Visualization in Malware and Forensics

By

Danny Quist

Presented At

The Digital Forensic Research Conference

DFRWS 2012 USA Washington, DC (Aug 6th - 8th)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

http:/dfrws.org

VISUALIZATION IN MALWARE AND FORENSICS

Danny Quist, Ph.D.

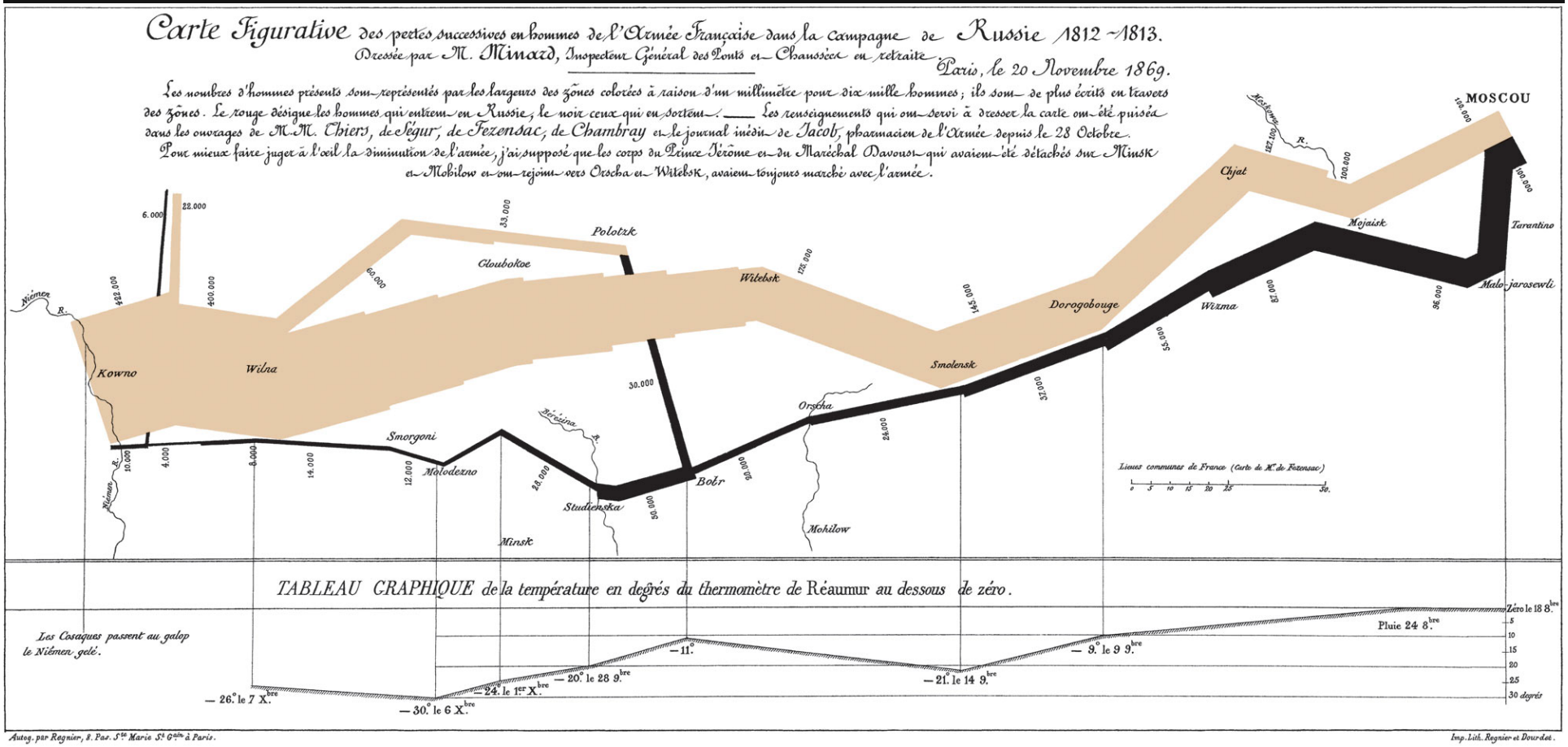
DFRWS 2012

August 7, 2012

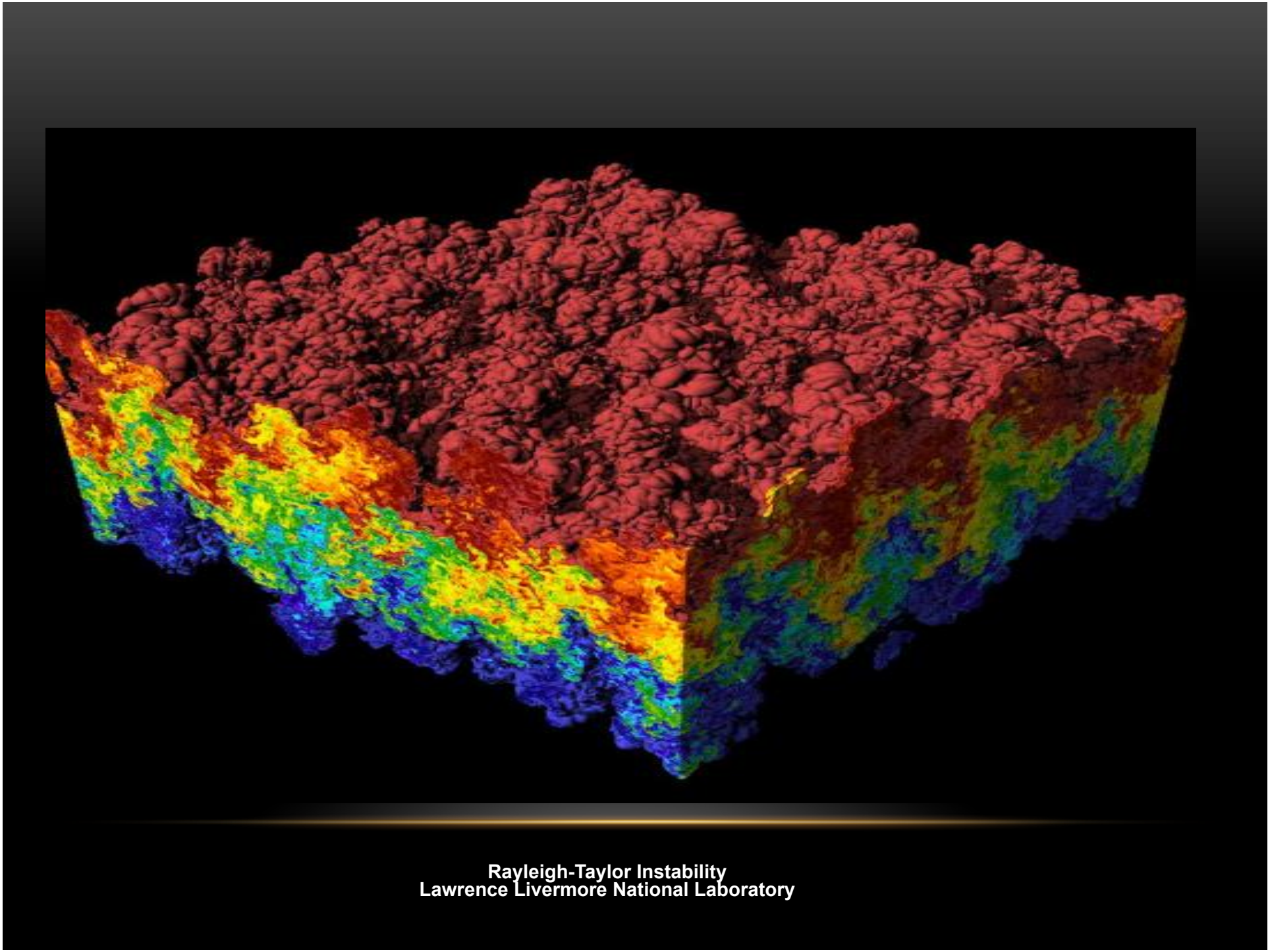
OVERVIEW

- Visualization Difficulties and Examples
- Malware and Reverse Engineering
- Steps for more effective visualization

THE GOAL:
MAKE DATA EASIER TO UNDERSTAND



Charles Minard's 1869 Graphic Showing Napoleon's 1812 Russian campaign, their movements, as well as the temperature



Rayleigh-Taylor Instability
Lawrence Livermore National Laboratory

VISUALIZATION OF TEXT DATA IS TERRIBLE

- Many resources dedicated to physical world visualization
- Representing text files is difficult
- Visualization must make a specific task easier
- Too many art projects in visualization

