Alternatives in Analysis

Security Analytics Project

Mark Ryan del Moral Talabis
Secure-DNA

Agenda

- High-level overview of the analysis techniques out there
- To help you get started with YOUR analysis and research by introducing you to existing tools
- Tip of the iceberg – this will be FAST..

The Security Analytics

Security

Security Analytics

GOAL: Look for new and alternative ways to analyze security data

Data Analysis

[redacted]

A Data-Centric World

- As security data collection tools continue to improve and evolve, the quantity of data that we collect increases exponentially
  - Honeypots and Honeynets
  - Malware Collectors
  - Honeyclients
  - Firewall
  - IDS/IPS
  - System/Network devices

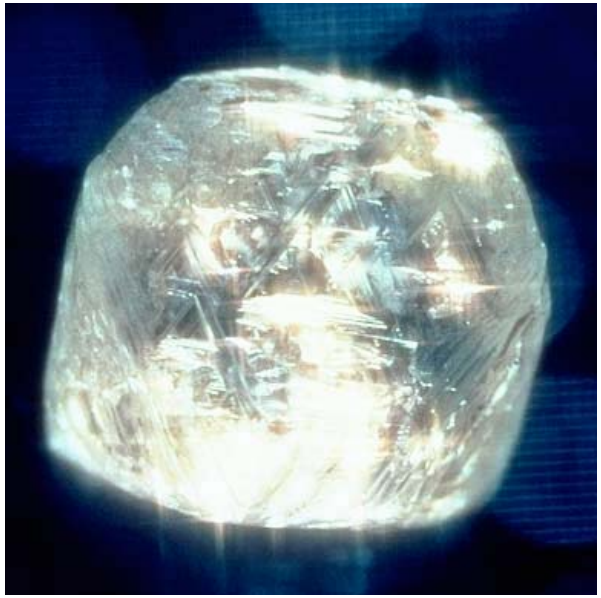
The Aftermath

- After the cool tools what remains are tons and tons of data to sift through!

The Value of Data

- Data is often only as valuable as what the analysis can shape it into.

The Value of Data



Analysis

[redacted]

Building our Arsenal

- Time to build up our arsenal of analysis
  - Tools
  - Techniques
- How? Where?

Looking Beyond Security

- Though security in itself is a unique field with unique needs, analysis techniques often span the boundaries of different disciplines

Looking Beyond Security

[redacted]

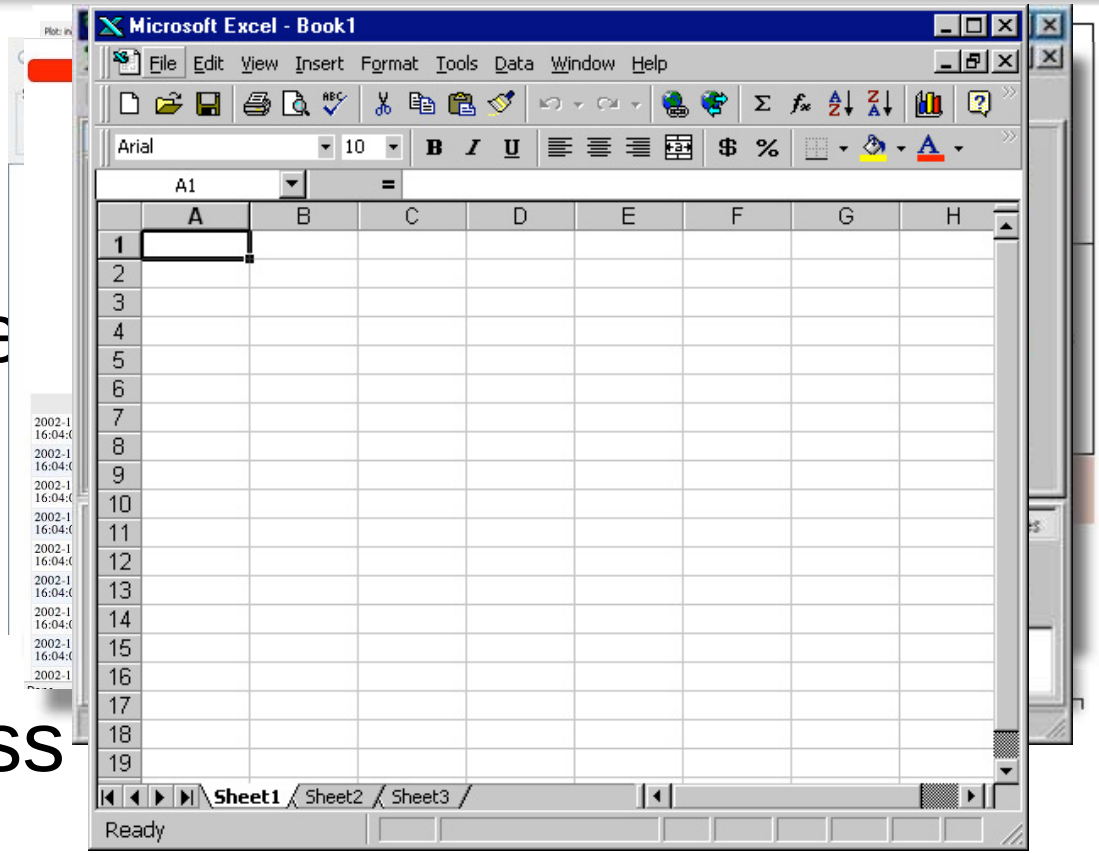
Our Analytics Arsenal

- Techniques
  - Data and Text Mining
  - Clustering
  - Machine Learning
  - Baselining
  - Visualization
  - Behavioral Analysis
  - Game Theory

