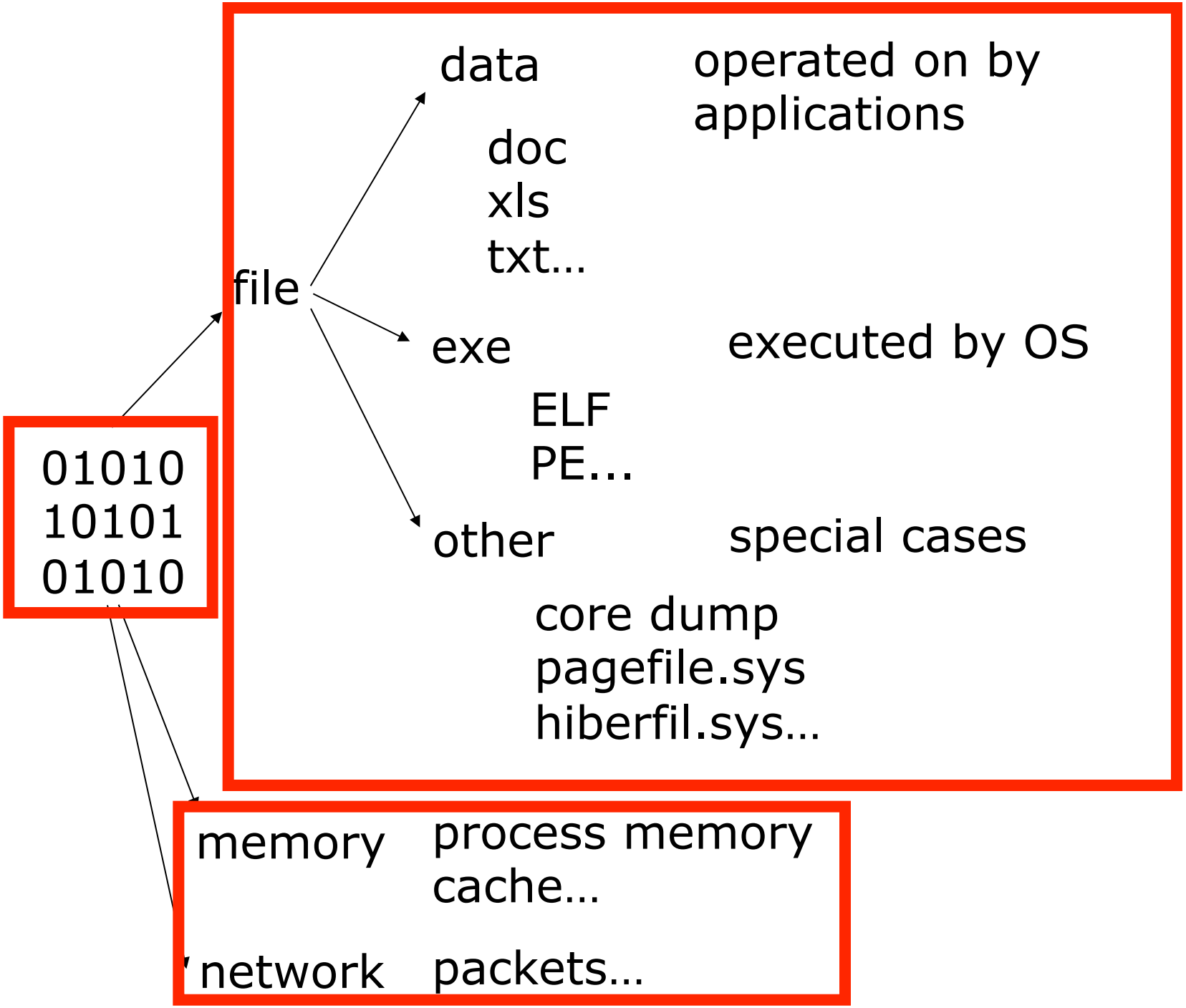
# Visual Reverse Engineering of Binary and Data Files

*Gregory Conti*
*Erik Dean*
*Matthew Sinda*
*Benjamin Sangster*

*United States Military Academy*
*West Point, New York*


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…

# 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

# Framework

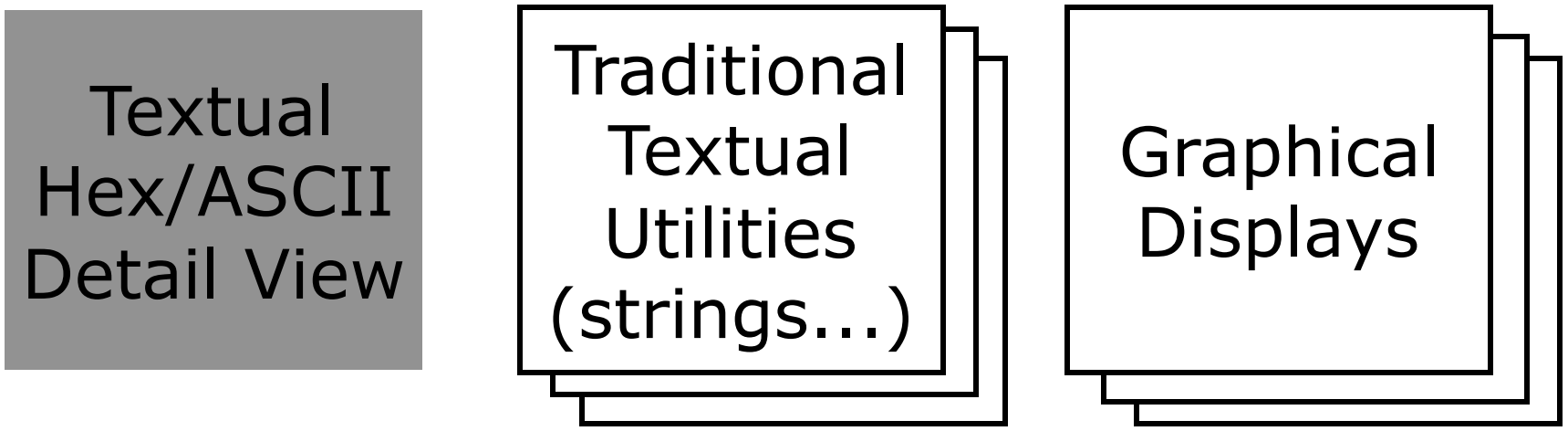
- File Independent Level
  - Entropy
  - Byte Frequency
  - N-Gram Analysis
  - Strings
  - Hex / Decimal / ASCII
  - Bit Plot  (2D/3D)
  - File Statistics
- File Specific Level
  - Complete or Partial Knowledge of File Structure
  - For Example, Metadata