STEPS FOR VISUALIZATION

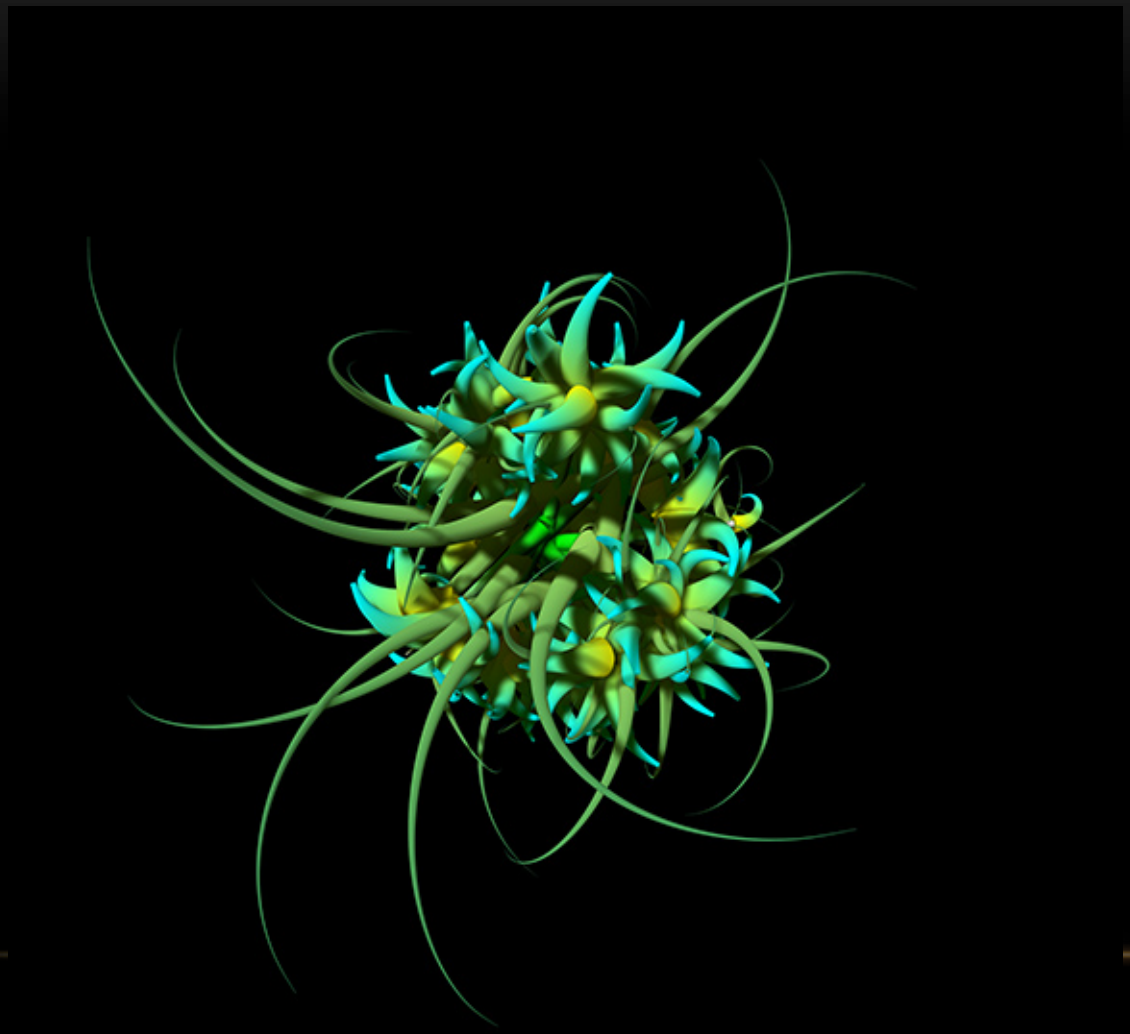
- Understand your users work flow
- Define the process you would like to improve
- Experiment with existing tools
- Build the visualization tool
- Measure whether it is effective
  (User Study)

FUNDAMENTAL VISUALIZATION QUESTION

How is my tool better than grep?

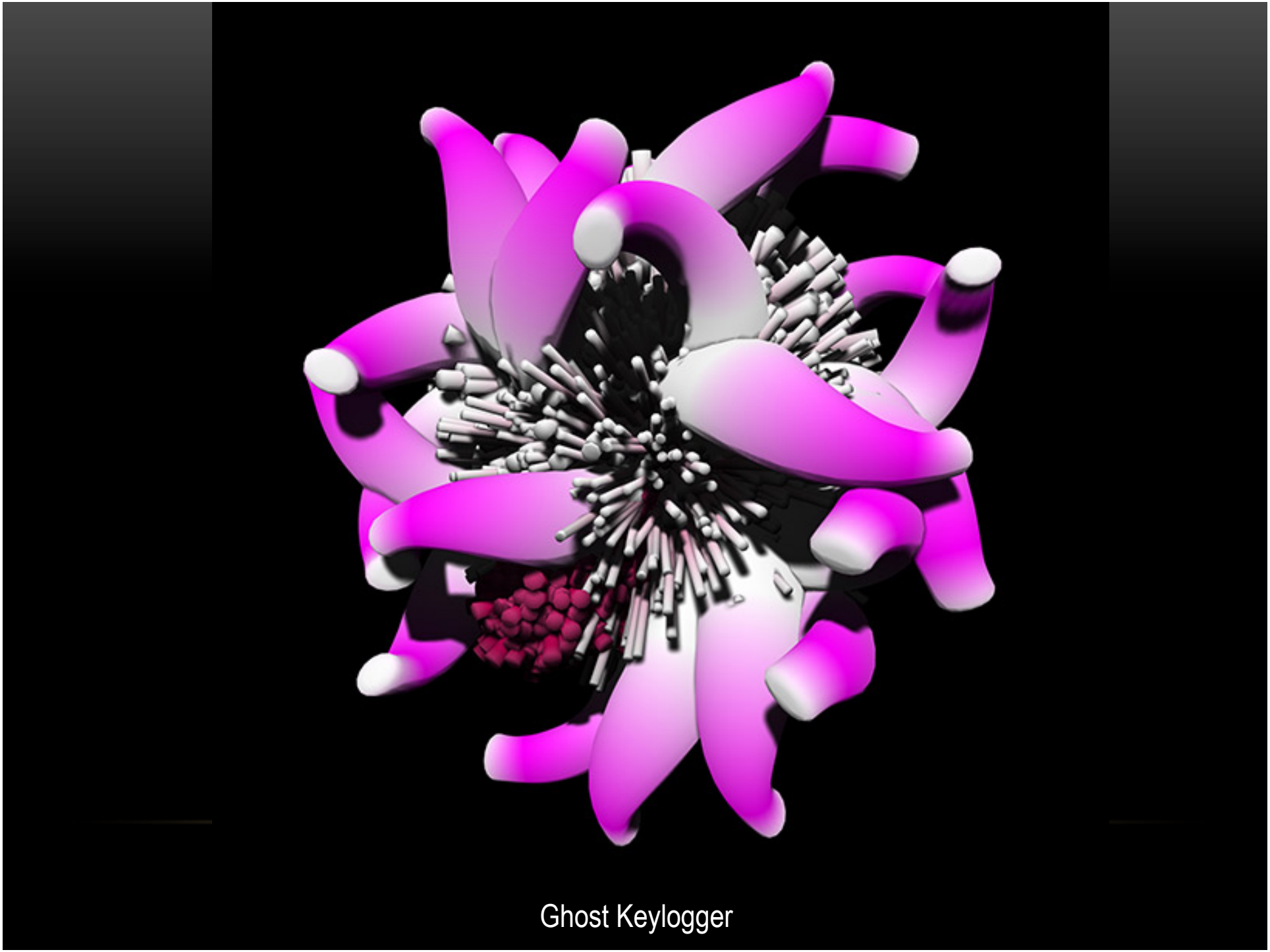
EXAMPLES OF VISUALIZATION IN SECURITY

MIT MEDIA LAB: ALEX DRAGULESCU



MyDoom Virus

NetSky Virus


Ghost Keylogger

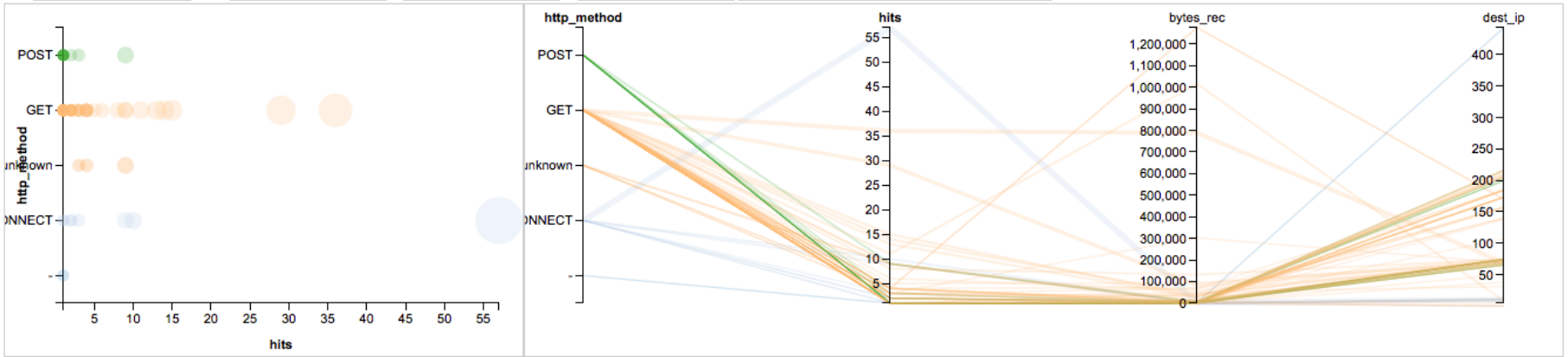
SPLUNK VISUALIZATIONS

Long uri path + string in Bluecoat Proxy

	src_ip	dest_ip	dest_host	uri_length	mean_uri_length	stdev_uri_length	uri
1	10.13.37.104	-	ots.optimize.webtrends.com	2556	131.739935	225.069347	/ots/ots/js-3.1/311121/1027ed543b58578e6e3b824071758d9bdeafd1265a7b24402f7551e927c3acb81cdbfa67bfd3446b42706edbe6b0608936758c58f0c7
2	10.13.37.1011	66.235.142.20	metrics.cnn.com	2356	131.739935	225.069347	/b/ss/aolturnercnnmoney-2010/1/H.24.1/s99911066957988?AQB=1&ndh=1&t=18%2F5%2F2012%2015%3A8%3A18%201%20360&ce=UTF-8&ns=cnn&p
3	10.13.37.1011	173.192.61.227	juggler.services.disqus.com	2308	131.739935	225.069347	/event.js?thread_slug=httpmoneycnncomgalleries2012retirement1206galleryretirement_guide_best_placesfortune&user_type=anon&referrer=http%3A%2
4	10.13.37.104	66.235.133.9	om.sears.com	2244	131.739935	225.069347	/b/ss/searscom/1/H.22.1/s96136986890317?AQB=1&ndh=1&t=18%2F5%2F2012%2015%3A8%3A47%201%20360&cdp=2&pageName=Subcategory%2
5	10.13.37.109	184.85.60.20	s7.addthis.com	2062	131.739935	225.069347	/static/r07/pinit005.html?url=http%253A%252F%252Flifeinc.today.msnbc.msn.com%252F_news%252F2012%252F06%252F18%252F12282421-gen-x-m