[redacted]

Data Analysis Tools

- R-Project
- Weka
- Yale (RapidMine
- Tanagra
- FlowTag
- Honeysnap
- Excel and Access
- Orange

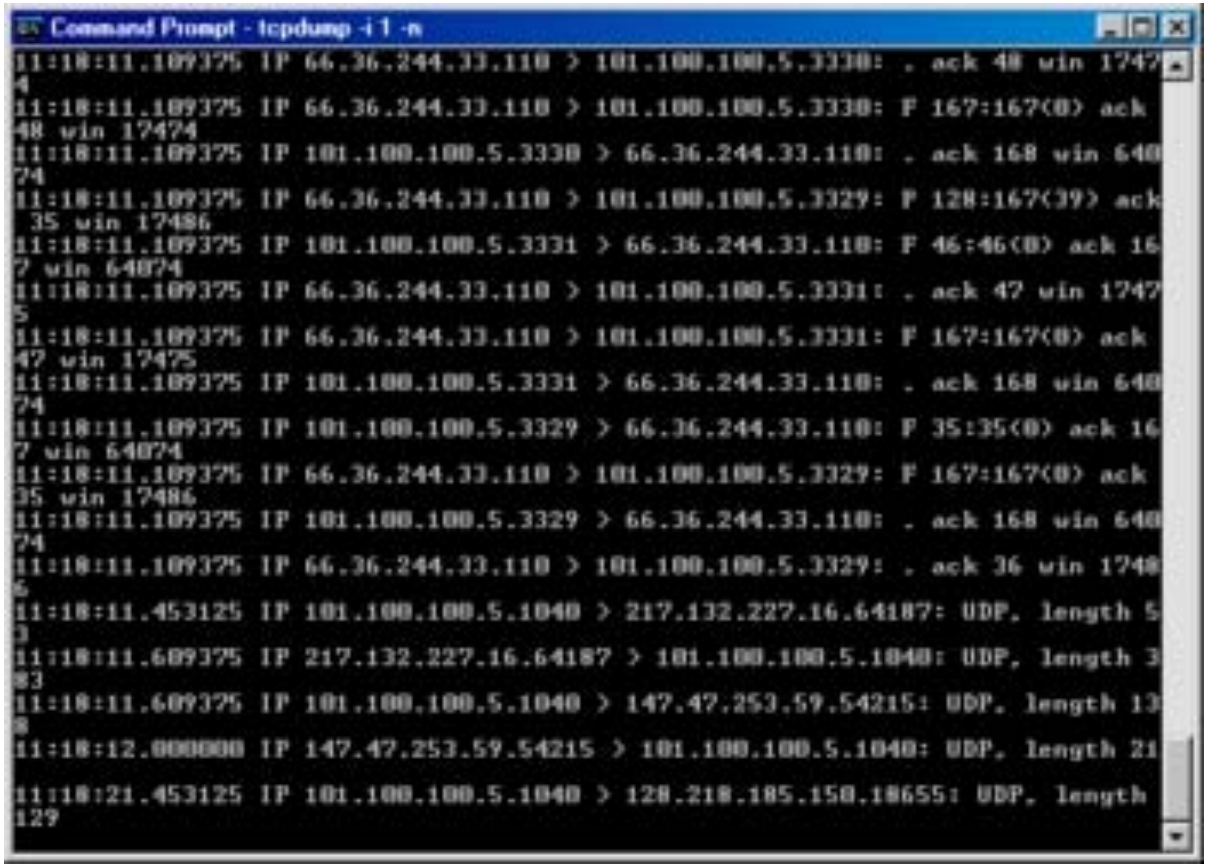


The Possibilities

Preprocessing and Data Cleansing

- Creating a ‘first-cut’ for further analysis
- New Stuff! **Honeysnap**
  - The Honeynet Project
  - Arthur Clune, UK Honeynet Project

What we're used to…



What it could be… (Honeysnap)

Login

Summary | Flow Details | Sebek Details | IRC Summary | IRC Details | IP Summary | IP Lookup |

Text
From
To
Command
IP Source
IP Destination
Port
Honeypot
Start time 1990/01/01 00:00 Choose
End time 2007/06/04 22:03 Choose
Search
Hide Search Form

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 >>>

Time	Honeypot	Source	Destination	Port	From	To	Command	Text
2002-11-29 16:04:08.120687	HS_Fake	80.117.14.44	192.168.100.28	7000	80.117.14.44	fargetta	pass	
2002-11-29 16:04:08.700647	HS_Fake	192.168.100.28	80.117.14.44	7000	welcome!psybnc@lam3rz.de	*	privnotice	psyBNC2.2.1
2002-11-29 16:04:08.700647	HS_Fake	80.117.14.44	192.168.100.28	7000	80.117.14.44	ahaa	user	"bobz" "192.168.100.28" □-:OwNz:-□
2002-11-29 16:04:08.700647	HS_Fake	80.117.14.44	192.168.100.28	7000	80.117.14.44	dj`bobz`	nick	
2002-11-29 16:04:08.780642	HS_Fake	192.168.100.28	80.117.14.44	7000	irc.psychoid.net	dj`bobz`	privnotice	psyBNC 2.2.1 Help (* = BounceAdmin only)
2002-11-29 16:04:08.780642	HS_Fake	192.168.100.28	80.117.14.44	7000	irc.psychoid.net	dj`bobz`	privnotice	BHELP SETLEAVEMSG - Sets your Leave-MSG when you leave
2002-11-29 16:04:08.780642	HS_Fake	192.168.100.28	80.117.14.44	7000	irc.psychoid.net	dj`bobz`	privnotice	BHELP DELOP - Deletes an added User who got Op
2002-11-29 16:04:08.780642	HS_Fake	192.168.100.28	80.117.14.44	7000	irc.psychoid.net	dj`bobz`	privnotice	BHELP LISTOPS - Lists all added Ops
2002-11-29	HS_Fake	192.168.100.28	80.117.14.44	7000	irc.psychoid.net	dj`bobz`	privnotice	BHELP LEAVEQUIT - If set to 1, parts all channels on

