

Alternatives in Analysis

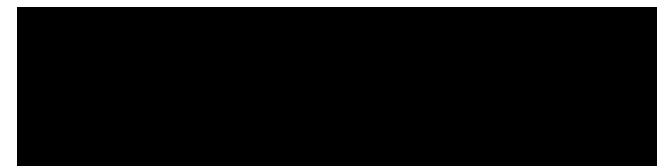
Security Analytics Project

Mark Ryan del Moral
Talabis

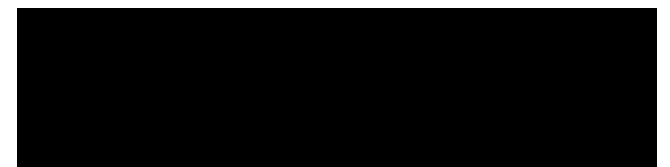
Secure-DNA

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..

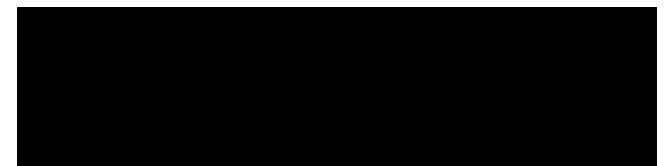


The Security Analytics



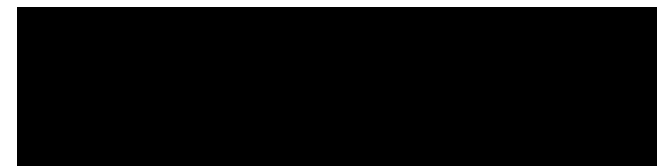
A Data-Centric World

- 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



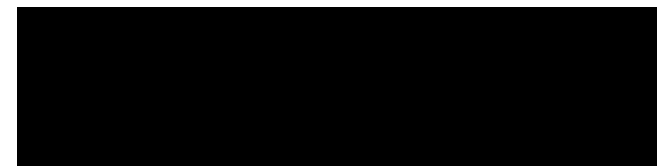
The Aftermath

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

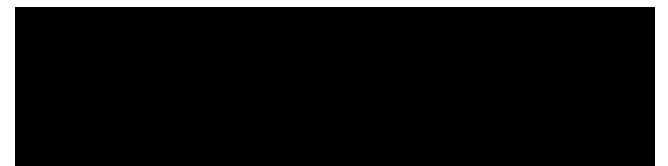
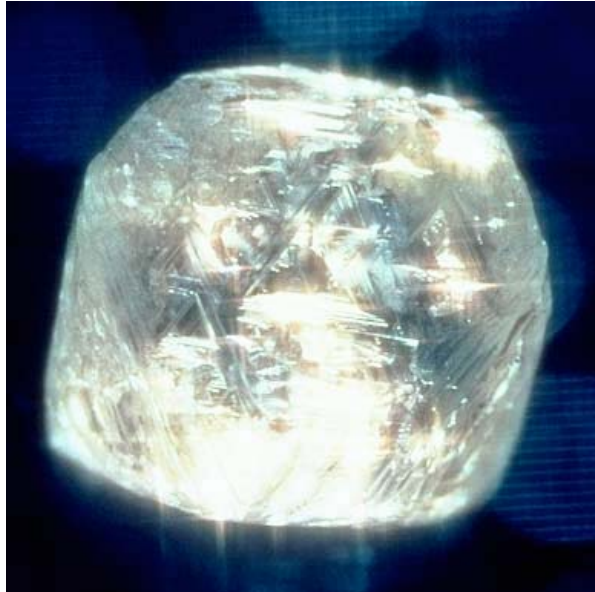


