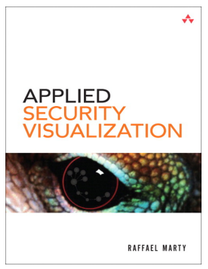
Cloud-based Log Analysis and Visualization

RMLL 2010, Bordeaux, France

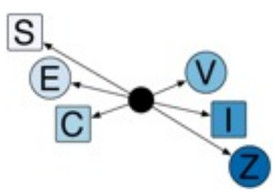
Raffael Marty - @zrlram

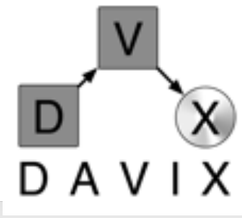
Raffael (Raffy) Marty

- Founder @ loggly
- Chief Security Strategist and Product Manager @ Splunk
- Manager Solutions @ ArcSight
- Intrusion Detection Research @ IBM Research
- IT Security Consultant @ PriceWaterhouse Coopers



Applied Security Visualization
Publisher: Addison Wesley (August, 2008)
ISBN: 0321510100





Agenda

- Introduction
- Visualization
- InfoViz Process
- Visualization **Tools**
- The Cloud
- Loggly
- Do it Yourself
  - AfterGlow
  - Google Visualization API
- Visualization **Use-Cases**
- Visualization **Resources**

Open Your Eyes

Security Is About Seeing

Goals

- Learn how you can
  - use **visualization** to help solve security problems
  - leverage the **cloud** to build security visualization tools

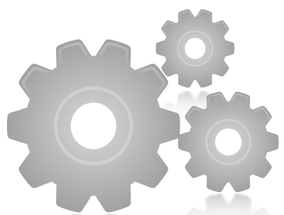
Information Visualization?

A picture is worth a thousand log records.

Explore and Discover

Answer a Question

Pose a New Question



Increase Efficiency

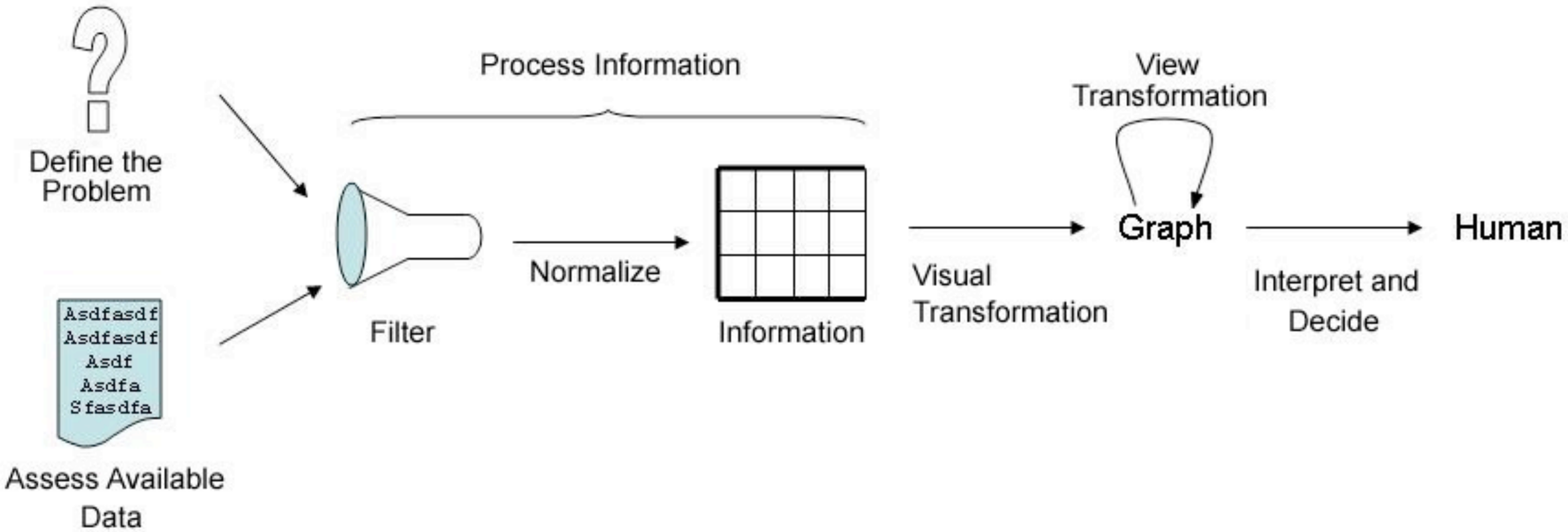
Communicate Information

Support Decisions

Inspire

Visualization and The Cloud

InfoViz Process



Collect

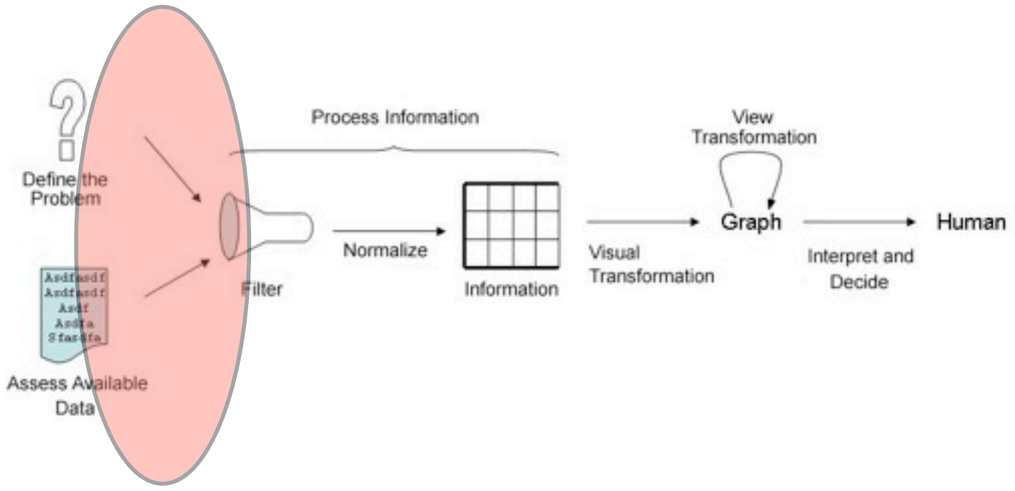
- large-scale data collection
- and processing

Process

- Your parsers
- Standard formats

Visualize

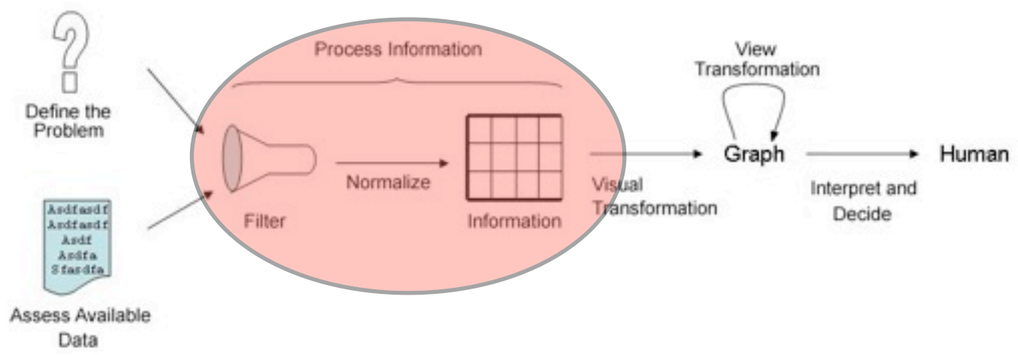
- Visualization Tools
- and Libraries


Define the Problem
Process Information
View Transformation
Graph
Human
Normalize
Visual Transformation
Interpret and Decide
Filter
Information
Assess Available Data

Collect

Log Management

- Log Collection and Centralization
- Log Storage
- Log Filtering
- Log Aggregation
- Log Search and Extraction
- Log Retention and Archiving


Define the Problem
Process Information
Normalize
Filter
Information
Visual Transformation
View Transformation
Graph
Interpret and Decide
Human
Asdfasdf
Asdfasdf
Asdf
Asdfa
Sfasdfa
Assess Available Data

Process

Standard Formats

- Multiple formats

```
Oct 13 20:00:43.874401 rule 193/0(match): block in on xl0: 212.251.89.126.3859 >: S
1818630320:1818630320(0) win 65535 <mss 1460,nop,nop,sackOK> (DF)

Oct 13 20:00:43 fwbox local4:warn|warning fw07 %PIX-4-106023: Deny tcp src
internet: 212.251.89.126/3859 dst 212.254.110.98/135 by access-group
"internet_access_in"

Oct 13 20:00:43 fwbox kernel: DROPPED IN=eth0 OUT= MAC=ff:ff:ff:ff:ff:ff:00:0f:cc:
81:40:94:08:00 SRC=212.251.89.126 DST=212.254.110.98 LEN=576 TOS=0x00 PREC=0x00
TTL=255 ID=8624 PROTO=TCP SPT=3859 DPT=135 LEN=556
```

