

Alternatives in Analysis

# Security Analytics Project

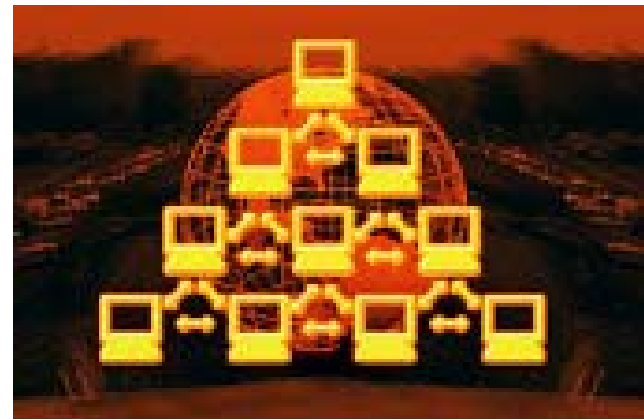
Mark Ryan del Moral  
Talabis  
Secure-DNA

# The Security Analytics Project



# 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
  - Lot's more...

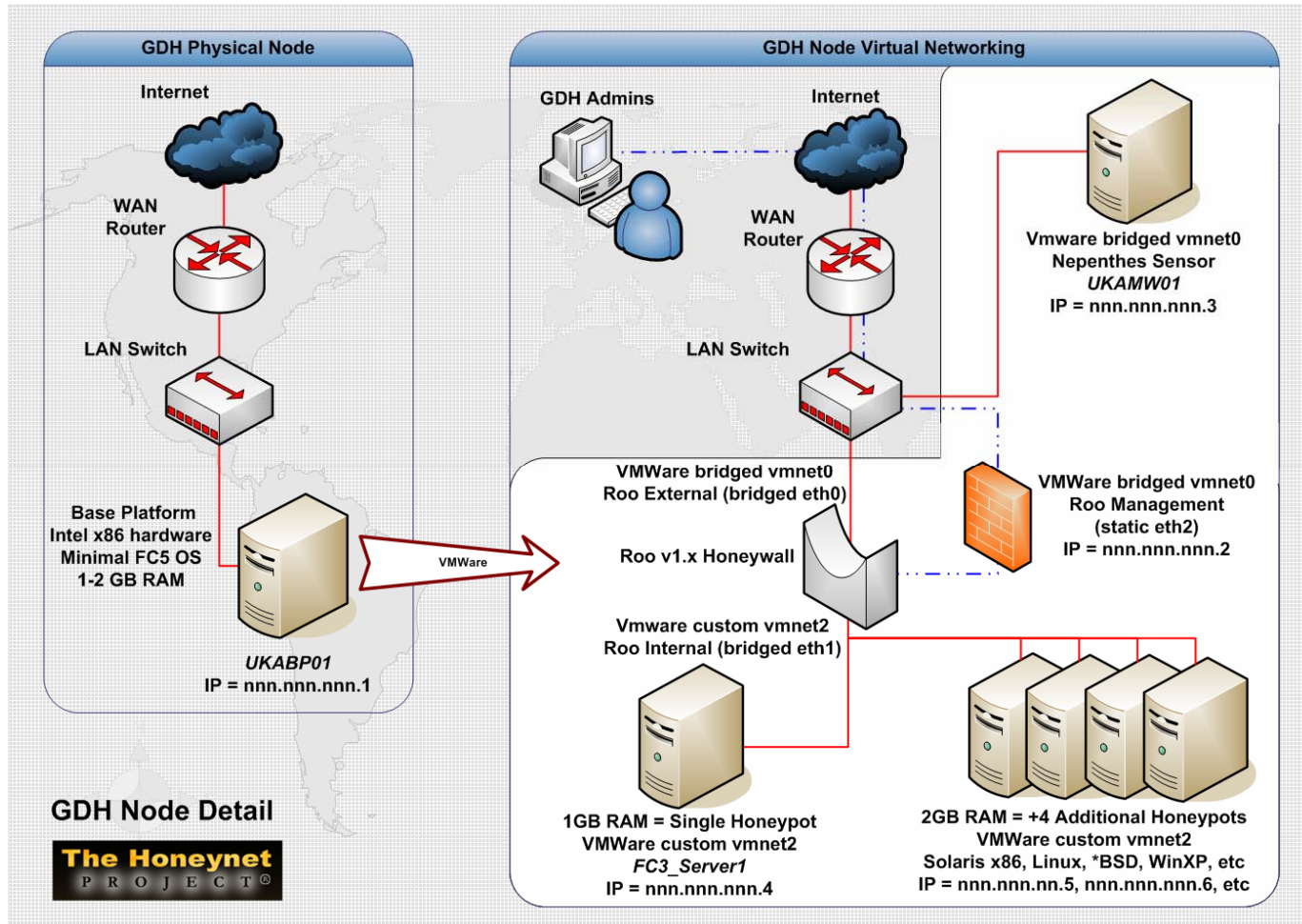


# New Stuff!

## Global Distributed HoneyNet

- High-interaction distributed honeynets with central data analysis
  - The HoneyNet Project
  - UK HoneyNet Project, David Watson

# Global Distributed Honeyynet



# 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.



# Looking Beyond Security

- Though security in itself is a unique field with unique needs, analysis techniques often span the boundaries of different disciplines
  - Mathematics
  - Economics
  - Statistics
  - Artificial Intelligence
  - Psychology and Sociology
  - Biology
  - Graphics design

# Techniques

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

# Data Analysis Tools

- R-Project
- Weka
- Tanagra
- FlowTag
- Honeysnap
- Excel and Access (yes.. From Microsoft)
- Orange
- SVMLite

# The Possibilities

# Techniques

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

# Preprocessing and Data Cleansing

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

# Honeysnap

The screenshot shows a web browser window with a single tab titled "Honeysnap". The browser's address bar is empty. The page content is as follows:

- A red horizontal bar at the top contains a navigation menu: [Summary](#) | [Flow Details](#) | [Sebek Details](#) | [IRC Summary](#) | [IRC Details](#) | [IP Summary](#) | [IP Lookup](#) |
- A [Login](#) link is located in the top right corner.
- The main heading is **Welcome to Honeysnap**.
- The current time is displayed as "Time is Mon Jun 4 22:15:00 2007".
- A status message reads: "Database has 61123 flows, 0 sebek records and 14737 irc messages from 1 honeypots".
- A red horizontal bar at the bottom contains a [Clear Search](#) button.

# Honeysnap and flows

Honeysnap Login

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

Detailed Search

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