# 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

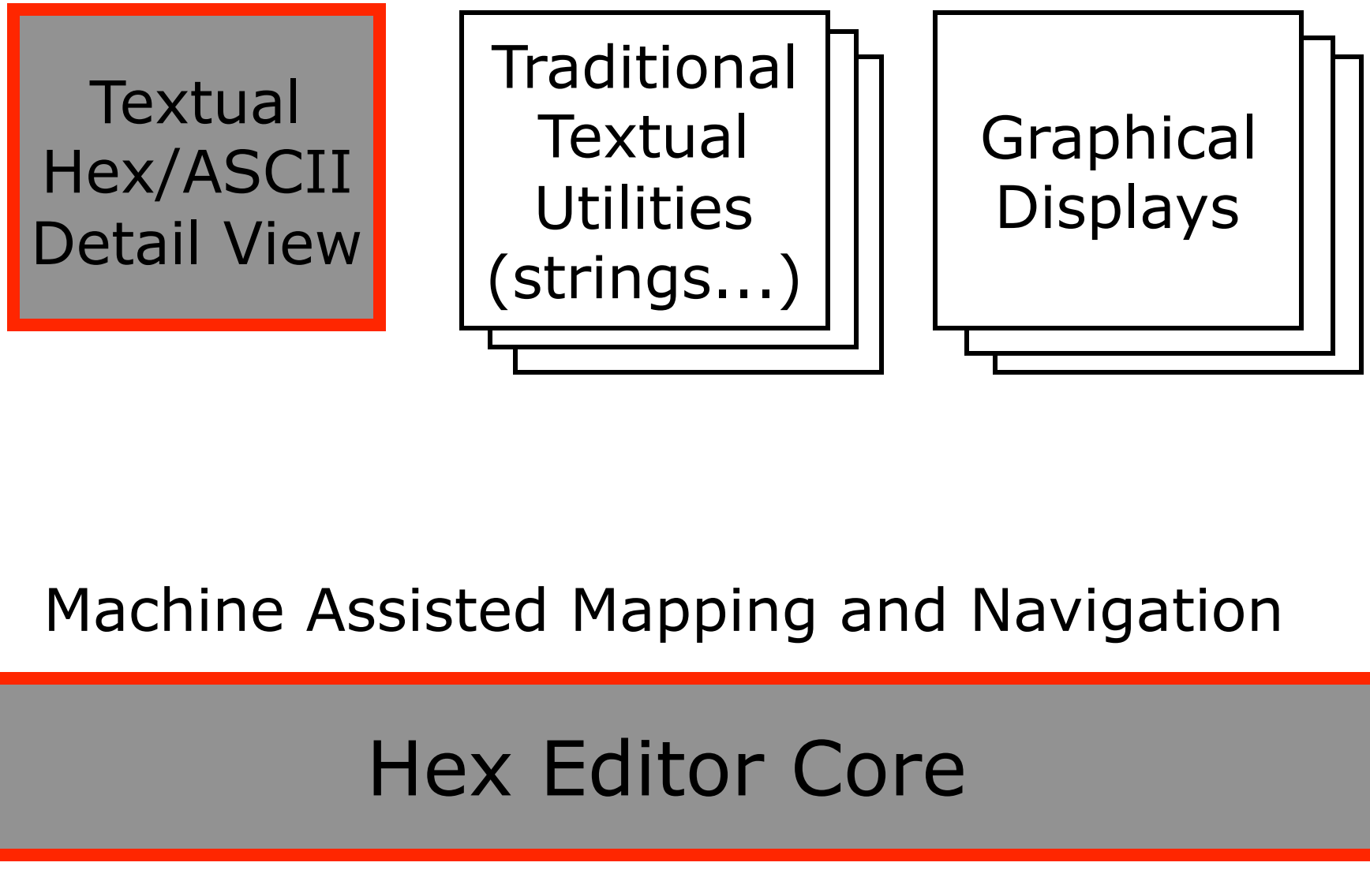
# Framework

- 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

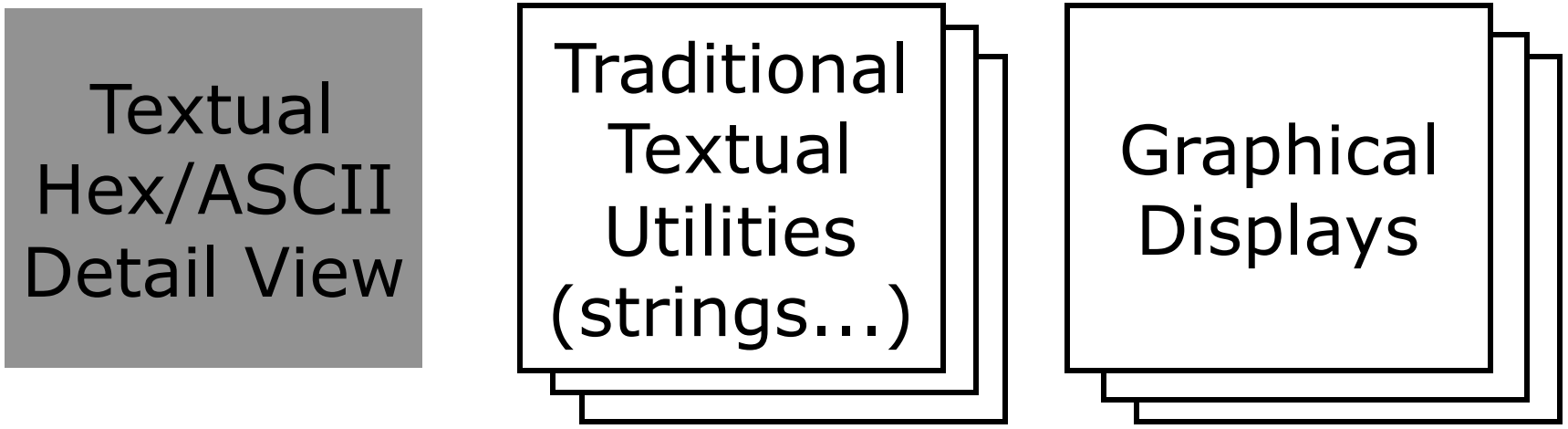



Machine Assisted Mapping and Navigation

Hex Editor Core


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

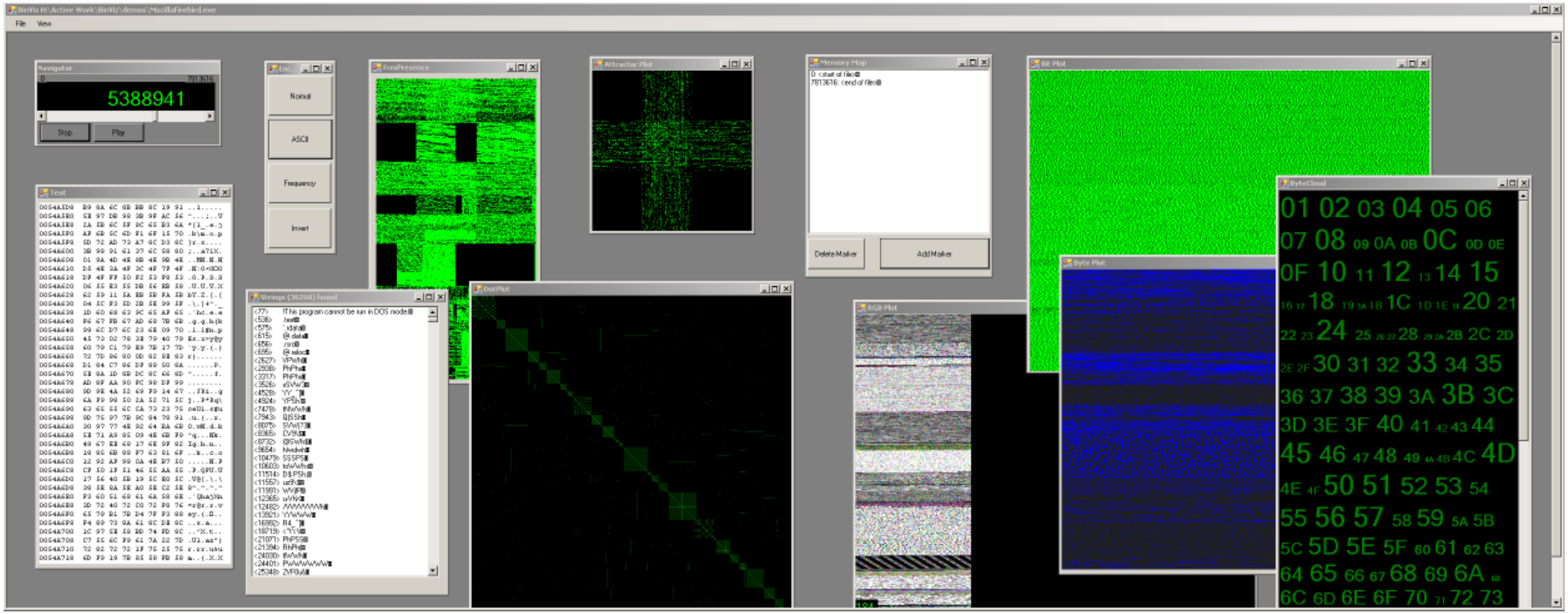

Textual
Hex/ASCII
Detail View
Traditional
Textual
Utilities
(strings...)
Graphical
Displays
Machine Assisted Mapping and Navigation
Hex Editor Core


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

# Towards a Visual Hex Editor

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

# System Overview



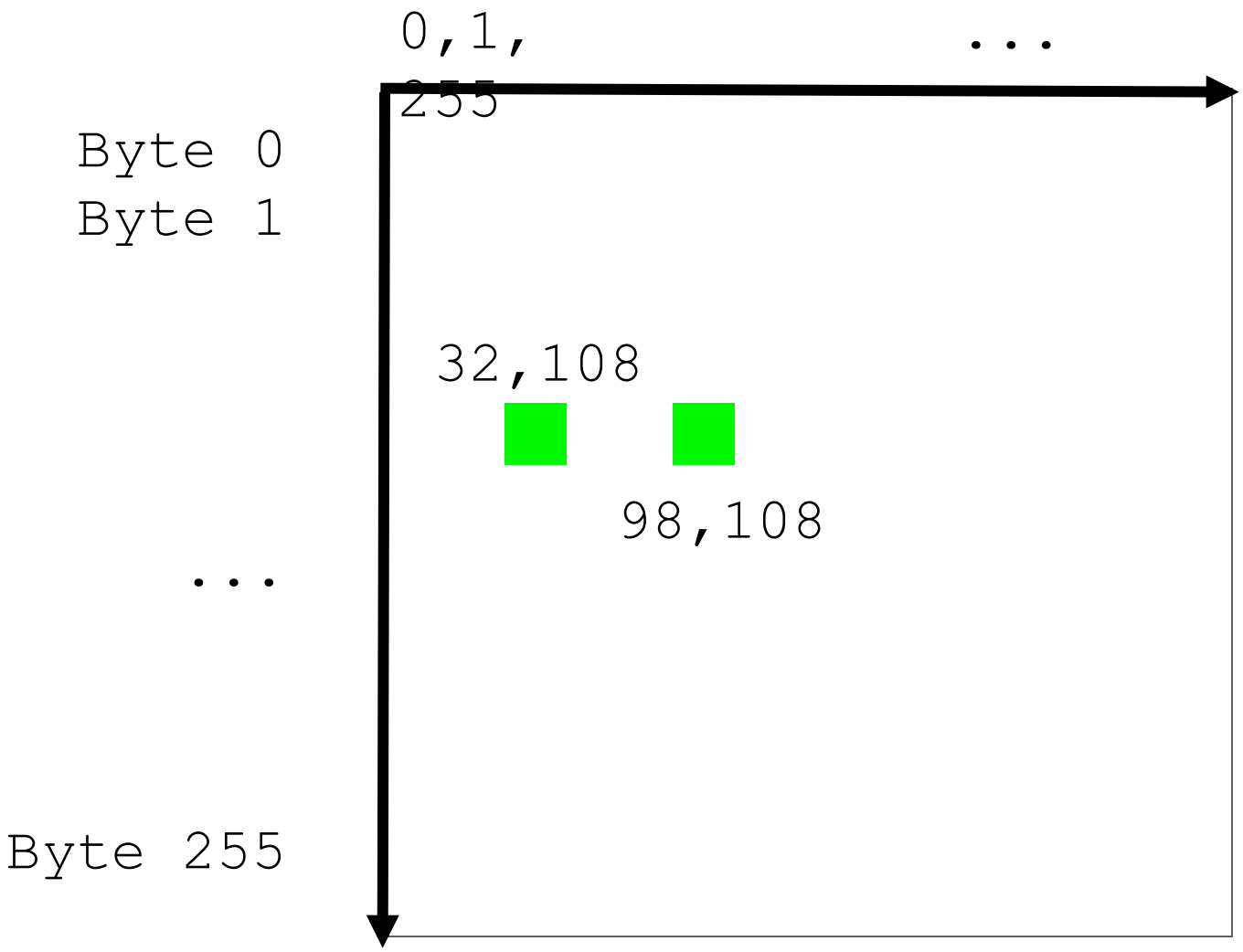

- **Textual**: Text/ASCII, Strings, ByteCloud
- **Graphical**: Bitplot, BytePlot, RGBPlot, BytePresence, ByteFrequency, Digram, Dotplot
- **Interaction**: VCR, Memory Map, Color Coding