The Value of Data

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

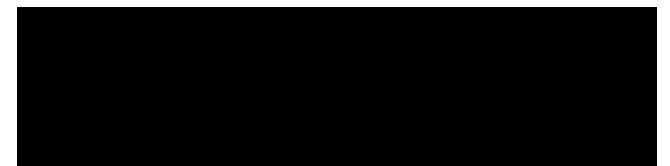


The Value of Data



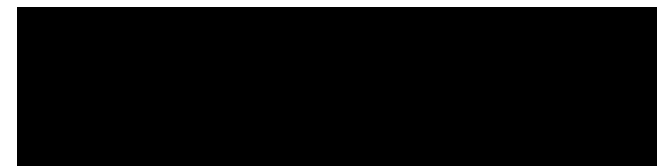
Building our Arsenal

- 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

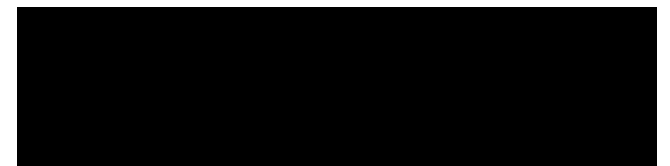


Looking Beyond Security



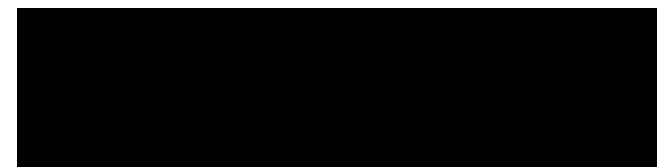
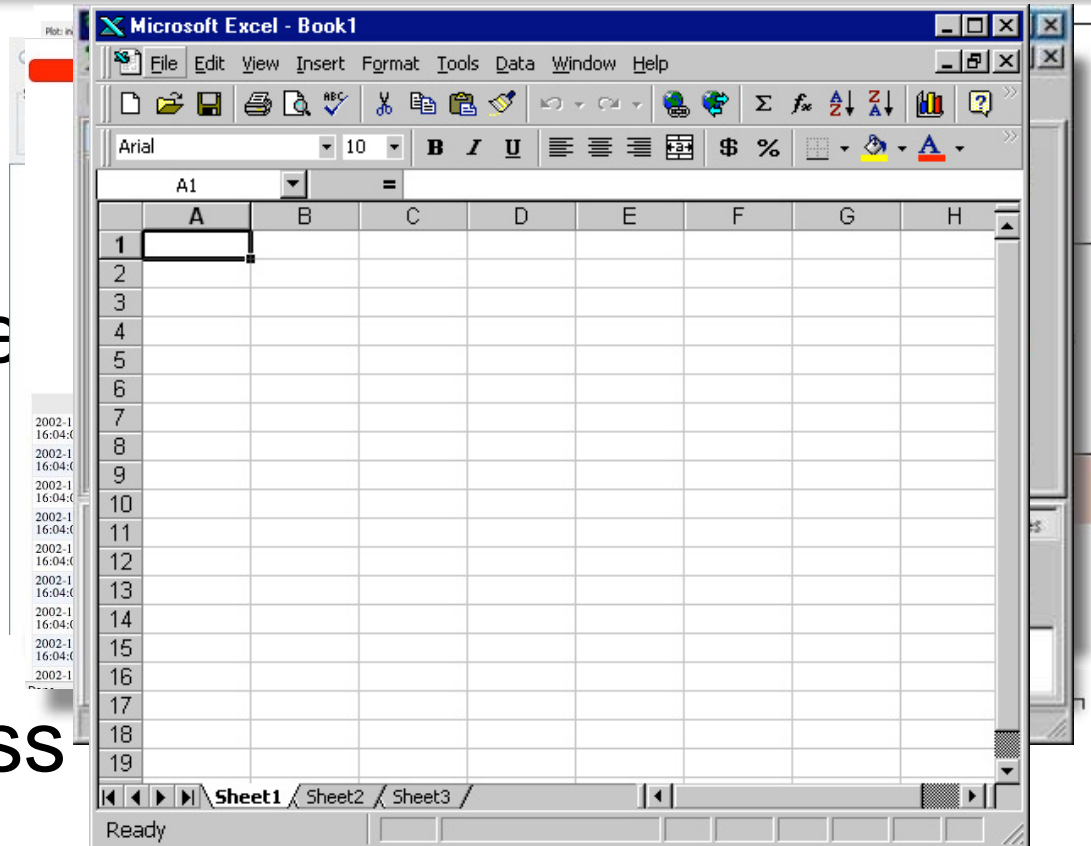
Our Analytics Arsenal

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



Data Analysis Tools

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



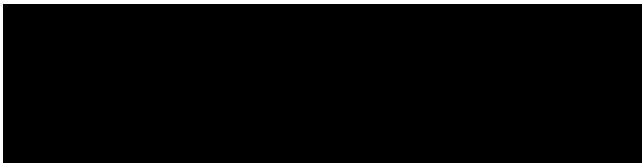
The Possibilities

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...

```
Command Prompt - tcpdump -i 1 -n
11:18:11.109375 IP 66.36.244.33.110 > 101.100.100.5.3330: . ack 48 win 17474
11:18:11.109375 IP 66.36.244.33.110 > 101.100.100.5.3330: F 167:167(0) ack 48 win 17474
11:18:11.109375 IP 101.100.100.5.3330 > 66.36.244.33.110: . ack 168 win 64074
11:18:11.109375 IP 66.36.244.33.110 > 101.100.100.5.3329: F 128:167(39) ack 35 win 17486
11:18:11.109375 IP 101.100.100.5.3331 > 66.36.244.33.110: F 46:46(0) ack 167 win 64074
11:18:11.109375 IP 66.36.244.33.110 > 101.100.100.5.3331: . ack 47 win 17475
11:18:11.109375 IP 66.36.244.33.110 > 101.100.100.5.3331: F 167:167(0) ack 47 win 17475
11:18:11.109375 IP 101.100.100.5.3331 > 66.36.244.33.110: . ack 168 win 64074
11:18:11.109375 IP 101.100.100.5.3329 > 66.36.244.33.110: F 35:35(0) ack 167 win 64074
11:18:11.109375 IP 66.36.244.33.110 > 101.100.100.5.3329: F 167:167(0) ack 35 win 17486
11:18:11.109375 IP 101.100.100.5.3329 > 66.36.244.33.110: . ack 168 win 64074
11:18:11.109375 IP 66.36.244.33.110 > 101.100.100.5.3329: . ack 36 win 17486
11:18:11.453125 IP 101.100.100.5.1040 > 217.132.227.16.64187: UDP, length 53
11:18:11.609375 IP 217.132.227.16.64187 > 101.100.100.5.1040: UDP, length 83
11:18:11.609375 IP 101.100.100.5.1040 > 147.47.253.59.54215: UDP, length 88
11:18:12.000000 IP 147.47.253.59.54215 > 101.100.100.5.1040: UDP, length 218
11:18:21.453125 IP 101.100.100.5.1040 > 128.218.185.150.10655: UDP, length 129
```



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
 End time

Hide Search Form

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

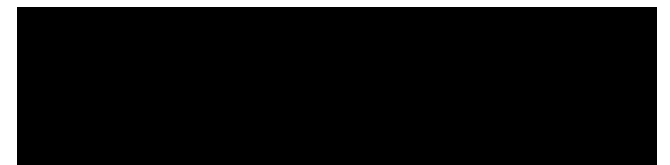
Time	Honeypot	Source	Destination	Port	From	To	Command	Text
2002-11-29 16:04:08.120687	HS_Fake	80.117.14.44	192.168.100.28	7000	80.117.14.44	fargetta	pass	
2002-11-29 16:04:08.700647	HS_Fake	192.168.100.28	80.117.14.44	7000	welcome!psybnc@lam3rz.de	*	privnotice	psyBNC2.2.1
2002-11-29 16:04:08.700647	HS_Fake	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 SETTLEAVEMSG - 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

