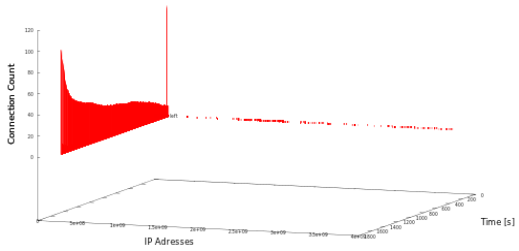
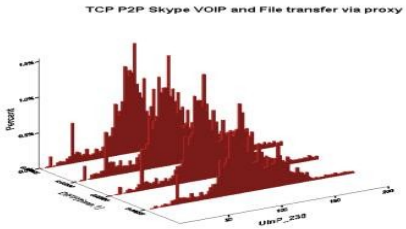
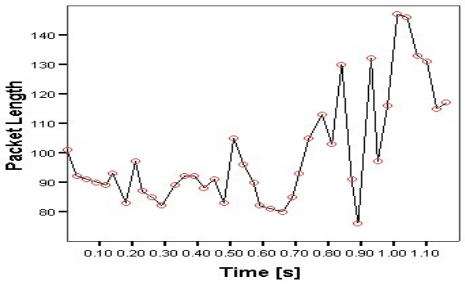
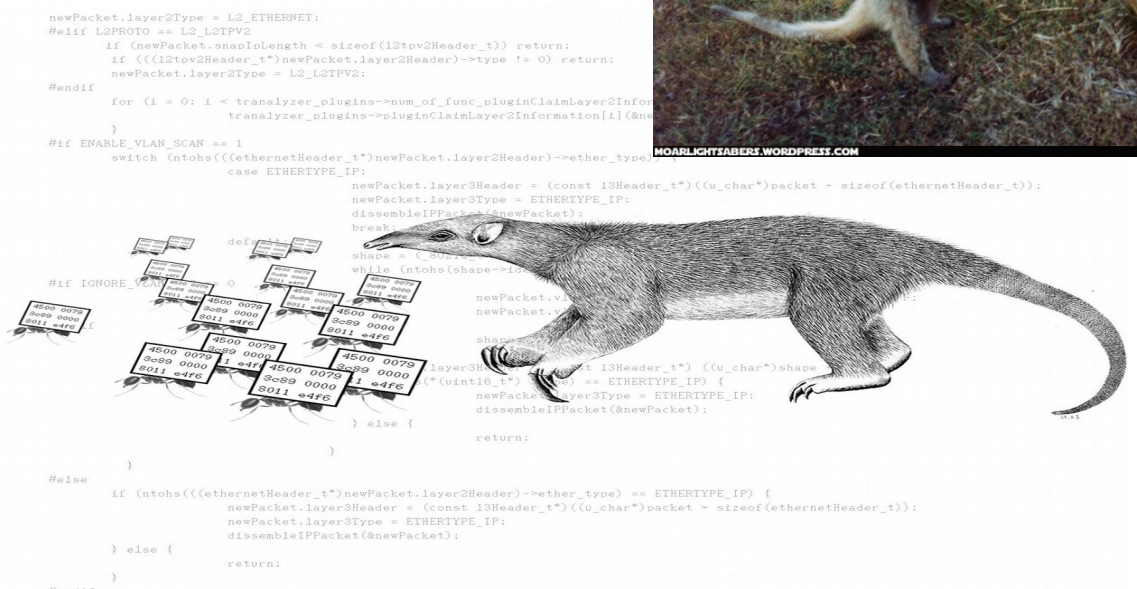
The Unknown Unknown
Traffic Mining









Stefan Burschka

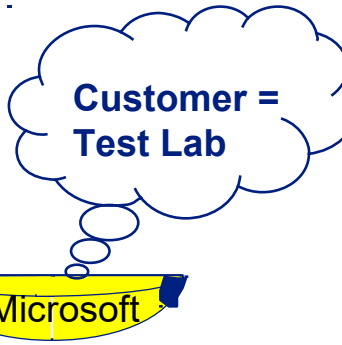
Darmstadt, 4.10.2016

Together ahead. RUAG

Controlled Flight into Terrain

- There is a union for the lifeform SW
- Seldom problem oriented KISS Engineering
- High Complexity, small fault tolerance