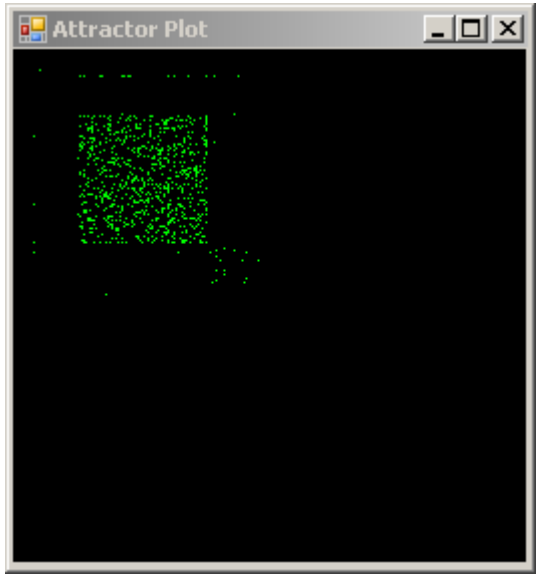
# Digraph View

**black hat**

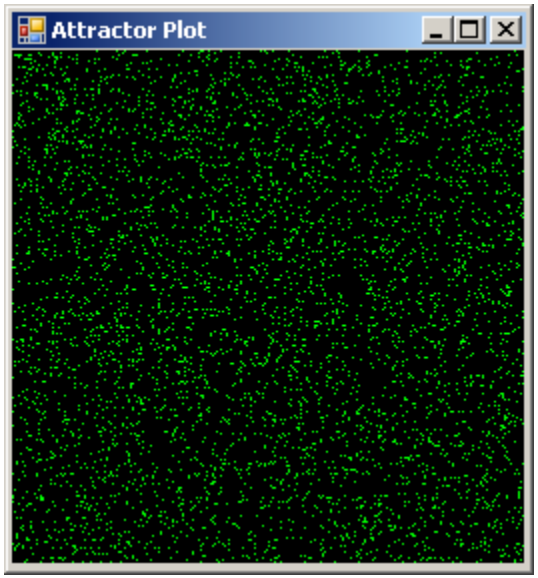
```
bl          (98,108)
 la         (108,97)
  ac        (97,99)
   ck       (99,107)
    k_      (107,32)
     _h     (32,104)
      ha    (104,97)
       at   (97,116)
```