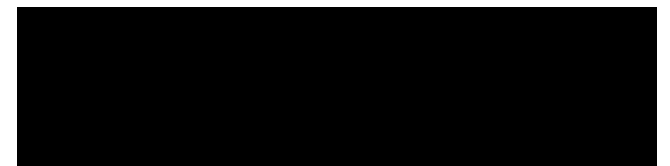
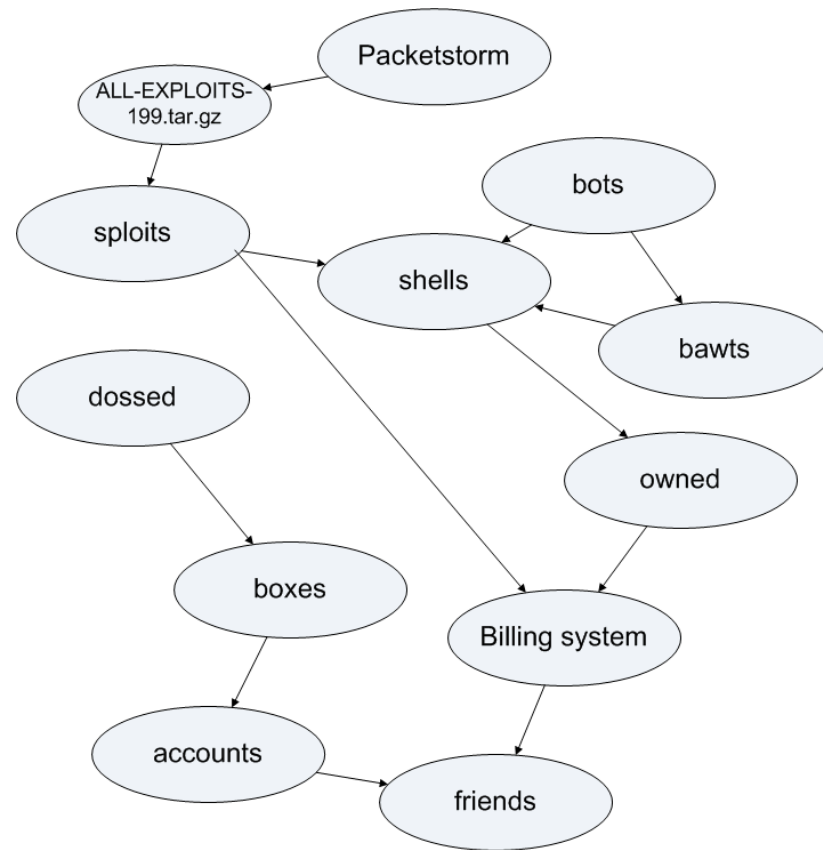
- Data mining is the process of automatically searching large volumes of data for patterns
- Text mining is the process of deriving high quality information from text.
- Applications:
 - Forensic Analysis
 - Log analysis
 - IRC analysis
- Sample research:
 - Topical Analysis of IRC hacker chatter through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

What we're used to...

:D1ck :yo
:J4n3 :i downloaded a file from packetstorm
:J4n3 :name was ALL-EXPLOITS-1999
:D1ck :yep?
:J4n3 :6 mb file
:J4n3 :ALL-EXPLOITS-199.tar.gz
:J4n3 :ALL-EXPLOITS-1999.tar.gz
:J4n3 :too many spoils in them
:J4n3 :it made 10 folders
:J4n3 :every folder contain different spoils
:D1ck :ok and?
:J4n3 :i mean to say u also download it
:J4n3! :yaar that synfbod 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! :;)

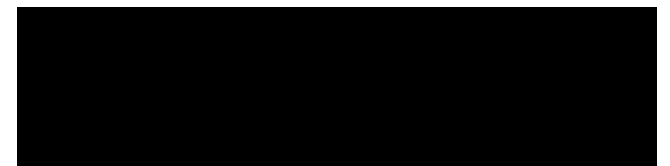


What it could be... (Word Vector)



Behavioral Analysis

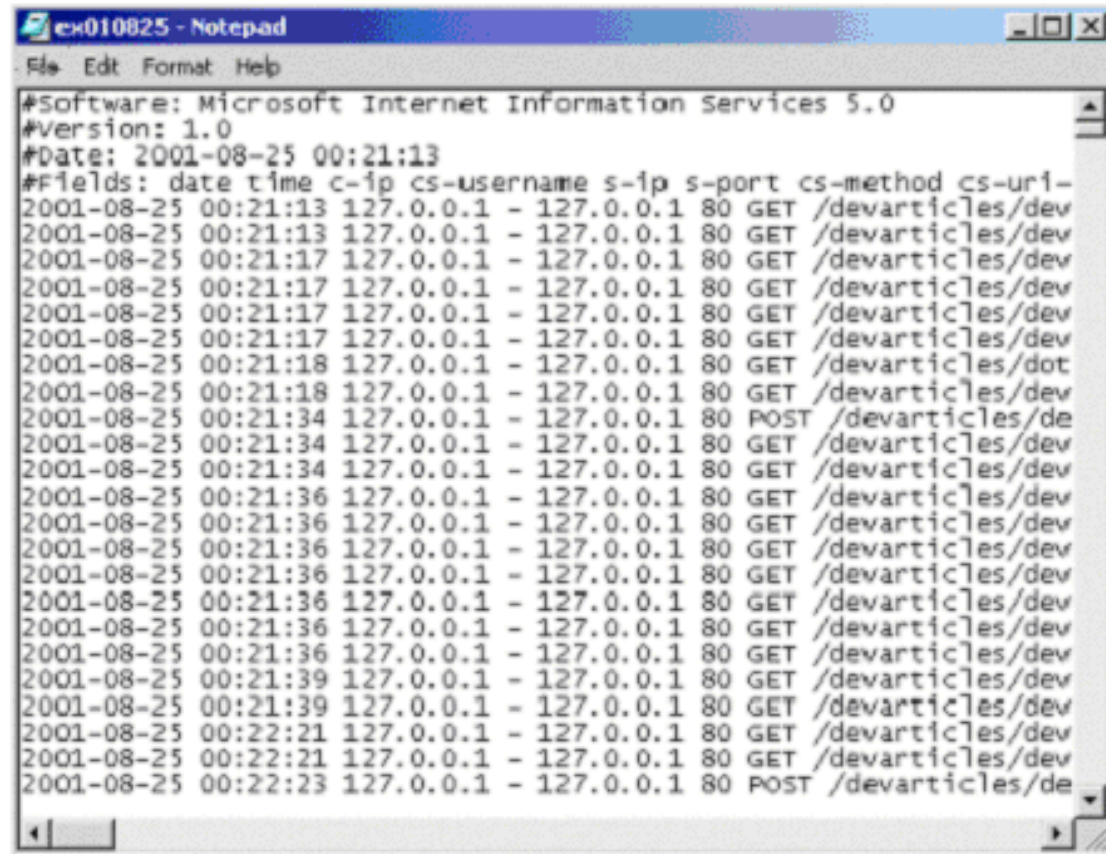
- Study of human behaviour
- Perfect for:
 - Analysis hacker behavior and motivation
- Sample research:
 - Study of hacker motivations through IRC hacker chatter