Data and Text Mining

- Data mining is the process of automatically searching large volumes of data for patterns
- Text mining is the process of deriving high quality information from text.
- Applications:
  - Forensic Analysis
  - Log analysis
  - IRC analysis
- Sample research:
  - Topical Analysis of IRC hacker chatter through text mining

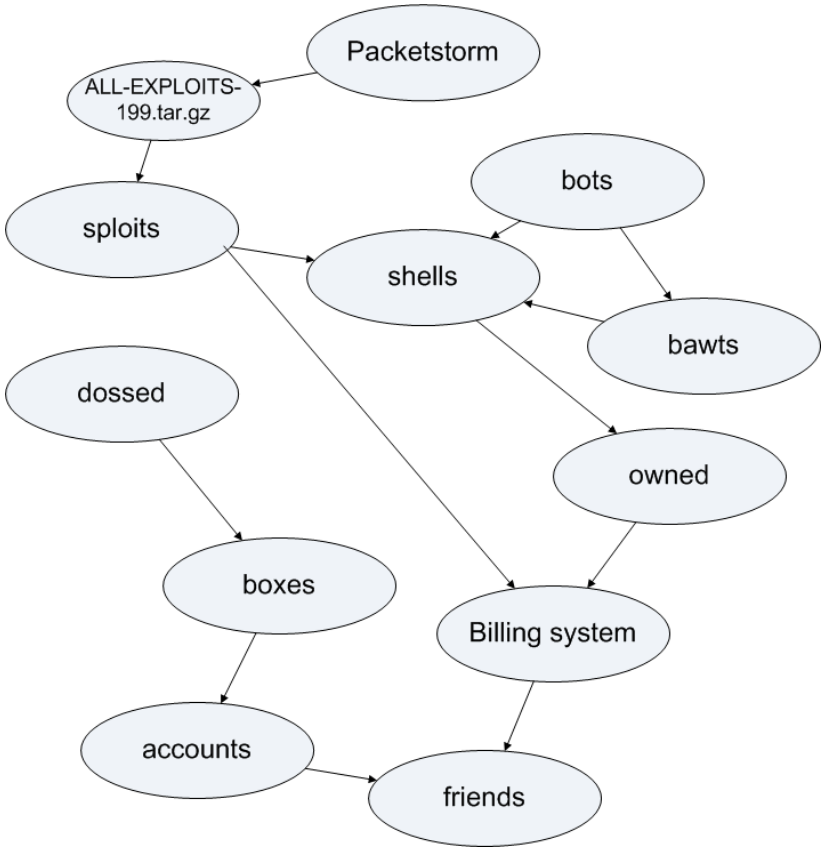
[redacted]

What we're used to...

:D1ck :yo
:J4n3 :i downloaded a fle from packetstorm
:J4n3 :name was ALL-EXPLOITS-1999
:D1ck :yep?
:J4n3 :6 mb fle
:J4n3 :ALL-EXPLOITS-199.tar.gz
:J4n3 :ALL-EXPLOITS-1999.tar.gz
:J4n3 :too many sploits in them
:J4n3 :it made 10 folders
:J4n3 :evey folder contain different sploits
:D1ck :ok and?
:J4n3 :i mean to say u also download it
:J4n3! :yaar that synf bod is tight
:J4n3! :u know some hackphreak guy took over deathace's nick 2 weeks ago
:J4n3! :with his bot with ip *
:D1ck! :YEP
:D1ck! :yep in know i dossed him 2 times
:D1ck! :he is linuxs ka guy
:D1ck! :;)

[redacted]

What it could be… (Word Vector)



[redacted]

