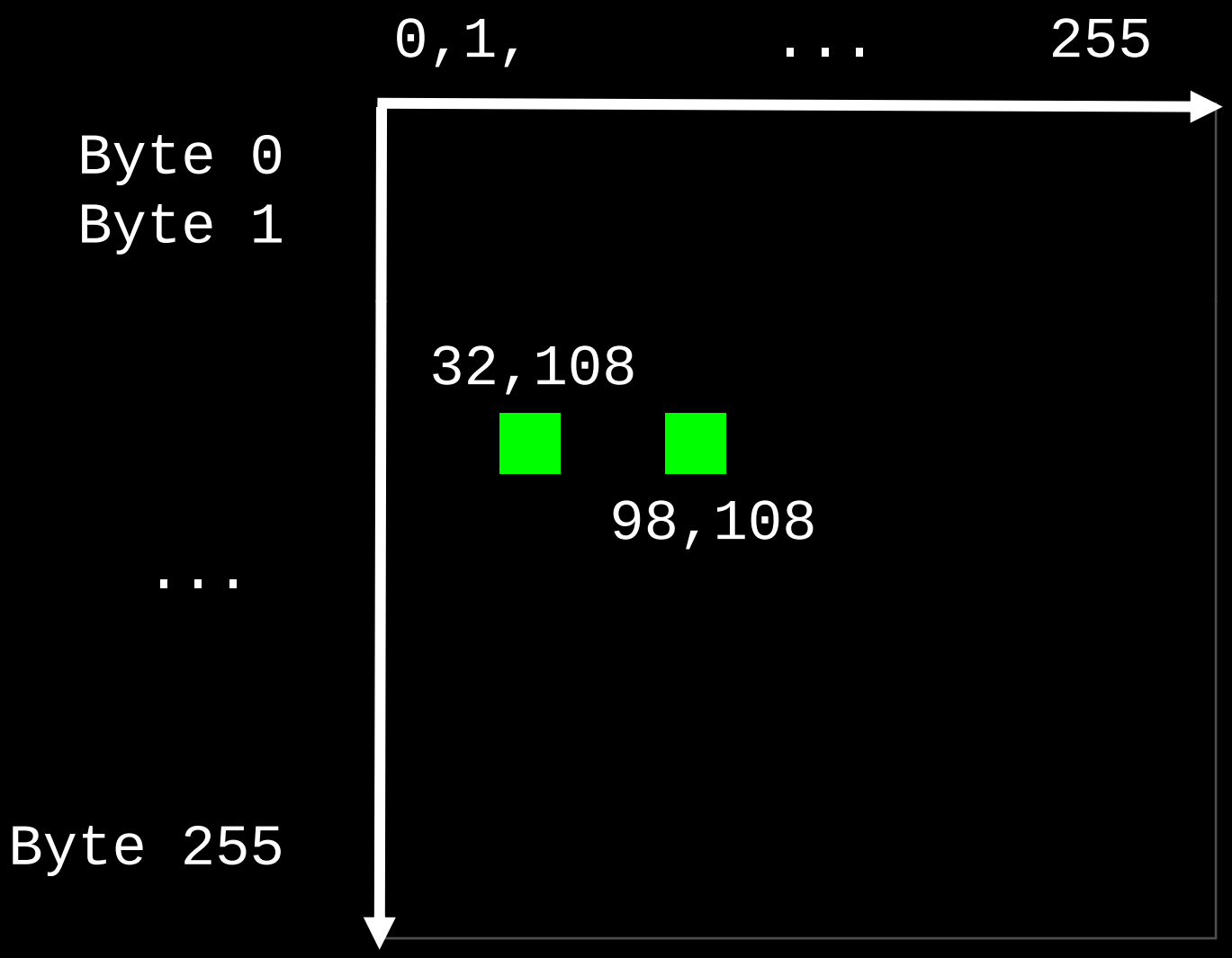
# Strange Attractors and TCP/IP Sequence Number Analysis

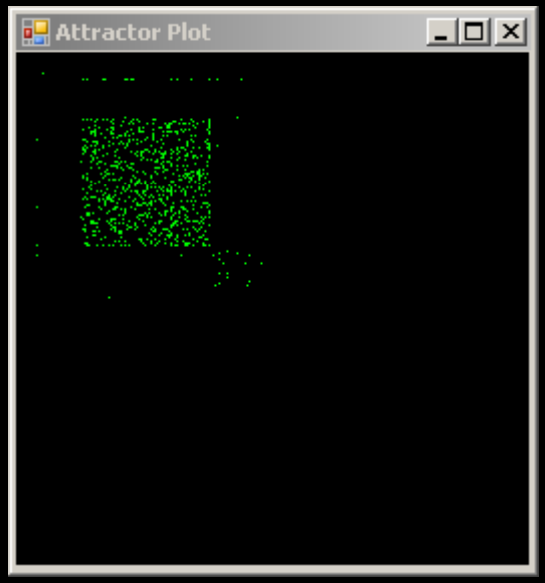
(Michal Zalewski)


- http://lcamtuf.coredump.cx/oldtcp/tcpseq.html
- http://lcamtuf.coredump.cx/newtcp/

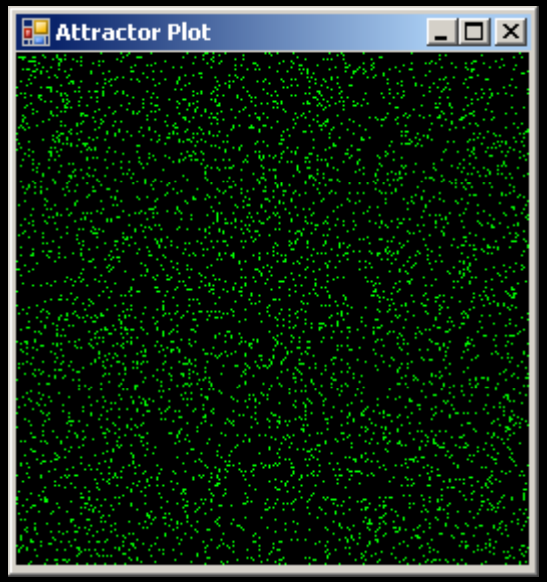

Digraph View
black hat
bl (98,108)
la (108,97)
ac (97,99)
ck (99,107)
k_ (107,32)
_h (32,104)
ha (104,97)
at (97,116)