# Digraph View

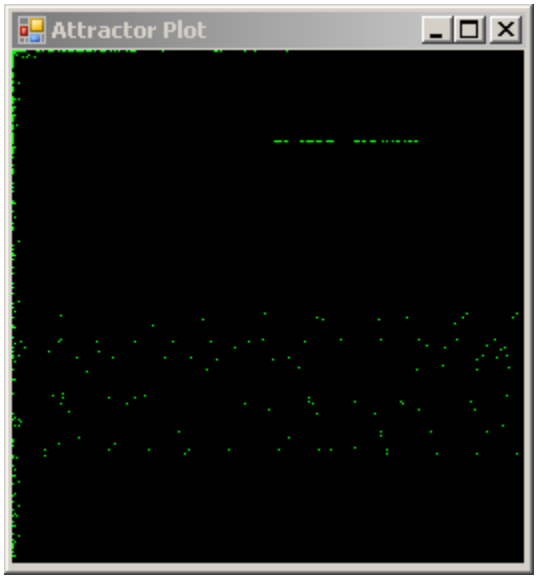




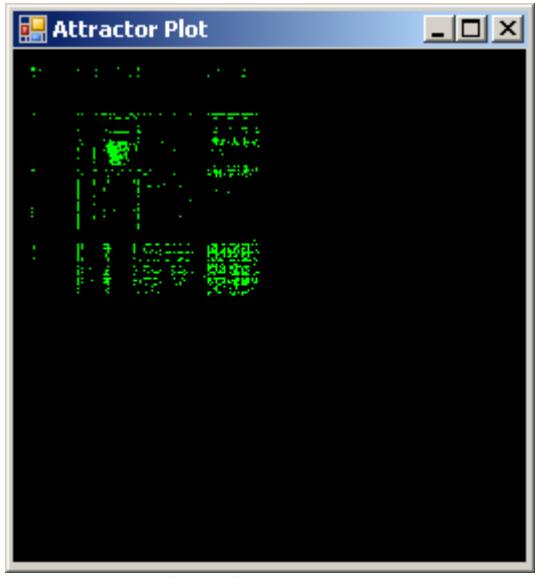

uuencoded



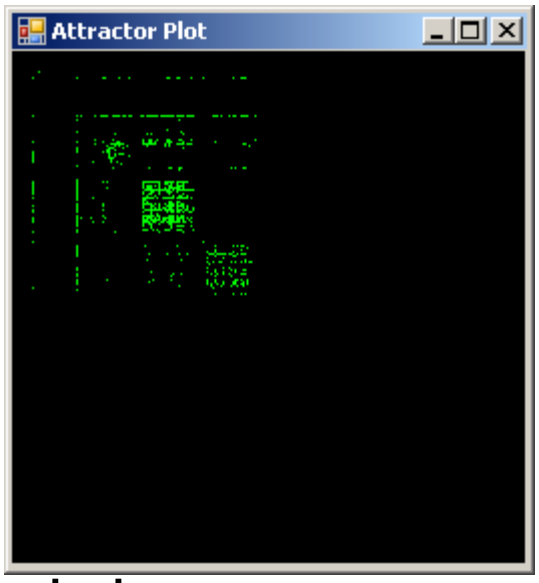

compression
encryption



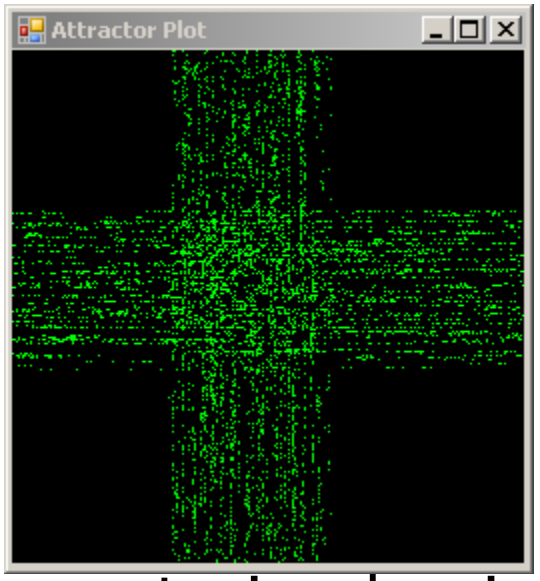

incrementing
words




slashdot.org

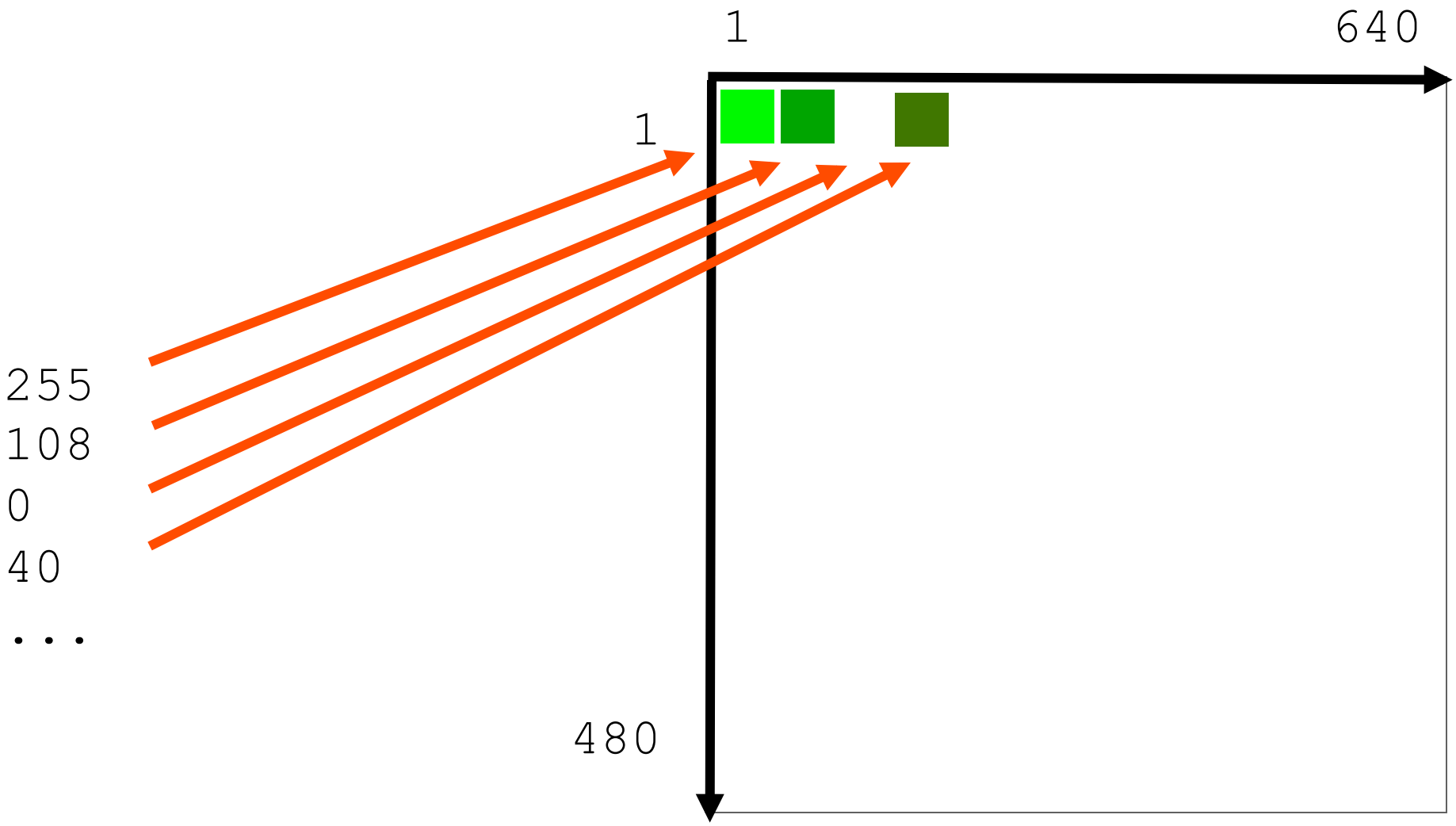



.txt

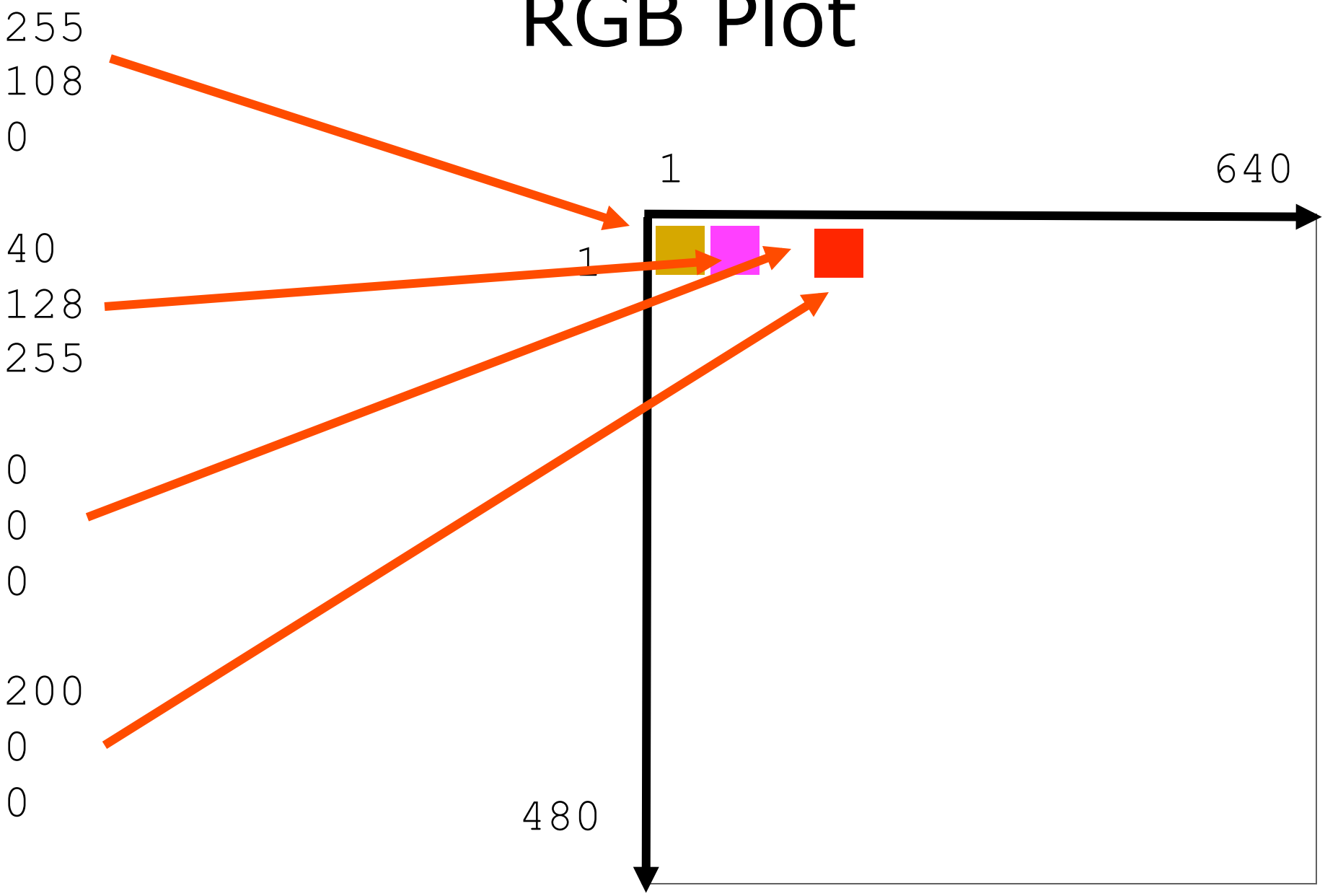



constrained pairs

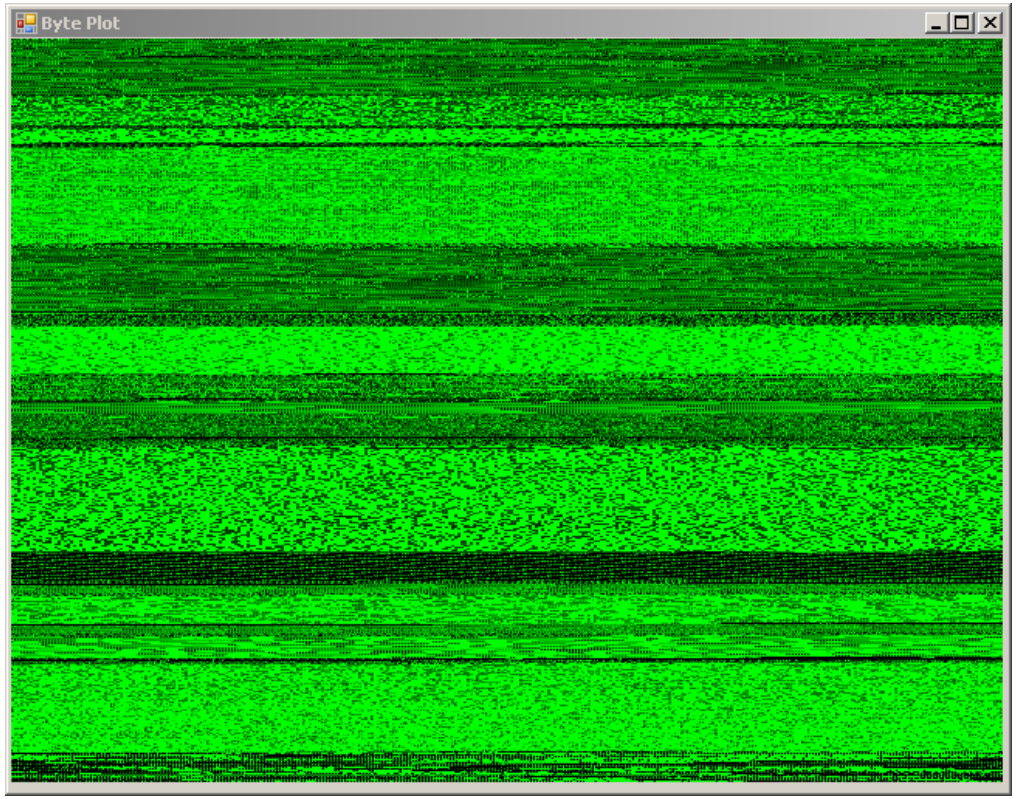