Parallel Coordinates Visualization of Bluecoat Proxy Data Shapes

size: hits color: http_method y: http_method x: hits bytes_rec dest_ip + Reset

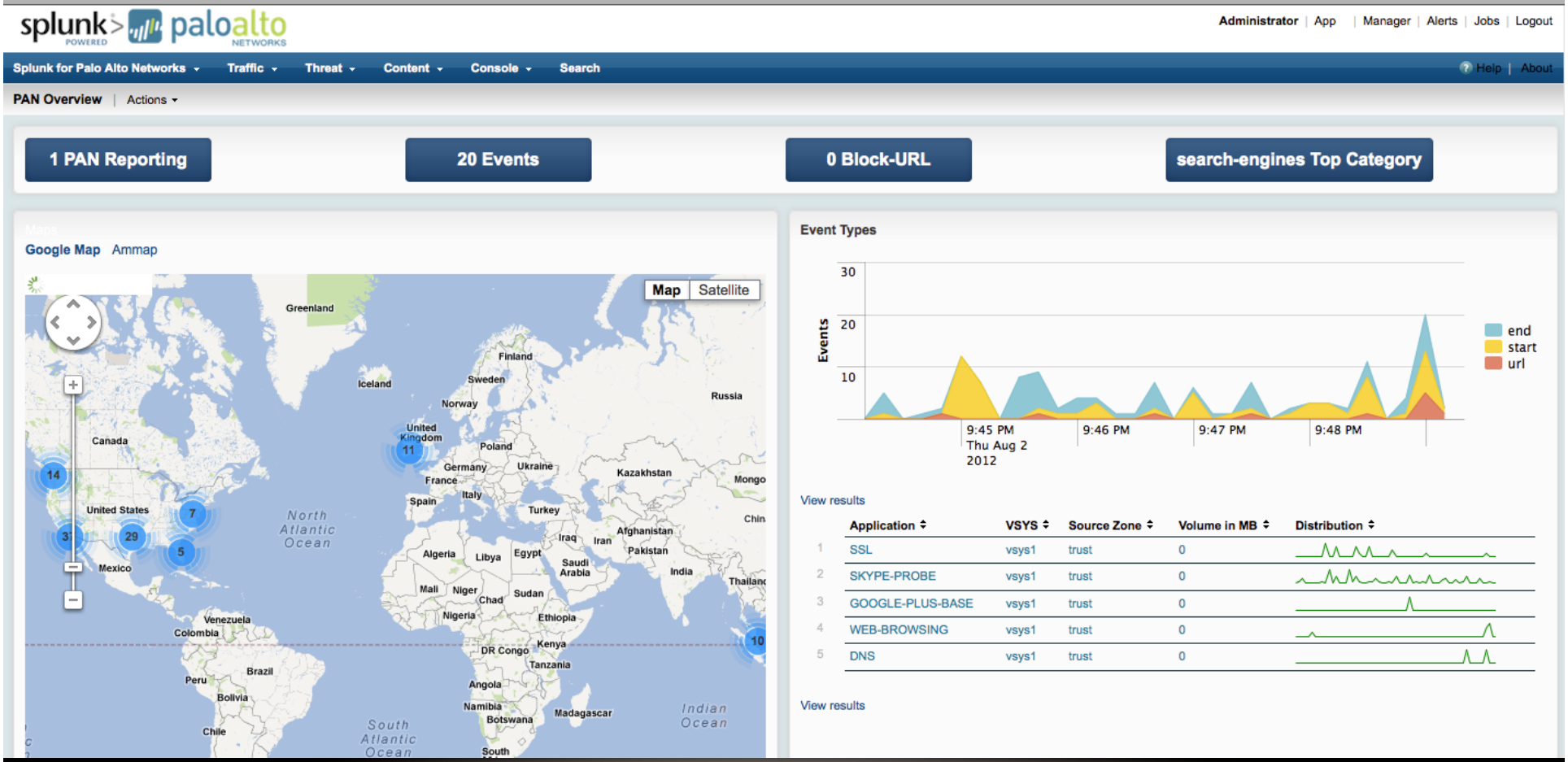


View results

Multidimensional Visualization of connection type, hits, bytes/s, unique destination IPs