Behavioral Analysis

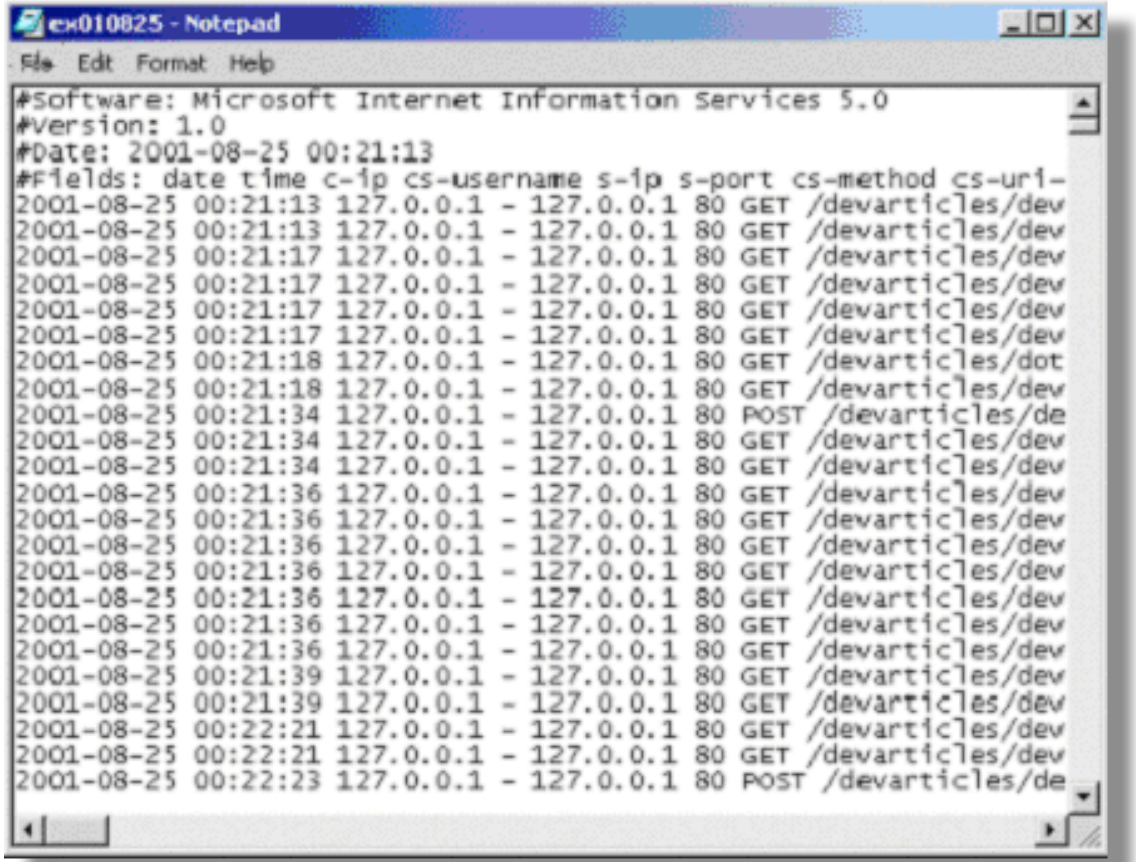
- Study of human behaviour
- Perfect for:
  - Analysis hacker behavior and motivation
- Sample research:
  - Study of hacker motivations through IRC hacker chatter

[redacted]

Clustering

- Classification of objects into different groups, so that the data in each group (ideally) share some common trait
- Perfect for:
  - Classification of Attacks
  - Malware Taxonomy
  - Finding deviations from logs
- Sample application:
  - Classifying Attacks Using K-M[redacted]

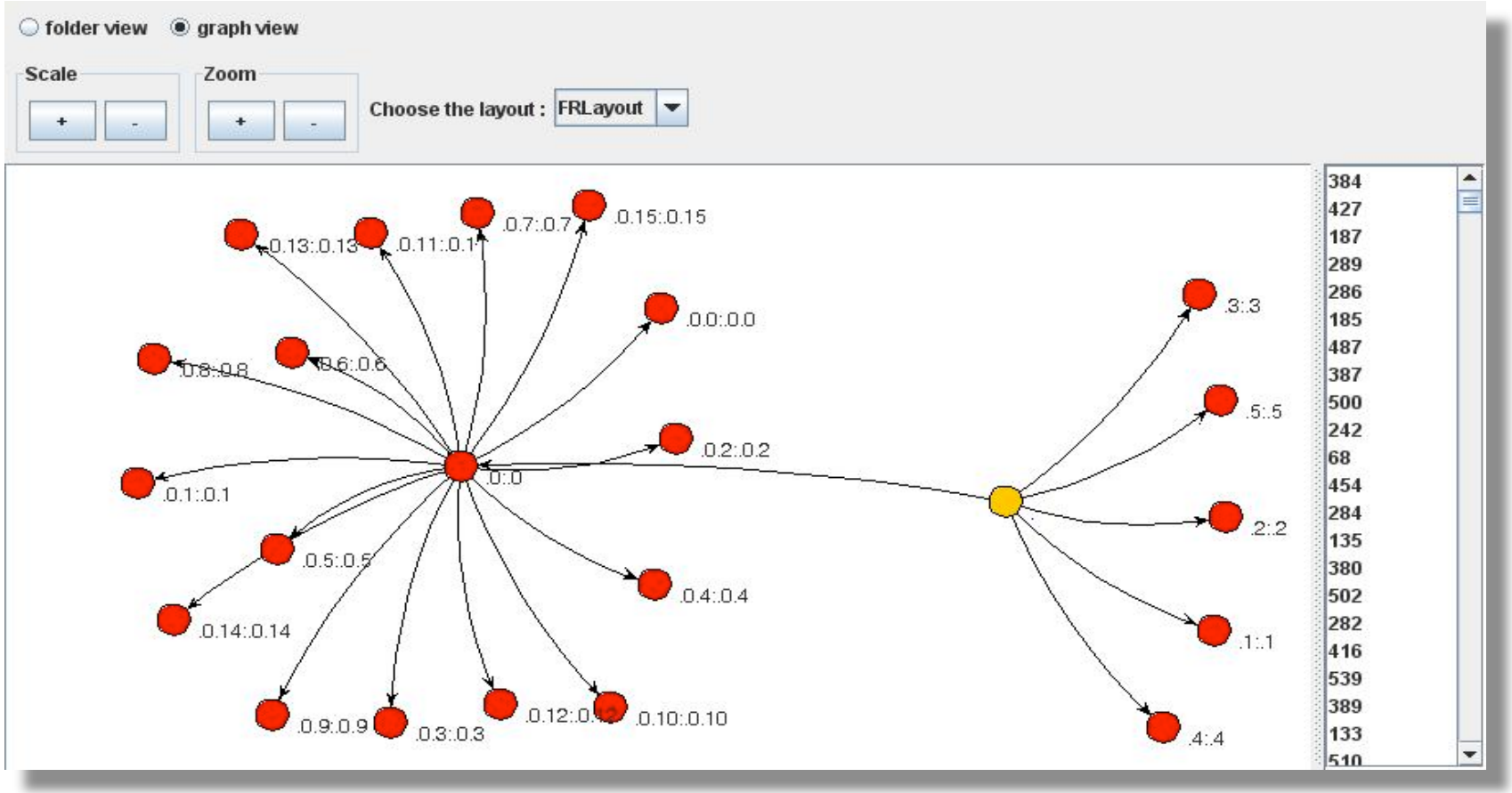
What we're used to…



```
#Software: Microsoft Internet Information Services 5.0
#Version: 1.0
#Date: 2001-08-25 00:21:13
#Fields: date time c-ip cs-username s-ip s-port cs-method cs-uri-
2001-08-25 00:21:13 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:13 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:17 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:17 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:17 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:17 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:18 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dot
2001-08-25 00:21:18 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:34 127.0.0.1 - 127.0.0.1 80 POST /devarticles/de
2001-08-25 00:21:34 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:34 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:36 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:36 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:36 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:36 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:36 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:36 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:36 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:39 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:39 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:22:21 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:22:21 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:22:23 127.0.0.1 - 127.0.0.1 80 POST /devarticles/de
```