# Digraph View

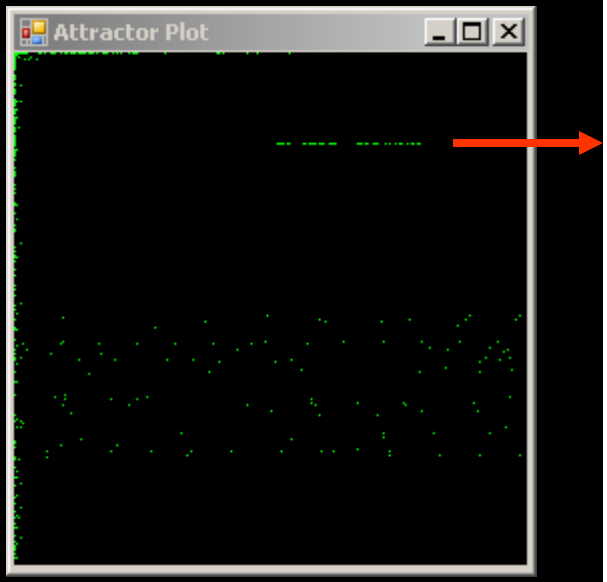





uuencoded

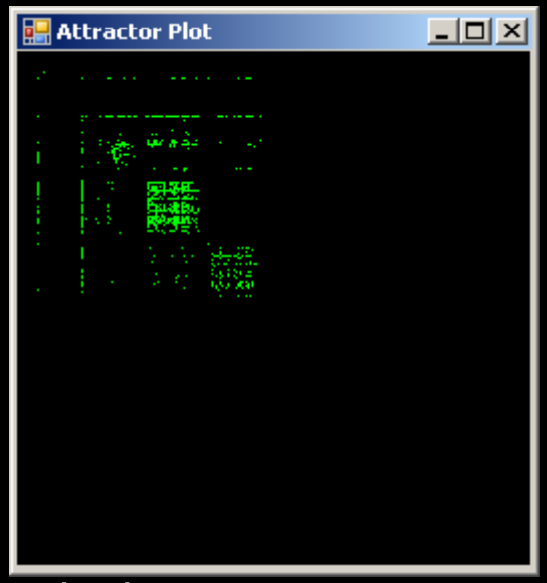



compression
encryption

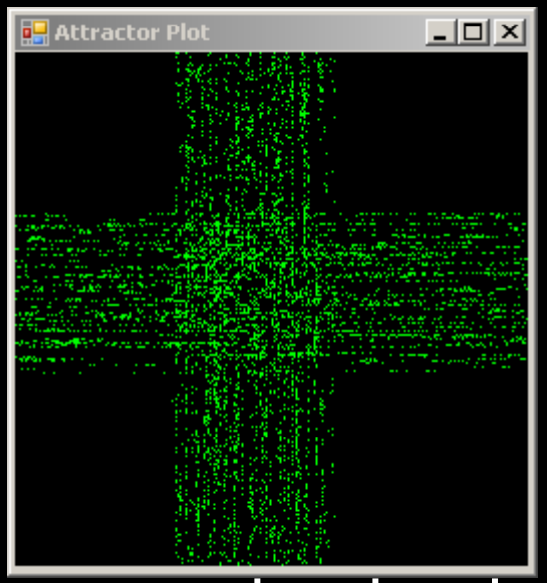



incrementing
words


slashdot.org

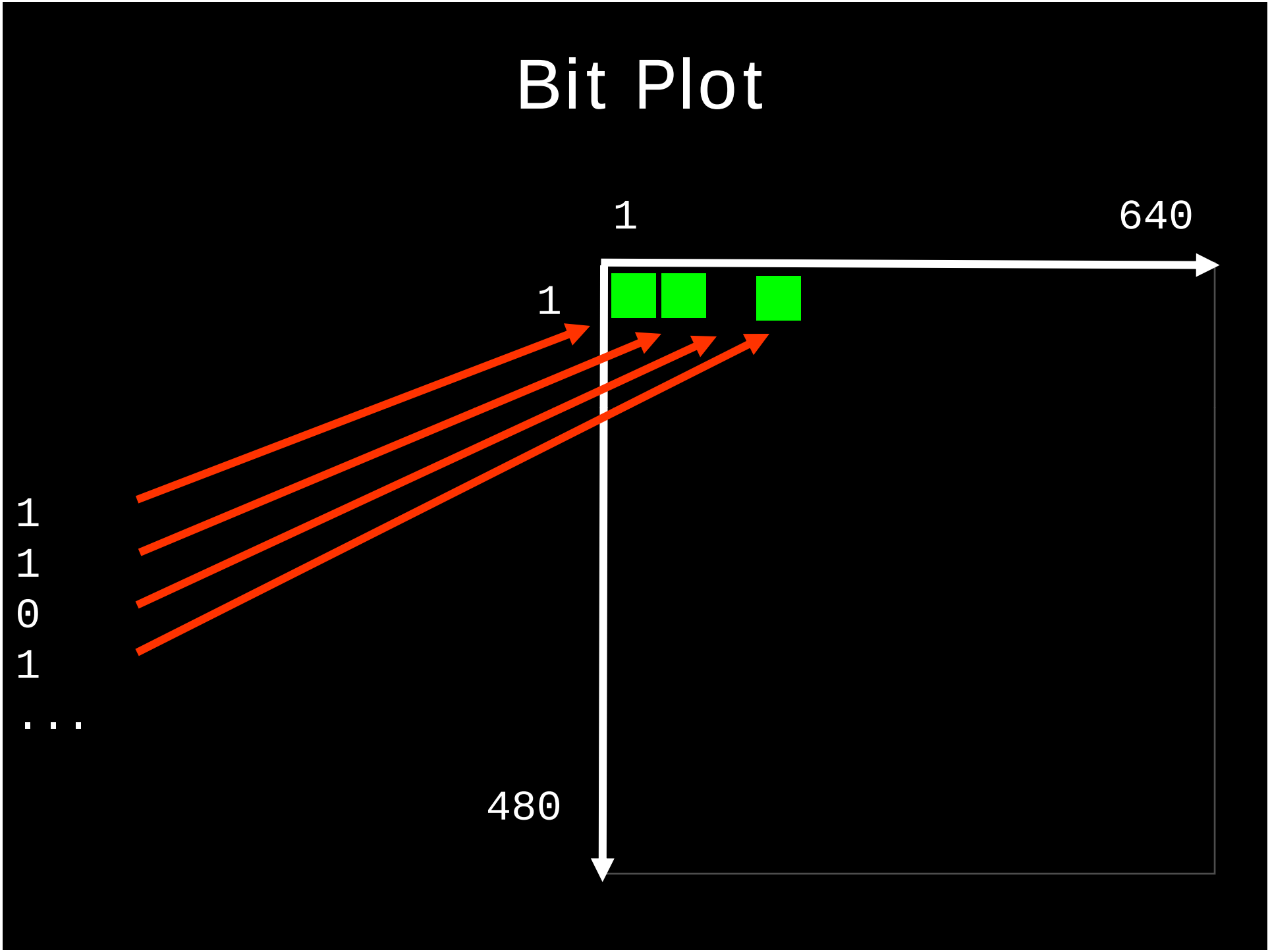



.txt

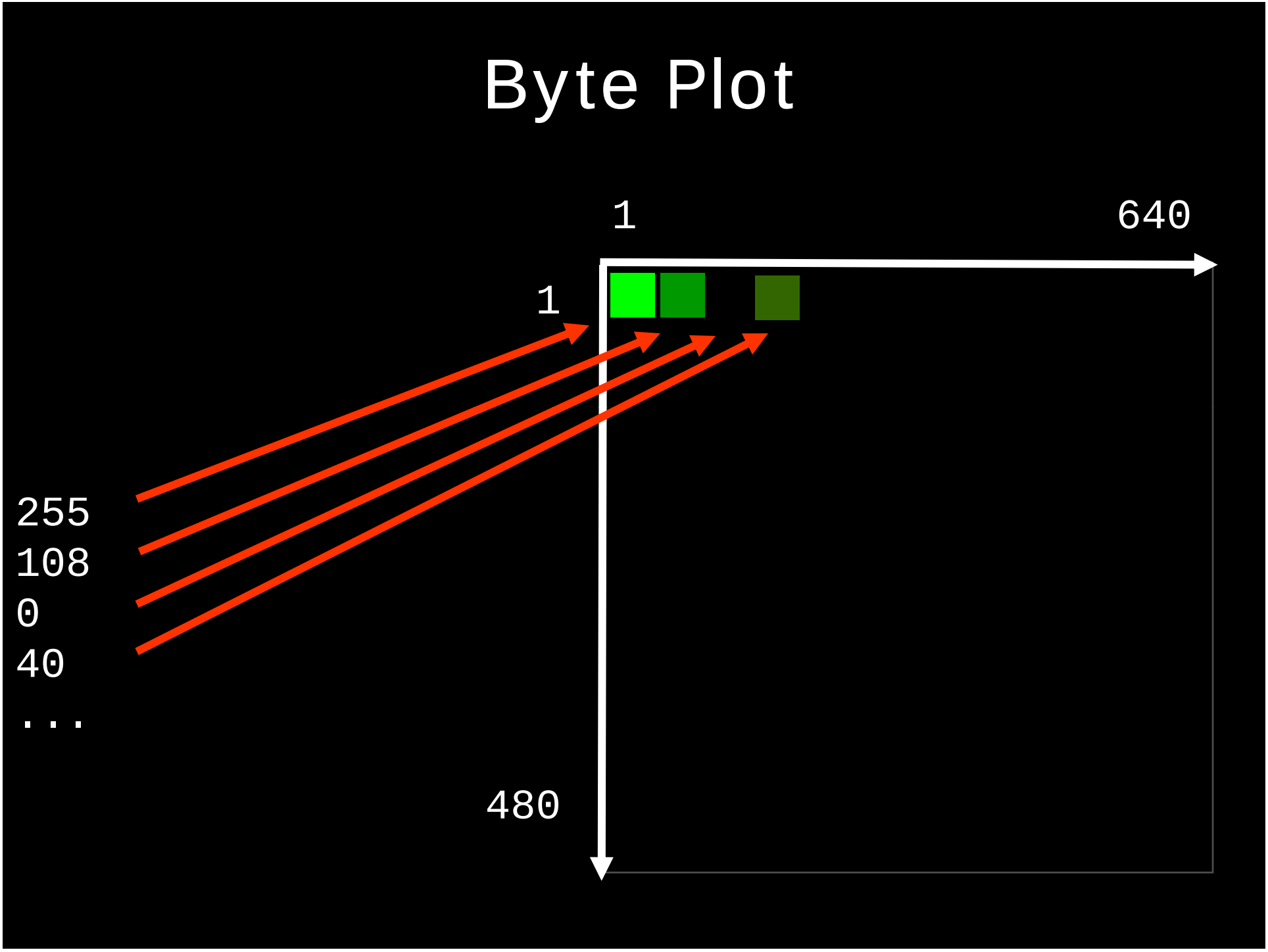



constrained pairs

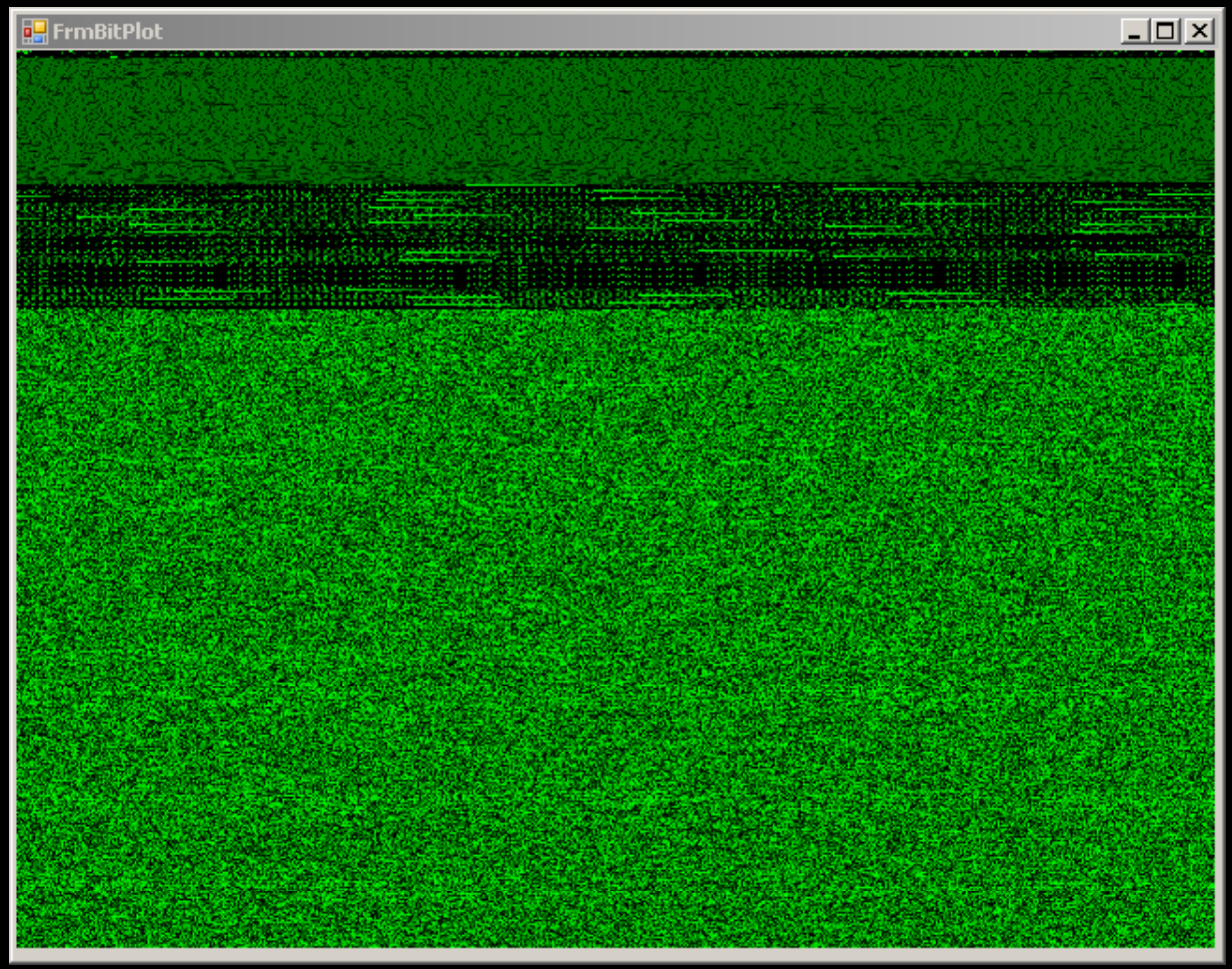
# Bit Plot



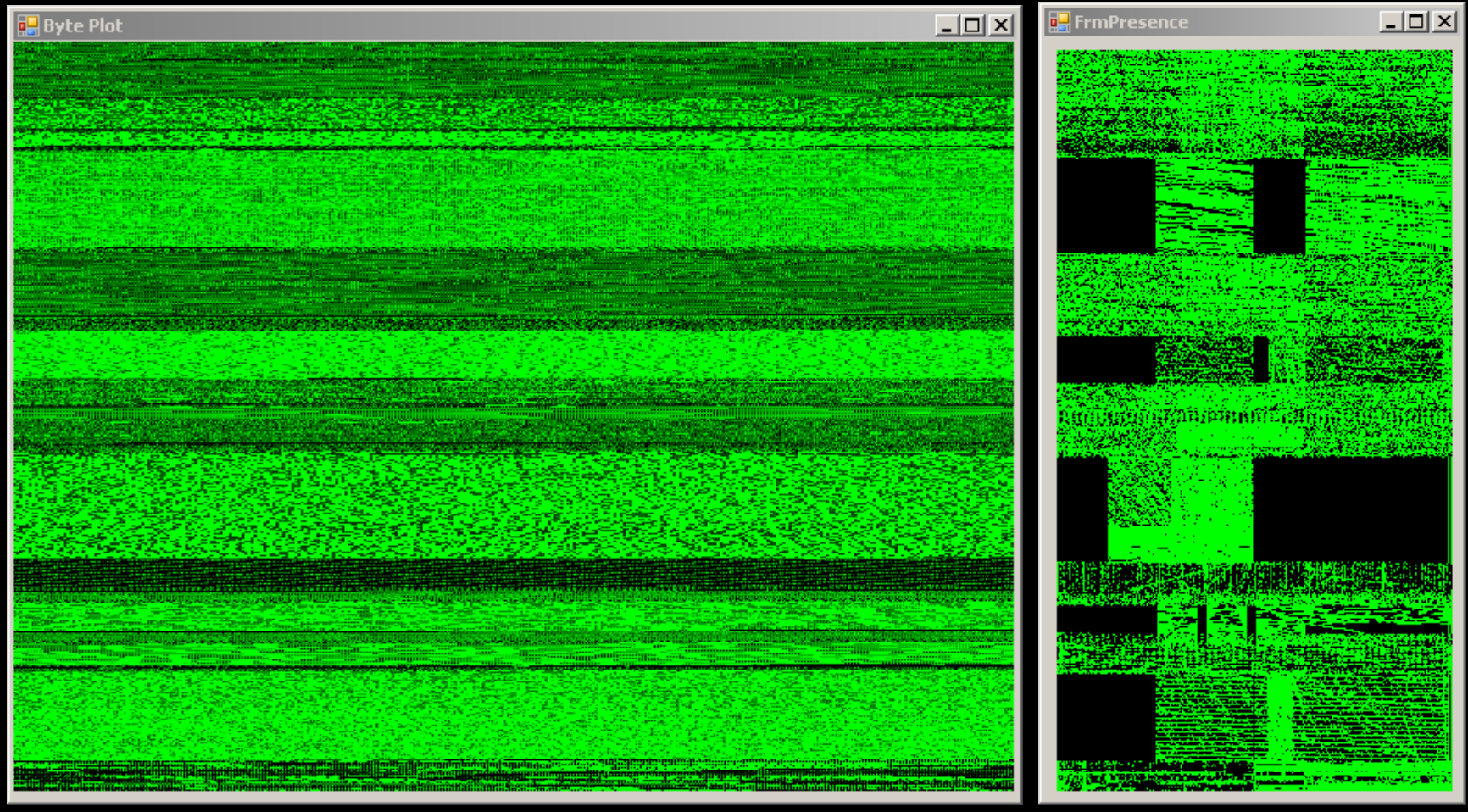