Starttime	Endtime	Honeypot	Source	Destination	Src Port	Dst Port	Packets	Bytes
2002-11-29 06:26:09.939592	2002-11-29 06:26:09.939592	HS_Fake	192.168.100.28	194.25.2.133	32789	53	1	44
2002-11-29 06:26:10.089582	2002-11-29 06:26:10.089582	HS_Fake	192.168.100.28	217.5.100.186	32789	53	1	44
2002-11-29 06:26:10.089582	2002-11-29 06:26:10.089582	HS_Fake	194.25.2.133	192.168.100.28	53	32789	1	245
2002-11-29 06:26:10.229573	2002-11-29 06:26:10.229573	HS_Fake	217.5.100.186	192.168.100.28	53	32789	1	187
2002-11-29 06:30:39.981424	2002-11-29 06:30:39.981424	HS_Fake	192.168.100.28	213.234.132.130	32789	53	1	44
2002-11-29 06:30:40.141413	2002-11-29 06:30:40.141413	HS_Fake	213.234.132.130	192.168.100.28	53	32789	1	168
2002-11-29 06:30:40.151412	2002-11-29 06:30:40.151412	HS_Fake	192.168.100.28	213.234.128.211	32789	53	1	52
2002-11-29 06:30:40.311401	2002-11-29 06:30:40.311401	HS_Fake	213.234.128.211	192.168.100.28	53	32789	1	120
2002-11-29 06:33:40.009306	2002-11-29 06:33:40.009306	HS_Fake	192.168.100.28	168.95.1.14	32789	53	1	45
2002-11-29 06:33:40.239291	2002-11-29 06:33:40.239291	HS_Fake	168.95.1.14	192.168.100.28	53	32789	1	92
2002-11-29 06:36:40.037184	2002-11-29 06:36:40.037184	HS_Fake	192.168.100.28	210.94.0.7	32789	53	1	45
2002-11-29 06:36:40.227172	2002-11-29 06:36:40.227172	HS_Fake	192.168.100.28	210.180.98.69	32789	53	1	45
2002-11-29 06:36:40.227172	2002-11-29 06:36:40.227172	HS_Fake	210.94.0.7	192.168.100.28	53	32789	1	123
2002-11-29 06:36:40.407160	2002-11-29 06:36:40.407160	HS_Fake	210.180.98.69	192.168.100.28	53	32789	1	123
2002-11-29 06:38:10.051128	2002-11-29 06:38:15.120789	HS_Fake	192.168.100.28	200.33.146.213	32789	53	2	94
2002-11-29 06:38:10.091126	2002-11-29 06:38:10.091126	HS_Fake	192.168.100.28	200.23.1.1	32789	53	3	114
2002-11-29 06:38:10.091126	2002-11-29 06:38:15.160786	HS_Fake	200.33.146.213	192.168.100.28	53	32789	2	285
2002-11-29 06:38:10.151122	2002-11-29 06:38:15.160786	HS_Fake	192.168.100.28	200.33.146.217	32789	53	4	164
2002-11-29 06:38:10.151122	2002-11-29 06:38:10.151122	HS_Fake	200.23.1.1	192.168.100.28	53	32789	3	438
2002-11-29 06:38:10.191119	2002-11-29 06:38:15.200783	HS_Fake	200.33.146.217	192.168.100.28	53	32789	4	1007