Clustering

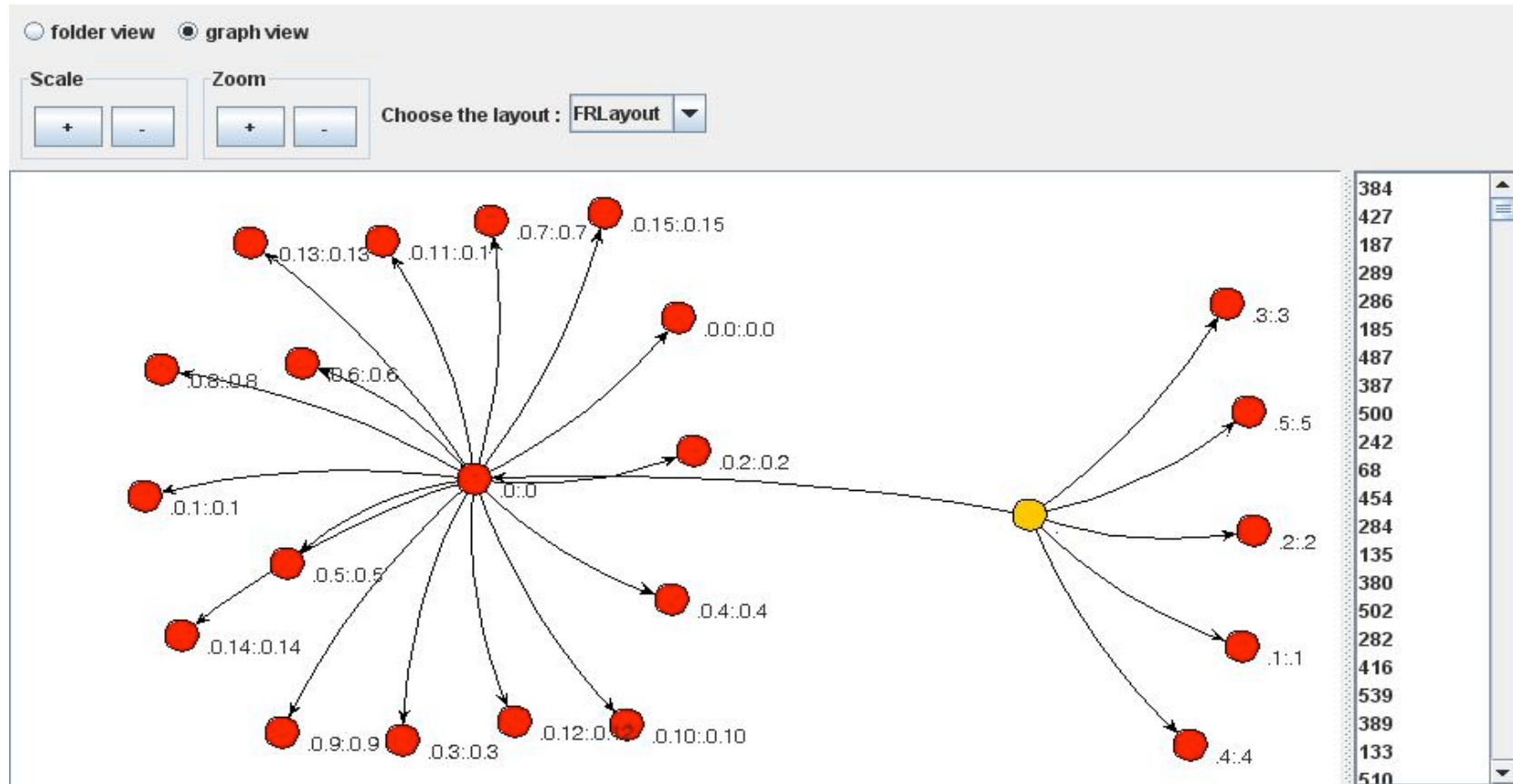
- Classification of objects into different groups, so that the data in each group (ideally) share some common trait
- Perfect for:
 - Classification of Attacks
 - Malware Taxonomy
 - Finding deviations from logs
- Sample application:
 - Classifying Attacks Using K-M

What we're used to...