# Byte Plot

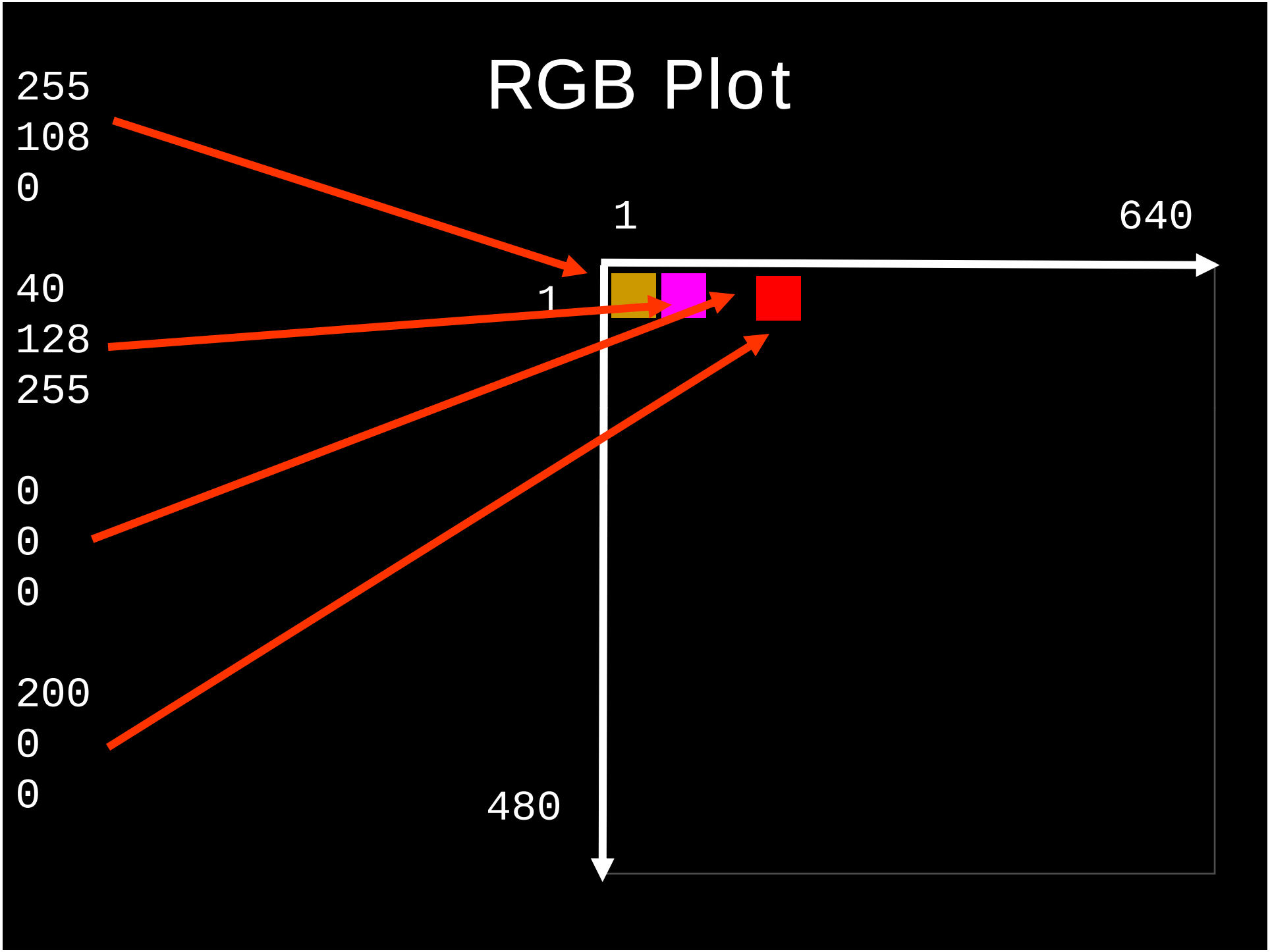


# Byte Plot Example
## (Word Document)



# Byte Presence




RGB Plot
255
108
0
40
128
255
0
0
0
200
0
0
1
640
1
480

# Dot Plots

- Jonathan Helfman's "Dotplot Patterns: A Literal Look at Pattern Languages."
- Dan Kaminsky, CCC & BH 2006

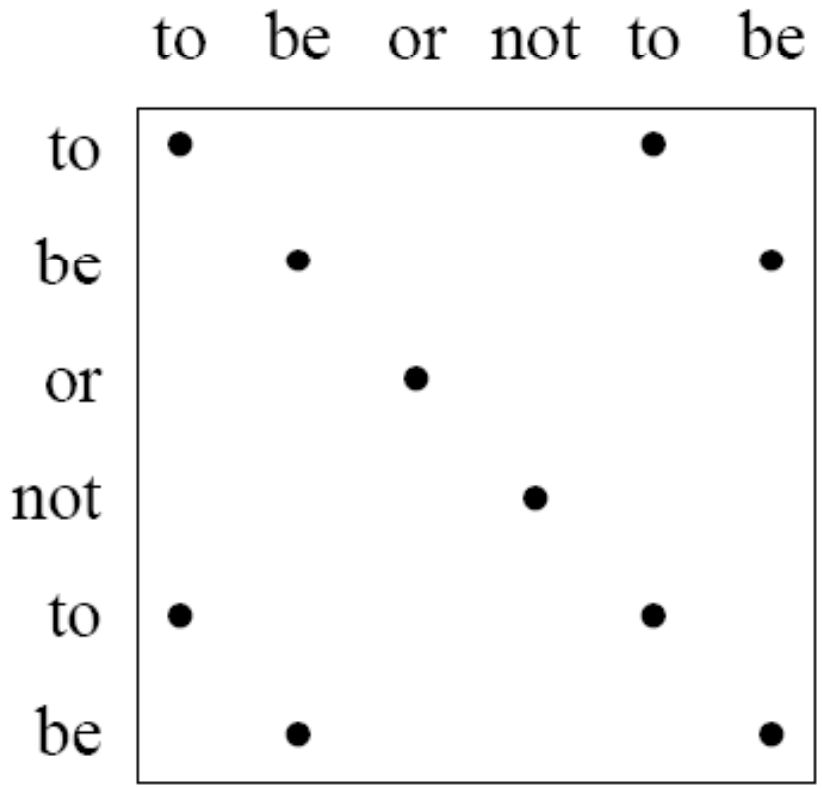
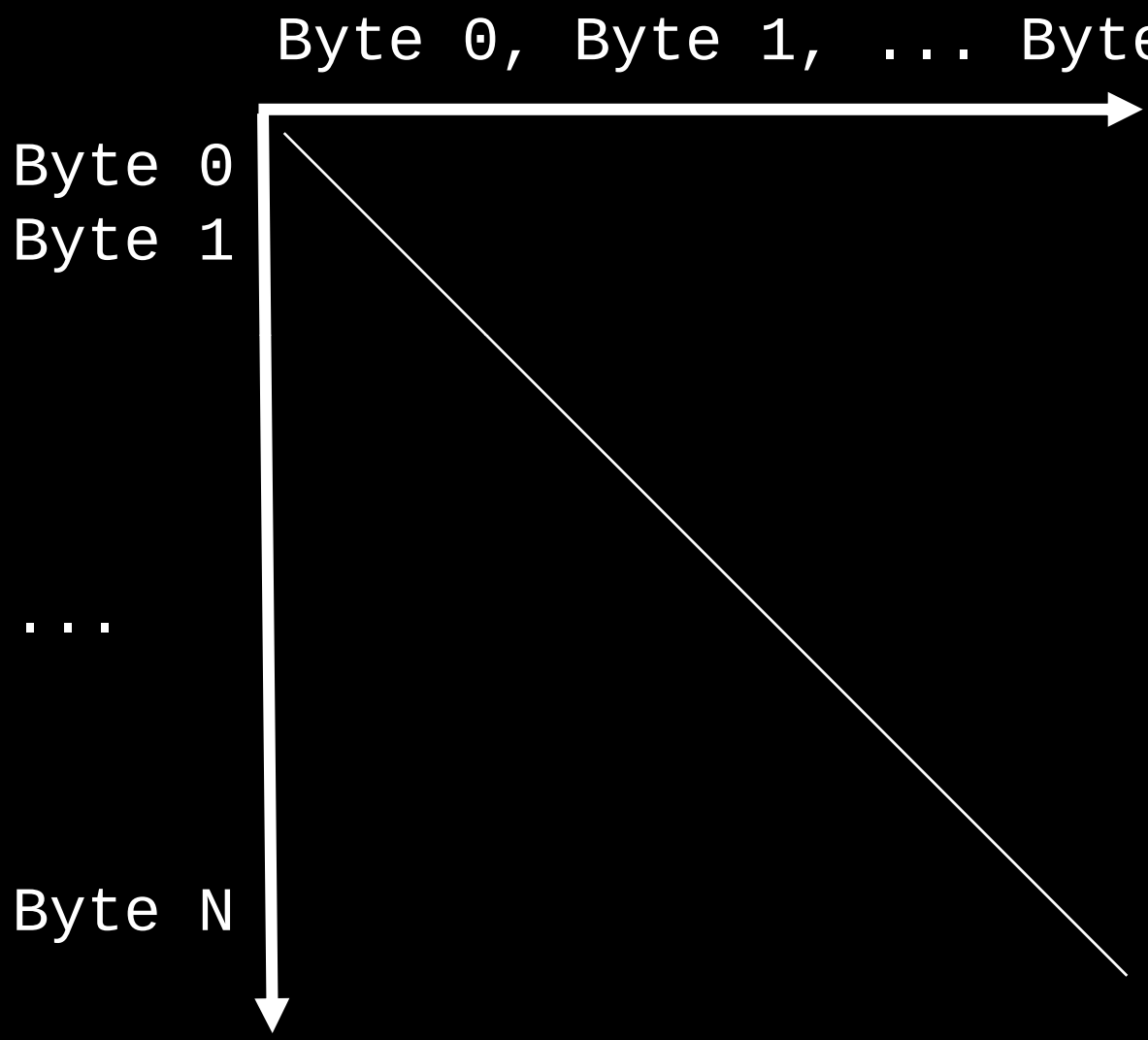