- Time- and economic pressure, info overload



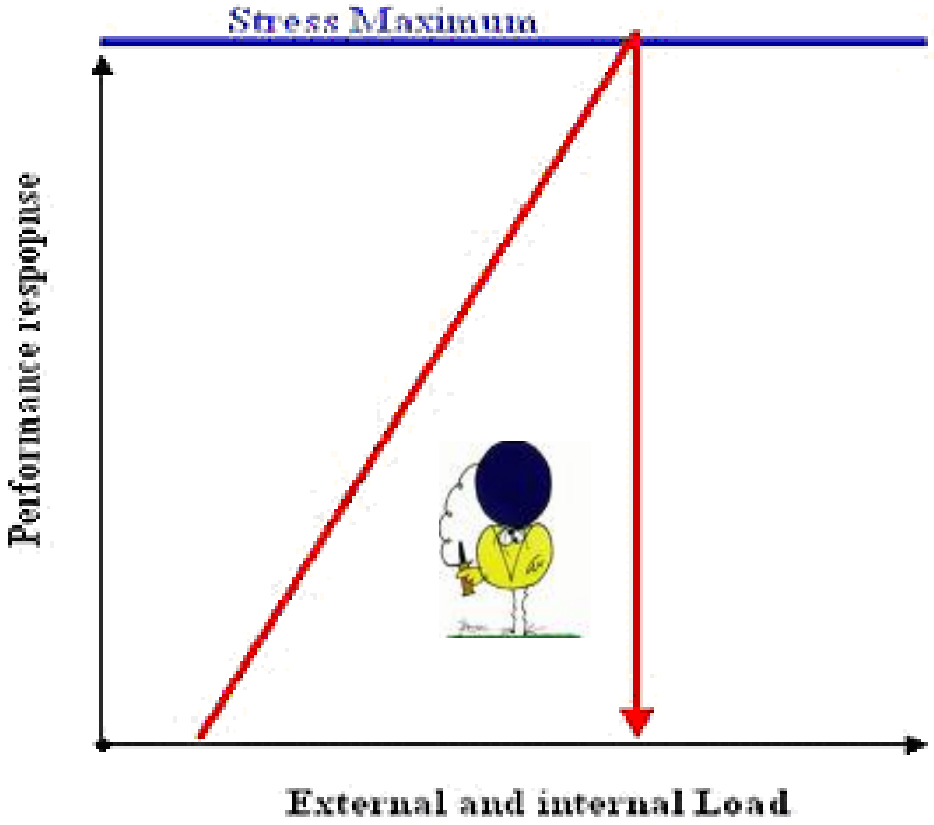
- Banana Principle: Conditioning of cutomers
- SW became a weapon

Together ahead. RUAG

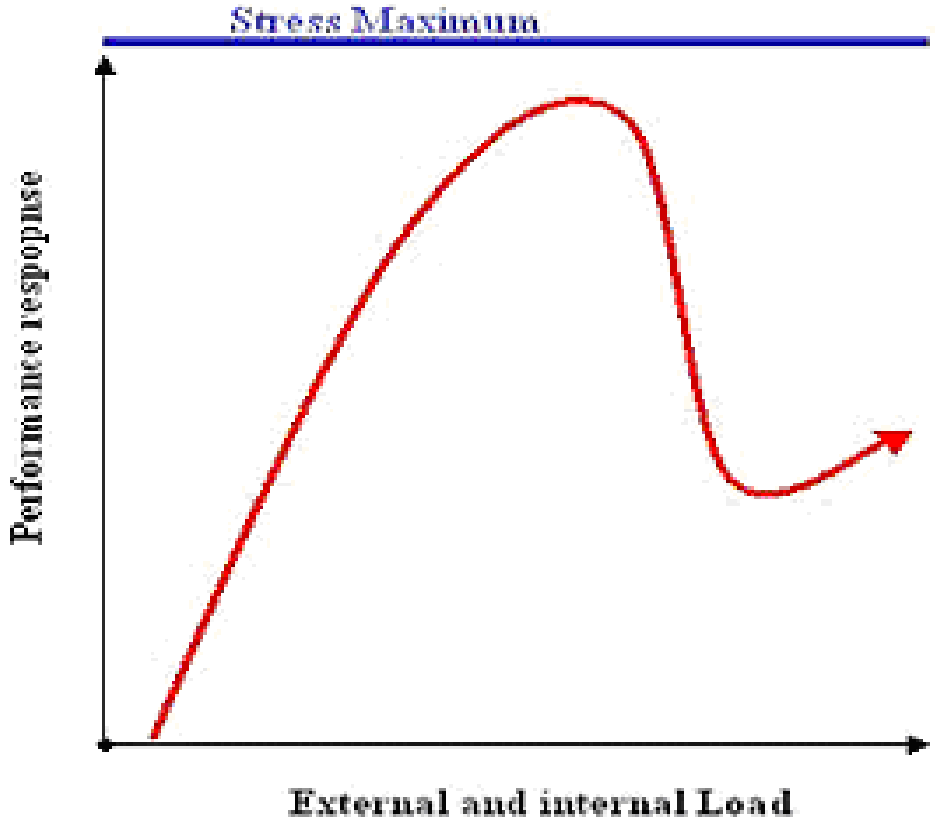
Result: Reaction to stress, faults, attacks

Safeguard: EU-IST FP5: IST-2001-32685 (2001-2004)

Stress Reaction of current digital and modern IT Systems



Stress Reaction of natural and fault-tolerant analog Systems



The Network is slow, The Network is insecure;
NO, it's not Microsoft, shut up, It wasn't me ...