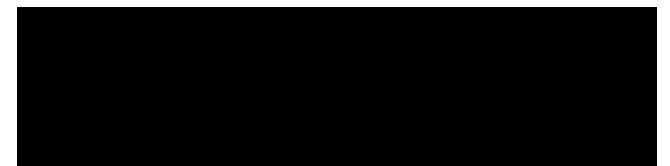
```
ex010825 - Notepad
File Edit Format Help
#Software: Microsoft Internet Information Services 5.0
#Version: 1.0
#Date: 2001-08-25 00:21:13
#Fields: date time c-ip cs-username s-ip s-port cs-method cs-uri-
2001-08-25 00:21:13 127.0.0.1 - 127.0.0.1 80 GET /devarticles/dev
2001-08-25 00:21:13 127.0.0.1 - 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
```

What it could be... (YALE)

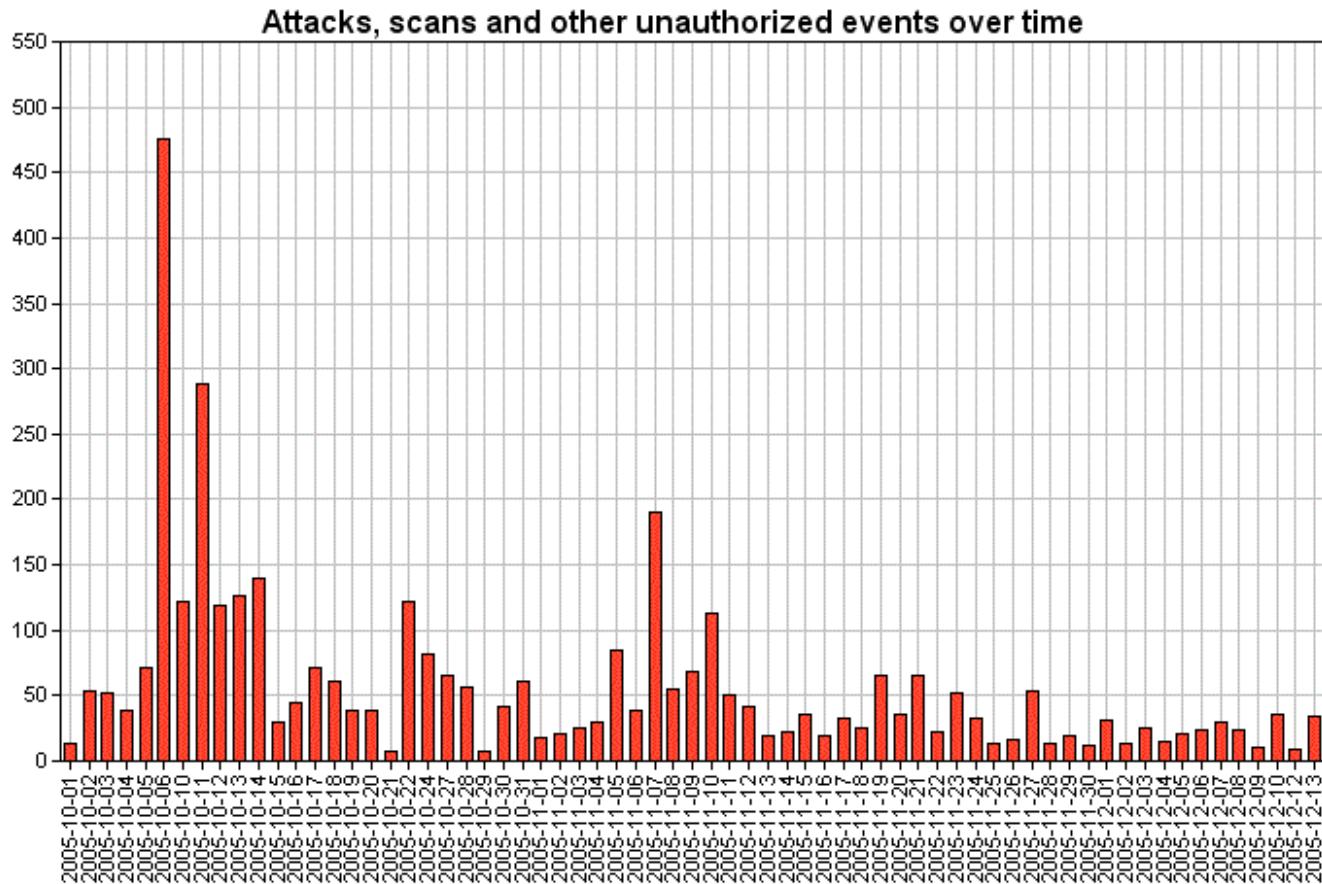


Statistics

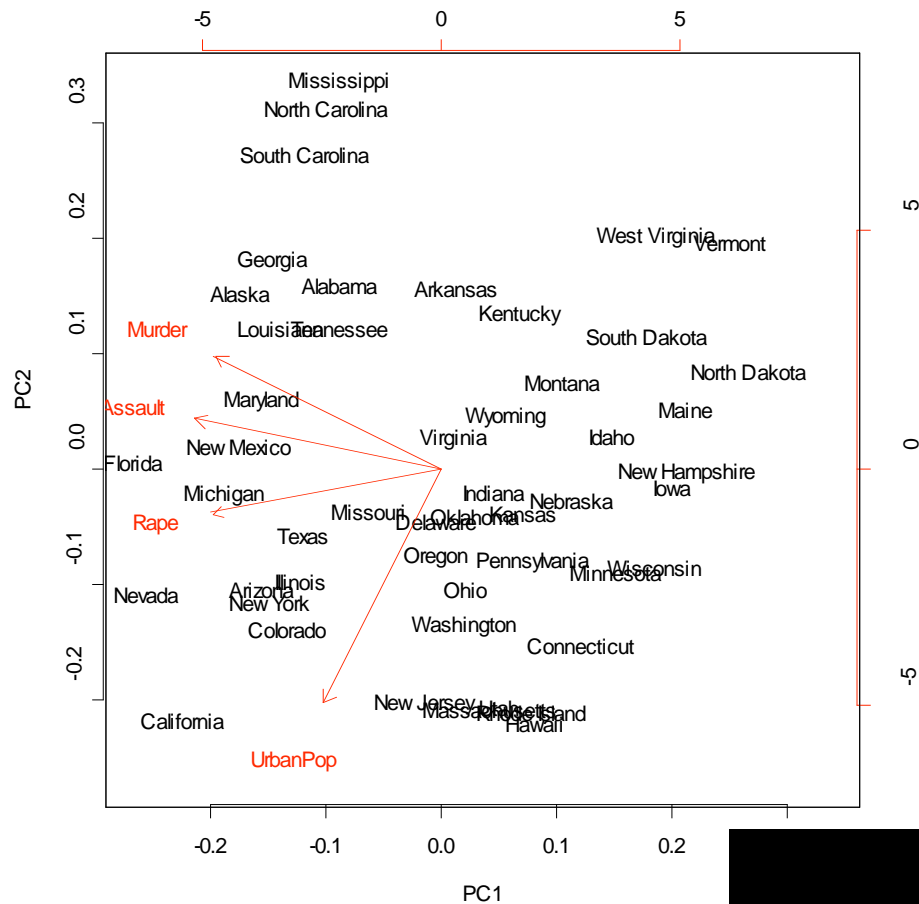
- Pertains to the collection, analysis, interpretation or explanation, and presentation of data.
- Perfect for:
 - Executives love stats
 - Baselines