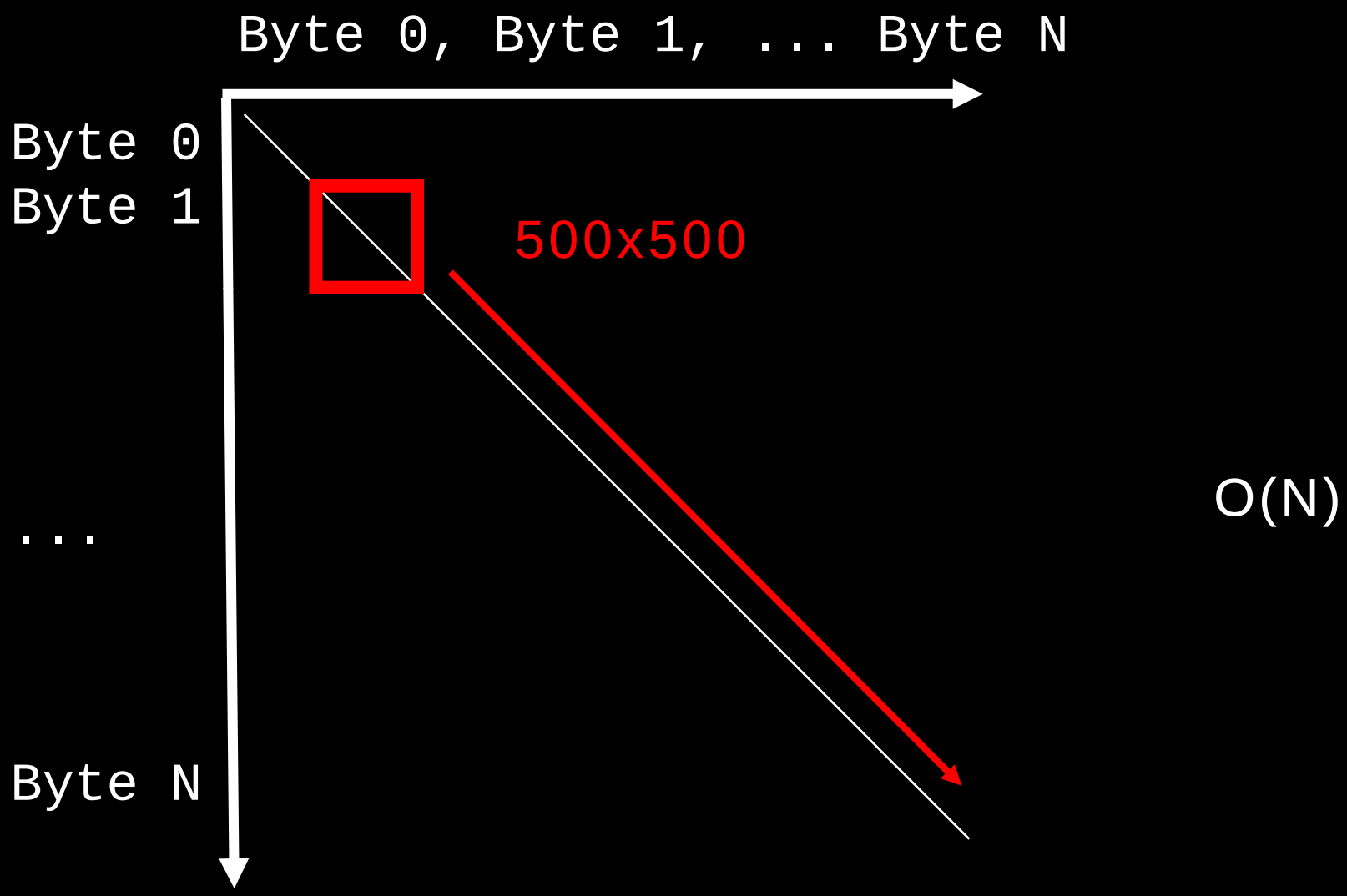

Figure 2: Six words of Shakespeare.

# DotPlots



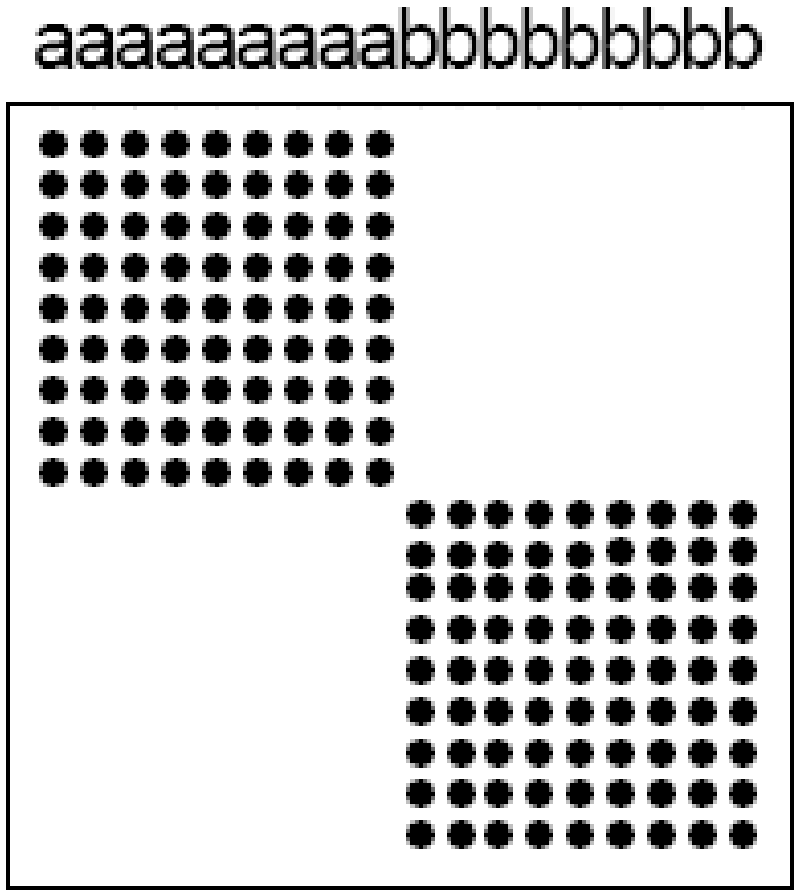

$O(N^2)$

# Dynamic DotPlots

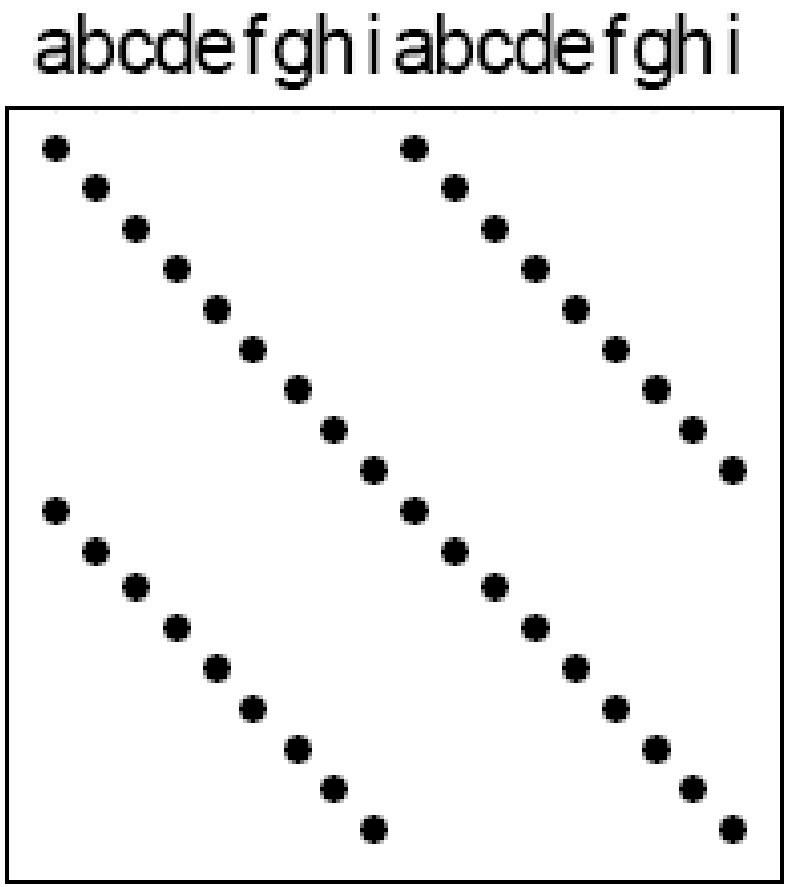



O(N)

# DotPlot Examples



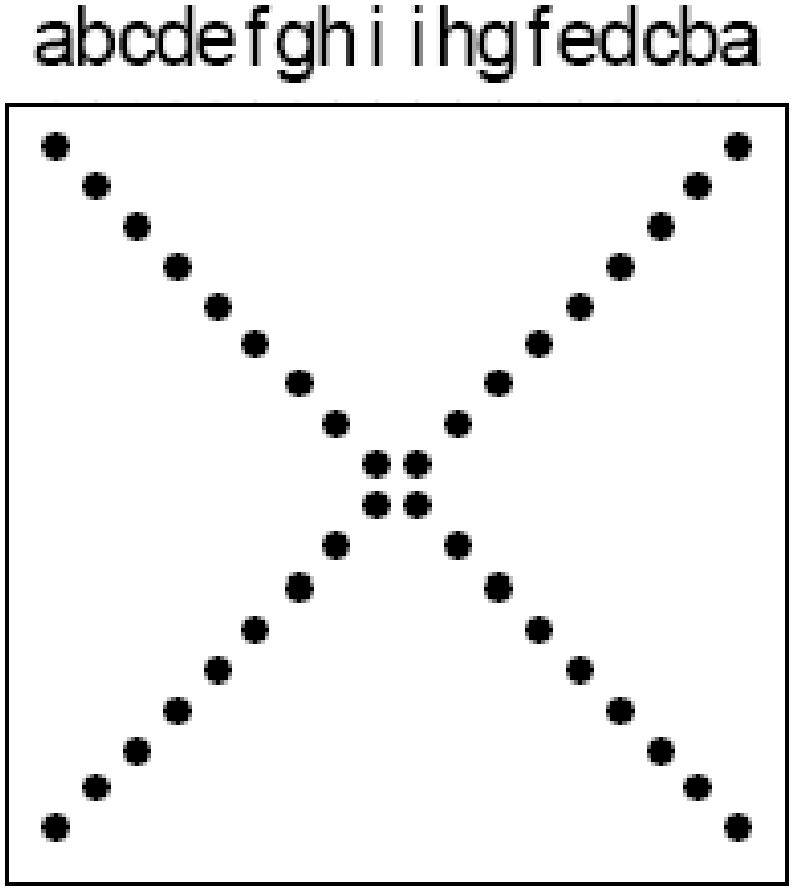

a) Squares.



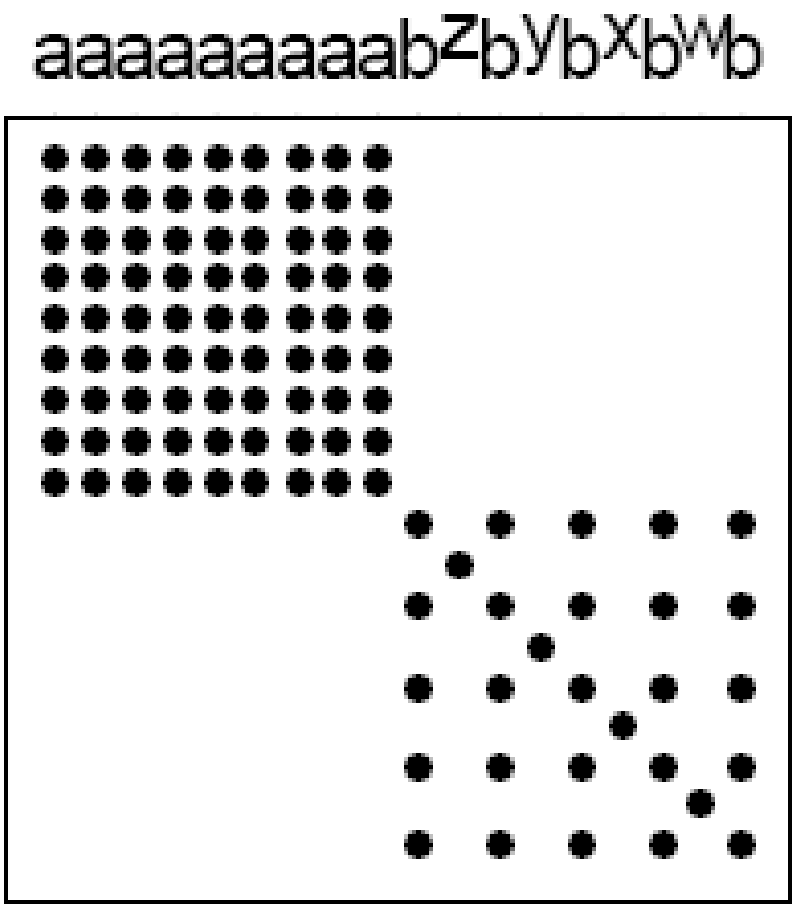

b) Diagonals.

Images: Jonathan Helfman, “Dotplot Patterns:  A Literal Look at Pattern Languages.”