# Byte Plot



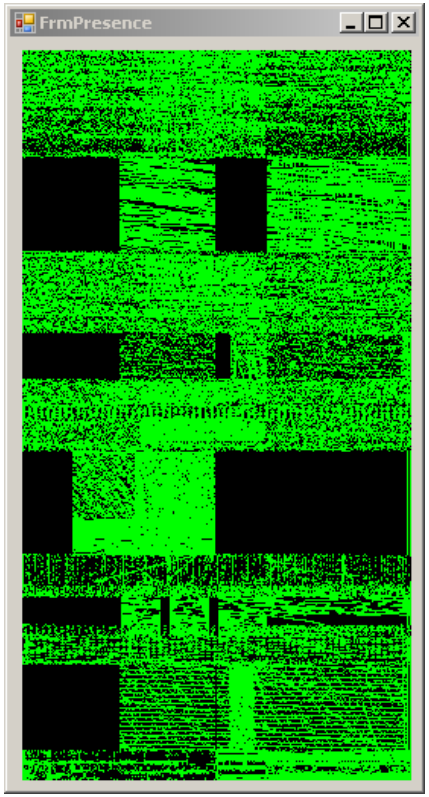
# RGB Plot



# Byte Presence



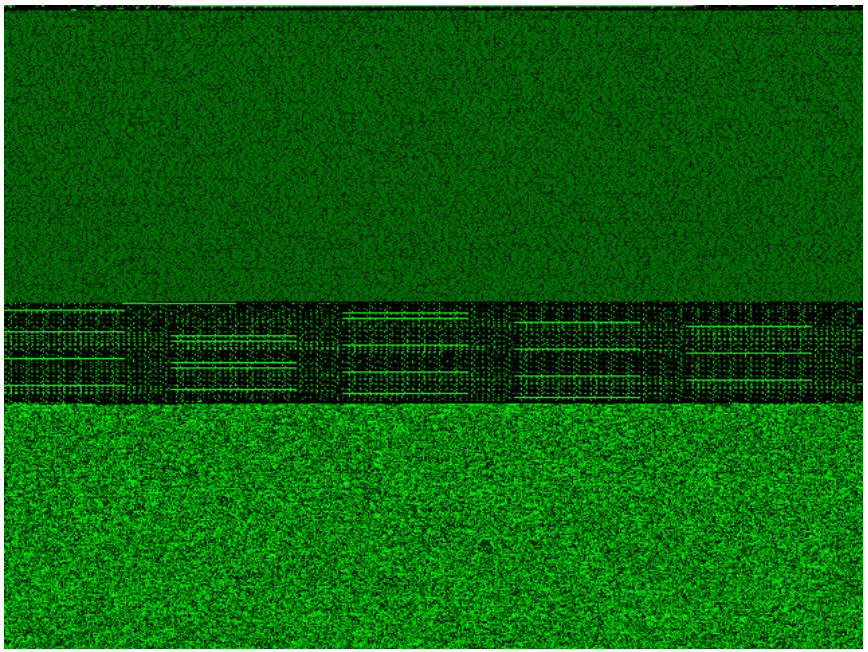

0 255

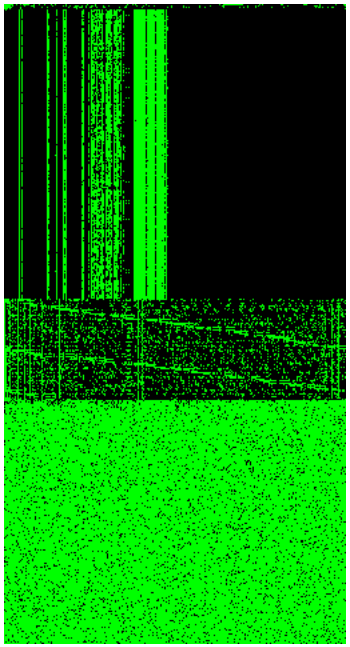


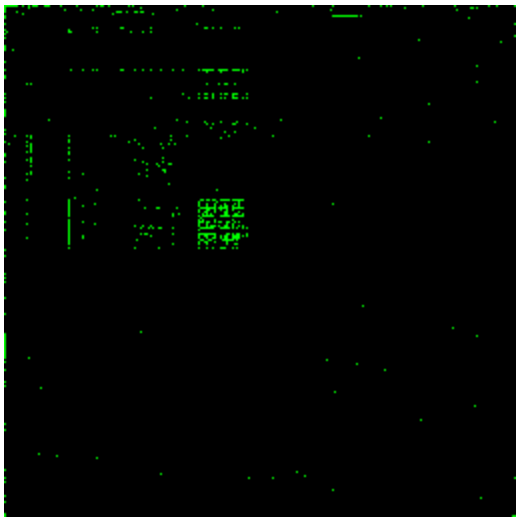
# Display Comparison

| | Pixels/Byte | 19" Monitor | Gain |
|---|---|---|---|
| **Textual Hex** | 300 pixels/byte | 4.4 KB | N/A |
| **Byte View** | 1 pixel/byte | 1.3 MB | 300x |
| **RGB View** | 3 bytes/pixel | 3.9 MB | 900x |