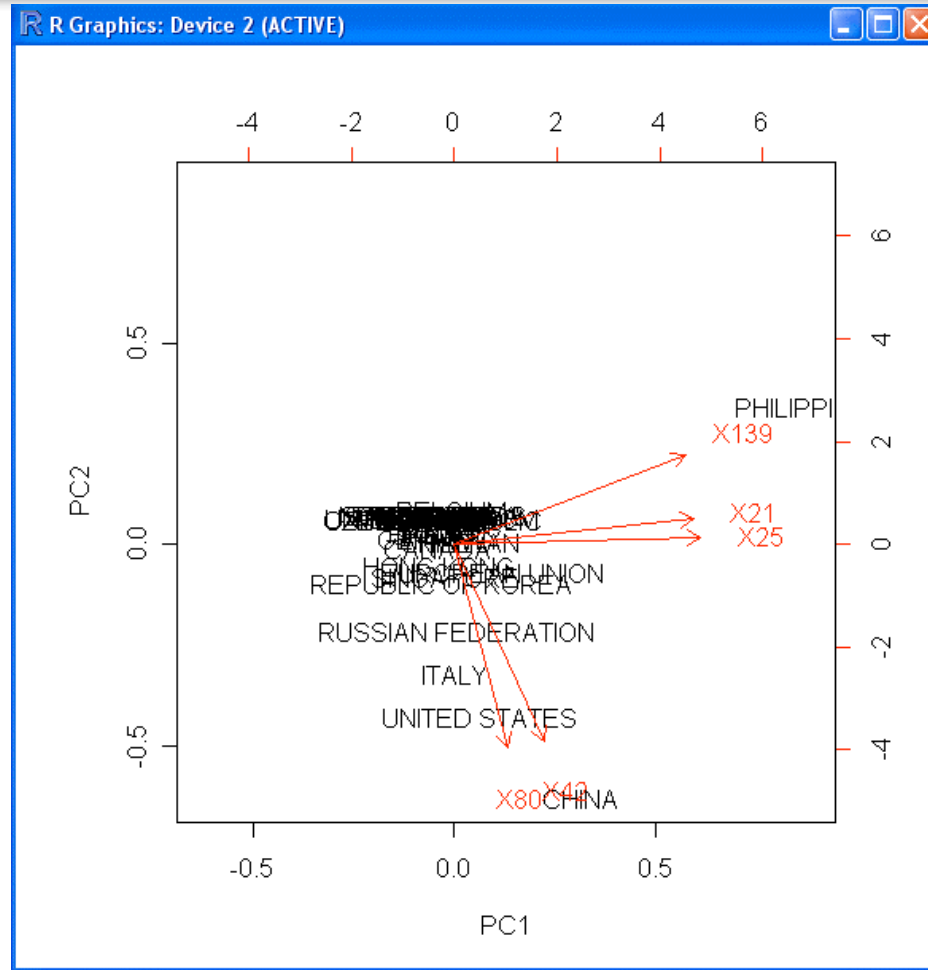
What we're used to...



From Criminology (R-Project)

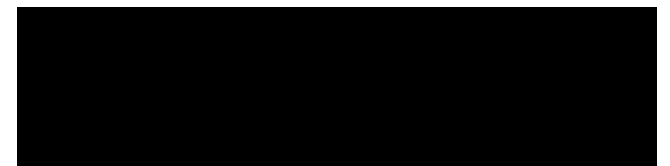


What it could be...