Manager (MBA)
Always right, DoR
License to Powerpoint

Production (poor Techie)
Knows, Always warned,
Always his fault: FUBAR
License to get fired

Finance (MBA)
Knows basic calculus
License to Excel

We didn't find the problem in 4 months, can you do the job in 2 weeks? (We supply only 40TB pcap data)

Together ahead. RUAG

Data (Traffic) Mining → Forensics

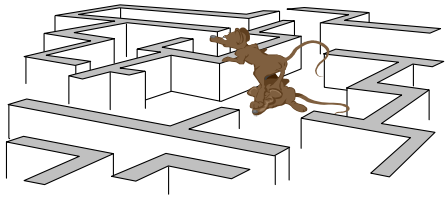
Input: Huge Vol, no Clues → Small Vol, Clues

Output: Small Vol, Clues, Ideas → Large Puzzle

TM: Almost Everything

→ Find what everybody missed
Find patterns, anomalies and the undetectable.

TM: Almost Nothing
193.5.230.129, 80



→ Tell me everything and especially
Owner, gender, color of hair and underpants, talks to pigs?
OS, type of computer ...

Together ahead. RUAG

Traffic Mining: find the unknown unknown

© RUAG RDNIZ
RUAG internal use only
Under signed NDA

Together ahead. RUAG

See the disaster now? Now you have context!

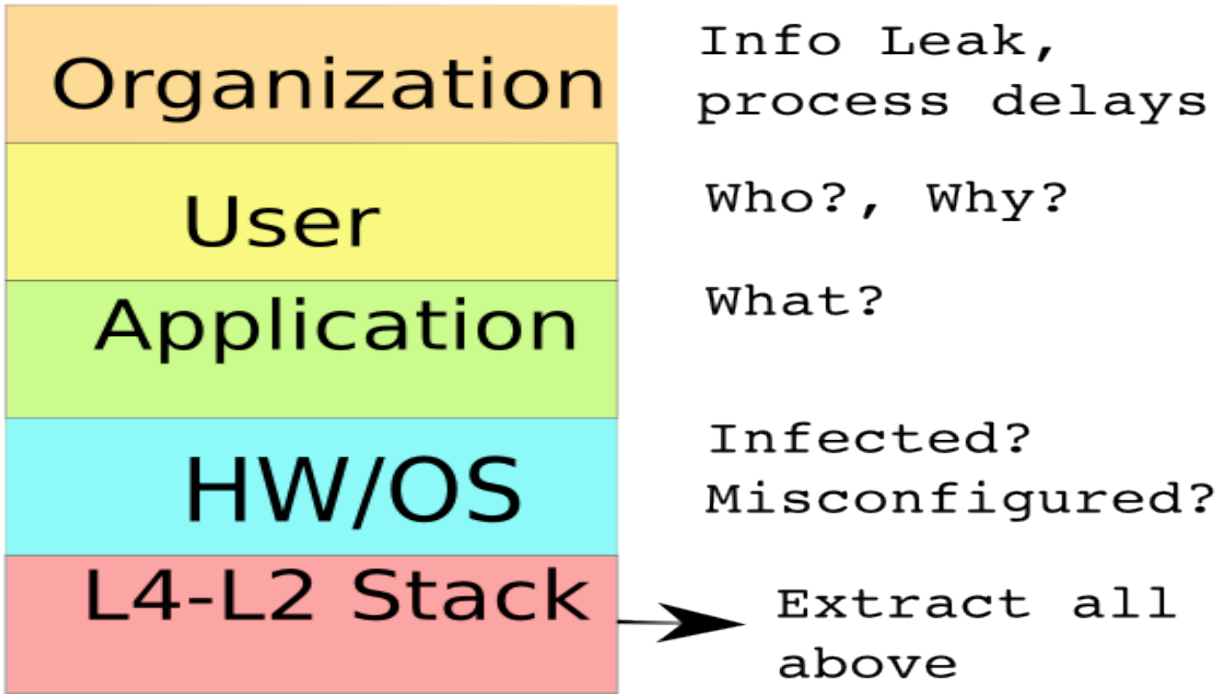
Together ahead. RUAG

Traffic Mining (TM): Find the right context

Hidden Knowledge: See, Understand, Invariants, Model

Application in

- Troubleshooting, Security (Classification, Encrypted TM)
- Network usage (VoiP, P2P traffic shaping, application/user profiling)
- Profiling & Marketing (User, Usage performance- & market- index)
- Law enforcement and Legal Interception (Indication/Evidence)