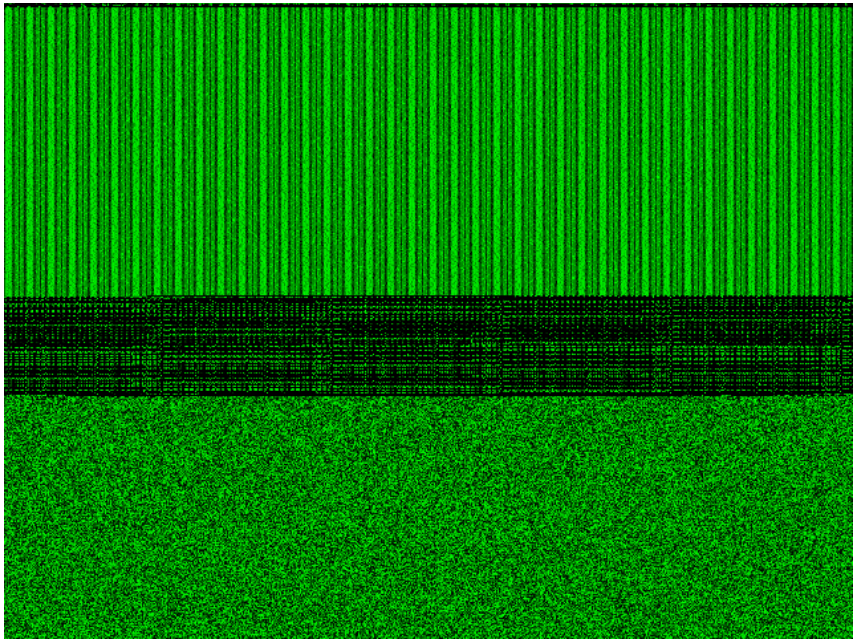
# Encryption

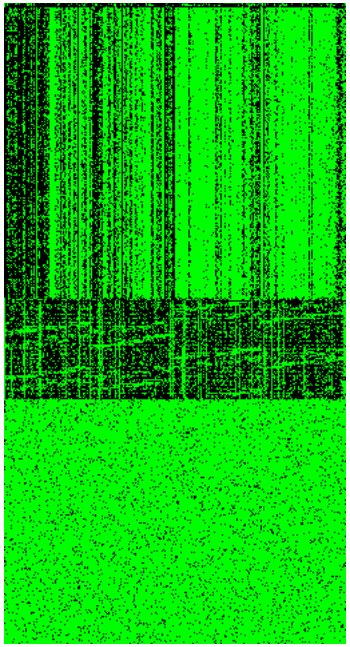


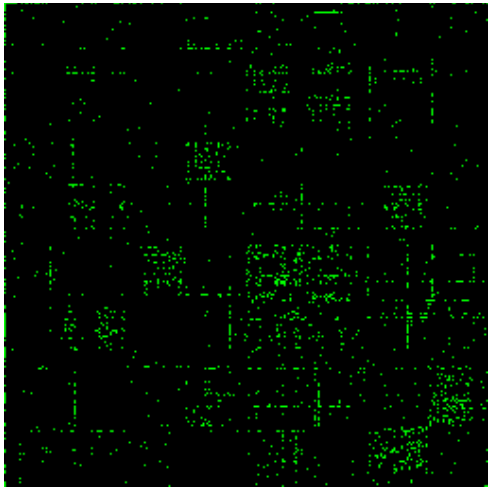




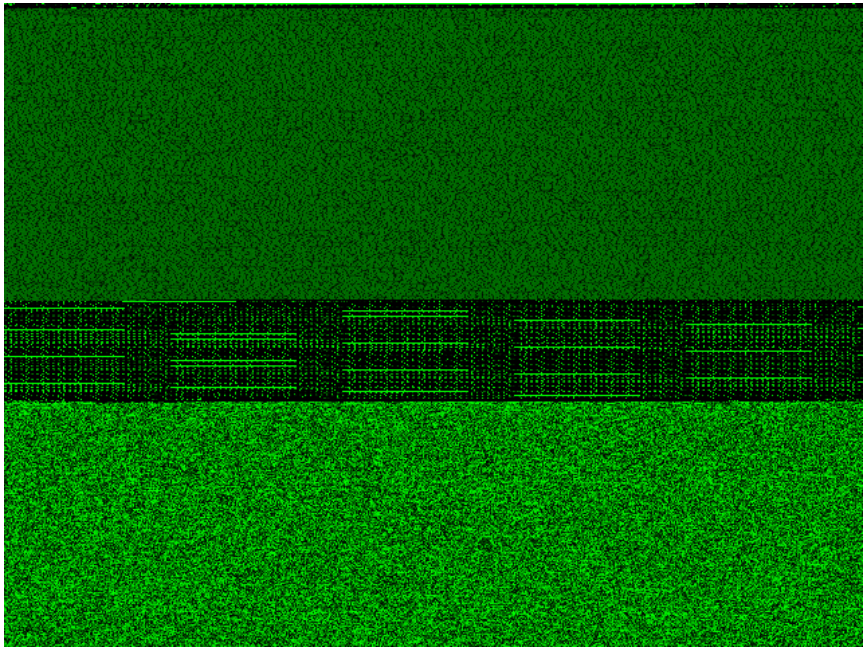
unencrypted

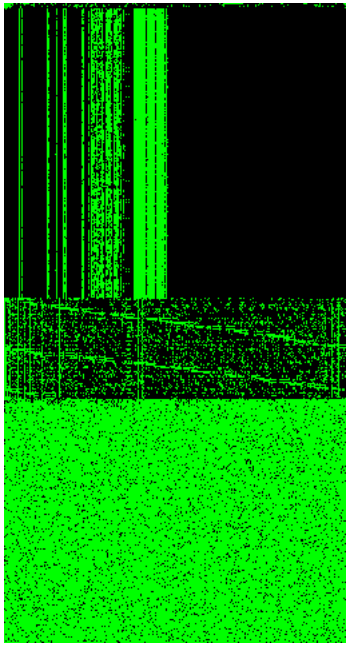


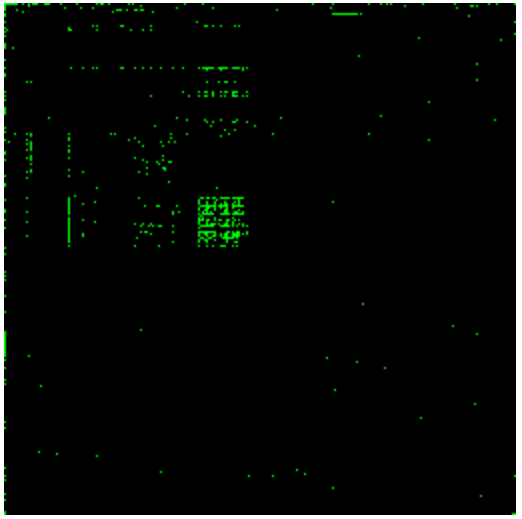




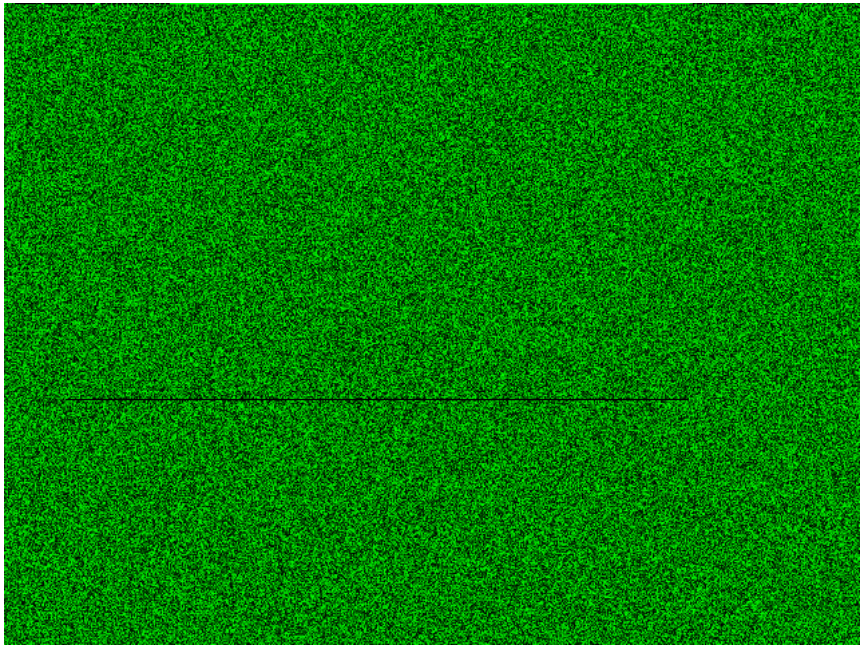
XOR

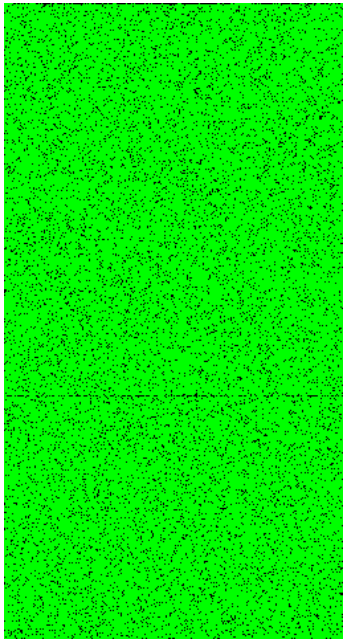


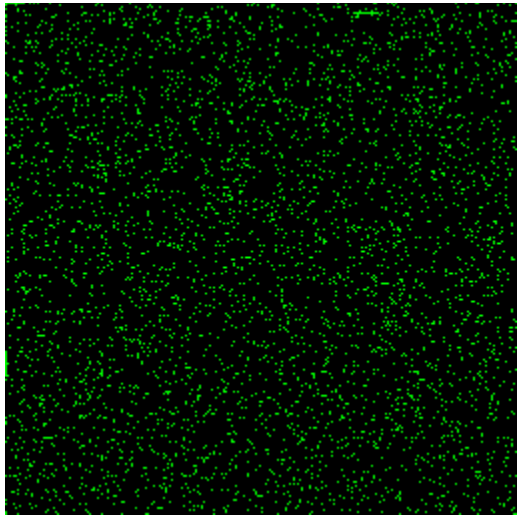




unencrypted



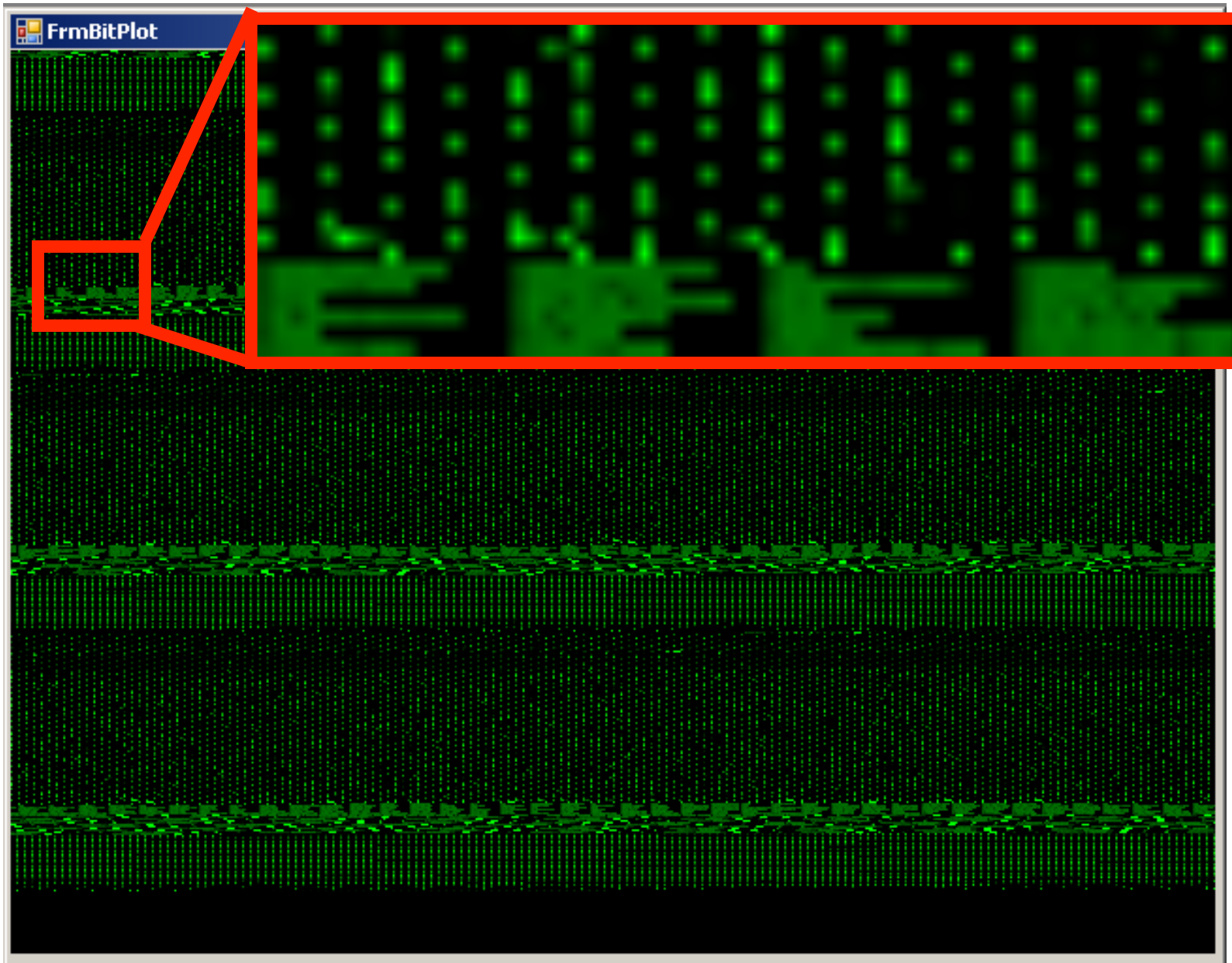




AES

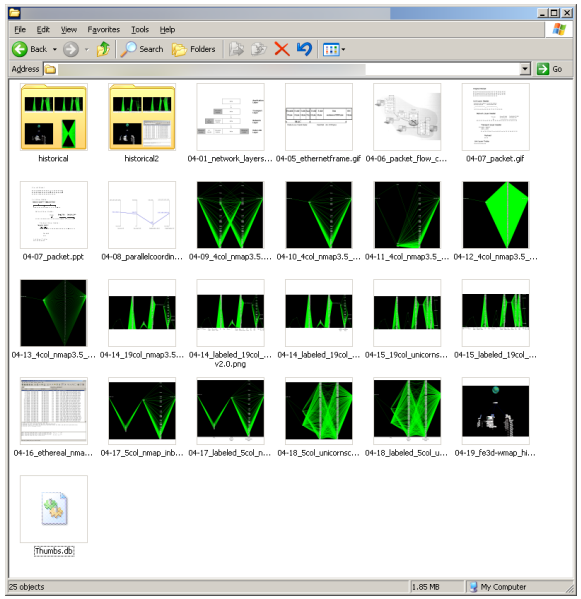
# Fixed Length Structure

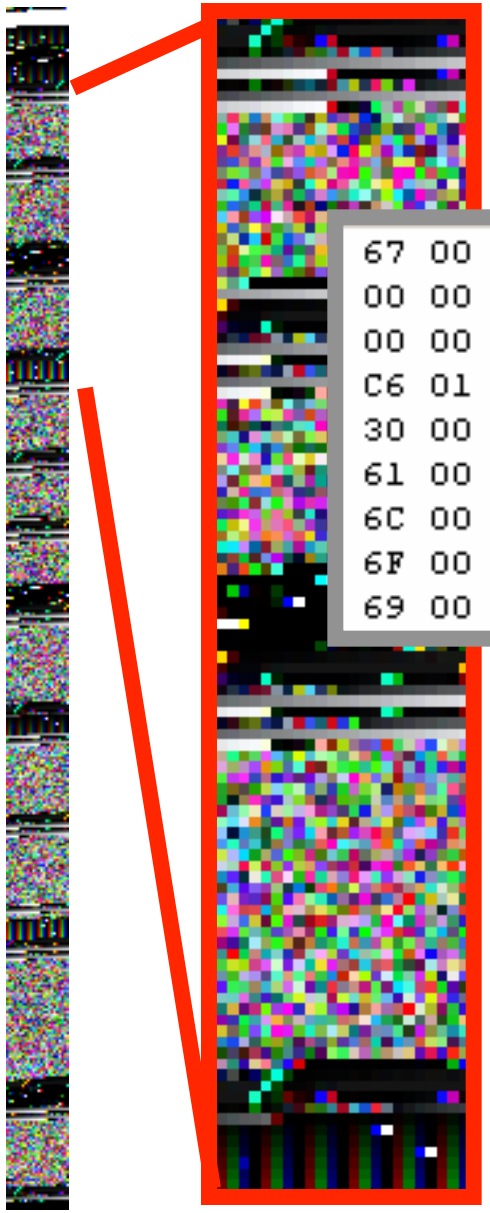
# Neverwinter Nights Database File



# Variable Length Structure

# Thumbs.db

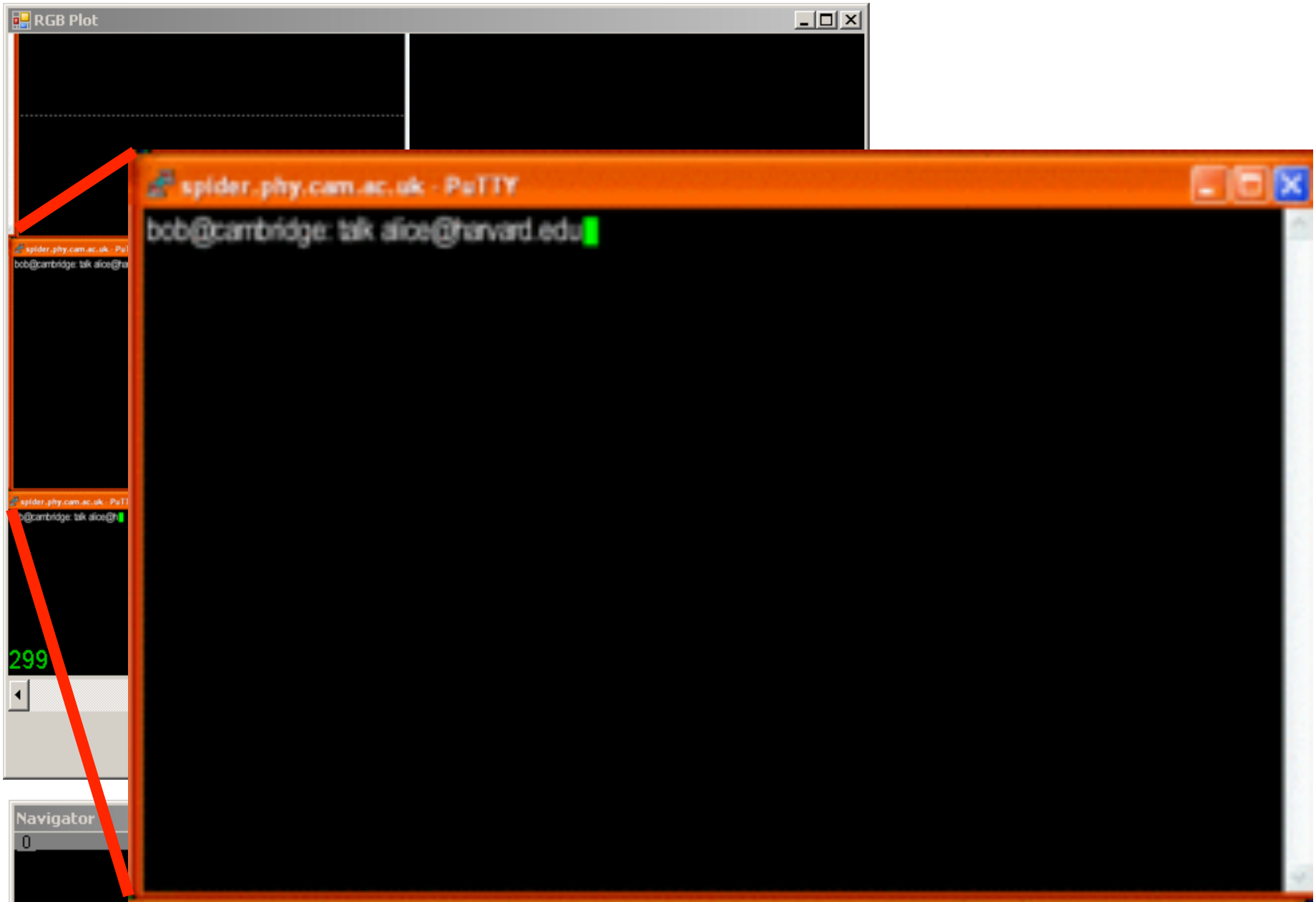