# DotPlot Examples

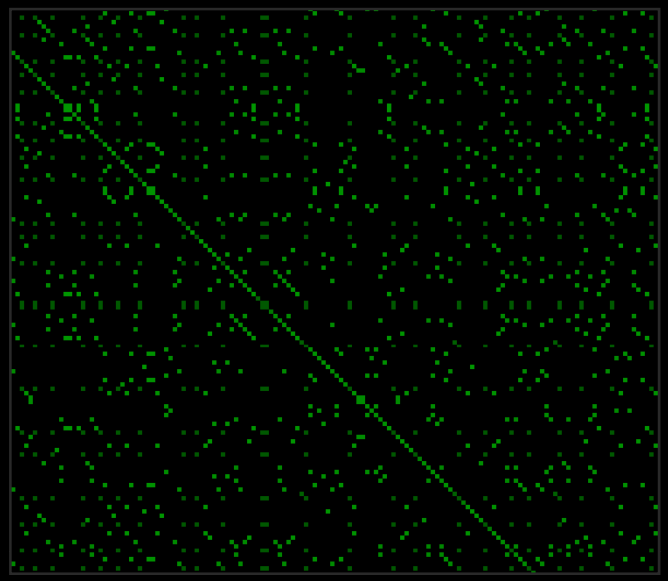



j) Palindrome.

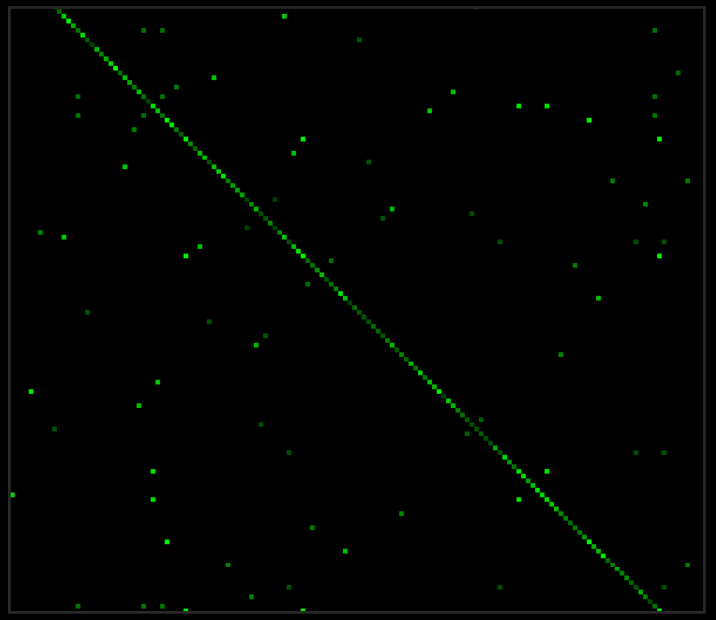



h) Density Variation.

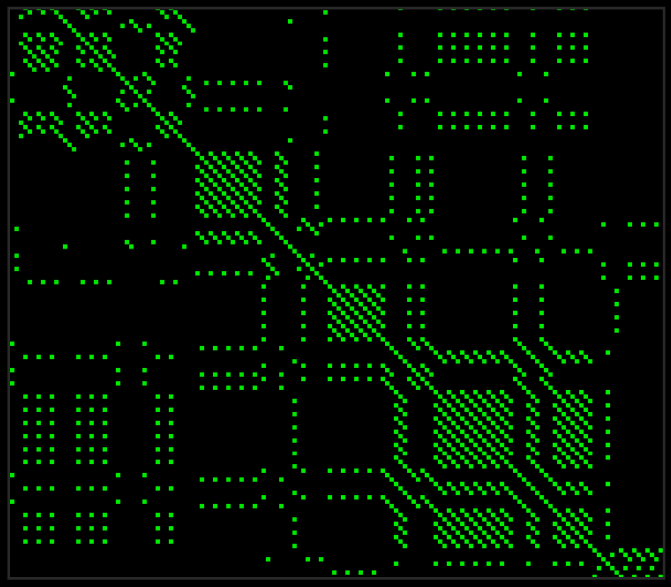
Images: Jonathan Helfman, “Dotplot Patterns:  A Literal Look at Pattern Languages.”