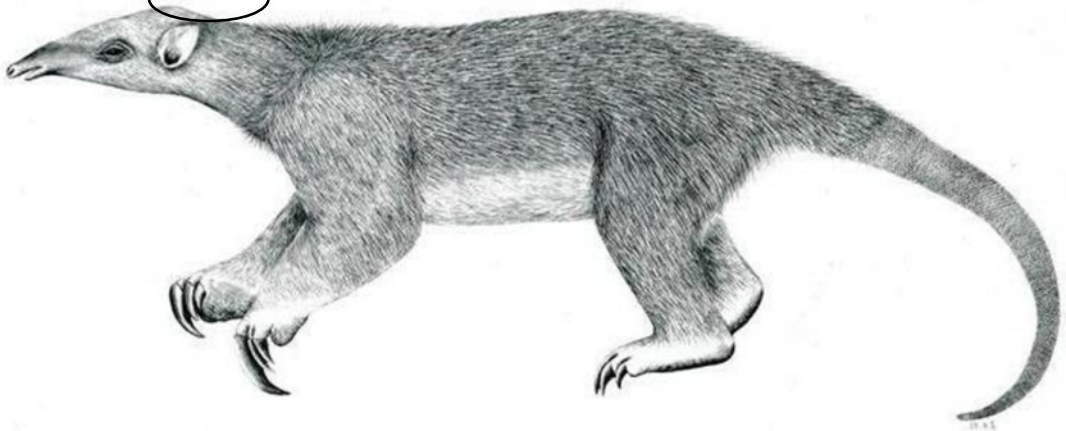


Together ahead. RUAG

Somebody throws the 40TB pcap at you,
asks questions and wants answers tomorrow.

Then you need a script friendly Swiss Anteater.

No Cisco, Huawei or ITEF, you are in control

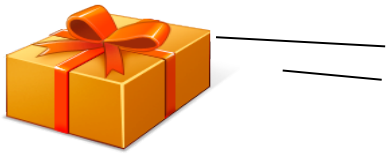


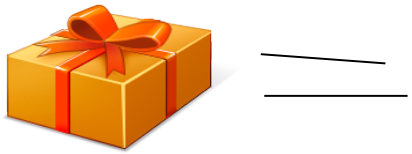
Together ahead. RUAG

Preprocessing/Context/Dim - Reduction Versatile Flow Compression

GONE TO NEVER-NEVER LAND

B





- Definition: (minimum 6-10 Tuple)

(sMAC, dMAC, ethtype, sctpStream), Vlan(s), srcIP, srcpPort, dstIP, dstPort, L4Protocol, …

- Or why not a bit more context and meaning?

srcWho, dstWho

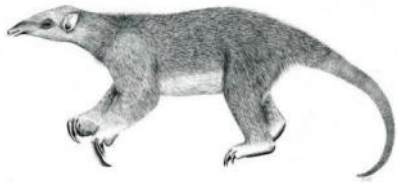
srcNetwork, dstNetwork

Bad, Good, Internal / External

A reliable Whothefuckisthat labelling

Together ahead. RUAG

Tranalyzer Flow Example



A

1196278772.439355		1196279184.642073		412.202718	0x9B420008			22
192.168.1.10		0x00000001		2119	182.236.128.34	0x8008036C		
80	6	00:0f:1f:cf:7c:45_00:00:0c:07:ac:0a_6387				http	6387	8272
464	5437587	0	579	15.494803	1.125660	-0.128590	-	
0.999829	1	87	128	128	0x00	0x42	0x0000	116 464
6231	4116	5437724		2253	63754	64831.988281		62501 65535
3342	2904	5713	0x18	0xF900	0x0000	0x03	0x00000000	
0x0000	-1.0	3	36578	1	… 0x00000401 www.google.com			

B

1196278772.409312		1196279184.642073		412.232761	0x9B430009			22
182.236.128.34		0x8008036C 80		192.168.1.10	0x00000001			
2119	6	00:d0:00:64:d0:00_00:0f:1f:cf:7c:45_8272				http	8272	6387
5437587	464	0	1380	20.066333	13190.574633		0.128590	
0.999829	1	3	63	63	0x00	0x42	0x0000	8146
5440245		109	116	464	8104	5840	5840.000000	655350
0	0	0	0x18	0x1B00	0x0000	0x03	0x00000000	
0x0000	-1.0	36578	3	1	… 0x00000400 www.google.com			
172.217.16.132								

Together ahead. RUAG

DPI/IDS

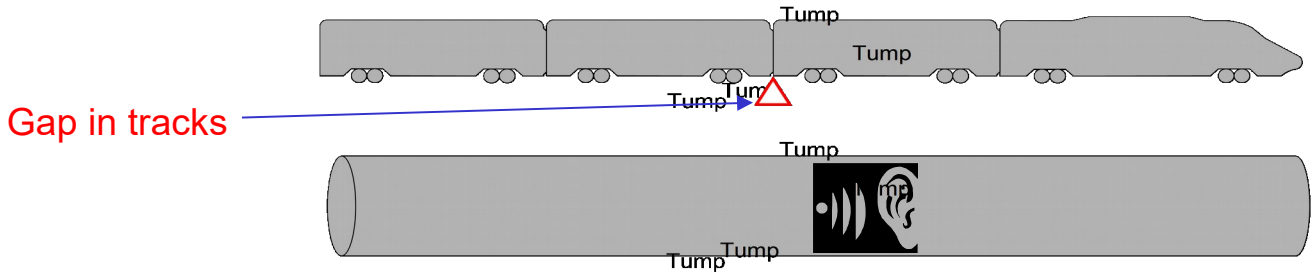
Ouch!! Encryption !
Which features to look at?