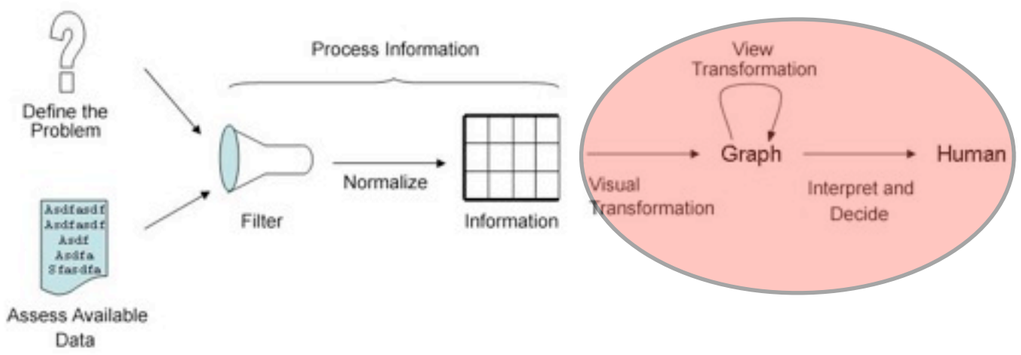
- Log Standards
  - CEE (cee.mitre.org)
  - IDMEF
  - SDEE
  - CBE
  - WELF
  - XDAS

Normalization

- **Parsers**

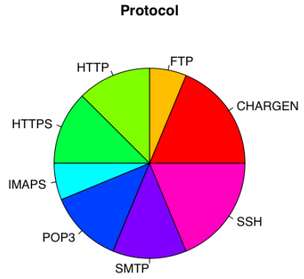
  "To analyze or separate (input, for example) into more easily processed components." (answers.com)

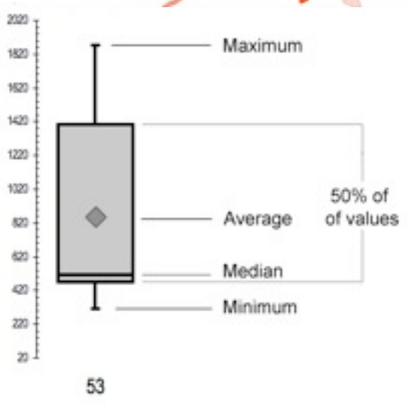
- Generate a common output format for vis-tools (e.g., CSV)
- For example
  - Regex `/(\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3})/g`
  - http://secviz.org/content/parser-exchange

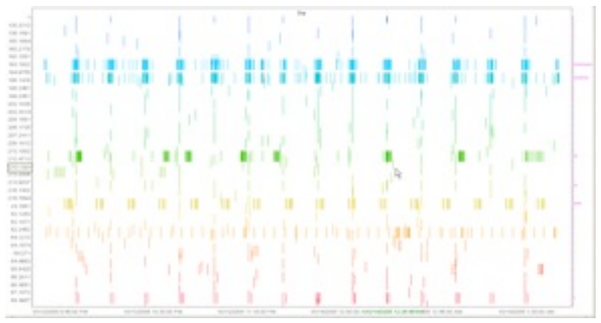

Define the Problem
Process Information
Asdfasdf
Asdfasdf
Asdf
Asdfa
Sfasdfa
Filter
Normalize
Information
View Transformation
Graph
Human
Visual Transformation
Interpret and Decide
Assess Available Data

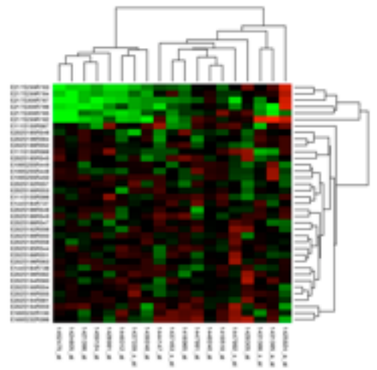
Visualize

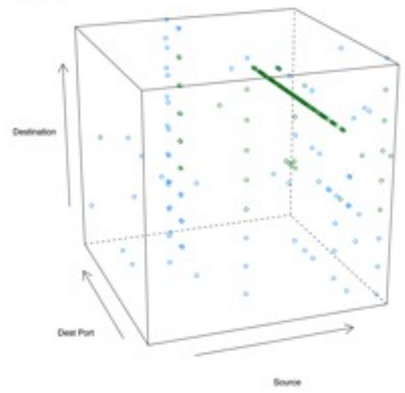
Choose Your Poison

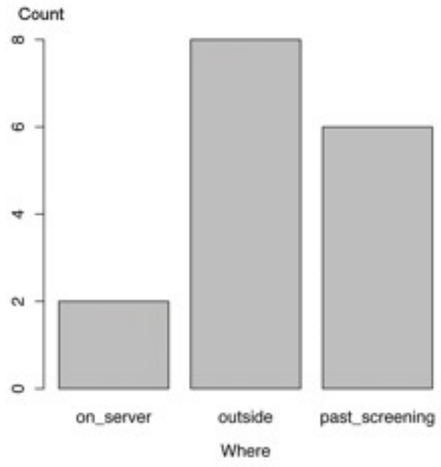


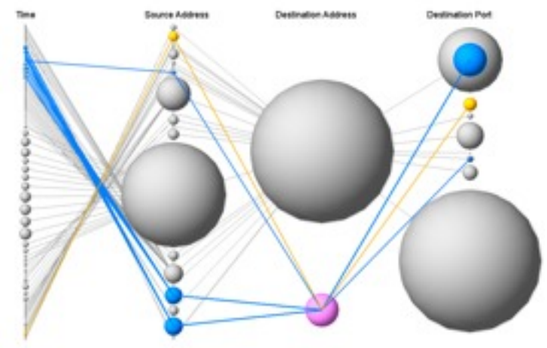


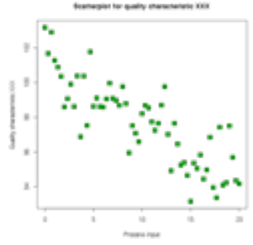


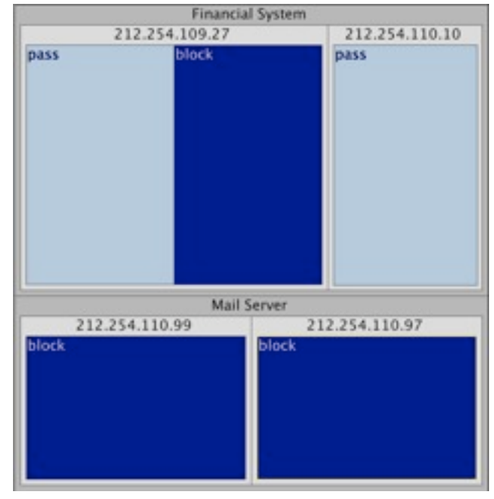










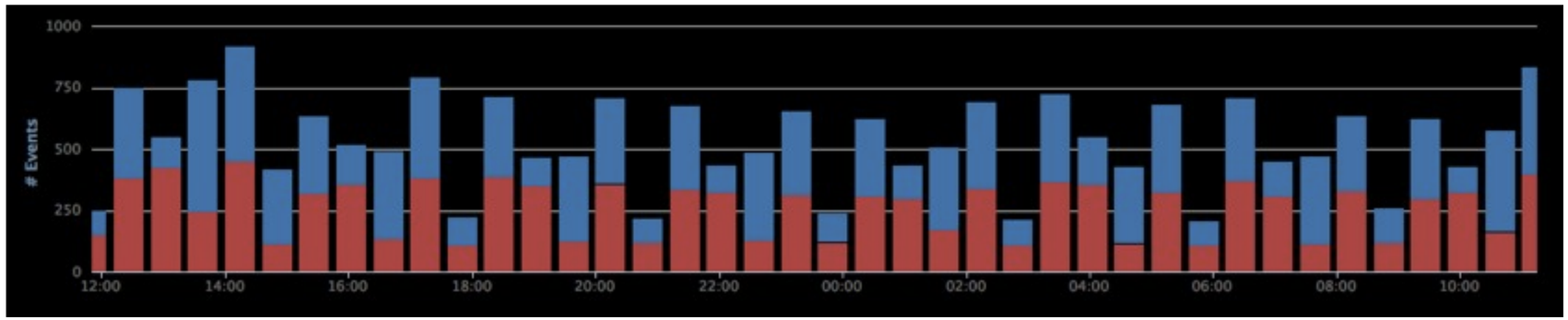


Reporting vs. Visualization

- Reporting Libraries
  - HighCharts
  - Flot
  - Google Chart API
  - Open Flash Chart
- Visualization Libraries
  - TheJIT
  - Graphael
  - Protovis
  - ProcessingJS
  - Flare