[redacted]

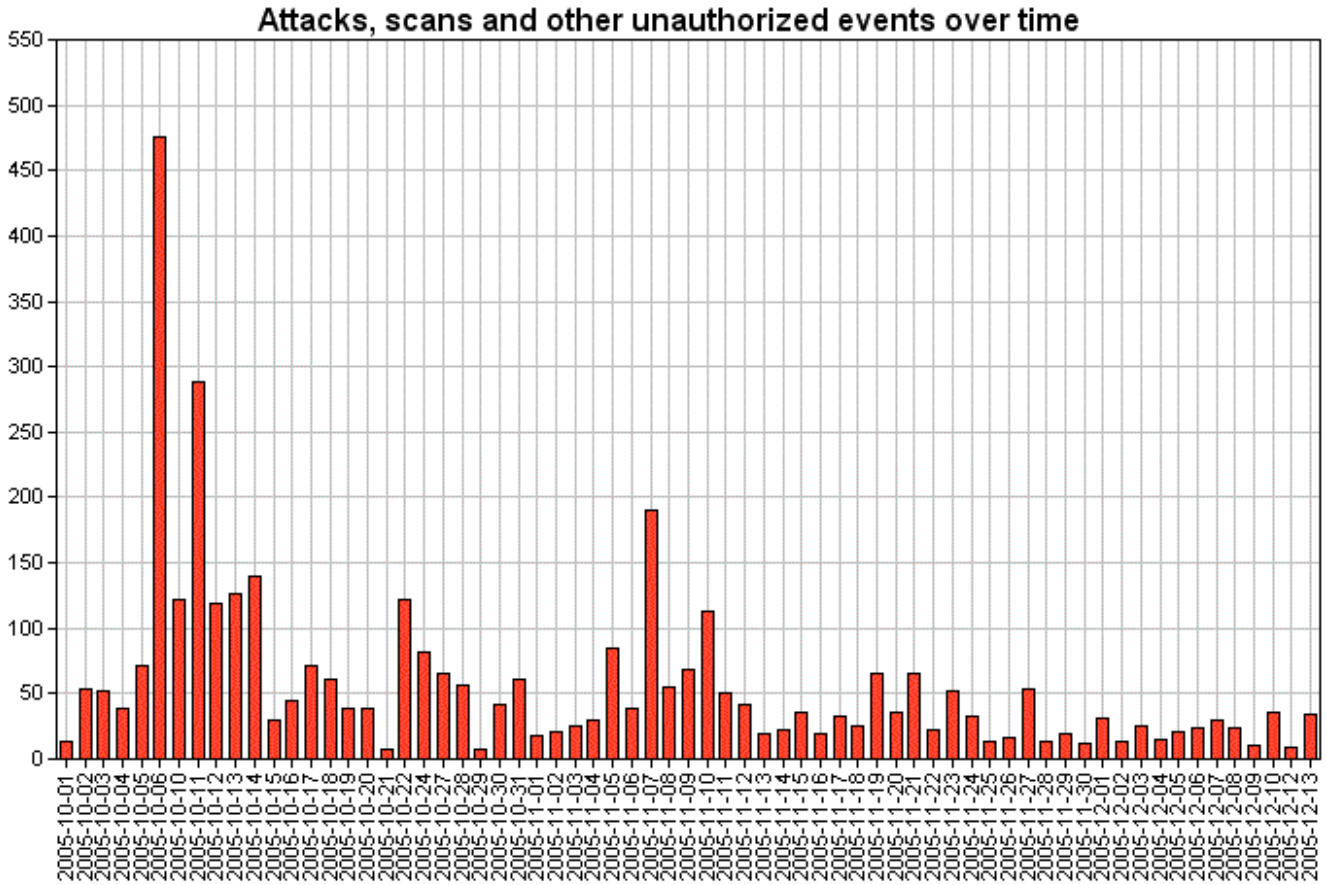
What it could be… (YALE)