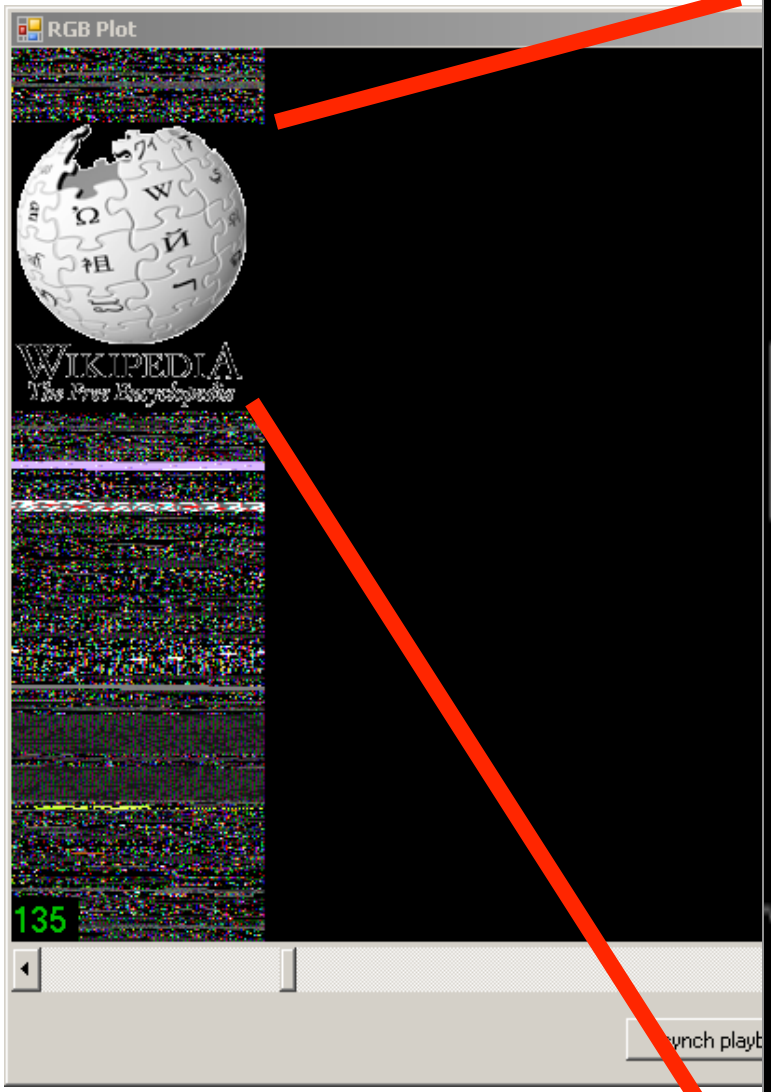
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 67 | 00 | 69 | 00 | 66 | 00 | 00 | 00 | g.i.f... |
| 00 | 00 | 64 | 00 | 00 | 00 | 47 | 00 | ..d...G. |
| 00 | 00 | 00 | BC | FF | EC | 64 | 75 | ......du |
| C6 | 01 | 30 | 00 | 34 | 00 | 2D | 00 | ..0.4.-. |
| 30 | 00 | 38 | 00 | 5F | 00 | 70 | 00 | 0.8._.p. |
| 61 | 00 | 72 | 00 | 61 | 00 | 6C | 00 | a.r.a.l. |
| 6C | 00 | 65 | 00 | 6C | 00 | 63 | 00 | l.e.l.c. |
| 6F | 00 | 6F | 00 | 72 | 00 | 64 | 00 | o.o.r.d. |
| 69 | 00 | 6E | 00 | 01 | 02 | 00 | 00 | i.n..... |

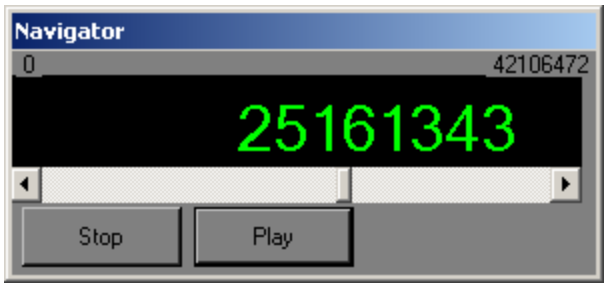
See http://www.acquisitiondata.com/white_papers/thumbsdbfiles.pdf for a well written white paper.