English Text

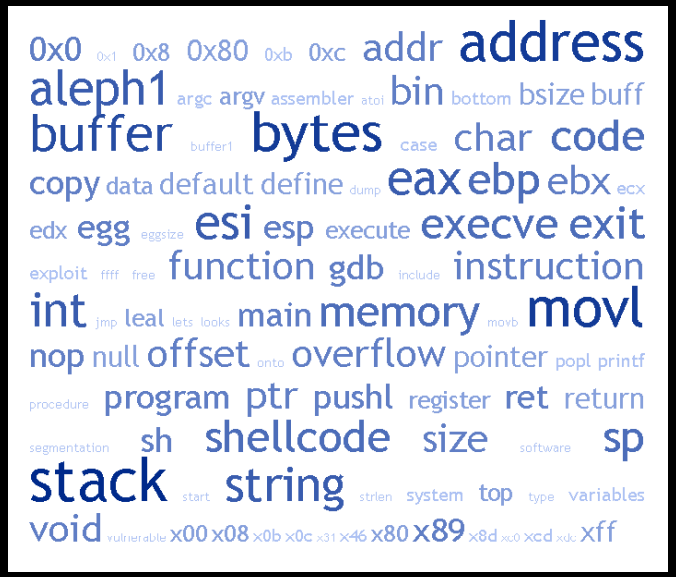


Compressed Audio



Bitmap Image

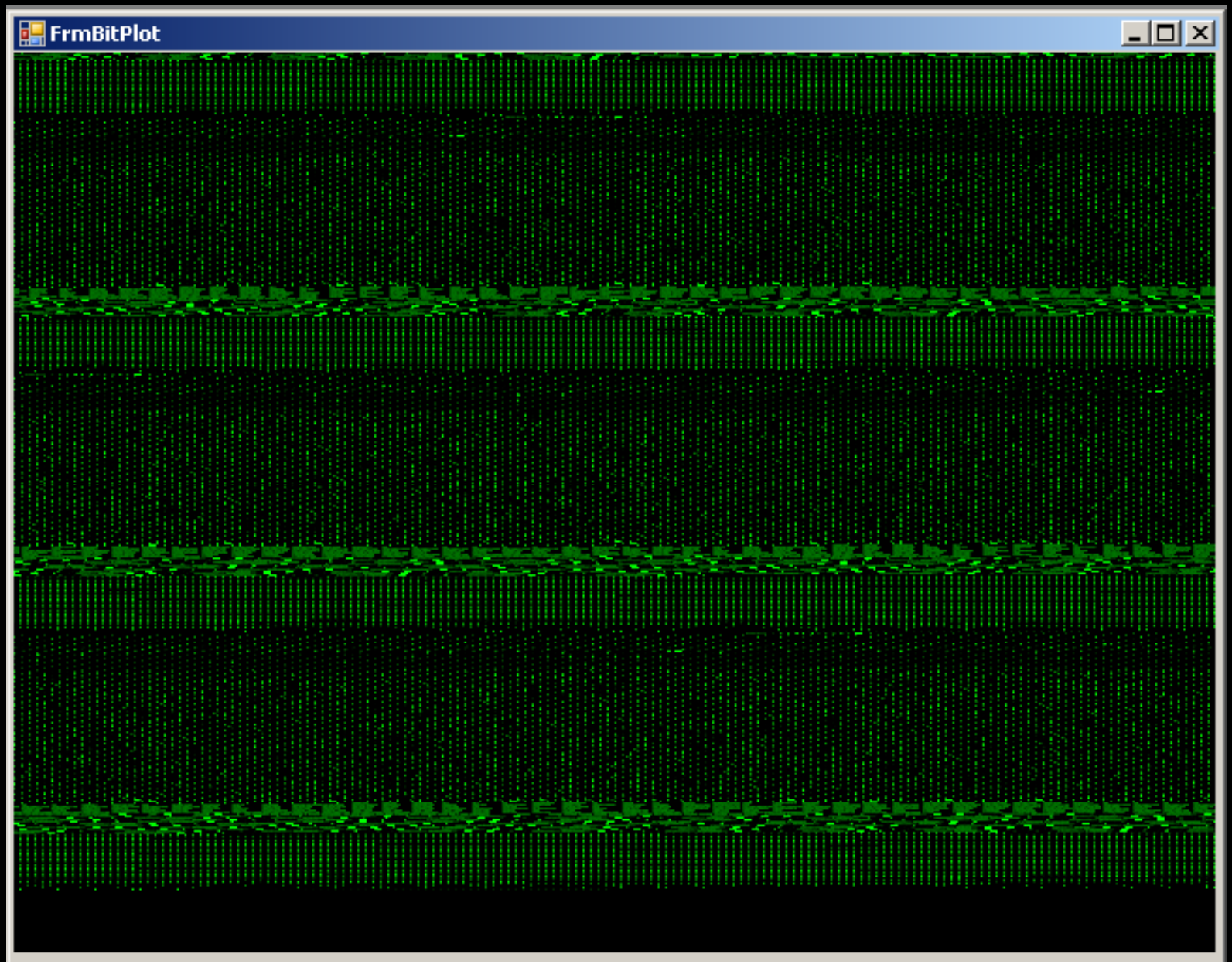
# Byte Clouds

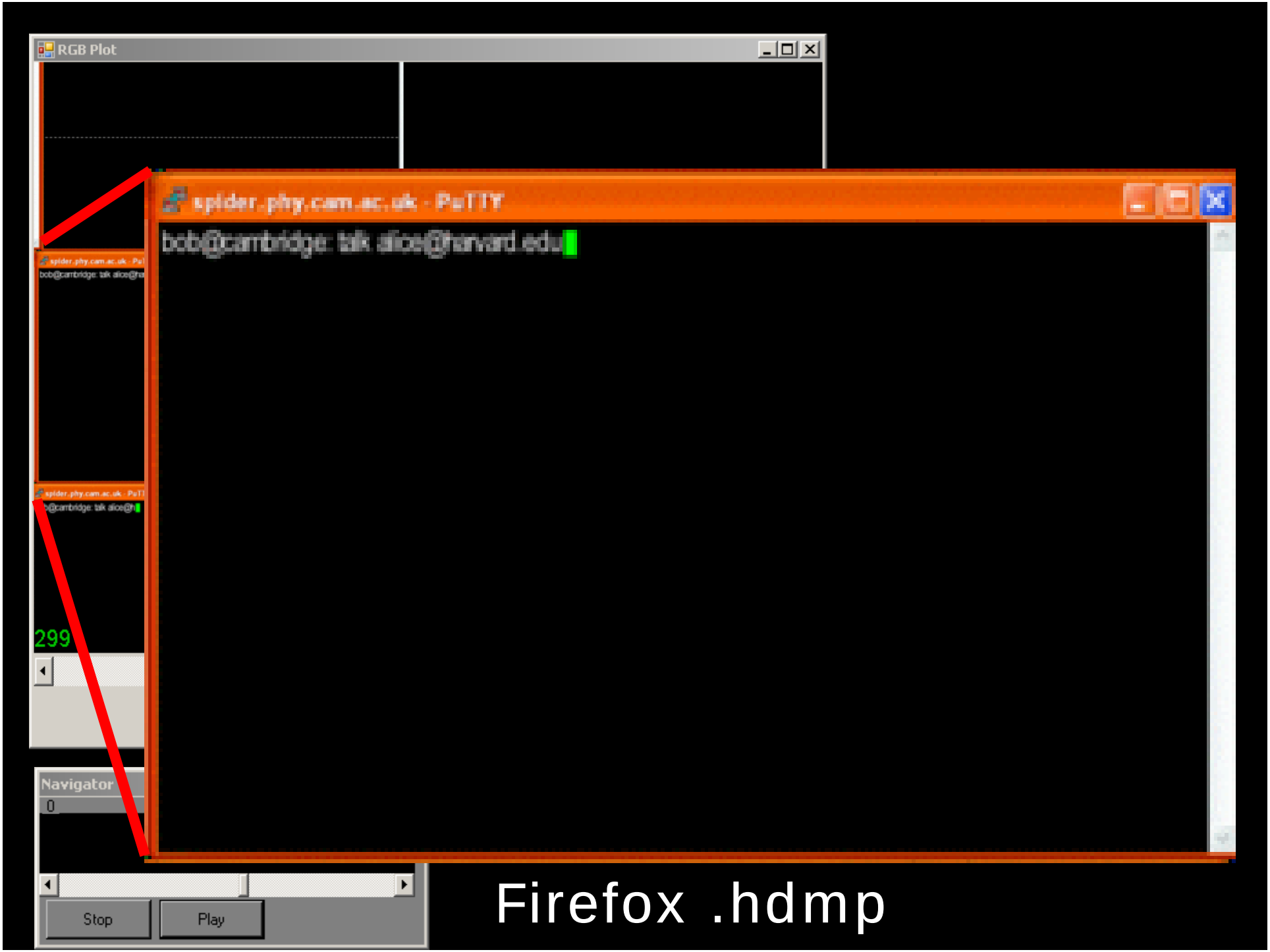



Tag Cloud
Smashing the Stack
for Fun and Profit
http://tagcrowd.com/

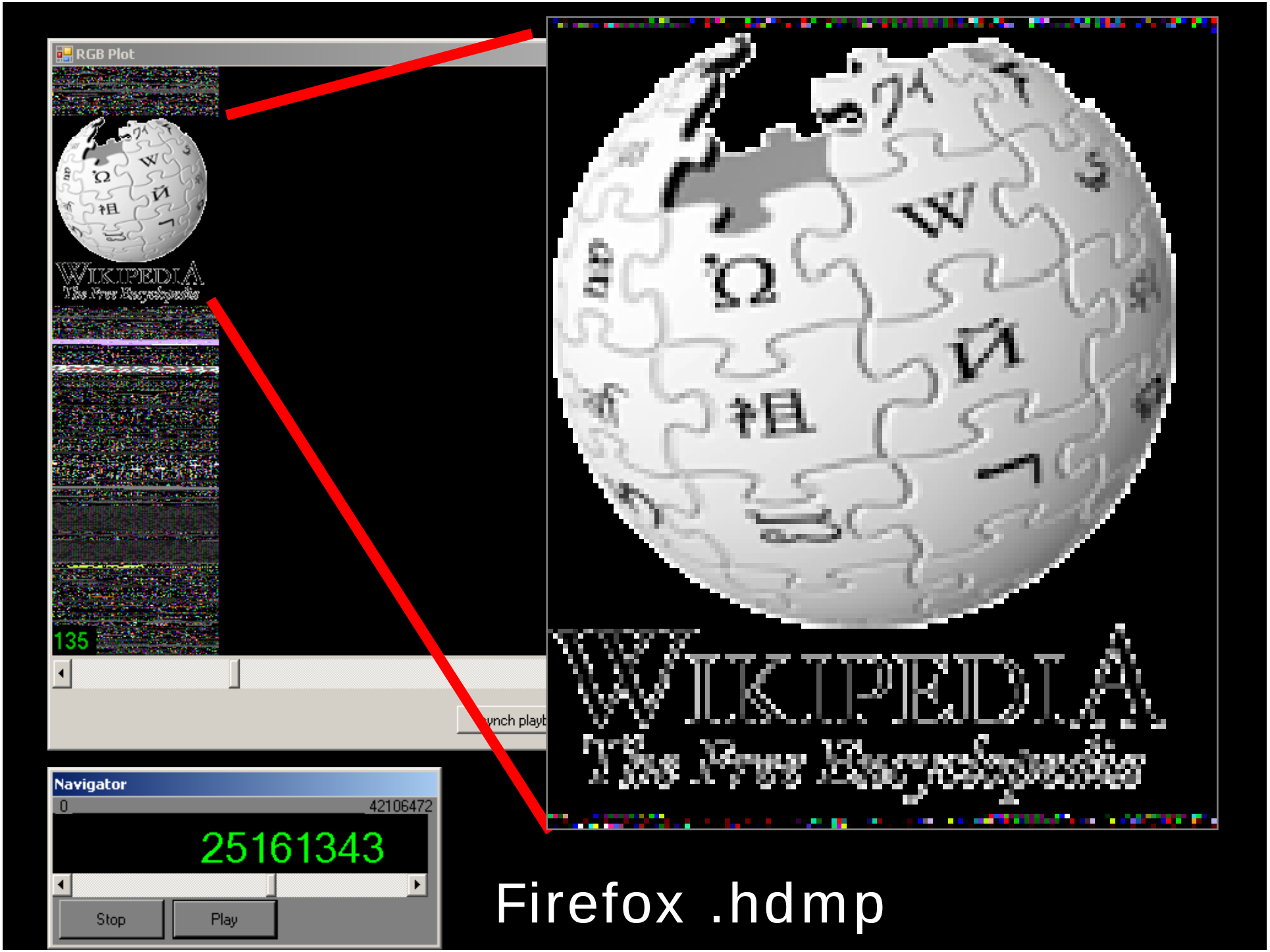

Byte Cloud

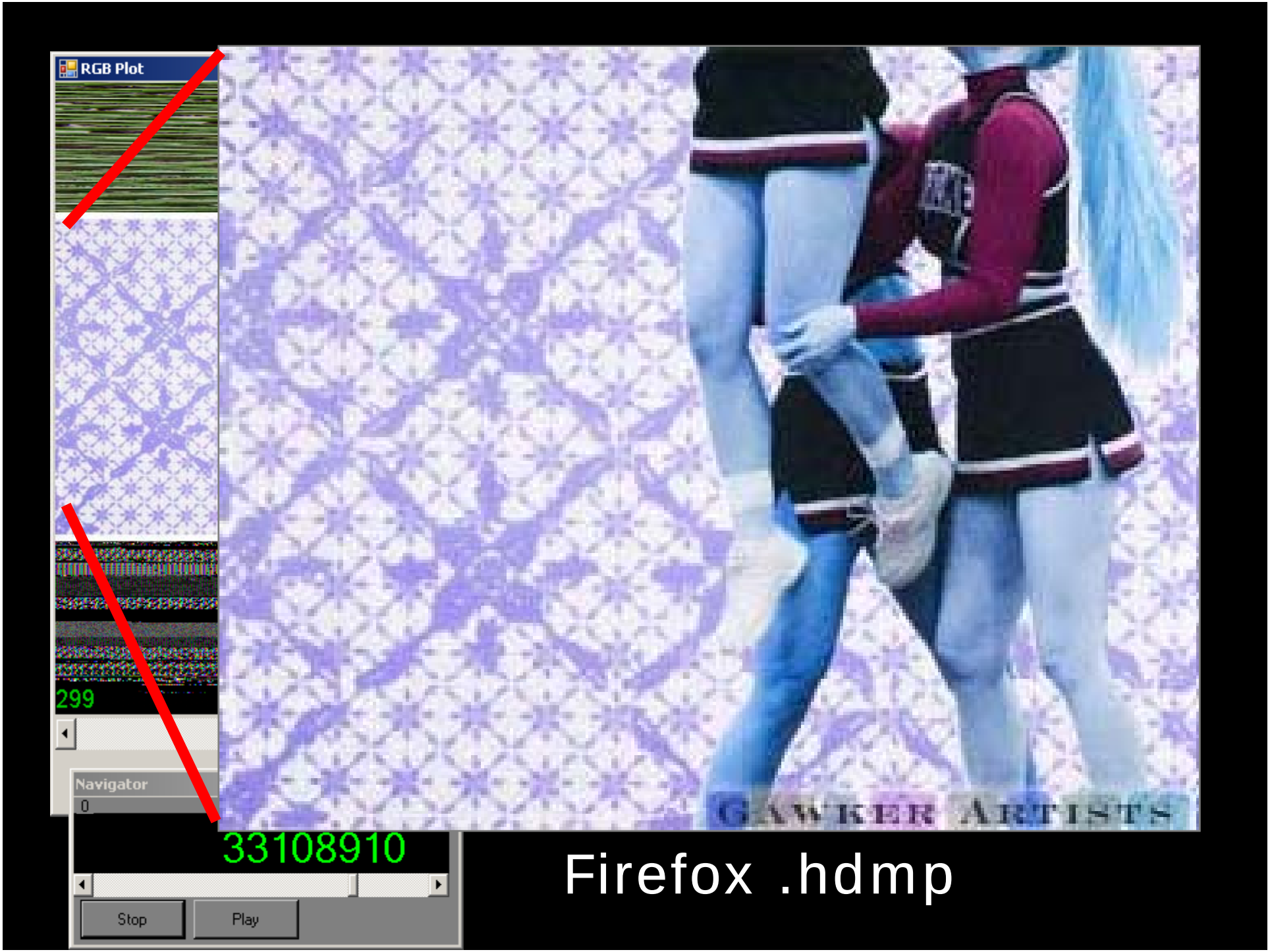
# Neverwinter Nights Database File



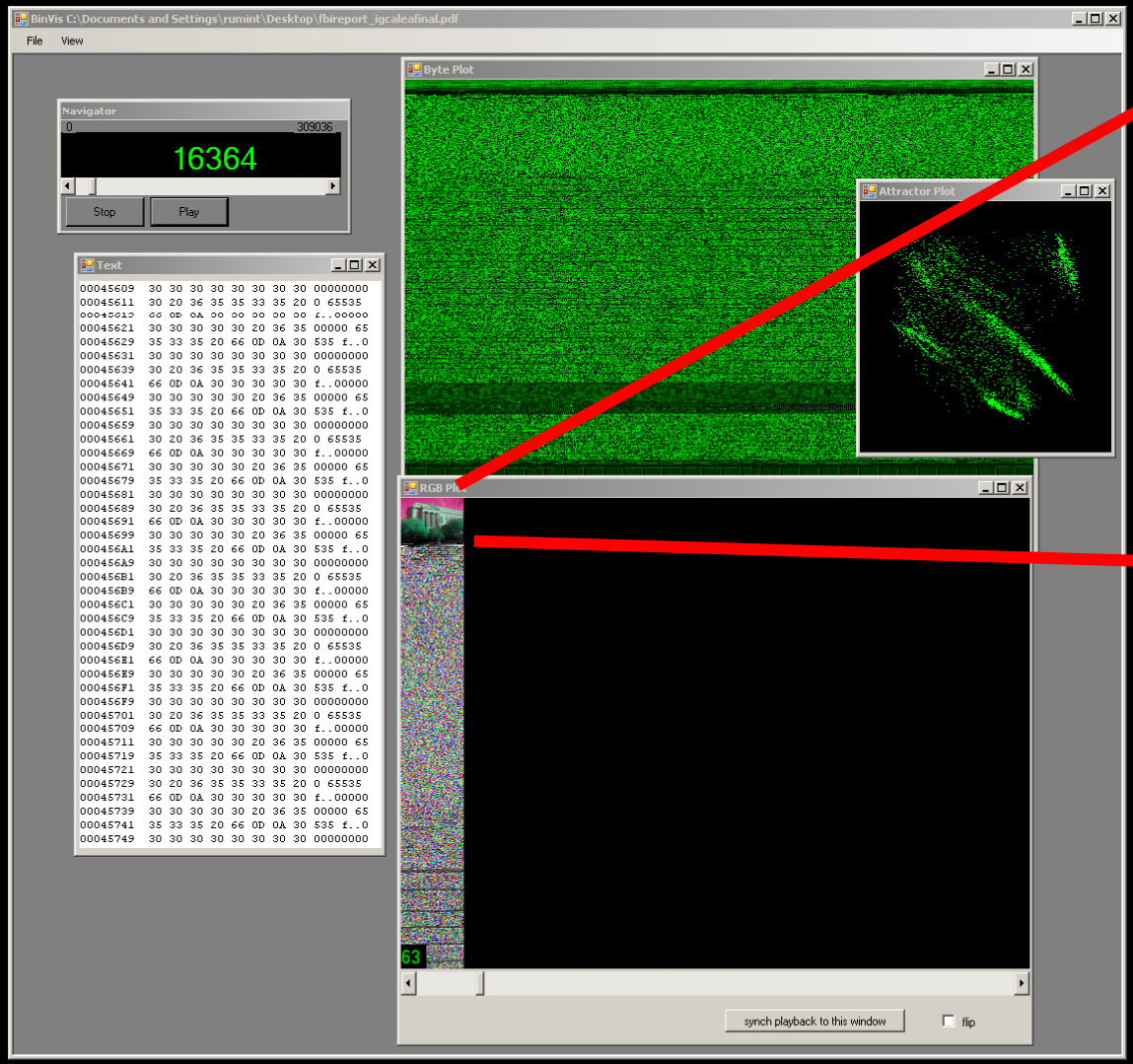



Firefox .hdmp


RGB Plot
135
Navigator
0
42106472
25161343
Stop
Play
Firefox .hdmp


RGB Plot
299
Navigator
0
33108910
Stop
Play
Firefox .hdmp


RGB Plot
319
synch playback to this window
flip
Navigator
0
42106472
24930574
Stop
Play
Firefox .hdmp




Redacted PDF...