Statistics

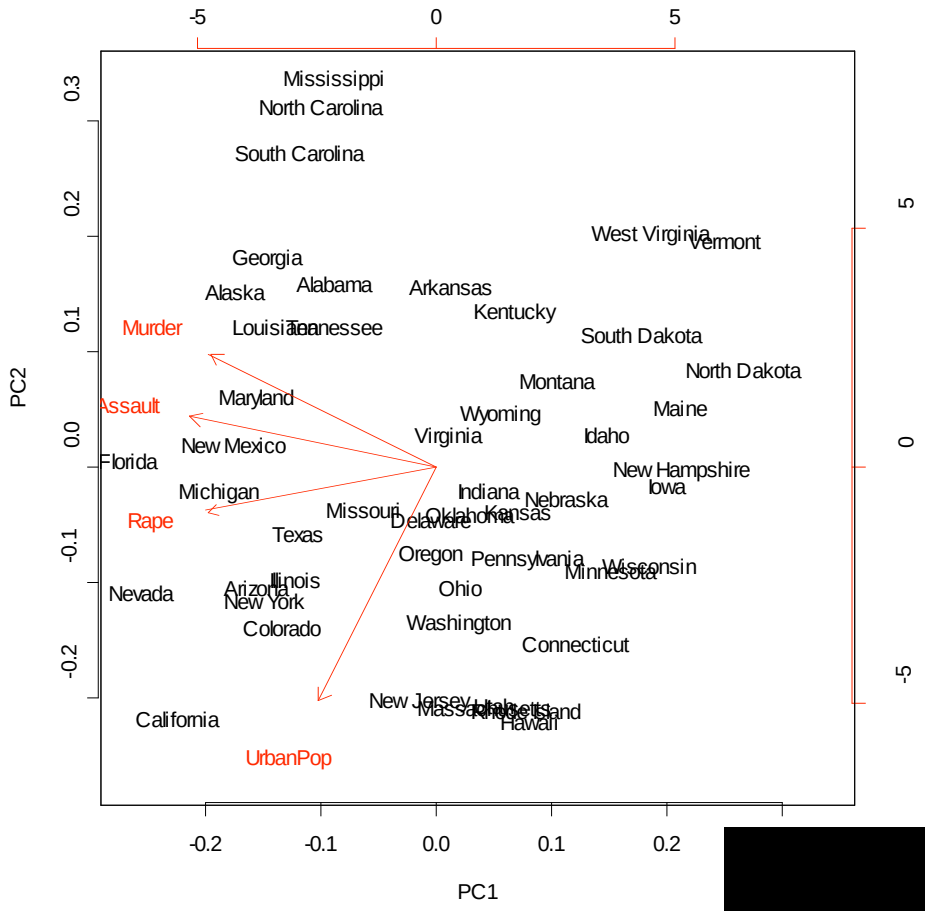
- Pertains to the collection, analysis, interpretation or explanation, and presentation of data.
- Perfect for:
  - Executives love stats
  - Baselines

What we're used to…

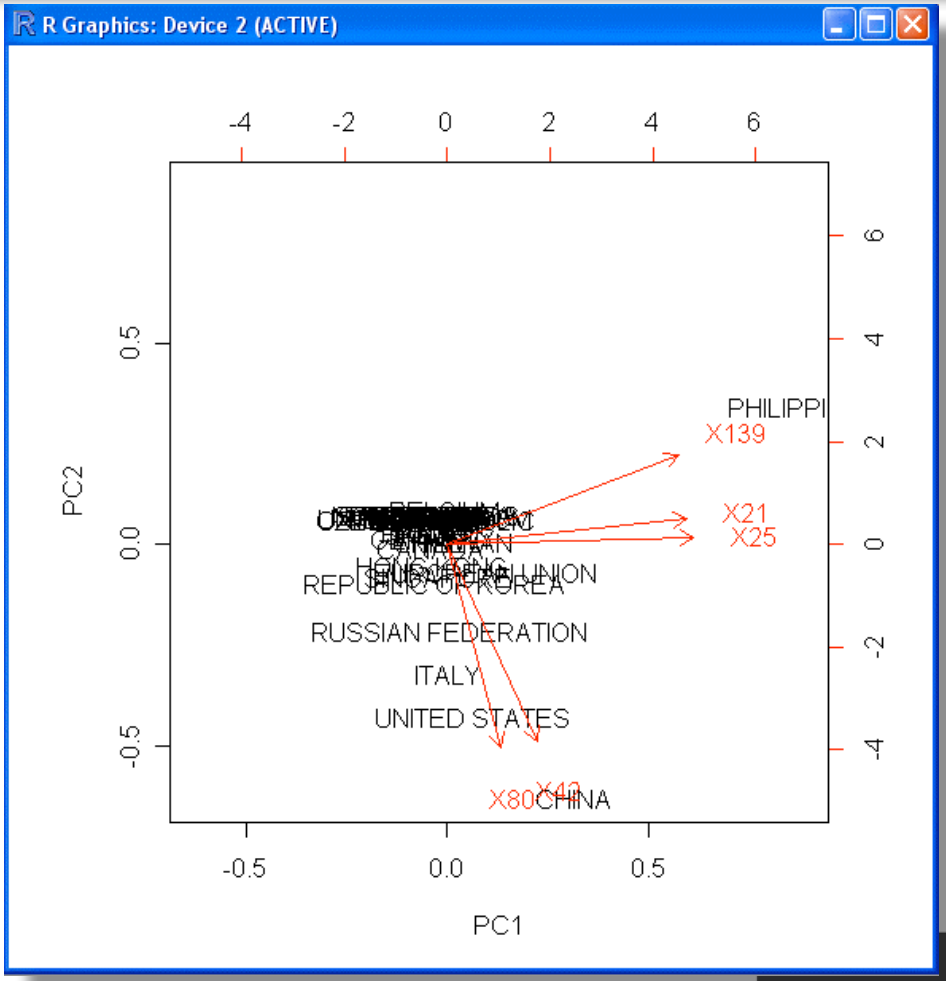


[redacted]