Thanks to Monzi Merza for these screenshots

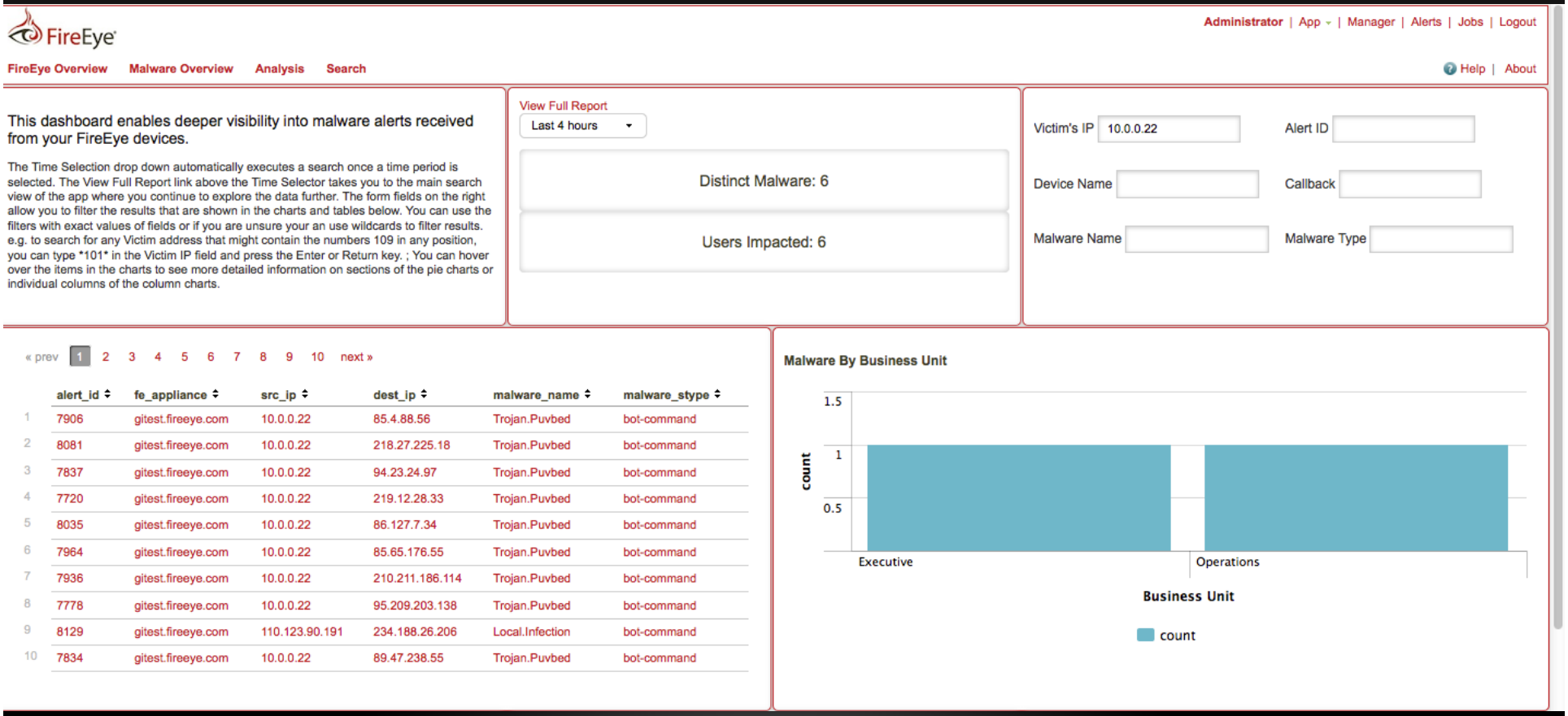
SPLUNK/PALO ALTO VISUALIZATIONS



Magnitudes correlated to geographical mapping, connection type and distribution

Thanks to Monzi Merza for these screenshots

FIREEYE VISUALIZATIONS



Number of malware infections by group, dead listing of activities

Thanks to Monzi Merza for these screenshots

FACEBOOK GEO PLOT AND CONNECTIONS

Paul Butler

VISUALIZING MALWARE

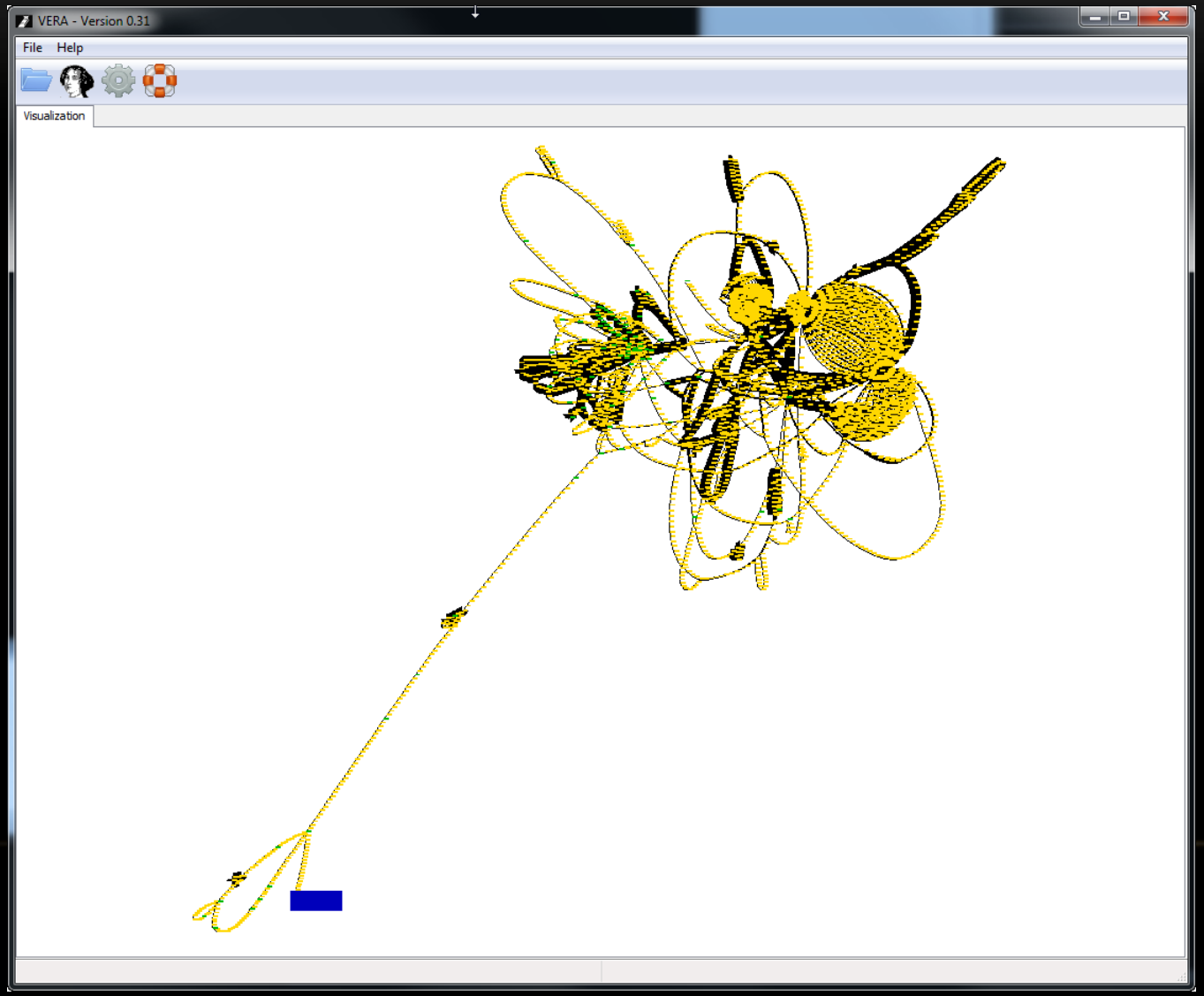
MALWARE: THE PROBLEMS

- Difficulties of malware analysis
  - Finding overall program structure
  - Finding relevant information
  - Unpacking or deobfuscating code
  - Quickly understanding program functionality
- Reverse engineering is difficult for the inexperienced
  - Training time is long without existing experience
  - Training is expensive
  - Some people don’t have aptitude for it
- Solution
  - Visualize the problem
  - Reduce cognitive load of analyst

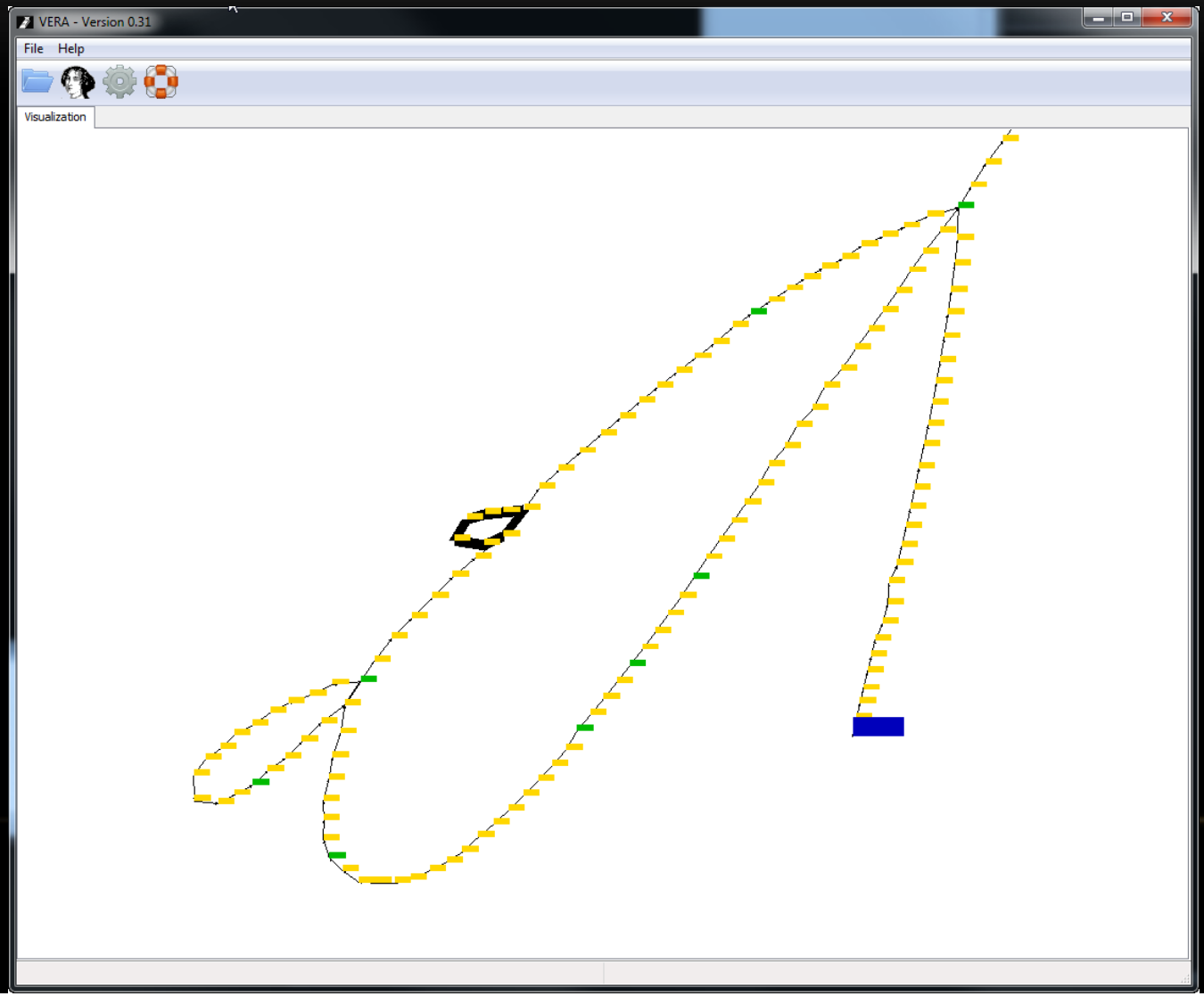
VERA

- Visualization of Executables for Reversing and Analysis
- Force directed graph of execution traces
- Helps with determining where to start the reverse engineering process
- Cuts down on RE time
- Makes unpacking easier