JavaScript vs. Flash vs. XYZ

HighCharts

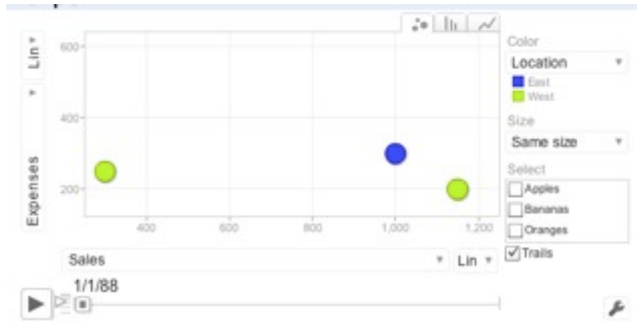


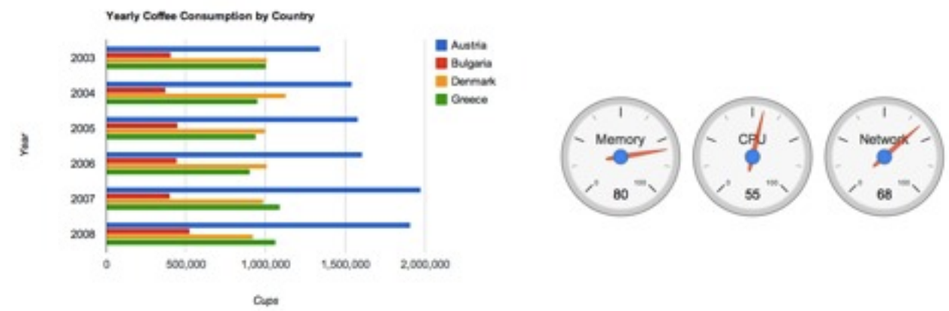
- Click-Through
- On load
  - near real-time updates
- Zoom

- AJAX data input via JSON

http://www.highcharts.com/

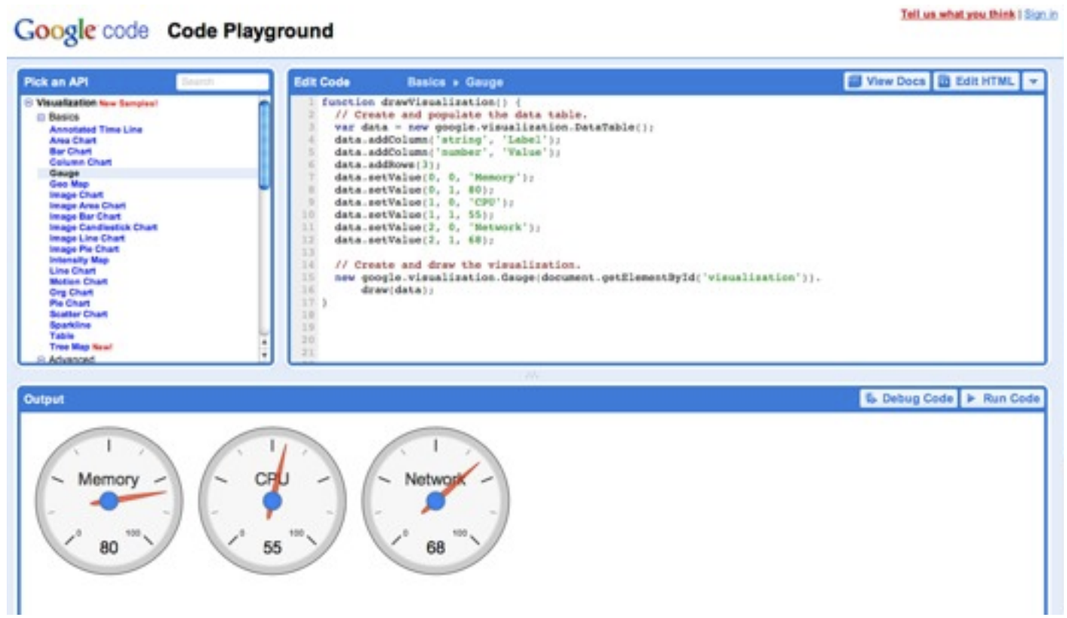
Google Visualization API





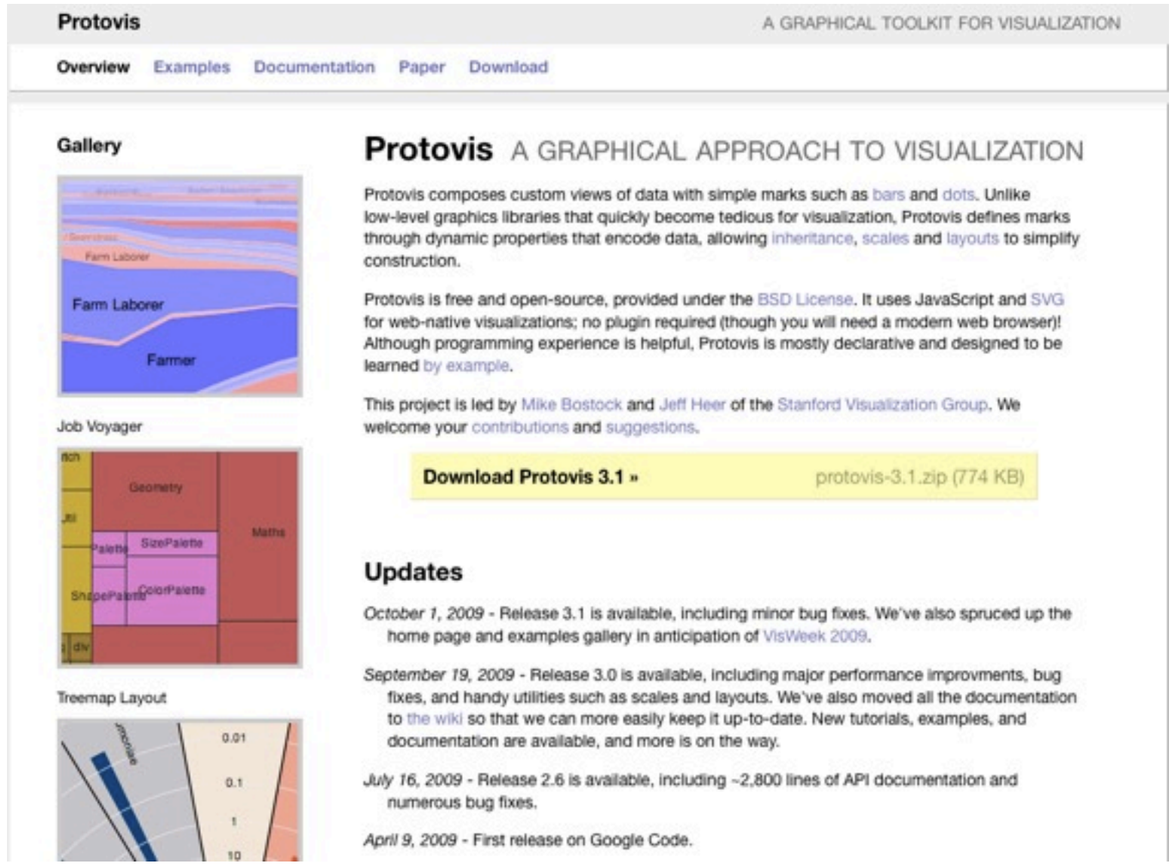
http://code.google.com/apis/visualization/interactive_charts.html

- JavaScript
- Based on DataTables()
- Many graphs
- Playground
  - http://code.google.com/apis/ajax/playground



ProtoVis

- JavaScript based visualization library
- Charting
- Treemaps
- BoxPlots
- Parallel Coordinates
- etc.