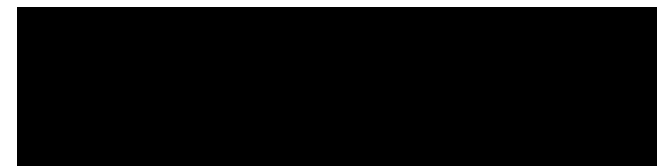
Genetics and Immune Concepts

- 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

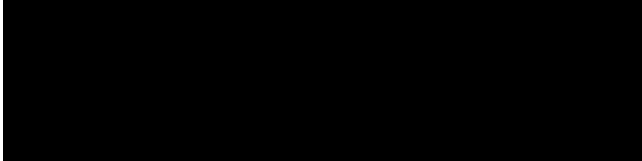
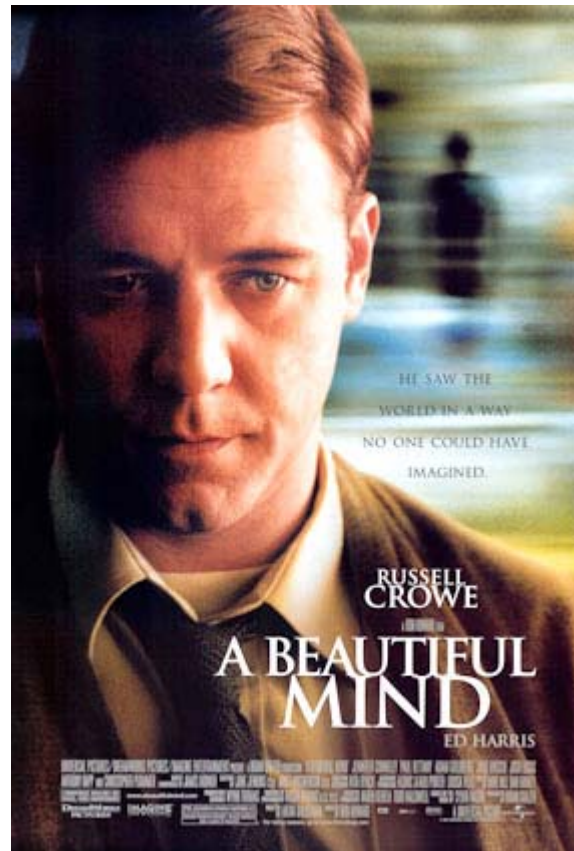


Economic Theories

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

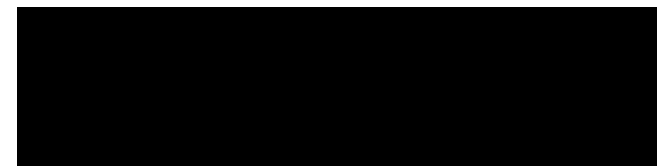


Game Theory



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...

The screenshot displays the Wireshark Network Analyzer interface. The main window shows a list of captured packets with columns for No., Len, Time, Source, Destination, Protocol, and Info. Packet 122 is selected, and its details are shown in the lower pane. The details pane shows the following information:

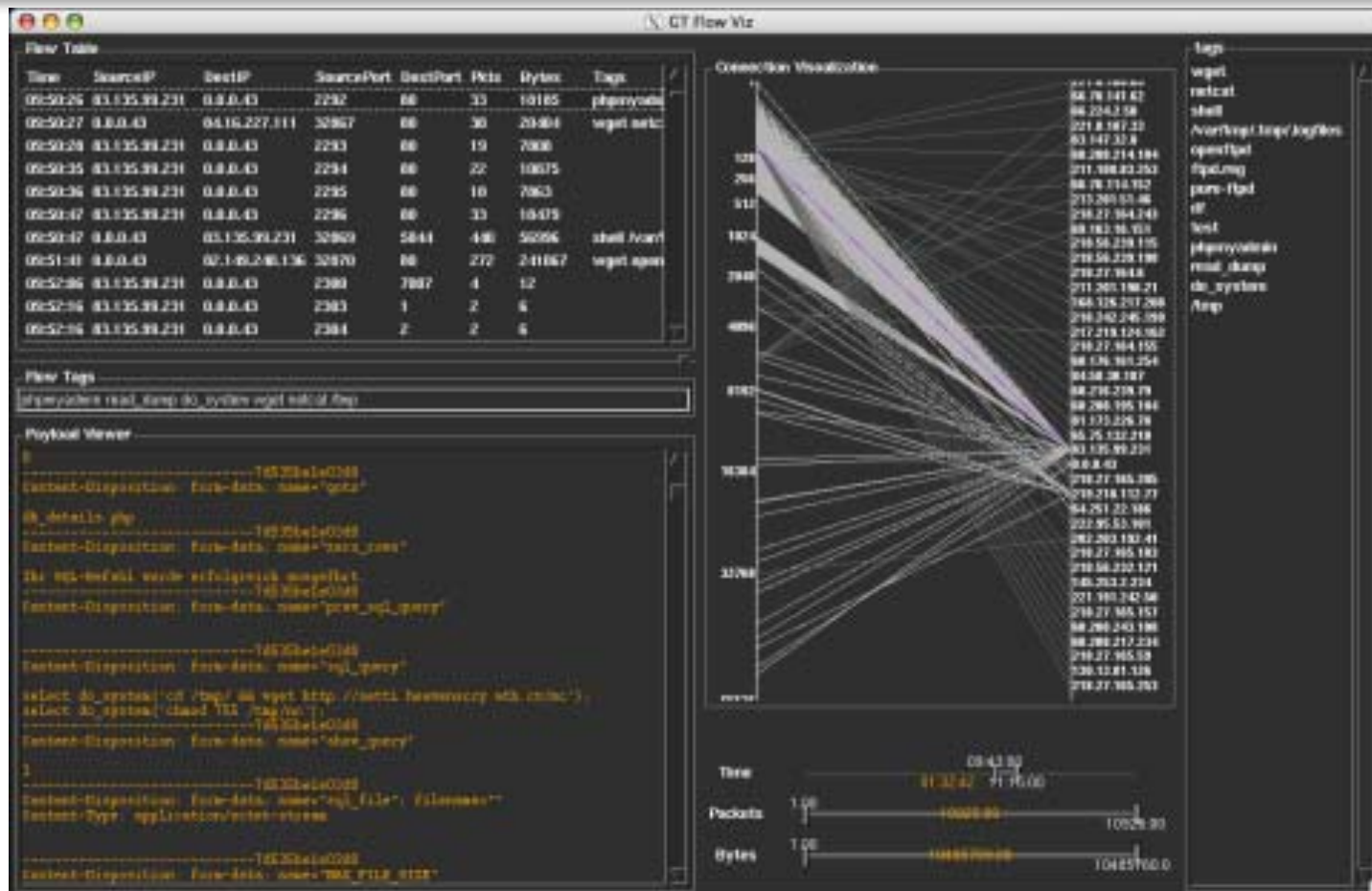
- Frame 122 (60 bytes on wire, 60 bytes captured)
- Ethernet II, Src: 00:c0:4f:c7:eb:c0 (00:c0:4f:c7:eb:c0), Dst: 00:00:0c:36:00:19 (00:00:0c:36:00:19)
- Internet Protocol, Src: 207.183.142.87 (207.183.142.87), Dst: 204.252.102.2 (204.252.102.2)
- Transmission Control Protocol, Src Port: 22587 (22587), Dst Port: 110 (110), Seq: 29, Ack: 134, Len: 6
 - Source port: 22587 (22587)
 - Destination port: 110 (110)
 - Sequence number: 29 (relative sequence number)
 - [Next sequence number: 35 (relative sequence number)]
 - Acknowledgement number: 134 (relative ack number)
 - Header length: 20 bytes
 - Flags: 0x0018 (PSH, ACK)