The
Packet

Together
ahead. RUAG

Encrypted TM: Packet Length Magic

Distinguish from by listening



Sound $\sim \vec{F} = d\vec{p}/dt = c$

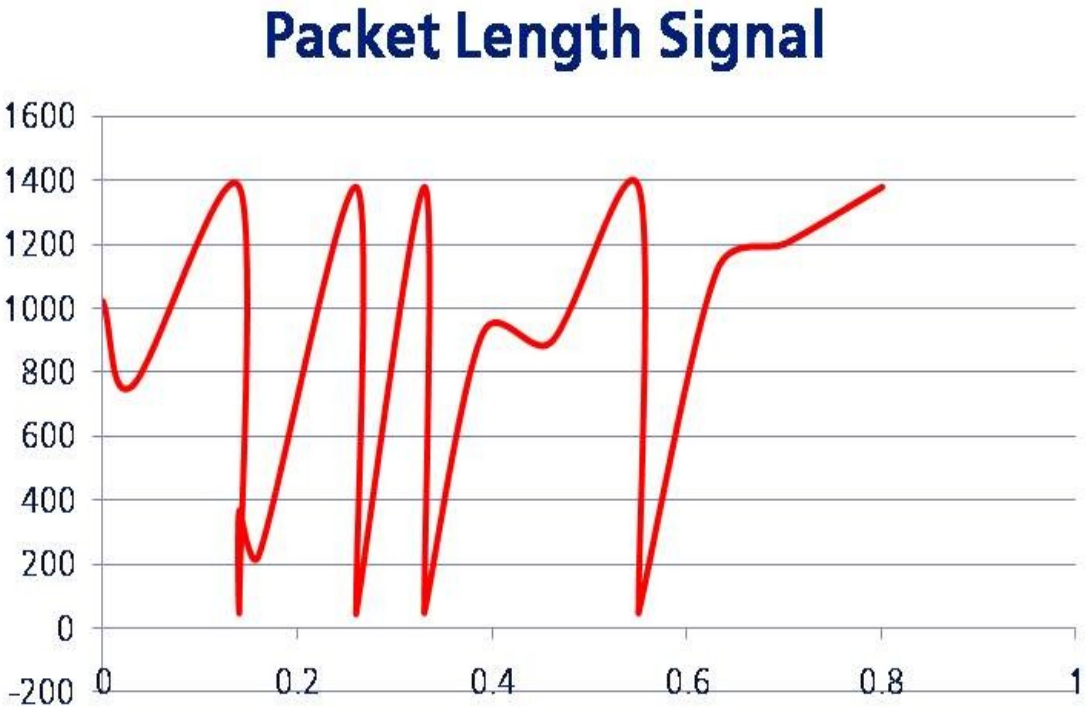
$dm/dt = dm/c$

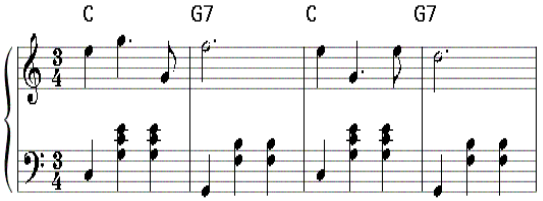
Packet Length

Packet Fire Rate (Interdistance)