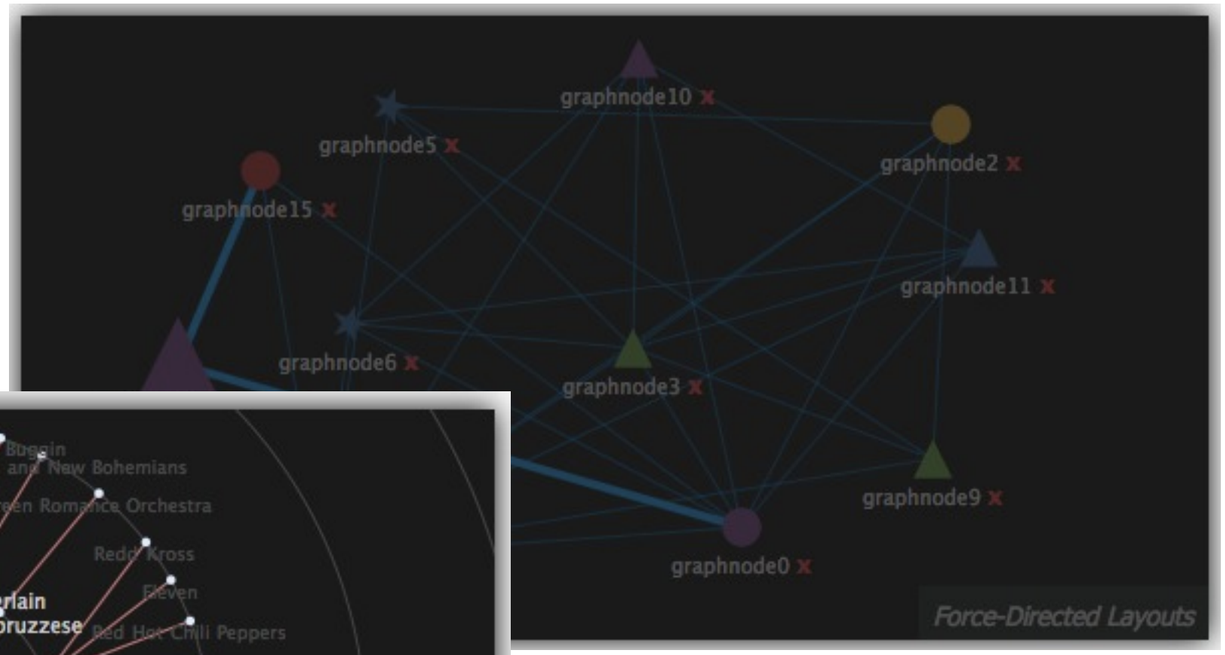


http://vis.stanford.edu/protovis/

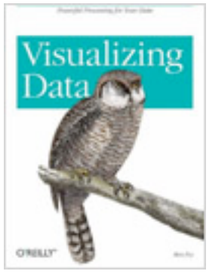
TheJIT

http://thejit.org/

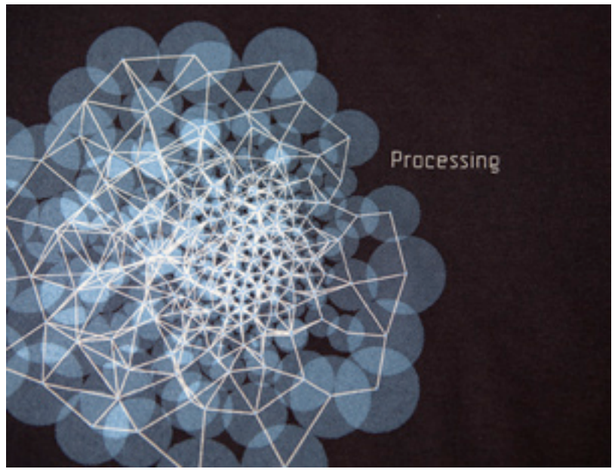
- JavaScript InfoVis Toolkit
- Interactive
- Link Graphs

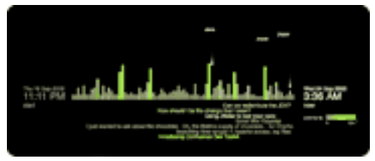


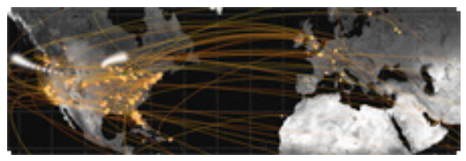
Processing



- Visualization library
- Java based
- Interactive (event handling)
- Number of libraries to
  - draw in OpenGL
  - read XML files
  - write PDF files
- Processing JS
  - JavaScript
  - HTML 5 Canvas
  - Web IDE





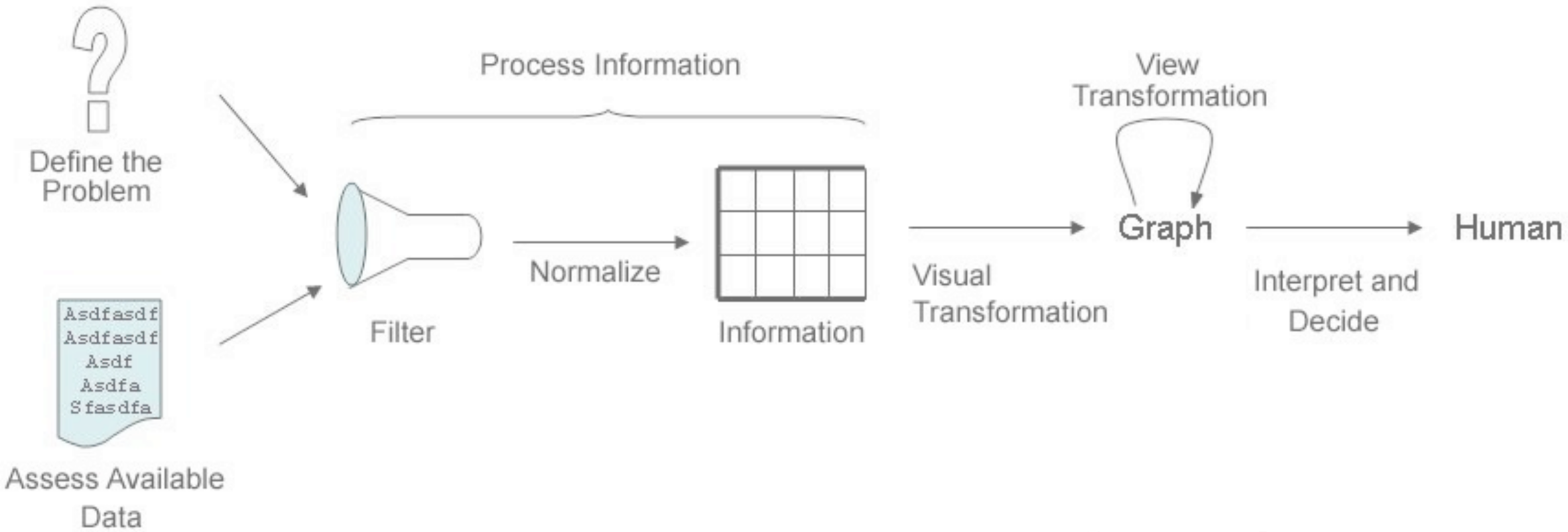


http://processingjs.org/

http://processing.org/

Building Your Own

Build Your Own



Loggly

Regexes

AfterGlow
Google Vis

Data Collection in the Cloud

The (public) Cloud

What it is

- multi–tenancy
- elastic
- “infinite” resources
- pay as you go
- self provisioning

It’s not

- private data center
- virtualization

Types

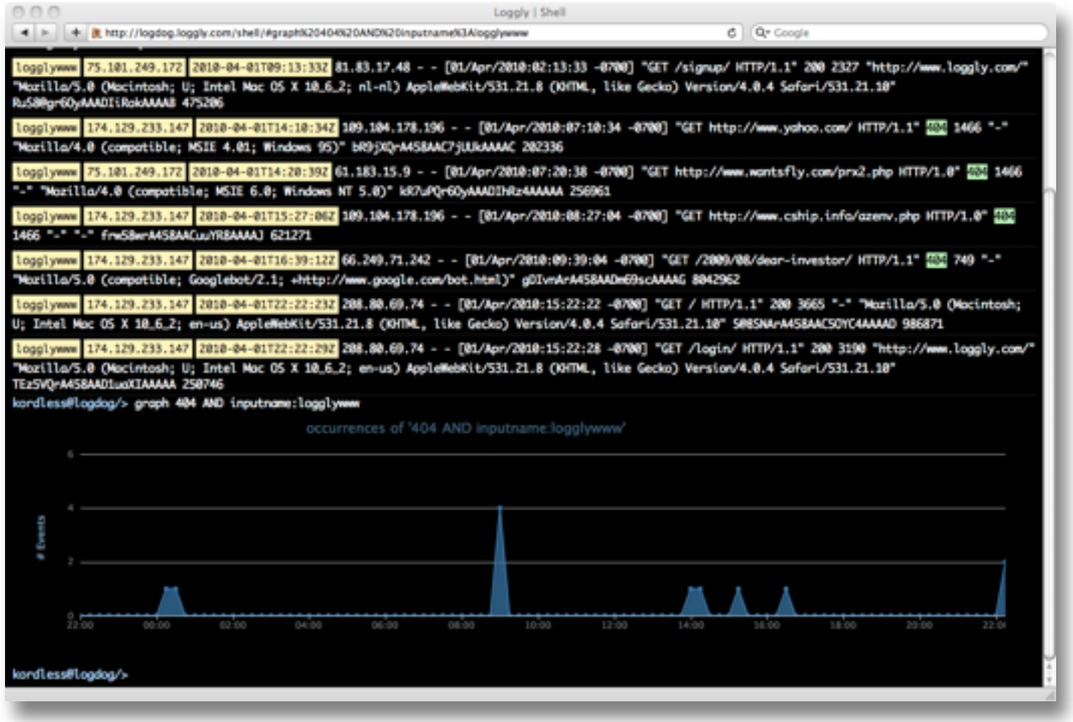
- SaaS – Software
- PaaS – Platform
- IaaS – Infrastructure