From Criminology (R-Project)



What it could be…



Genetics and Immune Concepts

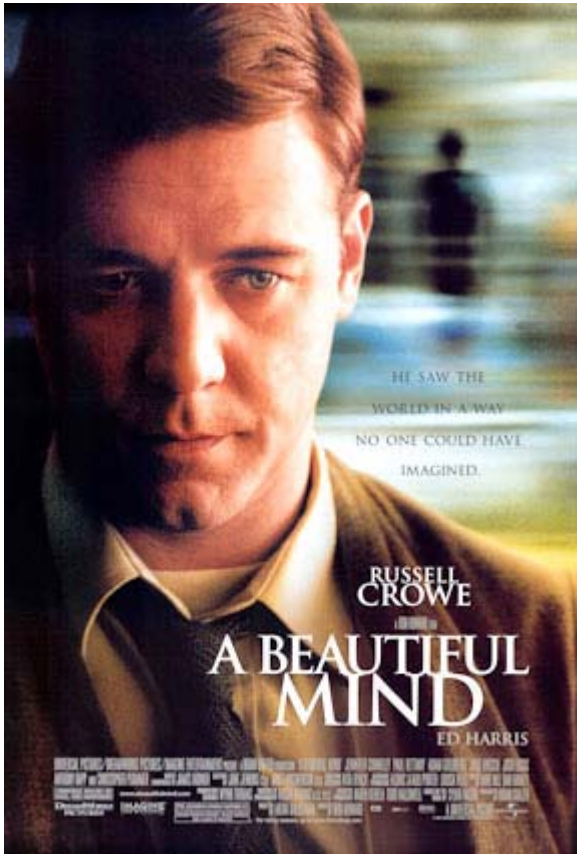
- Applications:
  - Analyzing and defending against attacks
  - Imitate defenses of the human body
- Sample research:
  - Code Breaking using Genetic Algorithm
  - Genetic Algorithm Approach for Intrusion Detection

Economic Theories

- Economics takes a lot from mathematics, statistics and other disciplines
- Perfect for:
  - All sorts of stuff
- Sample research:
  - Game Theory and Hacker Behaviour

[redacted]