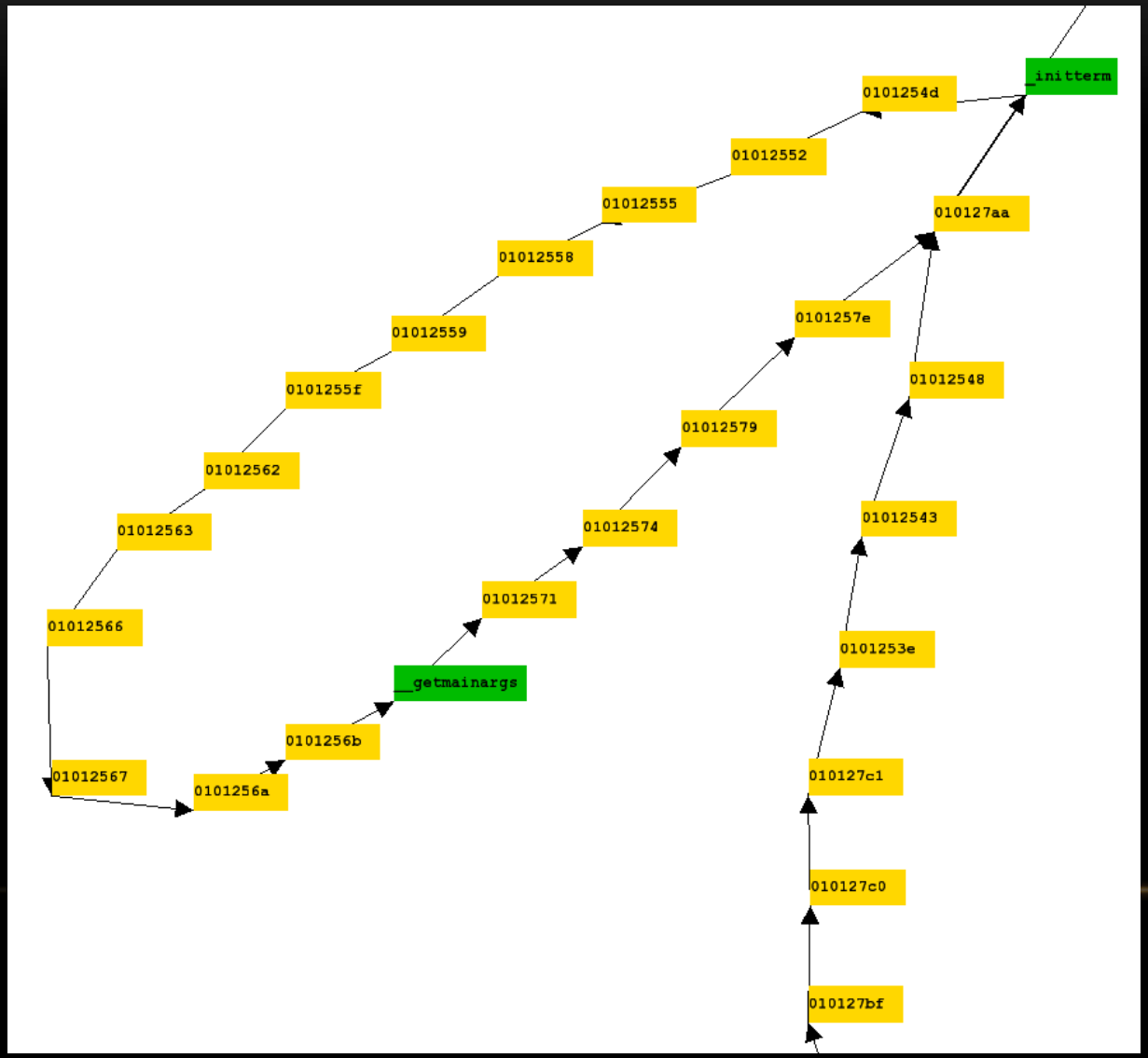
VERA - SCREENSHOTS



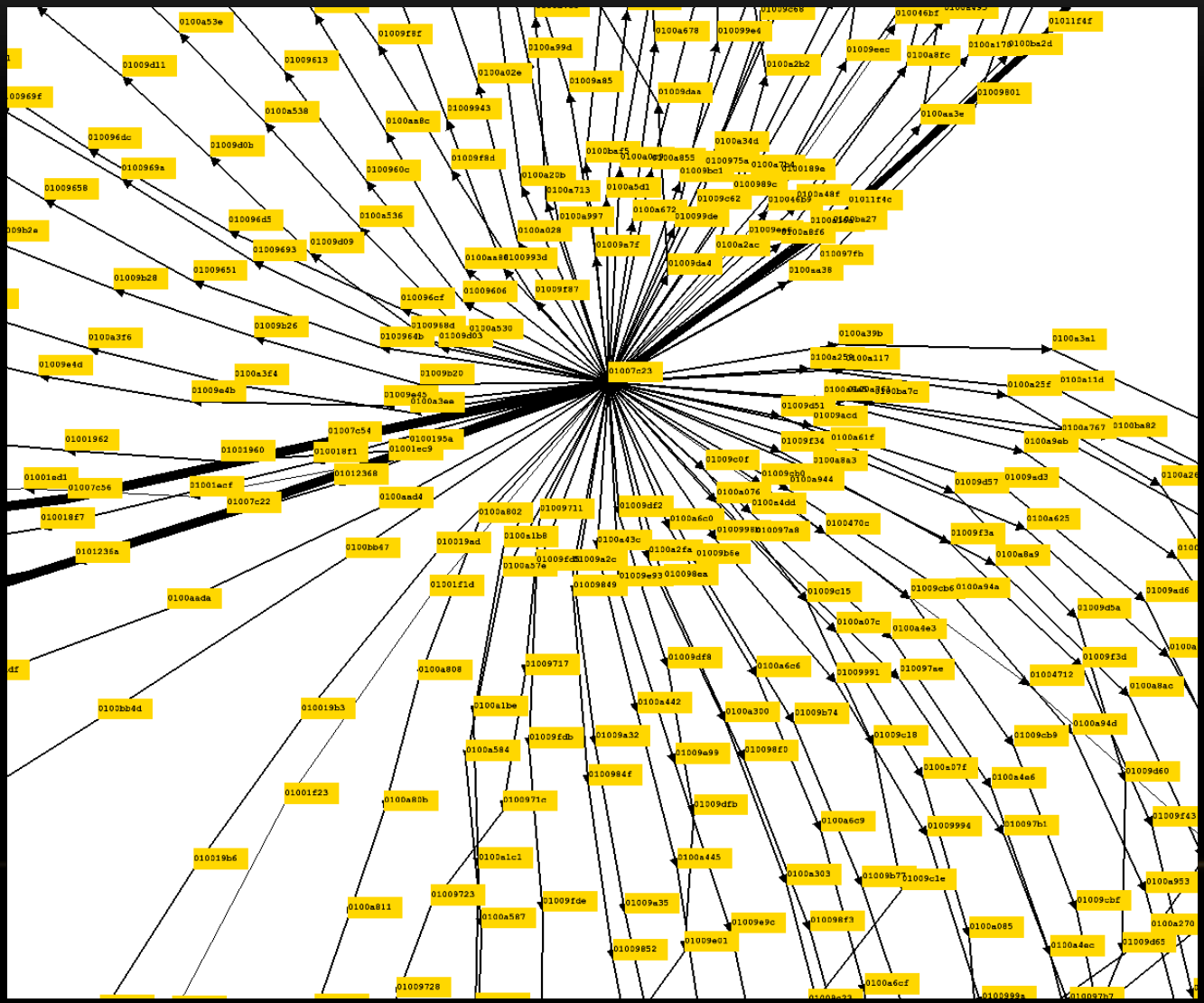
VERA - SCREENSHOTS



VERA - SCREENSHOTS



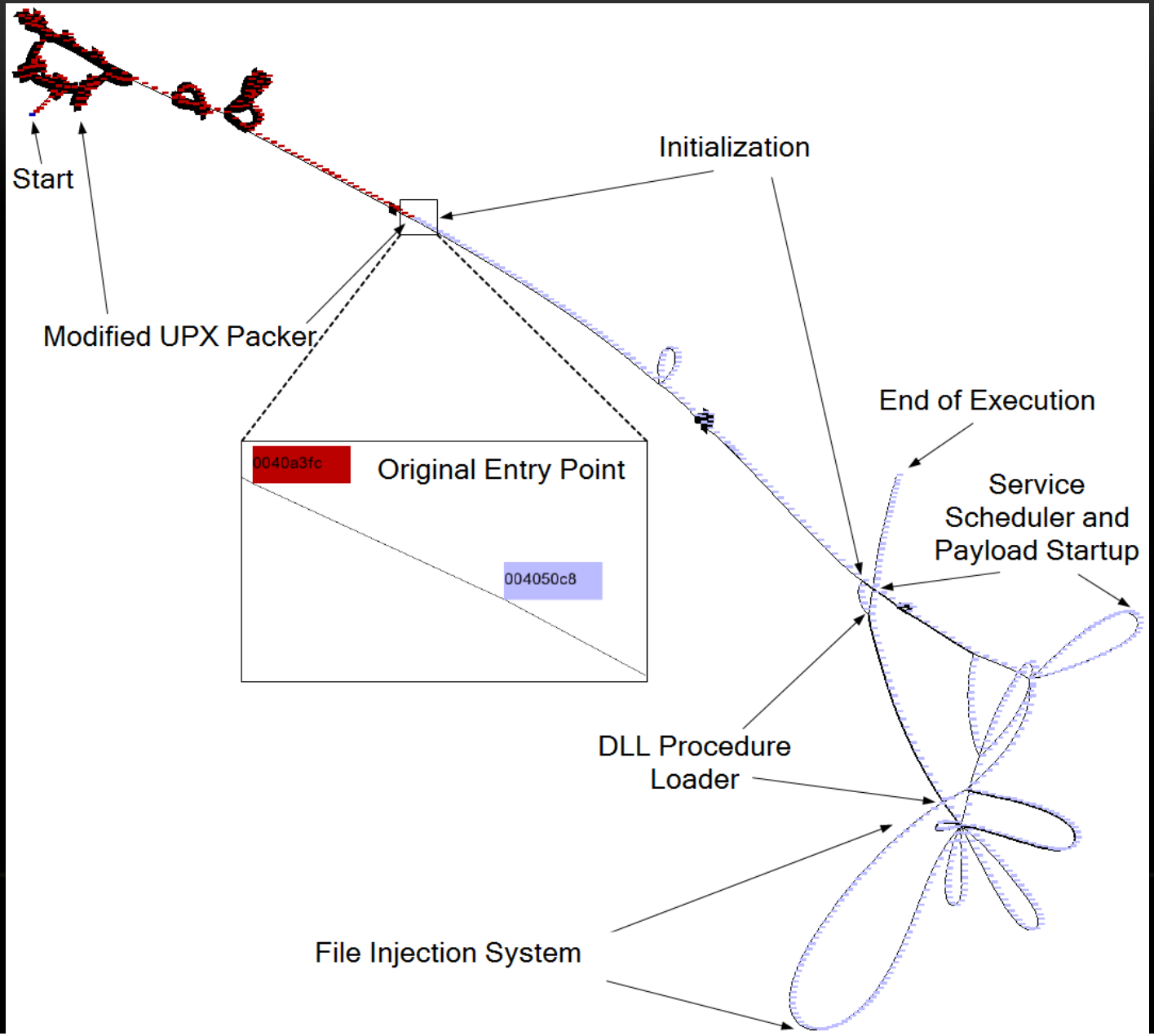
VERA - SCREENSHOTS



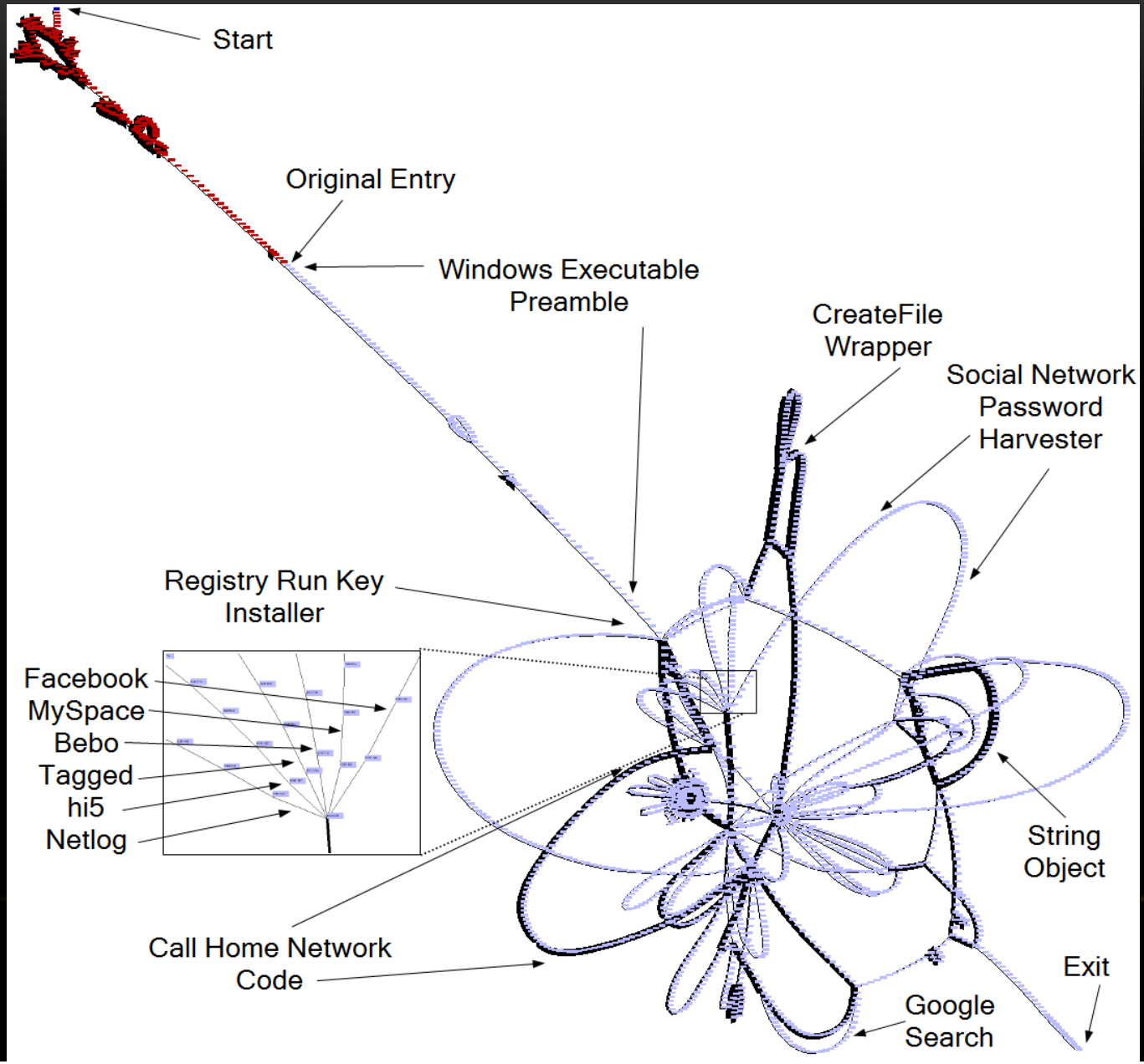
WHAT THE COLORS MEAN

- **Yellow** – Normal uncompressed low-entropy section data
- **Dark Green** – DLL / API / Section not present
- **Light Purple** – SizeOfRawData = 0
- **Dark Red** – High Entropy
- **Light Red** – Instructions not in the packed exe
- **Lime Green** – Operands don’t match

KOOBFACE INITIAL INSTALLATION



KOOBFACE SERVICE



MALWARE VISUALIZATION

Demonstration

DETERMINING EFFECTIVENESS

- User studies
  - Determine effectiveness of visualization
  - Quantify improvement of process
  - Often disproves effectiveness
  - Finding users can be difficult
- Testing criteria
  - Is this tool better than previous efforts?