# Weaknesses

- entire file may be extracted from bit/byte/RGB
  - May trigger AV or IDS
  - 8bit/byte steg
- Screams for big monitor

# Demos

# A Look to the Future…

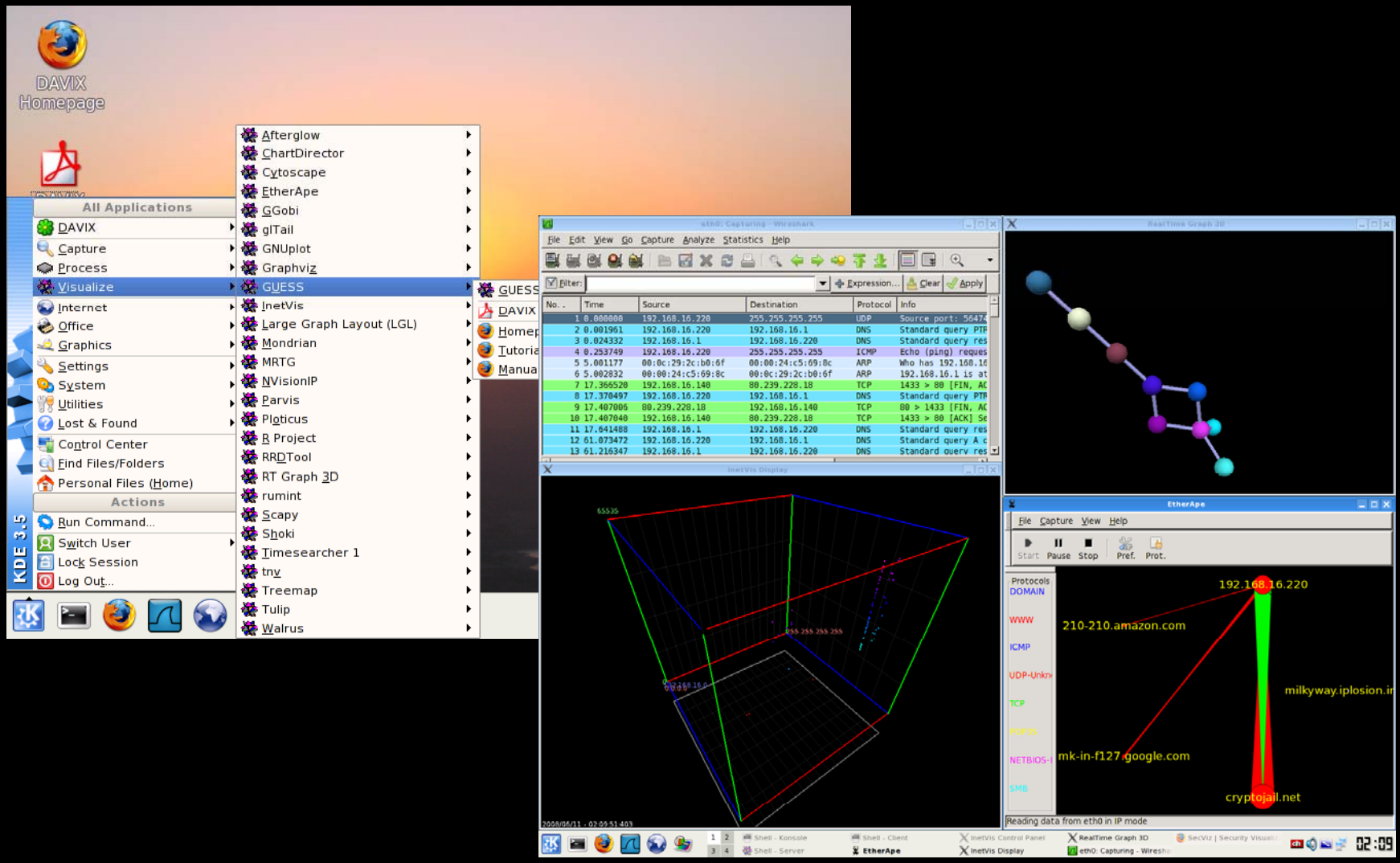
- Visual Front Ends for Offensive Tools
- Visual Cryptanalysis Support
- Human Insights Passed to Machine Processors
- User-centric Evaluation
- More Inspiration from General InfoVis Community
- Visual Fingerprints / Smart Books
- Web-based Visualization (AJAX)
- User-task Analyses
  - True Use Case Based Designs
  - Engagement of Users Beyond Students
- Examination of Full Range of Security Data
  - Merging Multiple Security Dataflows

# Future Work

- Plug-ins / Editable Config Files
  - Visualizations
  - Encodings
- Saving state
  - Memory Maps
- Improving Interaction
  - What works / What doesn’t
- Multiple Files / File Systems
- REGEX search
- Automated Memory Map Generation

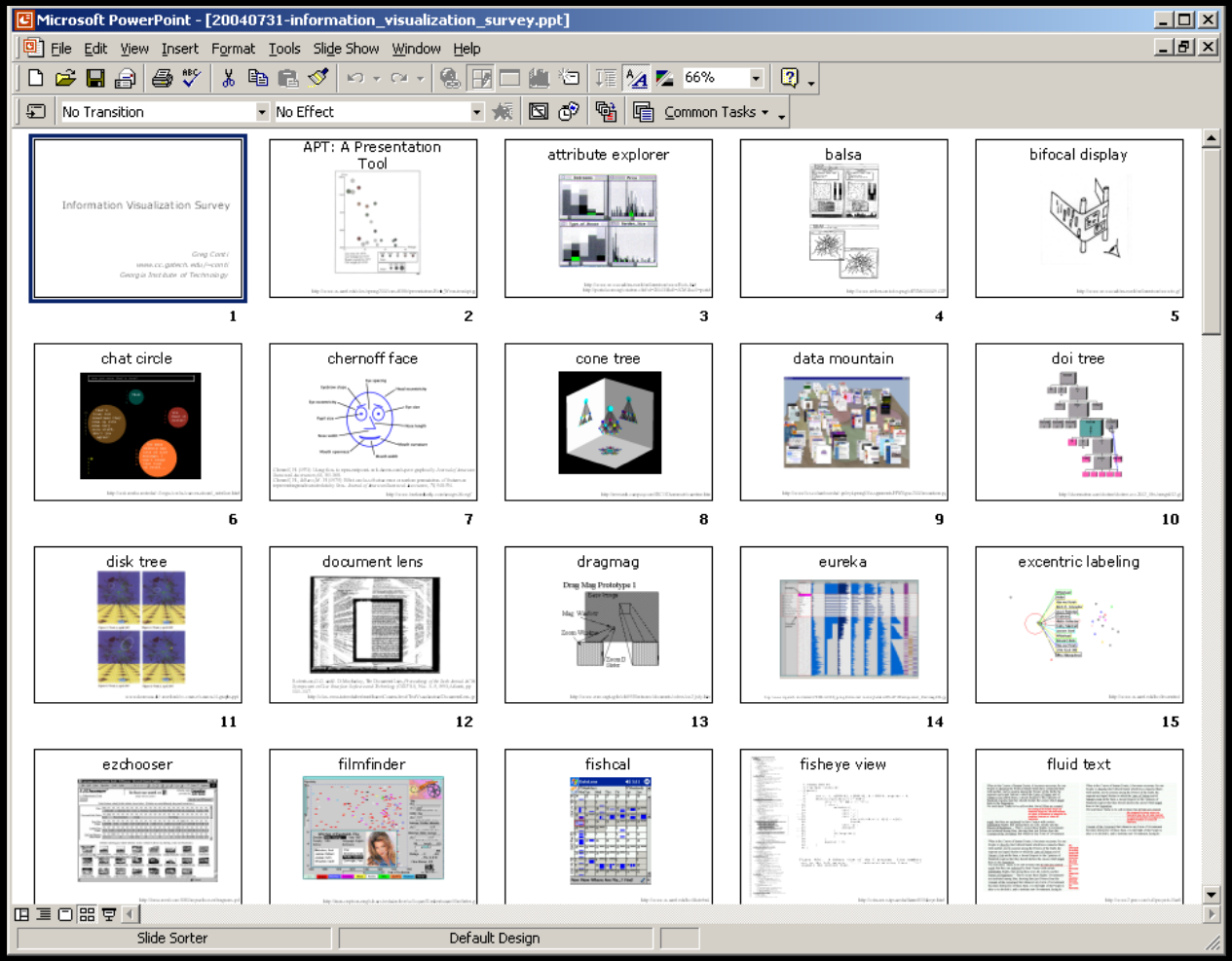
# DAVIX

(Jan Monsch and Raffy Marty)