Benefits

- No installation
- No elaborate configurations
- No maintenance
- Great scalability
- 7x24 availability

LaaS – Logging as a Service

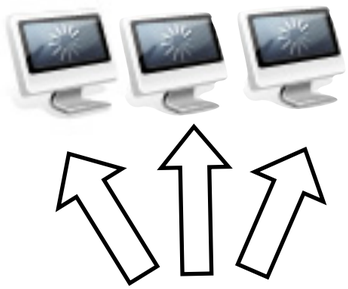
- **All your data in one place**
  - Loggly manages your data (index, store, archive, etc.)
- **Extremely fast search across all your data**
  - Data source agnostic (no parsers)
- **Data management**
  - access control
  - data segregation
  - data overview and summaries
- **API access**



Loggly Architecture

Data Sources

Clients



Loggly
user interface

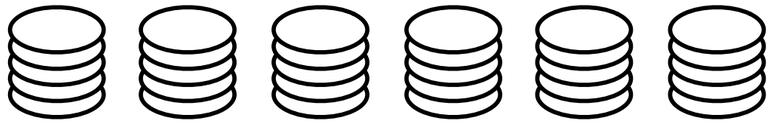
Proxies

API

Data collection
Data access

Indexers and Search Machines

Distributed
indexing and
processing



Distributed
data store

Loggly APIs

- URL format:

http://wiki.loggly.com/api-documentation

`http://<subdomain>.loggly.com/api/<resource>`

- RESTful API
  - Access through: /api/<**resource**>
  - JSON, XML, JSONP output
- Authentication
  - Basic auth
  - oAuth

HTTP Based

- **`GET`** – read
- **`POST`** – create
- **`PUT`** – update
- **`DELETE`** – delete

`http://loggly.loggly.com/api/search/?q=error`
User: guest / Password: loggly

syslog to:
logs.loggly.com:514

Search

http://[domain].loggly.com/api/search?q=404


```
{
    "data": [
        {
            "indexed": "2010-07-03T17:17:38.909Z",
            "ip": "75.101.249.172",
            "text": "Oct 13 20:00:38.018152 rule 57/0(match): pass in on xl1: 195.141.69.45.1030 > 62.2.32.250.53:  34388 [1au]
[|domain] (DF)",
            "inputname": "logglyweb",
            "timestamp": "2010-07-03 10:17:38"
        },
        {
            "indexed": "2010-07-03T17:17:37.879Z",
            "ip": "75.101.249.172",
            "text": "Oct 13 20:00:38.115862 rule 57/0(match): pass in on xl1: 195.141.69.45.1030 > 192.134.0.49.53:  49962 [1au]
[|domain] (DF)",
            "inputname": "logglyapp",
            "timestamp": "2010-07-03 10:17:37"
        },

        ...
```

Parser

Raw

```
Oct 13 20:00:38.018152 rule 57/0(match): pass in on xl1: 195.141.69.45.1030 > 62.2.32.250.53:  34388 [1au][|domain] (DF)
Oct 13 20:00:38.115862 rule 57/0(match): pass in on xl1: 195.141.69.45.1030 > 192.134.0.49.53:  49962 [1au][|domain] (DF)
Oct 13 20:00:38.157238 rule 57/0(match): pass in on xl1: 195.141.69.45.1030 > 194.25.2.133.53:  14434 [1au][|domain] (DF)
```

Regex / Parser

```
(.*) rule ([-\d]+\/\d+)\(.*?\): (pass|block) (in|out) on (\w+):
(\d+\.\d+\.\d+\.\d+)\.?(\d*) [<>]
(\d+\.\d+\.\d+\.\d+)\.?(\d*): (.*)
```

Normalized (CSV)

```
Oct 13 20:00:38.018152,57/0,match,pass,in,xl1,195.141.69.45,1030,62.2.32.250,53,34388 [1au][|domain] (DF)
Oct 13 20:00:38.115862,57/0,match,pass,in,xl1,195.141.69.45,1030,192.134.0.49,53,49962 [1au][|domain] (DF)
Oct 13 20:00:38.157238,57/0,match,pass,in,xl1,195.141.69.45,1030,194.25.2.133,53,14434 [1au][|domain] (DF)
```

Visualize

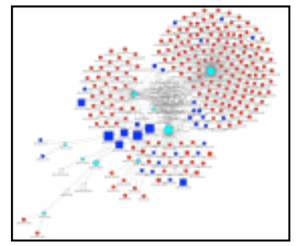
Parser

AfterGlow

CSV file

Grapher

Graph file



Configuration

```
color.source=“green” if ($fields[0] ne “d”)
cluster.target=regex_replace("(\\d\+)\\.")."/8"
threshold.event=5
size.target=$fields[1]
```

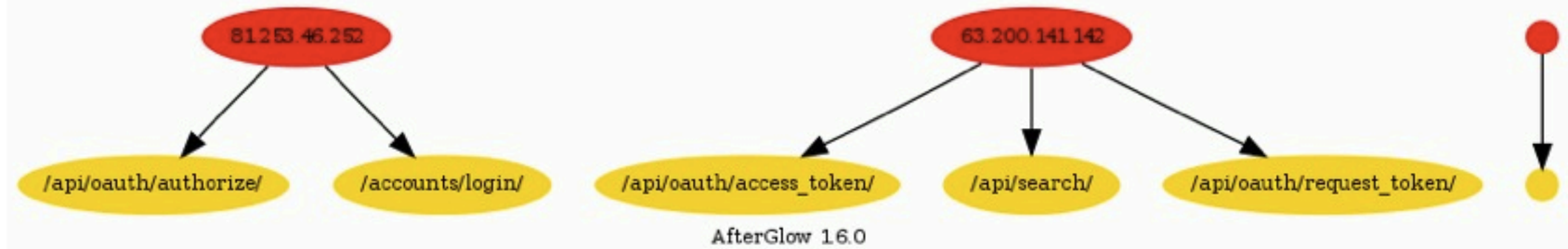
```
digraph structs {
  graph [label="AfterGlow 1.5.8",  fontsize=8];
  node [shape=ellipse, style=filled,
    fontsize=10, width=1, height=1,
    fixedsize=true];
  edge [len=1.6];

  "aaelenes" -> "Printing Resume" ;
  "abbe" -> "Information Encryption" ;
  "aanna" -> "Patent Access" ;
  "aatharuv" -> "Ping" ;
}
```

http://afterglow.sf.net

AfterGlow Cloud

loggly

AfterGlow – Loggly Mashup



Search Params

Subdomain logdog

Search inputname:logglyweb

Parser/Field Select Parser

Regex ^(\d+\.\d+\.\d+\.\d+)\s+\S+\s+\S+\s+\[.*?\]\s+\"GET (.*?)(?:\?[

Rows 20

Glow

AfterGlow Params

Split nodes? ☑

Split Mode One unique event node

Print node count ☐

Color Config

Filter Config

Threshold:

Source fanout:

Event fanout:

Cluster Config Select Clustering for Source | Select Clustering for Target

Node labels? ☐

Two nodes only (not three)? ☑

Default Edge Length 1.2

Which Viz Tool? neato

Loggly

↓

JSON

↓

CSV

↓

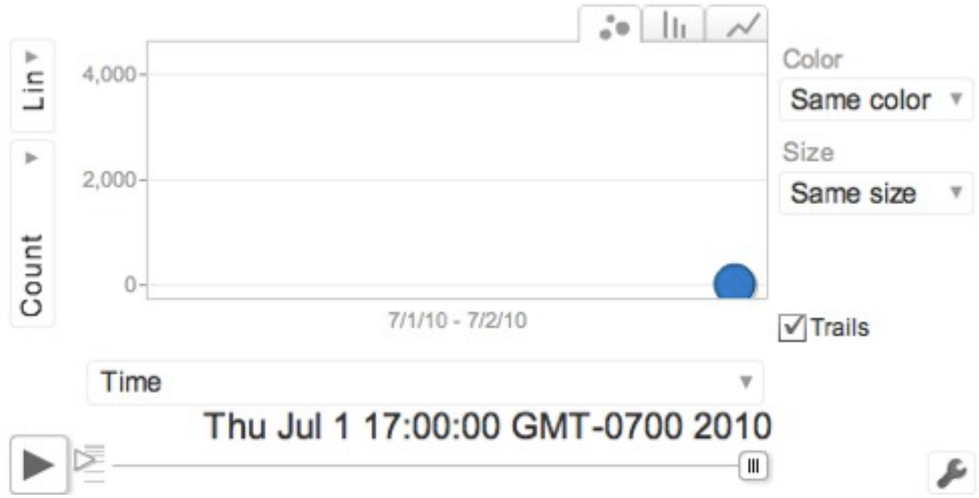
DOT

↓

Graph

Google Vis

- JSON to Graphs
- DataTable
  - used among all charts
- Interactivity through events

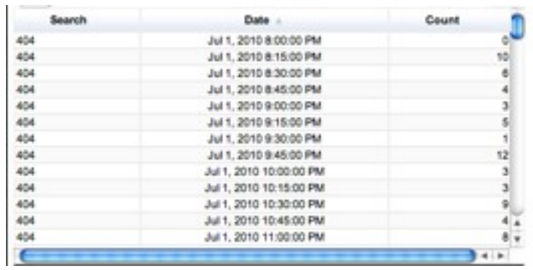


Search	Date	Count
404	Jul 1, 2010 8:00:00 PM	0
404	Jul 1, 2010 8:15:00 PM	10
404	Jul 1, 2010 8:30:00 PM	6
404	Jul 1, 2010 8:45:00 PM	4
404	Jul 1, 2010 9:00:00 PM	3
404	Jul 1, 2010 9:15:00 PM	5
404	Jul 1, 2010 9:30:00 PM	1
404	Jul 1, 2010 9:45:00 PM	12
404	Jul 1, 2010 10:00:00 PM	3
404	Jul 1, 2010 10:15:00 PM	3
404	Jul 1, 2010 10:30:00 PM	9
404	Jul 1, 2010 10:45:00 PM	4
404	Jul 1, 2010 11:00:00 PM	8

Google Vis Code

```
<script type="text/javascript">
      google.load('visualization', '1', {'packages':['motionchart', 'table', 'annotatedtimeline']});
      google.setOnLoadCallback(call);
      var trends = new Array();
      function call() {
          $.ajax({ url: "http://logdog.loggly.com/api/search/?q=404&facets=True&buckets=100",
                type:'GET',  dataType: 'jsonp', username: 'xxxxx', password: 'xxxxxx',
                success: function(data) {
                    trends = data.data
                    drawChart();
                }
          });
      }
      function drawChart() {
        var data = new google.visualization.DataTable();
        data.addColumn('string', 'Search');
        data.addColumn('datetime',   'Date');
        data.addColumn('number', 'Count');
        data.addRows(trends);

        var chart = new google.visualization.MotionChart(document.getElementById('chart_div'));
        chart.draw(data, {width: 600, height: 500, state:state});

        var view = new google.visualization.DataView(data);
        view.setRows(view.getFilteredRows([{column: 1, minValue: new Date(2007, 0, 1)}]));
        var table = new google.visualization.Table(document.getElementById('test_dataview'));
        table.draw(view, {sortColumn: 1});

        var time = new google.visualization.AnnotatedTimeLine(document.getElementById('timeline'));
        time.draw(timedata, {displayAnnotations: true});
      }
</script>
```

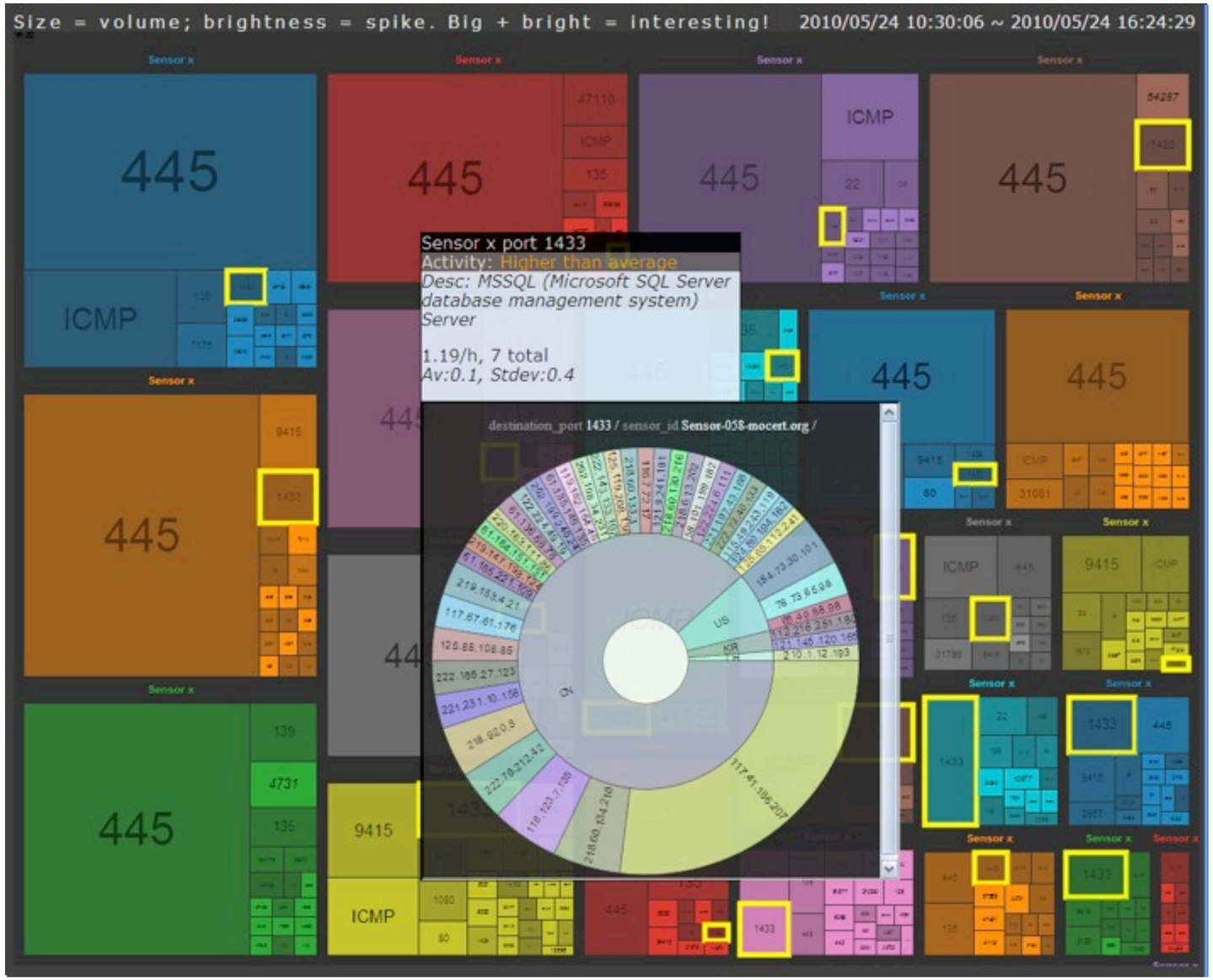
THIS CODE IS NOT FUNCTIONAL!