  - Does the tool effectively convey information?
  - How does the user feel about the tool?

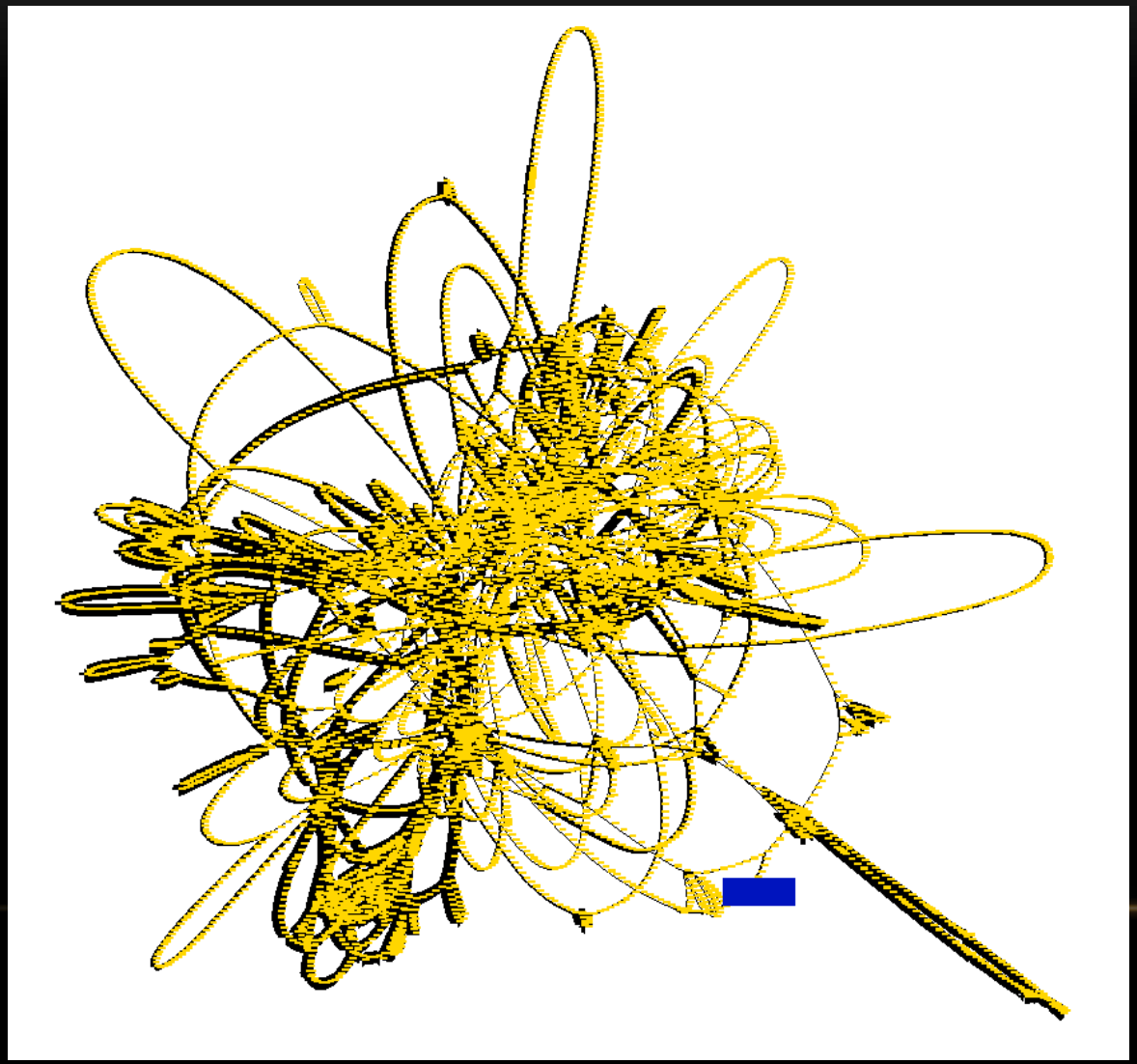
VERA USER STUDY

- Target of the tool: New reverse engineers
- Taught a week long reverse engineering course where unpacking and analysis was taught
- At the end of the week students were shown VERA, given user study
- Rated the effectiveness
- Questions asked
  - Where is the original entry point of a new packer student hasn't encountered before? (MEW and PECompact)
  - Identify portions of the executable
    - Packers
    - Initialization
    - Main loops
  - 3D vs. 2D

USER STUDY RESULTS

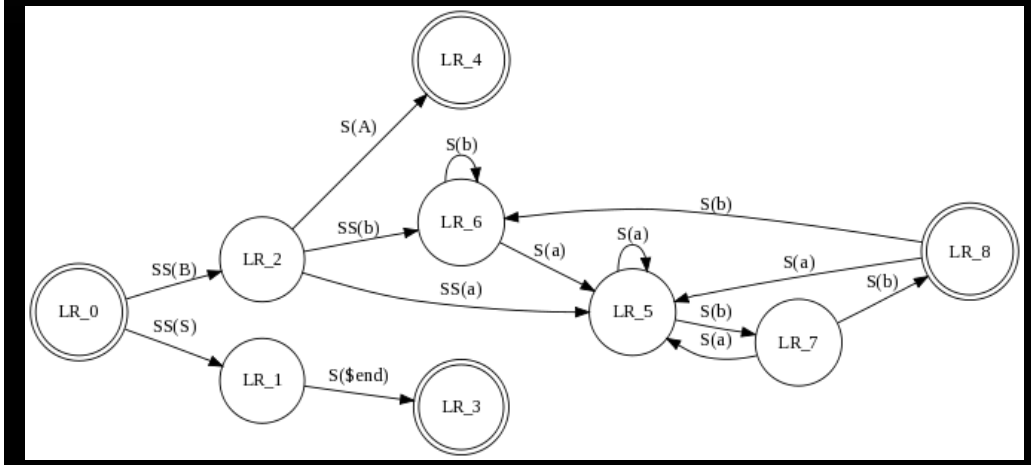
- Noticeable improvement in program structure discovery
  - Loops
  - Switch statements
  - If-statements
  - Packers
- Decrease in time to discover OEP
- 2D vs. 3D
  - Users more productive with 2D version
  - Users liked the 3D version more