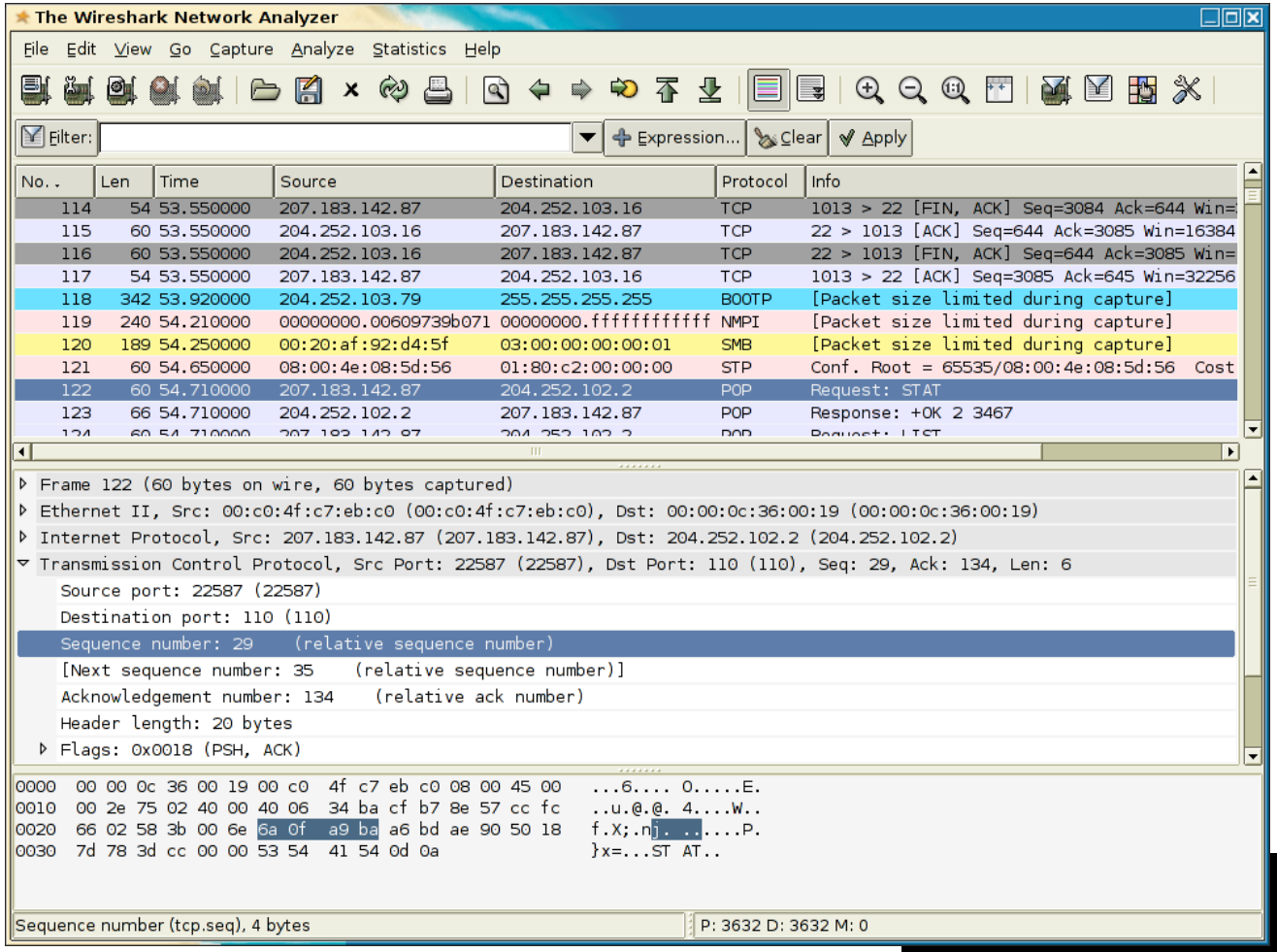
Game Theory


HE SAW THE
WORLD IN A WAY
NO ONE COULD HAVE
IMAGINED.
RUSSELL
CROWE
A BEAUTIFUL
MIND
ED HARRIS

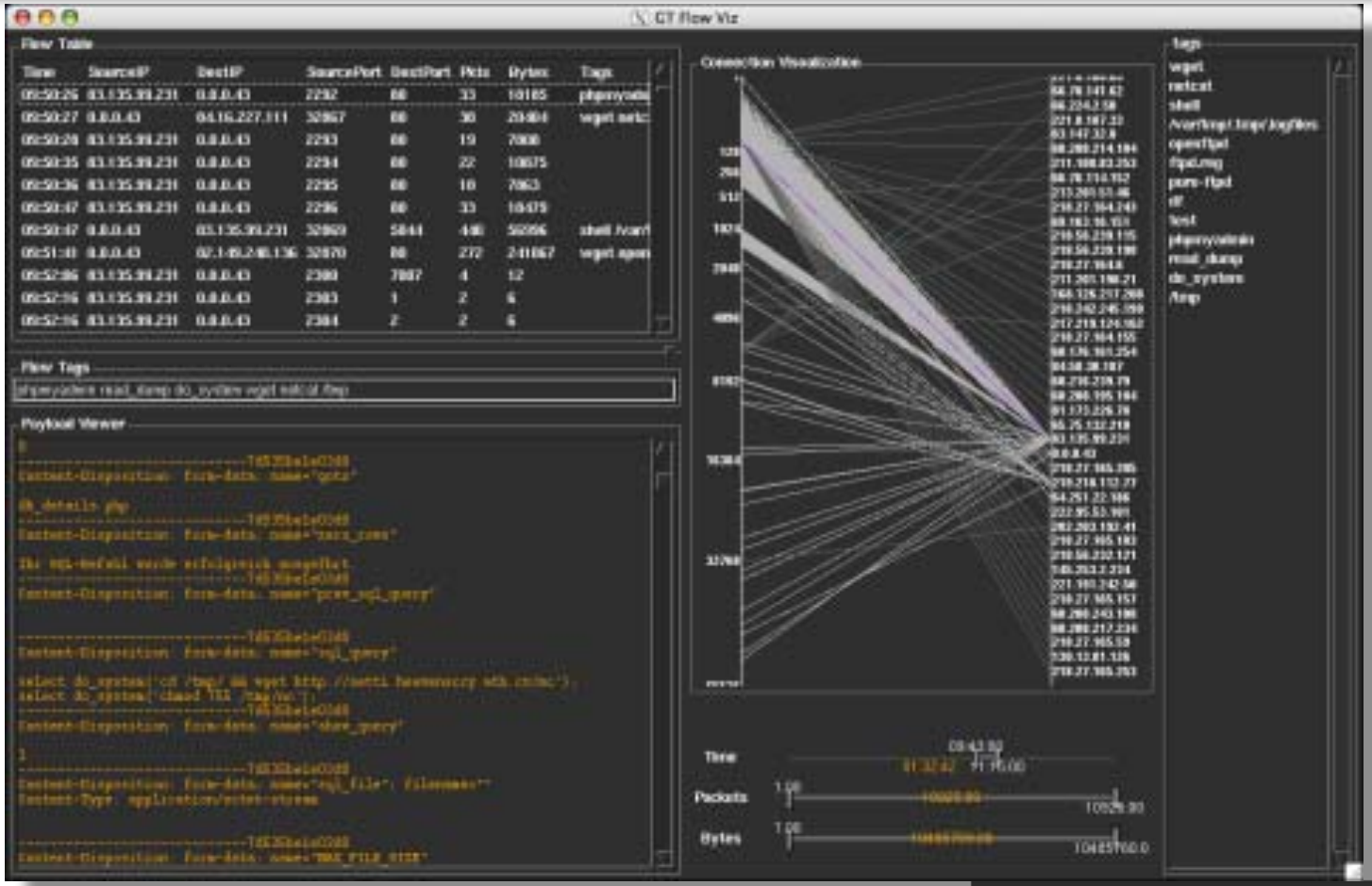
Visualization

- Picture paints a thousand words
- Perfect for:
  - Attack detection and analysis
- New Stuff! **FlowTag**
  - Visual tagging
  - Chris Lee, Georgia Tech

What we're used to…



What it could be… (FlowTag)



Conclusion

- High level overview of analysis tools and techiniqes
- Made you aware that there are a lot of things to use out there
- To produce good results techniques and tools could be used together

A Collaborative Effort

- A forum where people from different fields can share data and techniques
- Diversity is the Key! Everyone is welcome!
- Feel free to talk to me more about this stuff at: ryan@secure-dna.com

Thank You

Secure-DNA

Machine Learning

- Machine learning is concerned with the design and development of algorithms and techniques that allow computers to "learn"
- Useful for:
  - Predicting Attacks
  - Self-learning IDS
- Sample research:
  - Predicting attacks using Supp[redacted] Machines

Predicting Attacks