Visualization Use-Cases

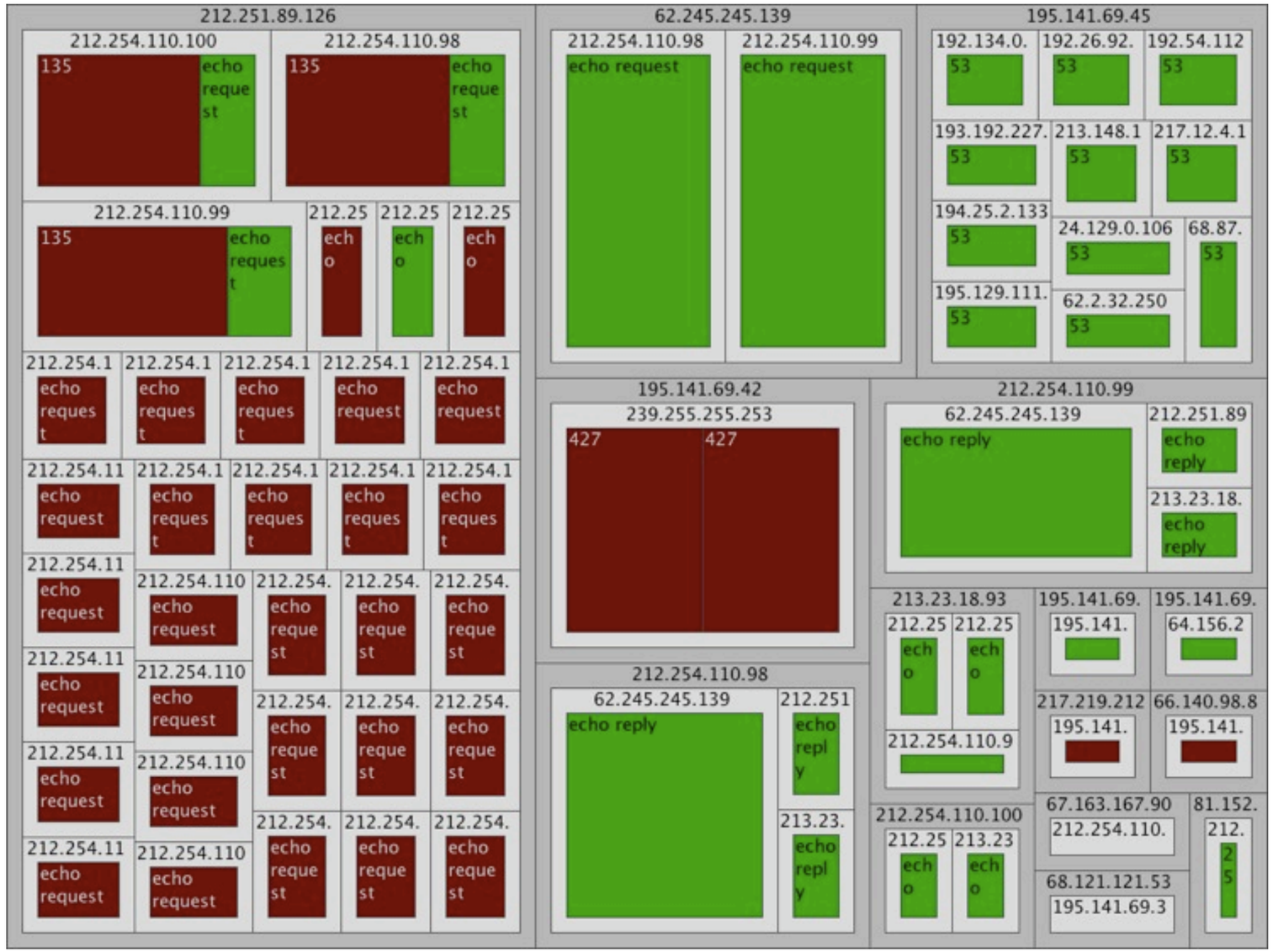
NetFlow Visualization

- Treemap
- Protovis.JS
- Size: Amount
- Brightness: Variance
- Color: Sensor
- Shows: Scans – bright spots