PROBLEM – THE YARNBALL



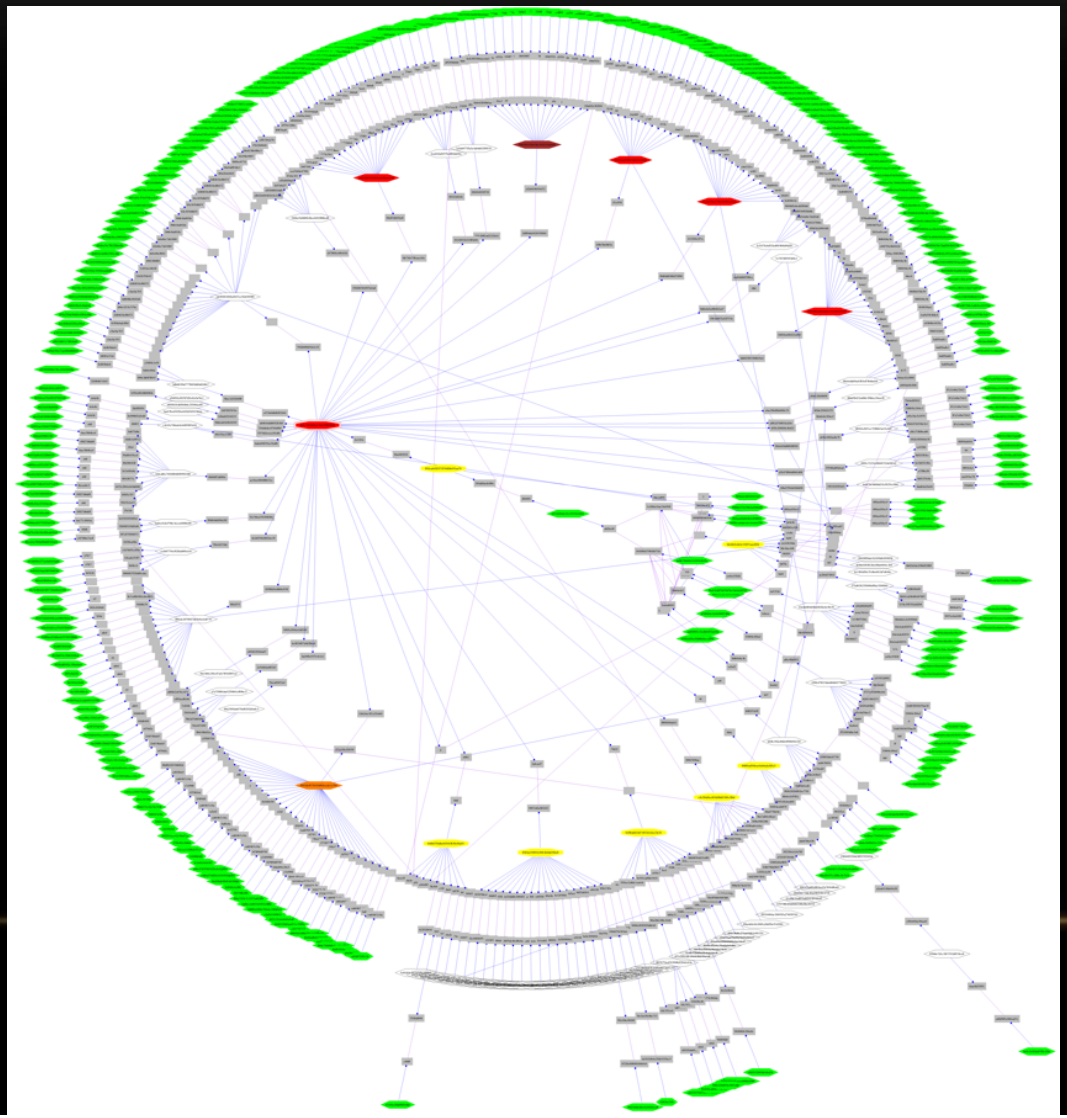
VISUALIZATION TOOL RECOMMENDATIONS

GRAPHVIZ

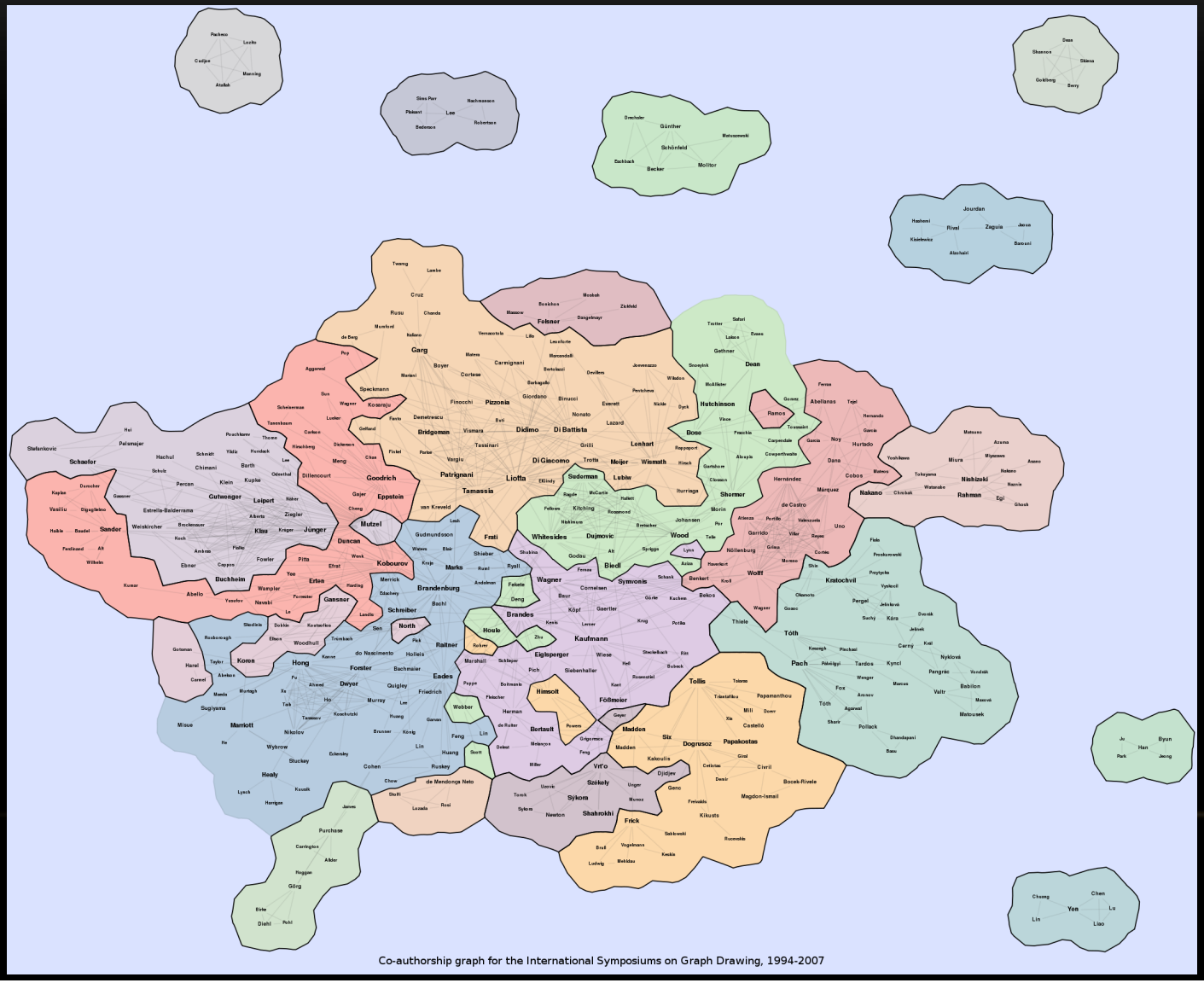


- Useful for drawing small graphs
- Tends to get bogged down and crash on larger graphs (>1000 vertices)
- Easy to use format, well known, nice looking visualization
- Graphviz.org