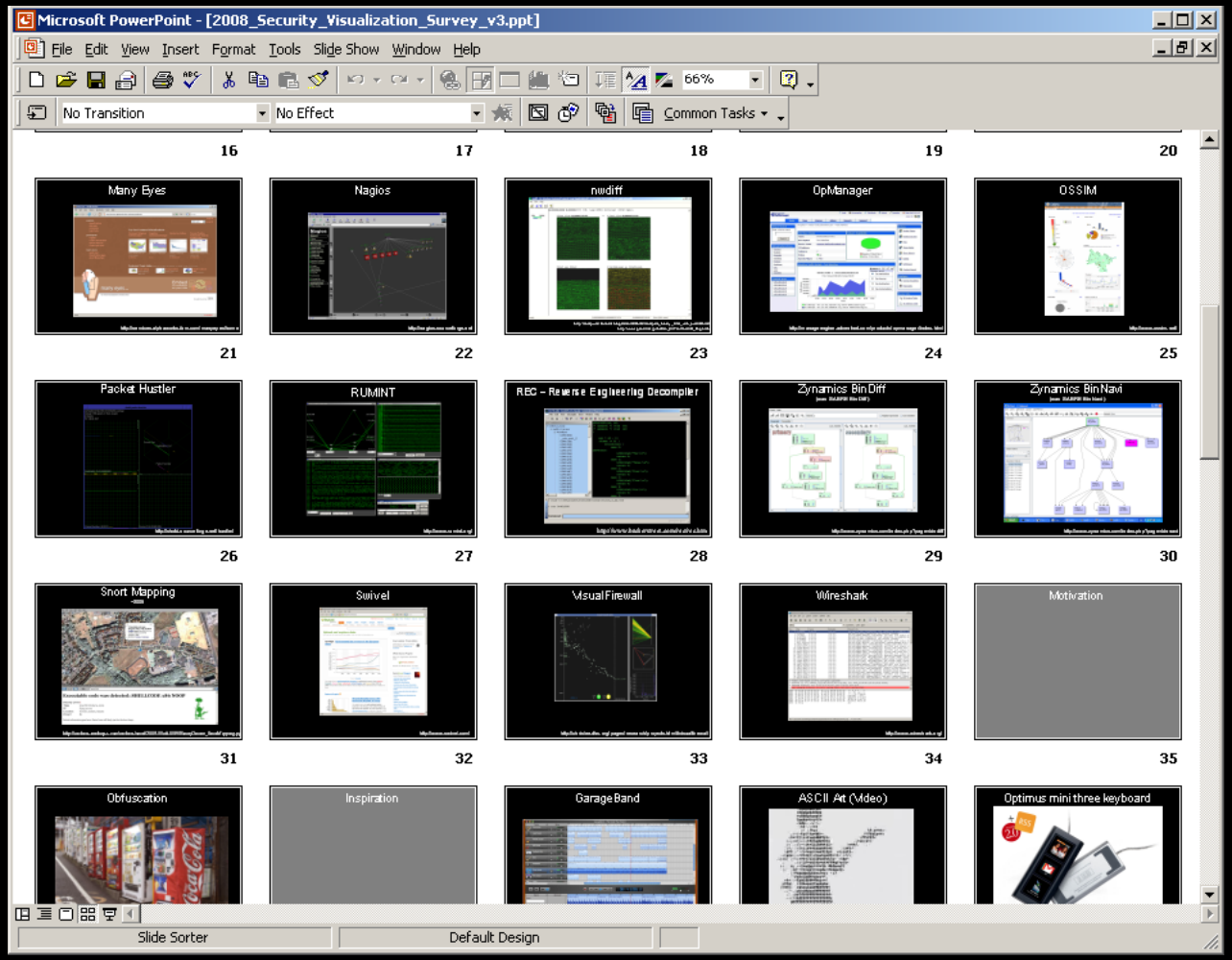



http://www.secviz.org/node/89

# InfoVis Survey



# Security Visualization Survey



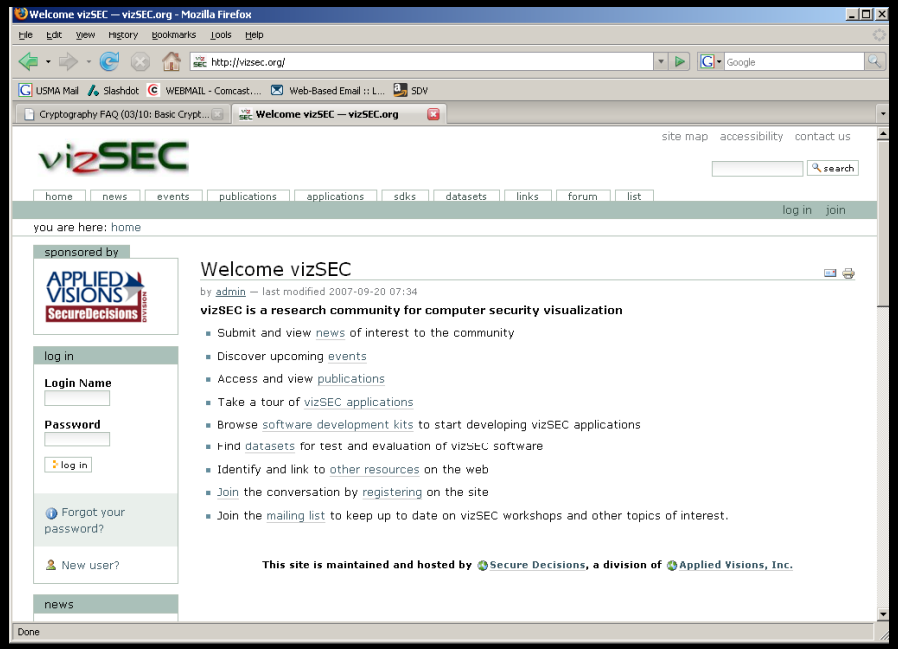
# Communities

## http://secviz.org/


"The place to share, discuss, challenge, and learn about security visualization."

Raffy Marty
Splunk

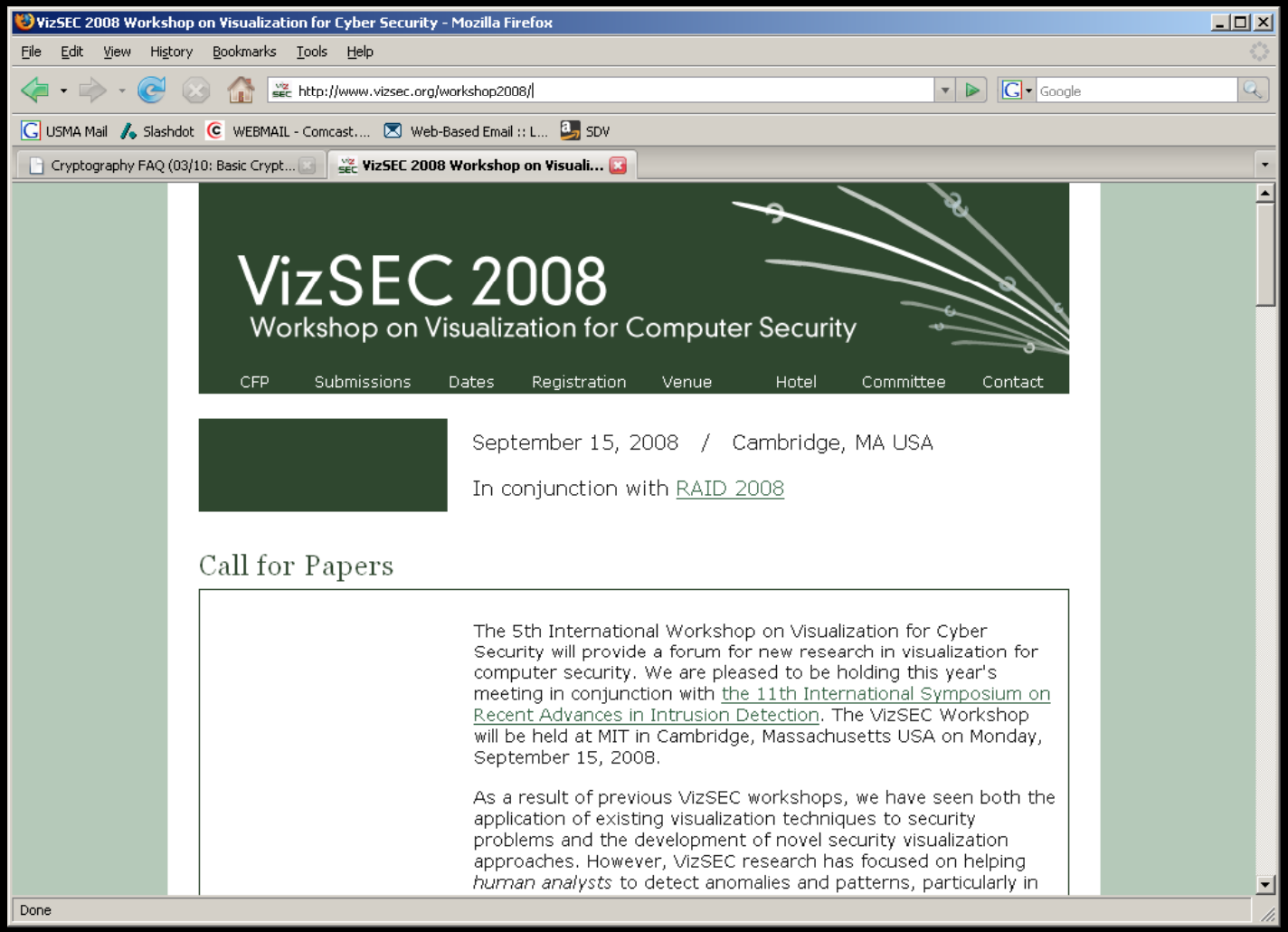
## http://vizsec.org/




"vizSEC is a research community for computer security visualization."

John Goodall
Secure Decisions

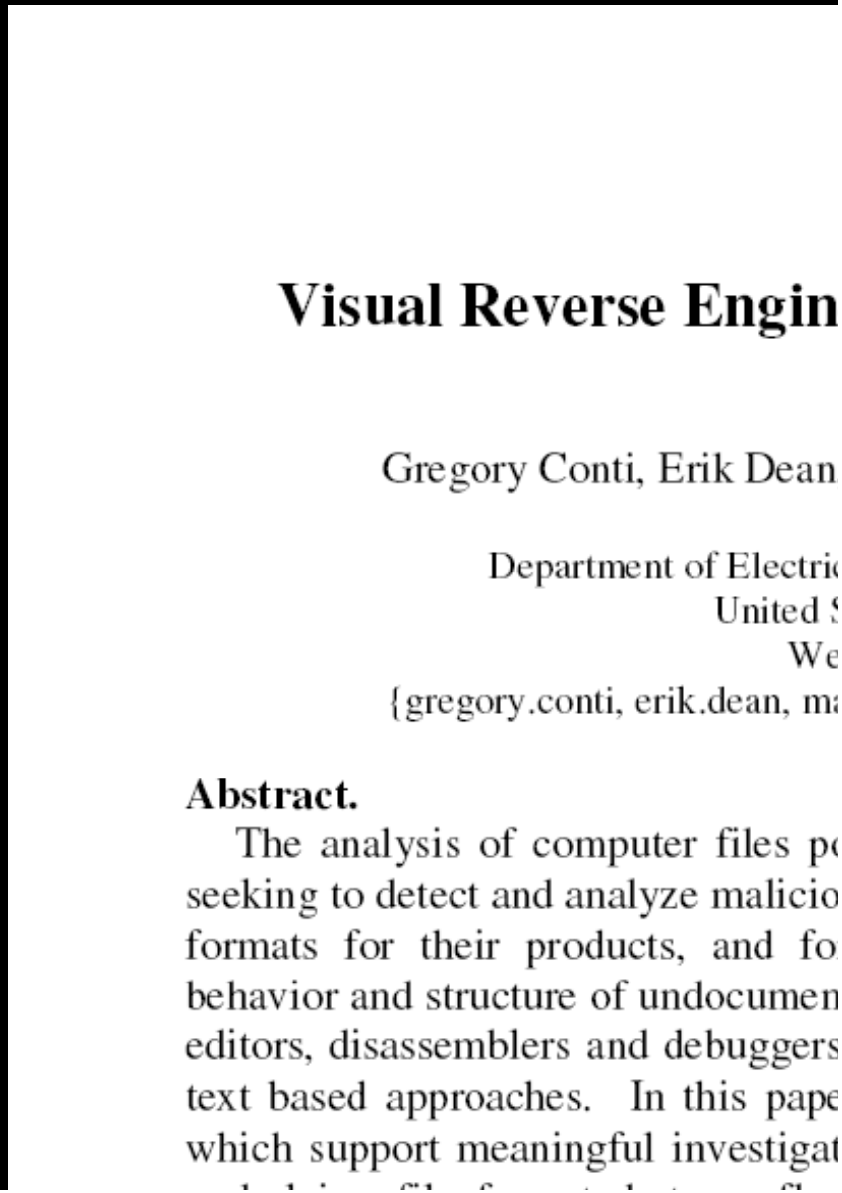
# VizSEC 2008




http://www.vizsec.org/workshop2008/

# More Information

- "Visual Reverse Engineering of Binary and Data Files." Gregory Conti, Erik Dean, Matthew Sinda, Benjamin Sangster. VizSEC 2008.
  - Available September
- Security Data Visualization (No Starch Press)
- Applied Security Visualization (Addison-Wesley)



**Visual Reverse Engin**

Gregory Conti, Erik Dean

Department of Electri
United
We
{gregory.conti, erik.dean, m

**Abstract.**

The analysis of computer files p
seeking to detect and analyze malicio
formats for their products, and fo
behavior and structure of undocumen
editors, disassemblers and debuggers
text based approaches. In this pape
which support meaningful investigat

# Acknowledgements

Damon Becknell, Jon Bentley, Jean Blair, Sergey Bratus, Chris Compton, Tom Cross, Ron Dodge, Carrie Gates, Chris Gates, Joe Grand, Julian Grizzard, Toby Kohlenberg, Oleg Kolesnikov, Frank Mabry, Raffy Marty, Brent Nolan, Gene Ressler, Ben Sangster, Matt Sinda, and Ed Sobiesk

"In fact, master reversers like Fravia recommend cracking while intoxicated with a mixture of strong alcoholic beverages.

While for health reasons we cannot recommend this method, you may find that a relaxing cup of hot tea unwinds your mind and allows you to think in reverse."

-from *Security Warrior*