Clear Search

# Honeysnap and IRC chat

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 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

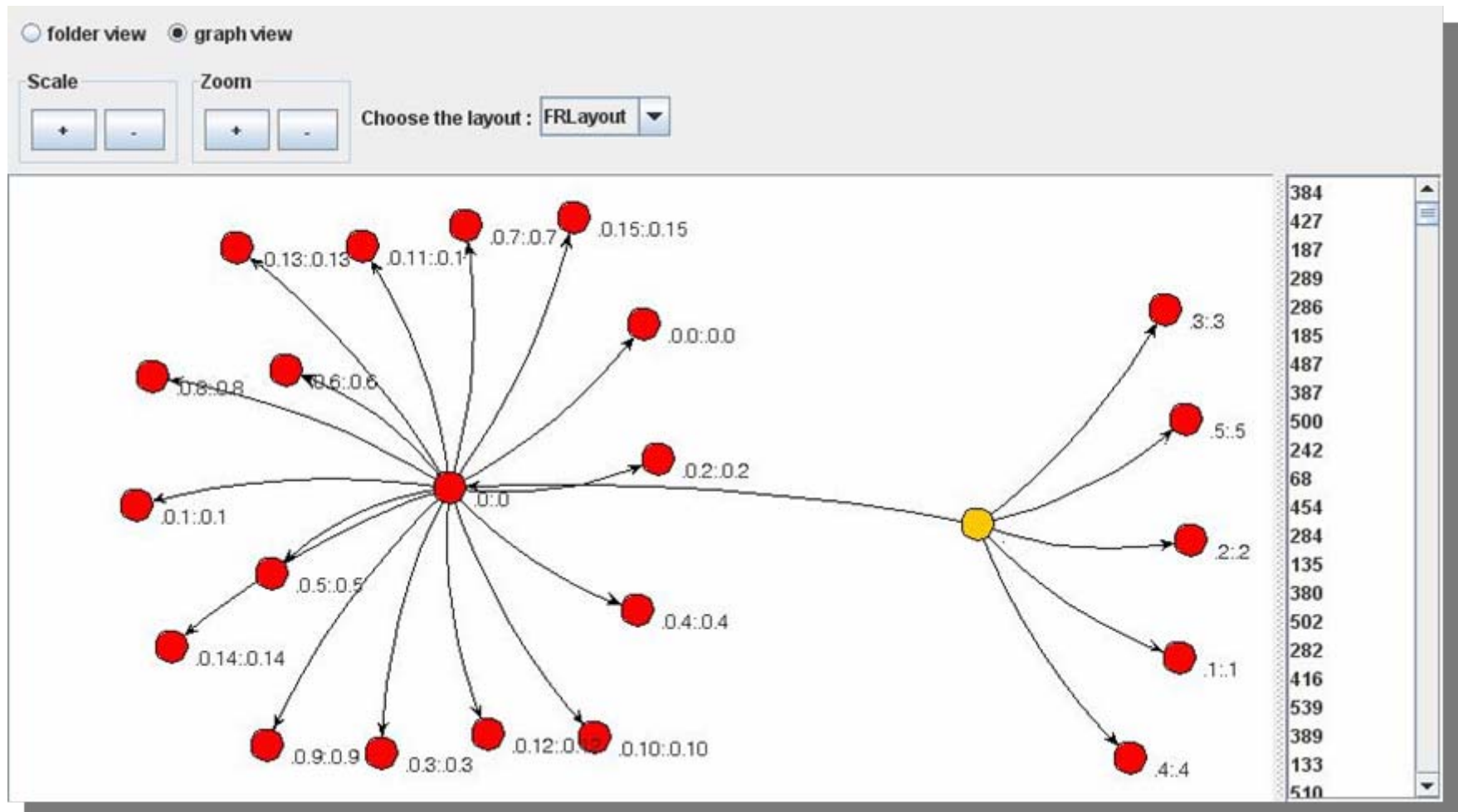
# Data and Text Mining

- 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

# 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-Means

# Classifying Attacks



# Machine Learning

- Machine learning is concerned with the design and development of algorithms and techniques that allow computers to "learn"
- Perfect for:
  - Predicting Attacks
  - Self-learning IDS
- Sample research:
  - Predicting attacks using Support Vector Machines

# Predicting Attacks

Form1

## Honeynet Intrusion Pattern Analyzer (HIPA) v 0.1

Ryan Talabis Ateneo de Manila University

Refresh

[DATE(timestamp)]	inet_nfoa(ip_src)	inet_nfoa(ip_dst)	dst_port	sig_name	tcp_dport	inet_nfoa(ip_src)	timestamp
5/25/2006							
5/26/2006							
5/27/2006							
5/28/2006							
5/29/2006							

inet_nfoa(ip_src)	timestamp	dst_port	sig_name	tcp_dport	inet_nfoa(ip_src)	timestamp
0.0.0.0	6/2006 8:58:13		WEB-MISC Chunked	42	69.128.153.189	/2006 12:41:41 A