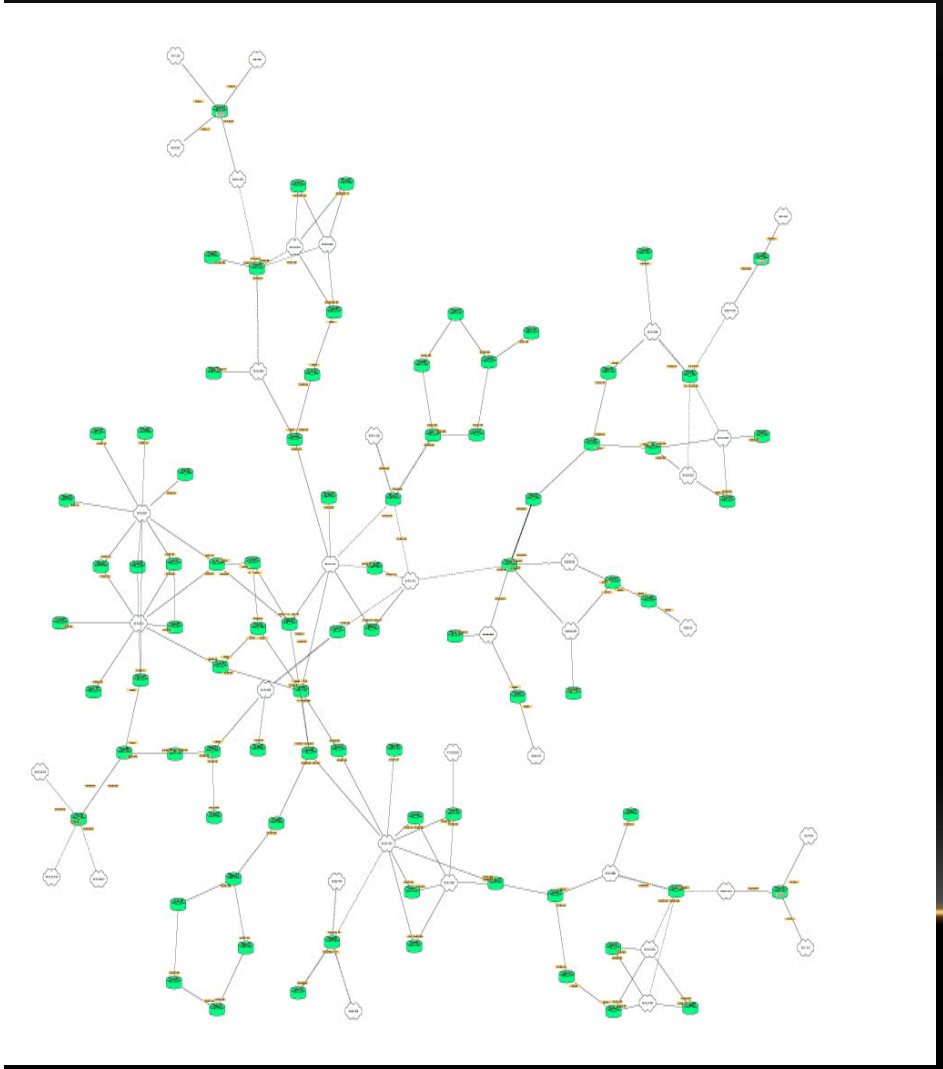
GRAPHVIZ – CIRCULAR GRAPH LAYOUT



GRAPHVIZ RELATIONSHIP MAPPING

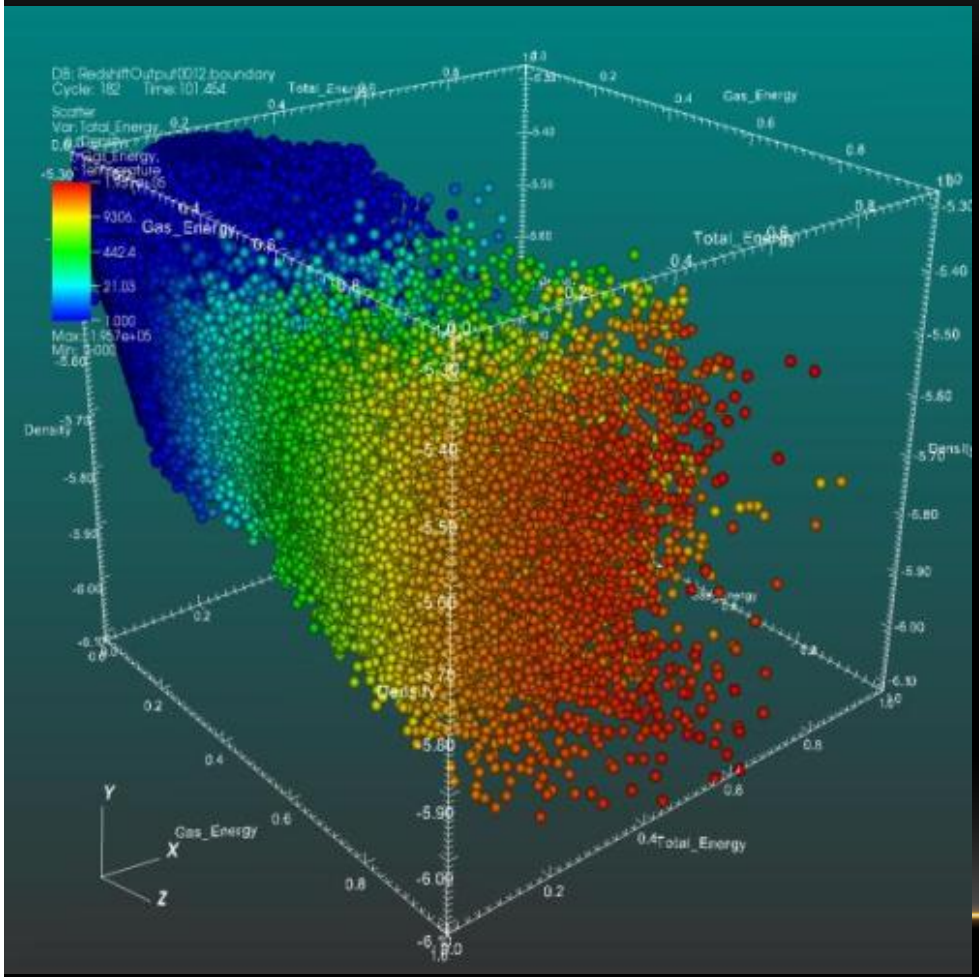


OREAS GOVISUAL DIAGRAM EDITOR



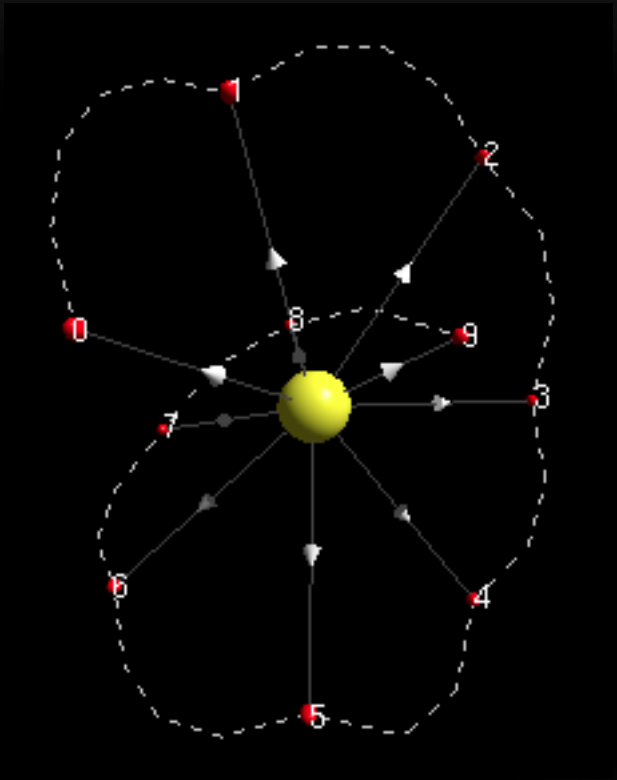
- Good for larger graph layout (>10,000 vertices)
- Has trouble rendering large graphs
- Open Source graph layout: Open Graph Display Framework (OGDF)
- Oreas.com
- Ogdf.nets

VISIT



- Primarily used for modeling 3D environments
- Good for similarity metrics, mapping entropy, etc.
- Texture and surface mapping
- https://wci.llnl.gov/codes/visit/home.html

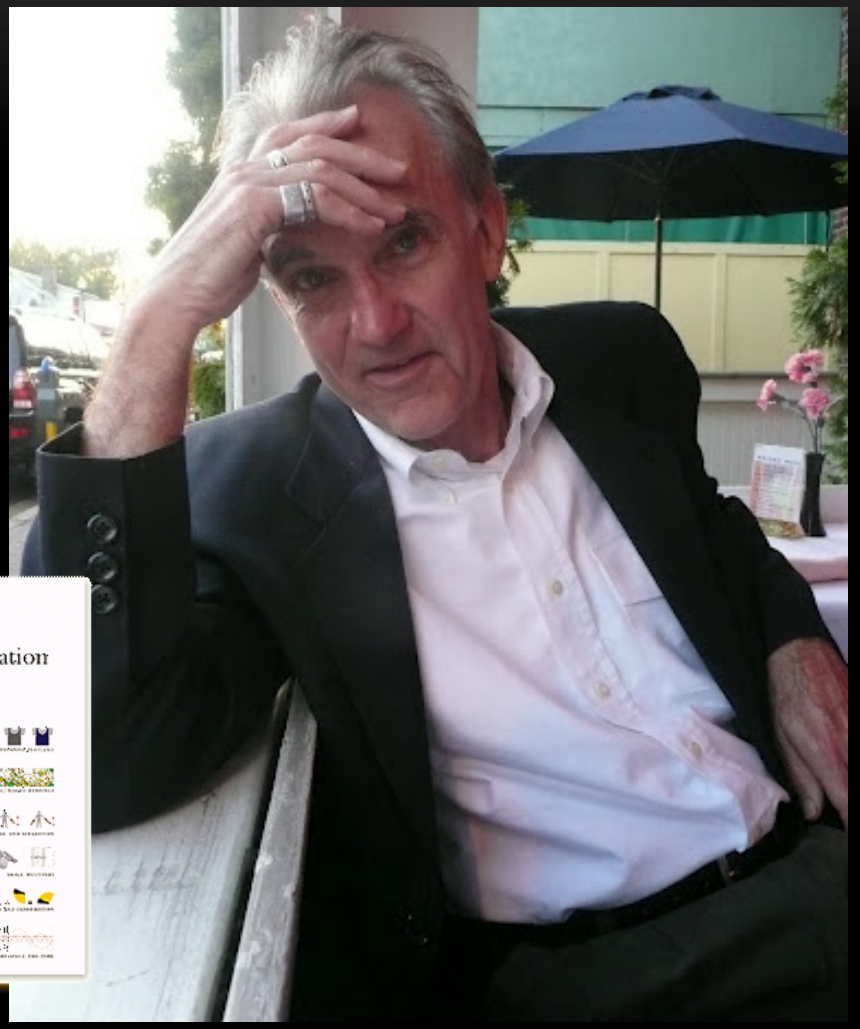
UBIGRAPH



- 3D Visualization Demonstrated Previously
- Extensible, easy to integrate with Python, C, or any other XML-RPC library
- Fast, lots of nodes
- Only available for OS X, Linux (no Windows)
- Ubietylab.net/ubigraph

VISUALIZATION

Edward Tufte
wrote some books
you should read



Edward Tufte

CONCLUSION

- Visualization, when done properly, adds to the comprehension of complex tasks
- Difficulty is making a compelling visualization
- Perform a user-study

SHAMELESS PLUG

Vizsec 2012

- Seattle Washington
- October 15
- In conjunction with VisWeek
- Vizsec.org