Together ahead. RUAG

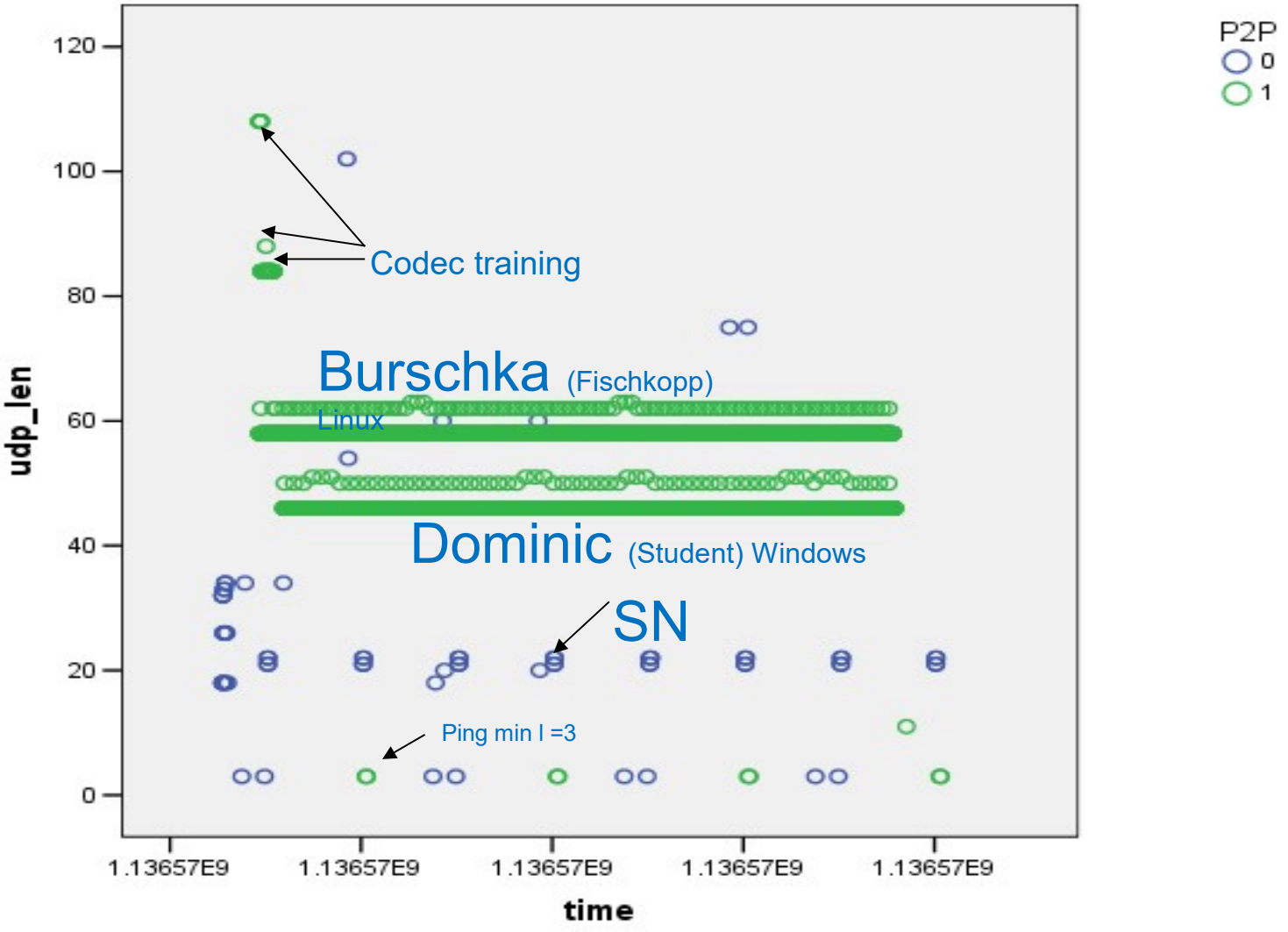
Packet Signal: Encrypted VoIP Mining



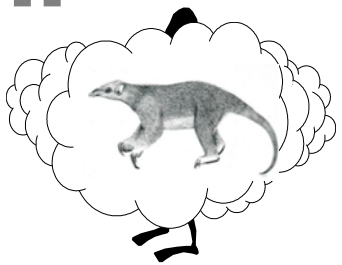


Together ahead. RUAG

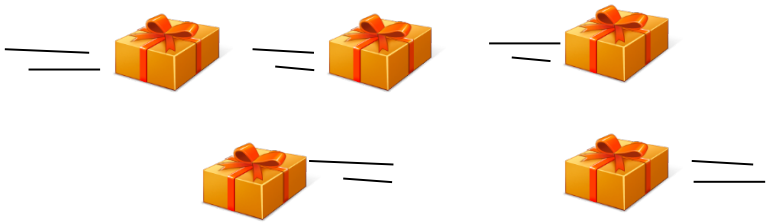
Exercise: Multiple Flow Packet Length Signal See the features?



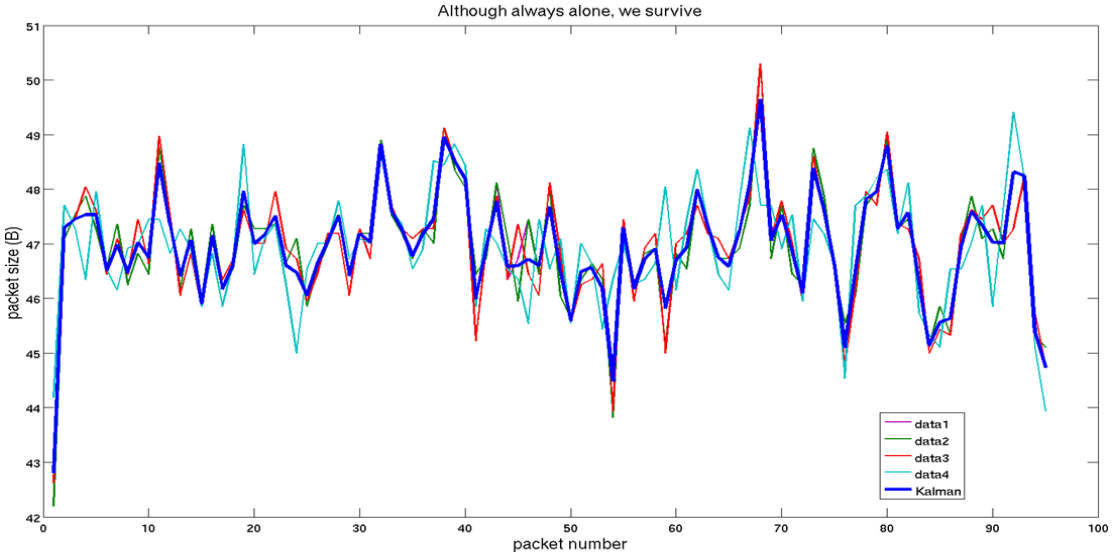
Encrypted Content Estimation



- SSH Command & Parameter Estimation
- IPSEC Tunnel: Application / User Profiling
- Covert Channels