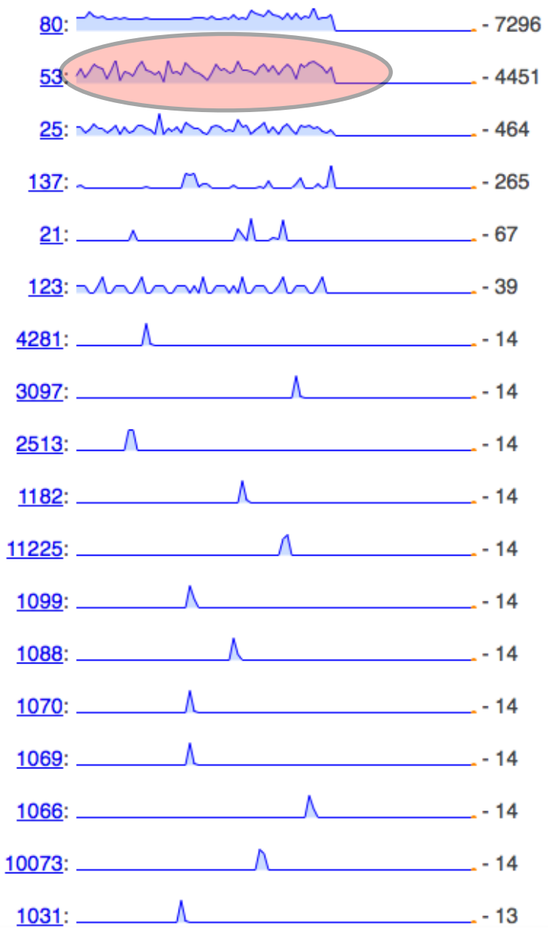
- Thanks to Chris Horsley

Firewall Treemap

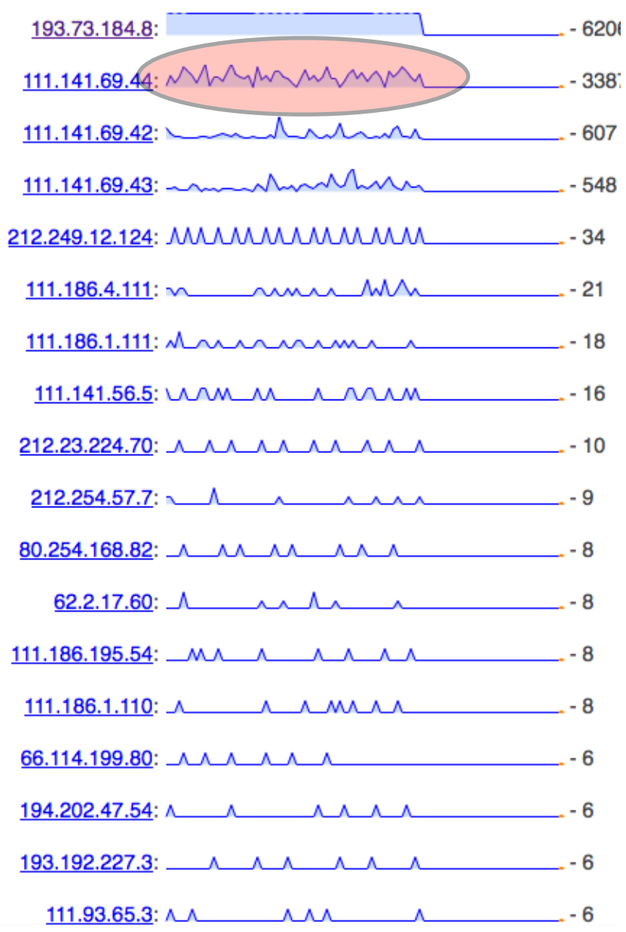


Firewall Log

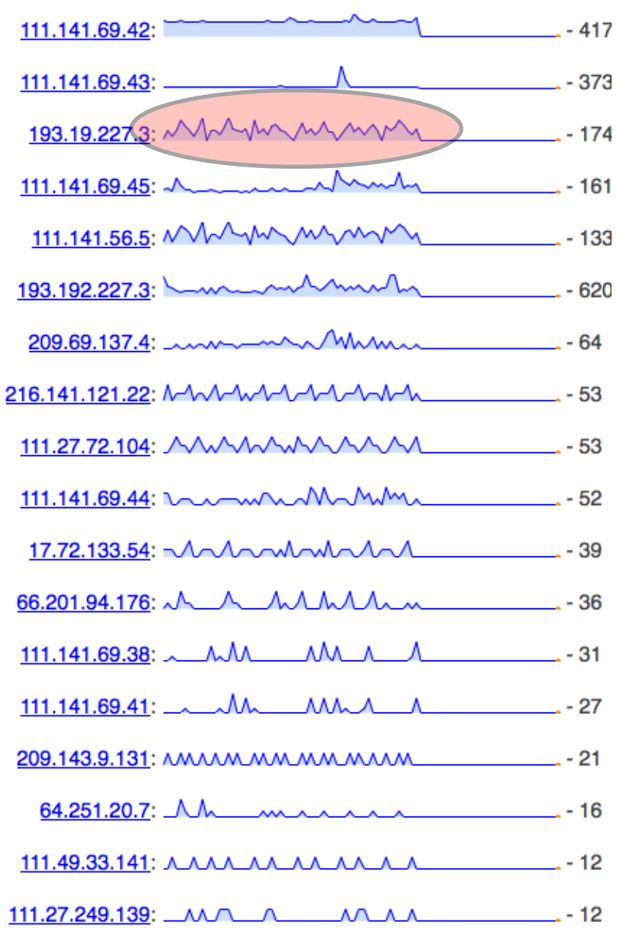
Port



Source IP



Destination IP

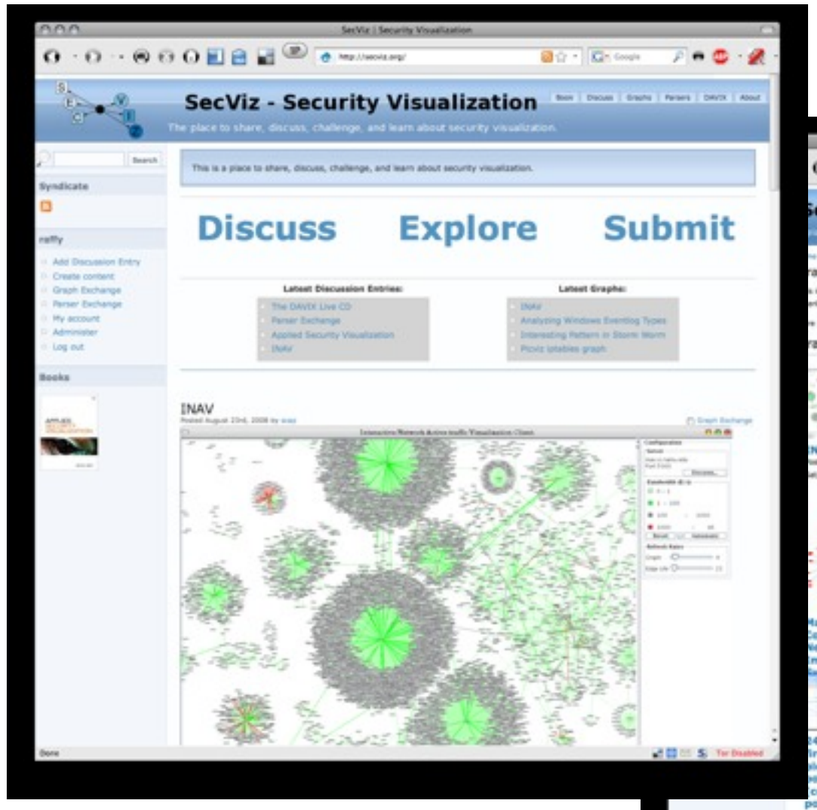


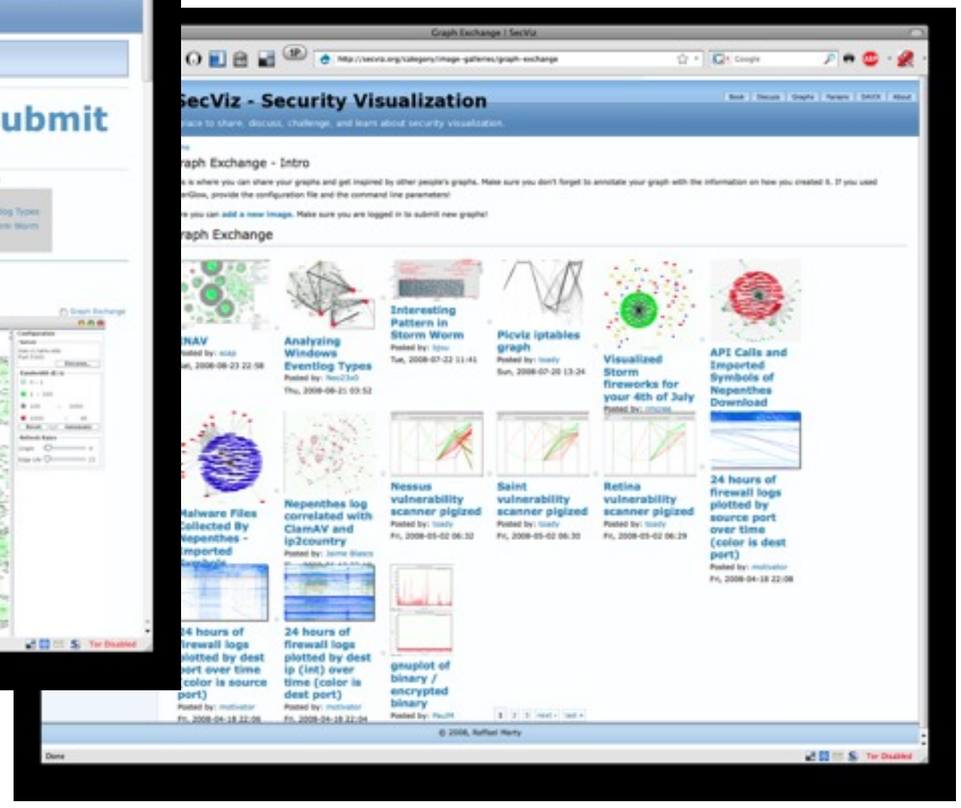
Visualization Resources

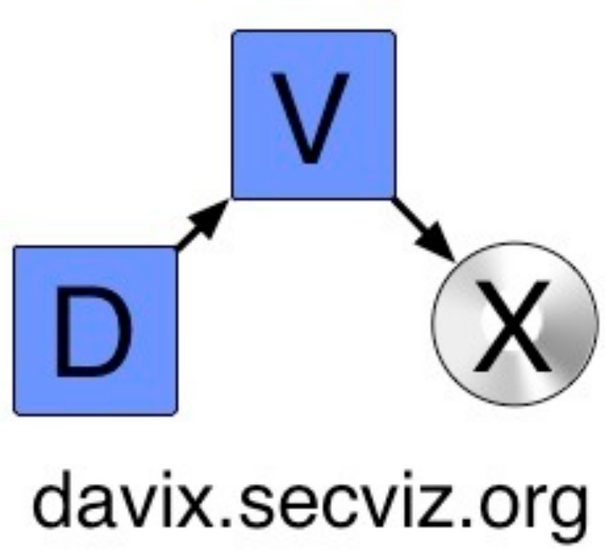
http://secviz.org

Share, discuss, challenge, and learn about security visualization.

- List: secviz.org/mailinglist
- Twitter: **@secviz**

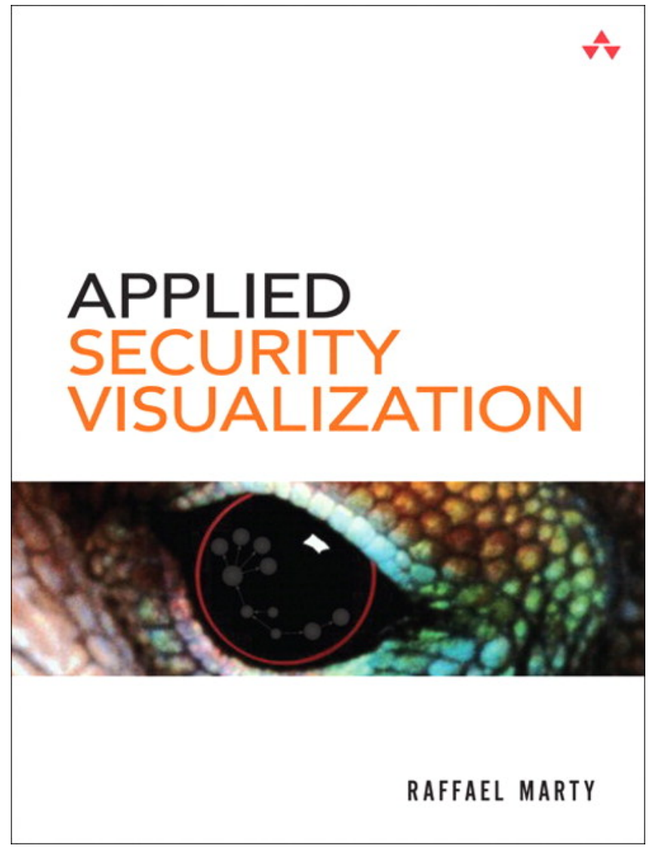






Applied Security Visualization

- Bridging the gap between security and visualization
- Hands-on, end to end examples
- Data processing and analysis



Chapters

- Visualization
- Data Sources
- From Data to Graphs
- Perimeter Threat
- Compliance
- Insider Threat
- Visualization Tools

Addison Wesley (August, 2008)
ISBN: 0321510100

Thank You!

raffael.marty@loggly.com
@zrlram