The packet bytes pane shows the following hex and ASCII data:

```
0000 00 00 0c 36 00 19 00 c0 4f c7 eb c0 08 00 45 00  ...6... 0....E.
0010 00 2e 75 02 40 00 40 06 34 ba cf b7 8e 57 cc fc  ..u.@.@. 4....W..
0020 66 02 58 3b 00 6e 6a 0f a9 ba a6 bd ae 90 50 18  f.X;nj.....P.
0030 7d 78 3d cc 00 00 53 54 41 54 0d 0a             }x=...ST AT..
```

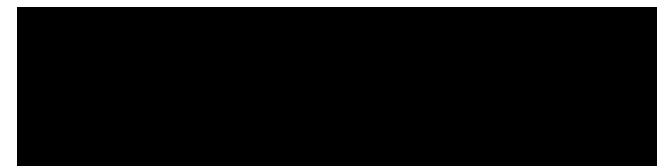
At the bottom of the details pane, it indicates: Sequence number (tcp.seq), 4 bytes | P: 3632 D: 3632 M: 0

What it could be... (FlowTag)



Conclusion

- High level overview of analysis tools and techniques
- 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 Support Machines

Predicting Attacks

Form1

Honeynet Intrusion Pattern Analyzer (HIPA) v 0.1

Ryan Talabis Ateneo de Manila University

Refresh

(DATE(timestamp))	(inet_ntoa(ip_src))	timestamp	dst_port	sig_name	tcp_dport	inet_ntoa(ip_src)	timestamp
5/25/2006	0.0.0.0	6/2006 8:58:13		WEB-MISC Chunked	42	69.128.153.189	/2006 12:41:41 A
5/26/2006	10.1.0.1	6/2006 8:58:13	22	WEB-FRONTPAGE	42	69.128.153.189	/2006 12:41:43 A
5/27/2006	10.1.0.16	6/2006 4:32:02	25	WEB-FRONTPAGE	42	69.128.153.189	/2006 12:41:44 A
5/28/2006	10.1.0.17	6/2006 1:11:07	42	SHELLCODE x86 NOP	42	69.128.153.189	/2006 12:41:45 A
5/29/2006	10.1.0.27	6/2006 8:58:15	67	EXPLOIT WINS overflow	42	69.128.153.189	/2006 12:41:46 A
	10.1.0.29	/2006 12:14:10	68	ICMP PING NMAP	42	69.128.153.189	/2006 12:41:48 A
	10.1.0.30	6/2006 5:42:49	80	SNMP public access	42	69.128.153.189	/2006 12:41:50 A
	10.1.0.34	6/2006 5:40:06	137	SNMP request udp	42	69.128.153.189	/2006 12:41:53 A
	10.1.0.36	6/2006 8:00:05	138	ICMP PING CyberKit	42	69.128.153.189	/2006 12:42:10 A
	10.1.0.37	6/2006 2:58:01	161	WEB-IIS WebDAV search	42	203.87.152.38	6/2006 8:08:36 A
			162	WEB-MISC WebDAV search			

timestamp	sig_name	inet_ntoa(ip_src)	inet_ntoa(ip_dst)	ip_ver	ip_hlen	ip_tos
/2006 12:41:44 AM	IP Packet detected	203.87.152.38	69.128.153.189	4	5	0
/2006 12:41:44 AM	IP Packet detected	69.128.153.189	203.87.152.38	4	5	0
/2006 12:41:44 AM	IP Packet detected	203.87.152.38	69.128.153.189	4	5	0
/2006 12:41:44 AM	IP Packet detected	69.128.153.189	203.87.152.38	4	5	0
/2006 12:41:44 AM	IP Packet detected	203.87.152.38	69.128.153.189	4	5	0
/2006 12:41:44 AM	SHELLCODE x86 NOP	69.128.153.189	203.87.152.38	4	5	0
/2006 12:41:44 AM	SHELLCODE x86 NOP	69.128.153.189	203.87.152.38	4	5	0
/2006 12:41:45 AM	EXPLOIT WINS overflow	69.128.153.189	203.87.152.38	4	5	0
/2006 12:41:45 AM	EXPLOIT WINS overflow	69.128.153.189	203.87.152.38	4	5	0
/2006 12:41:45 AM	IP Packet detected	203.87.152.38	69.128.153.189	4	5	0

ALL
Number of Events: 233

TIME
Attack Start: 5/26/2006 12:41:02 AM
Attack Stop: 5/26/2006 12:42:10 AM
Total Duration: 68.000 sec
Avg time per event: 0.292 sec
Avg time between tcp port 42 events: 0.540 sec
Avg time between tcp port 80 events: 0.075 sec

RELATED PORTS
tcp: 42 (126) 80 (106)
udp:

SIGNATURES
(http_inspect) BARE BYTE UNICODE ENCODING (2)
(http_inspect) OVERSIZE REQUEST-URI DIRECTORY (1)
EXPLOIT WINS overflow attempt (9)
SHELLCODE x86 NOP (84)
WEB-FRONTPAGE /_vti_bin/ access (2)
WEB-FRONTPAGE rad fp30reg.dll access (2)
WEB-MISC Chunked-Encoding transfer attempt (2)
WEB-MISC WebDAV search access (1)

PROTOCOL

Get Stats Save Profile Load Data