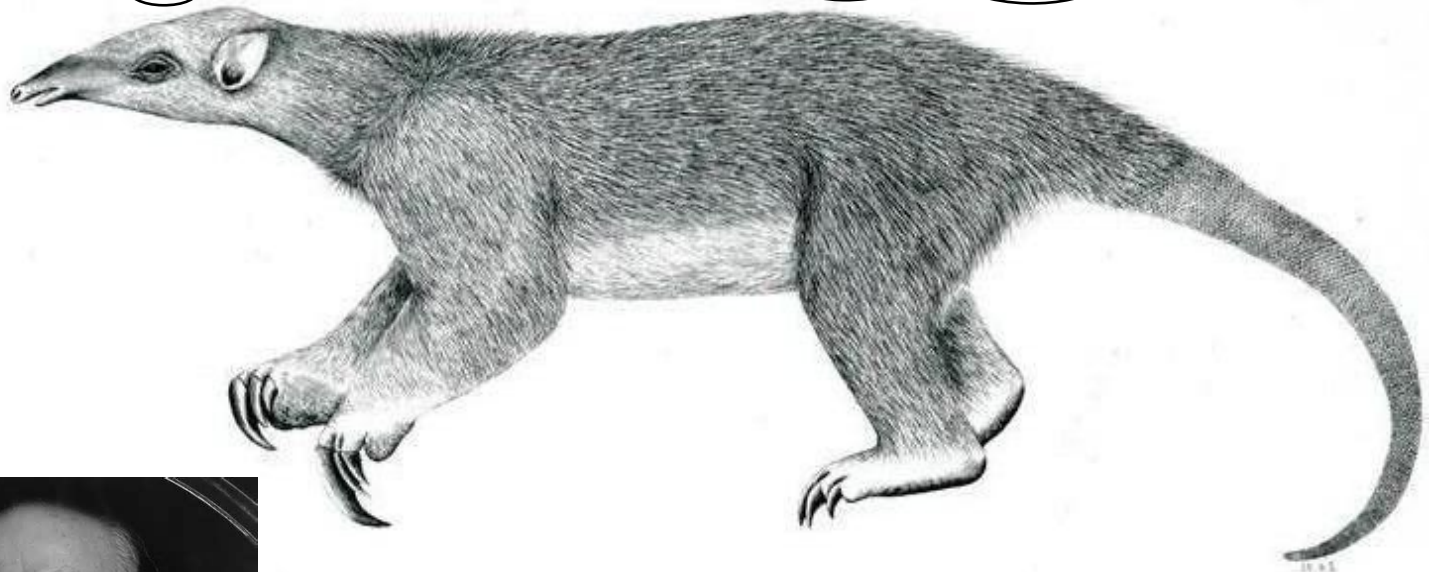


- Data leakage
- Skype Content Guessing: CCC 2011 (Datamining for Hackers)