# Demo
(Firefox hdmp)




Firefox .hdmp

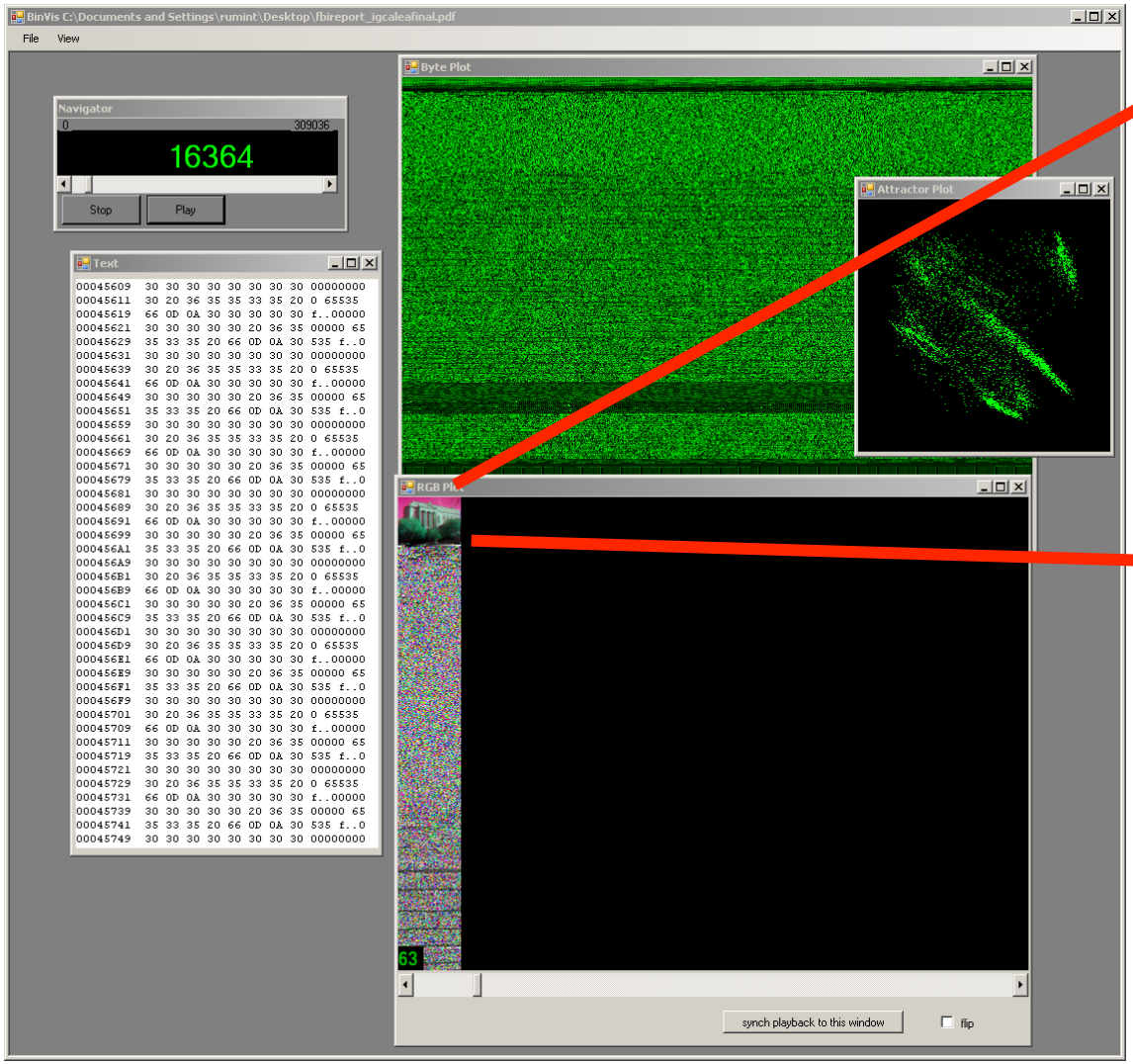



Firefox .hdmp


Firefox .hdmp


Firefox .hdmp




Redacted PDF...

# Weaknesses

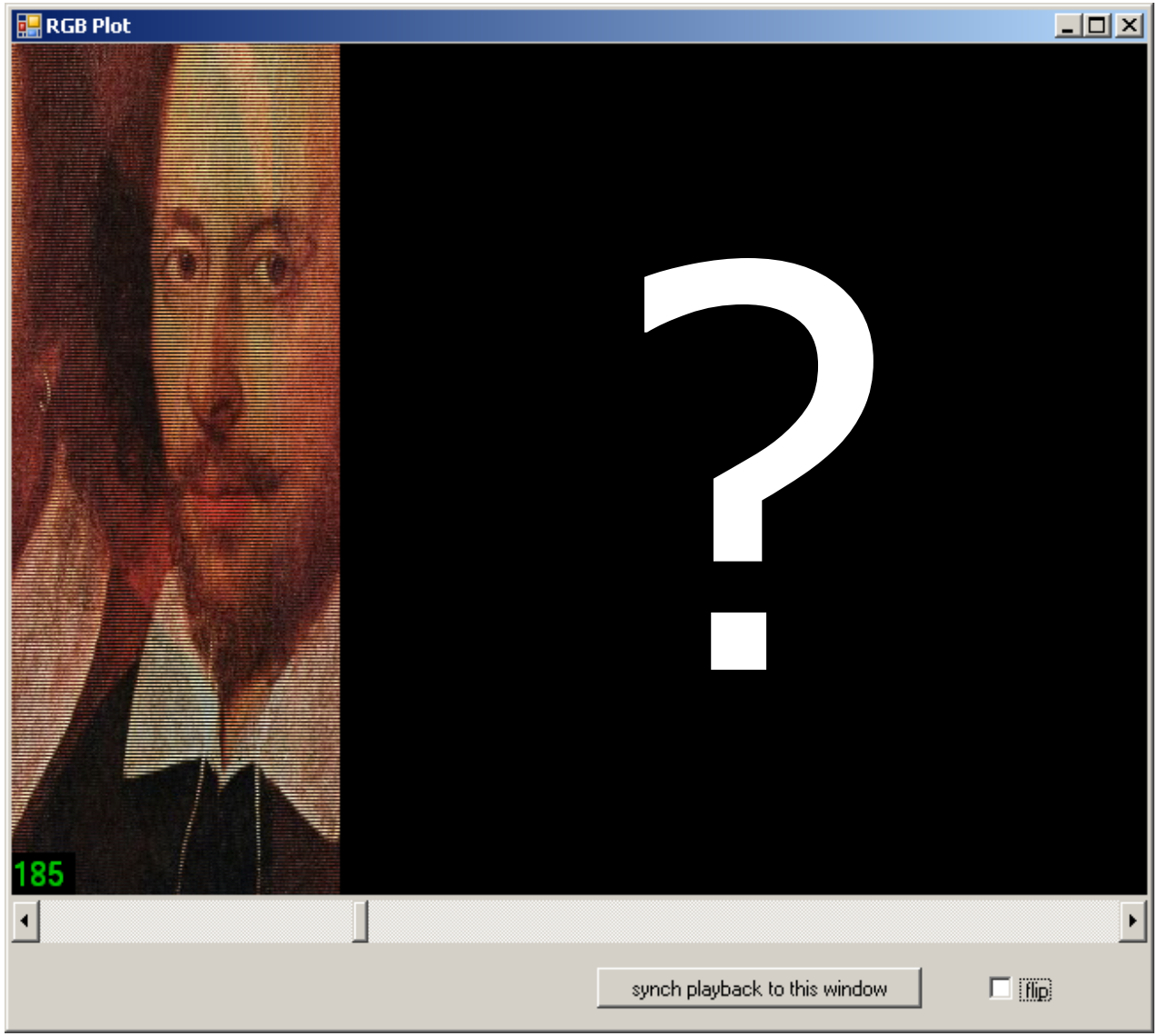
- entire file may be extracted from bit/byte/RGB
  - May trigger AV or IDS
  - 8bit/byte steg
- Screams for big monitor
- Better memory management
  - ~300MB+
- Unicode

# 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

# 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, Dino Schweitzer, Matt Sinda, and Ed Sobiesk




*Gregory Conti* *gregory.conti@usma.edu*|
*Erik Dean* *erik.dean@usma.edu*