10.1.0.1	6/2006 8:58:13	22	WEB-FRONTPAGE	42	69.128.153.189	/2006 12:41:43 A
10.1.0.16	6/2006 4:32:02	25	WEB-FRONTPAGE	42	69.128.153.189	/2006 12:41:44 A
10.1.0.17	6/2006 1:11:07	42	SHELLCODE x86 NI	42	69.128.153.189	/2006 12:41:45 A
10.1.0.27	6/2006 8:58:15	67	EXPLOIT WINS ove	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 r	42	203.87.152.38	6/2006 8:08:36 A
		162	WEB-MISC WebDAV			

timestamp	sig_name	inet_nfoa(ip_src)	inet_nfoa(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 NI	69.128.153.189	203.87.152.38	4	5	0
/2006 12:41:44 AM	SHELLCODE x86 NI	69.128.153.189	203.87.152.38	4	5	0
/2006 12:41:45 AM	EXPLOIT WINS ove	69.128.153.189	203.87.152.38	4	5	0
/2006 12:41:45 AM	EXPLOIT WINS ove	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:

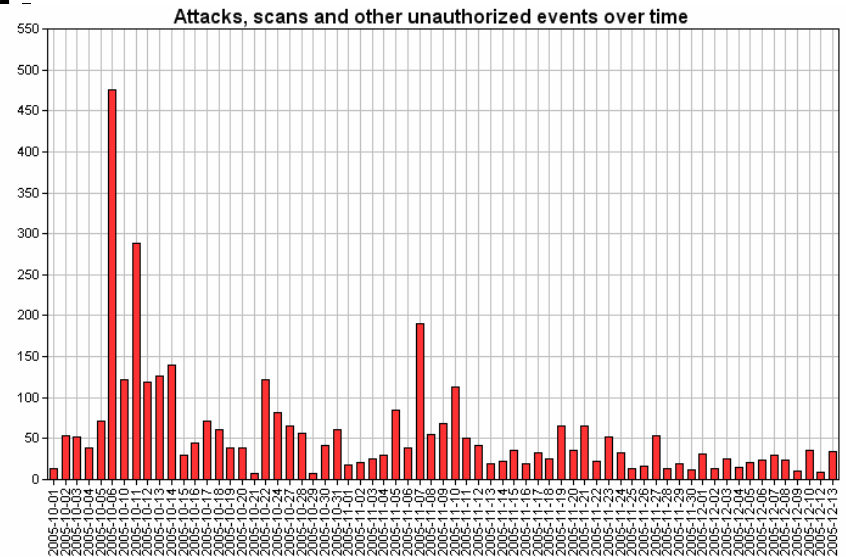
SIGNATURES  
(http\_inspect) BARE BYTE UNICODE ENCODING (2)  
(http\_inspect) OVERSIZE REQUEST-URI DIRECTORY (1)  
EXPLOIT WINS overflow attempt (3)  
SHELLCODE x86 NOOP (84)  
WEB-FRONTPAGE /\_vti\_bin/ access (2)  
WEB-FRONTPAGE rad lp30reg.dll access (2)  
WEB-MISC Chunked-Encoding transfer attempt (2)  
WEB-MISC WebDAV search access (1)

PROTOCOL

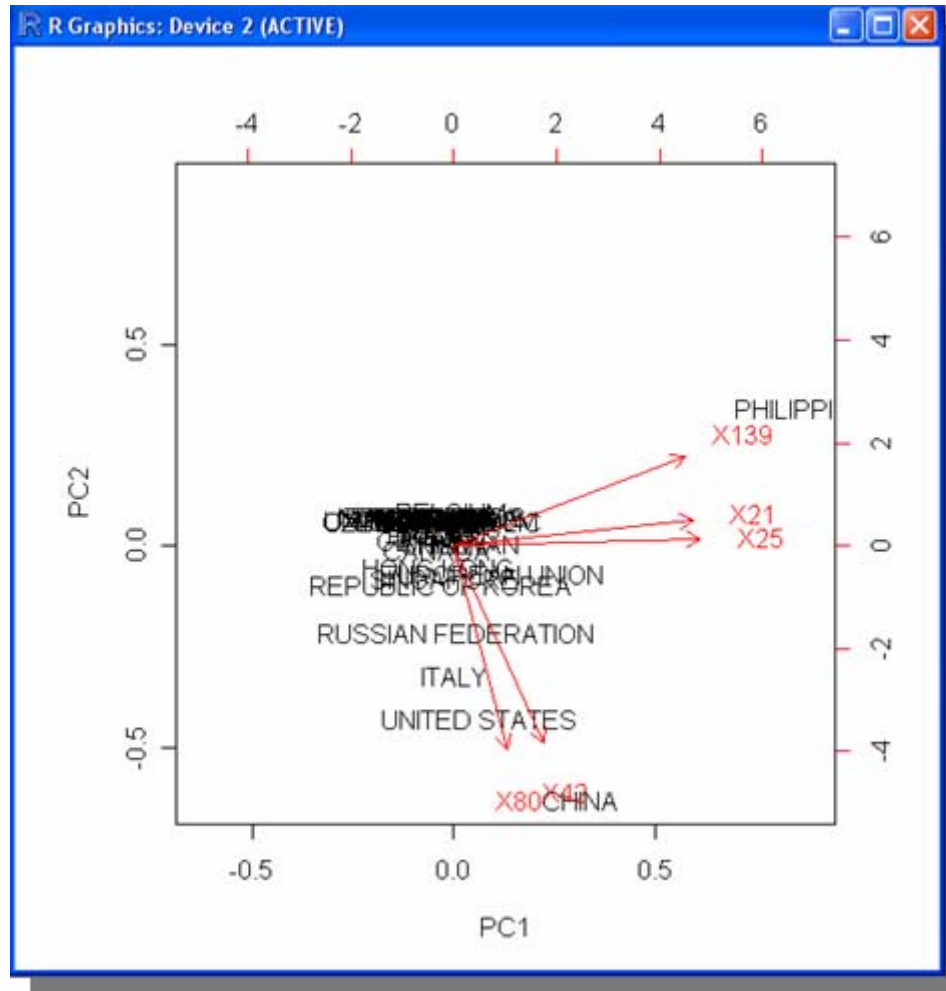
Get Stats Save Profile Load Data

# Statistics

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



# Detecting Patterns of Attack



# Behavioral Analysis

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

# 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

# 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

# Visualization

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



# FlowTag

- Video Demo

# Techniques

- Summary of Uses:
  - Detection
  - Classification
  - Patterns and Trends
  - Prediction
  - Motivation and Behaviors
  - Preprocessing and Data cleansing
- To produce good results techniques could be used together

# A Collaborative Effort

- A forum where people from different fields can share data and techniques

**Thank You**