Together ahead. RUAG

Anomaly Detection
Artificial Intelligence
The data set pitfalls

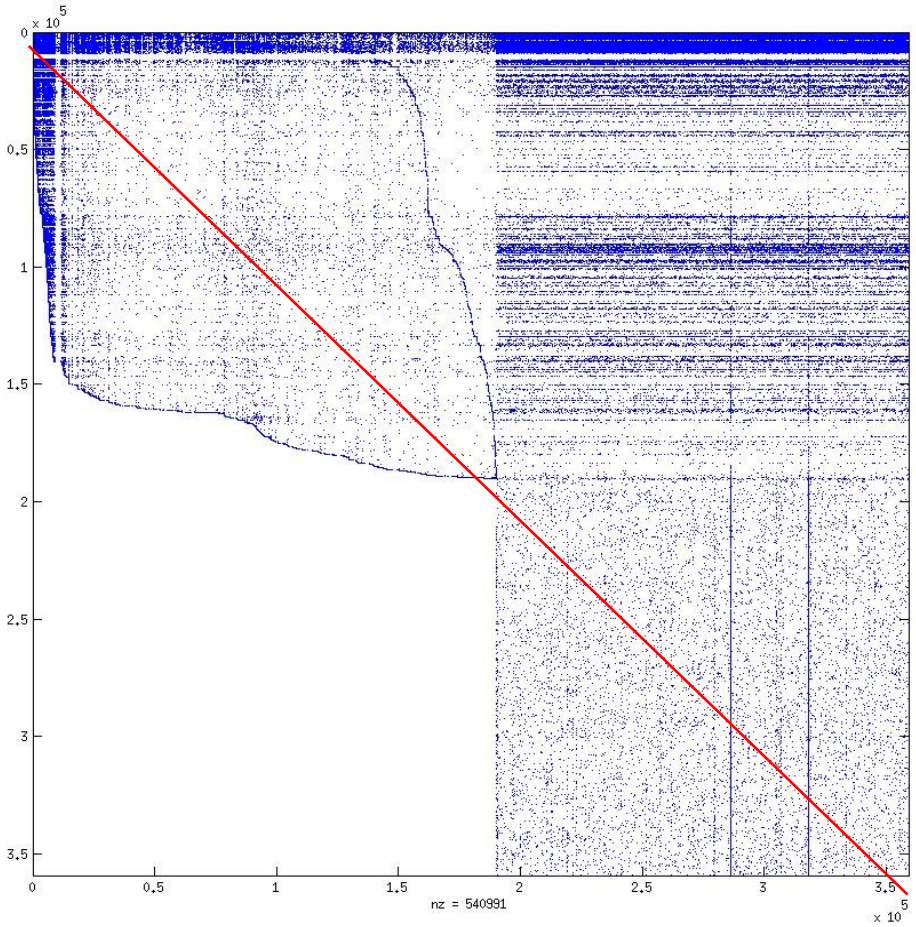


I only believe in data that I acquired myself

Together ahead. RUAG

TCP Connection Matrix
The famous complete dataset

Destination IP



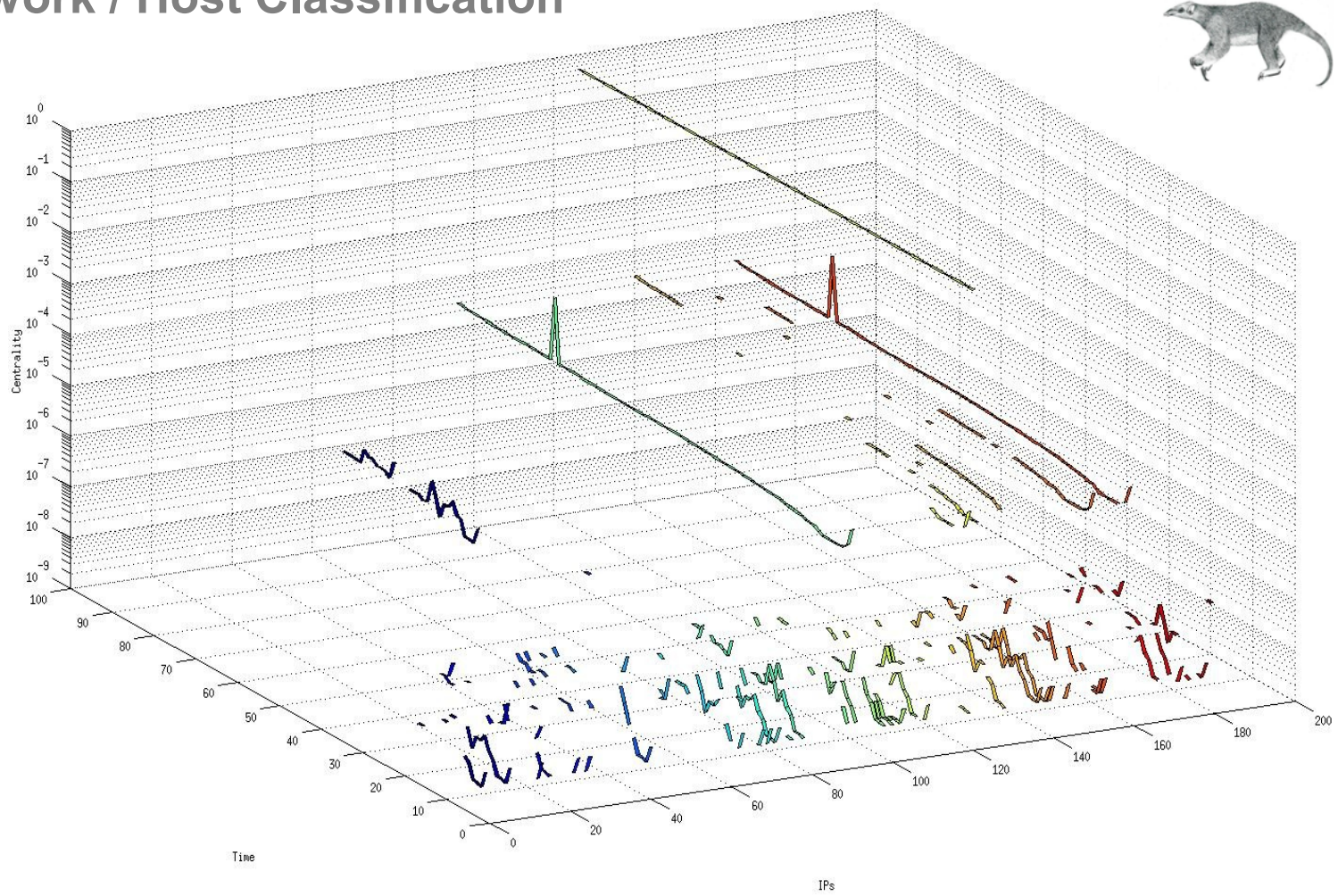
Source IP

Works also on Linux cmd line:
awk '{if($16) ….}' …. or sort | uniq ….

ahead. RUAG

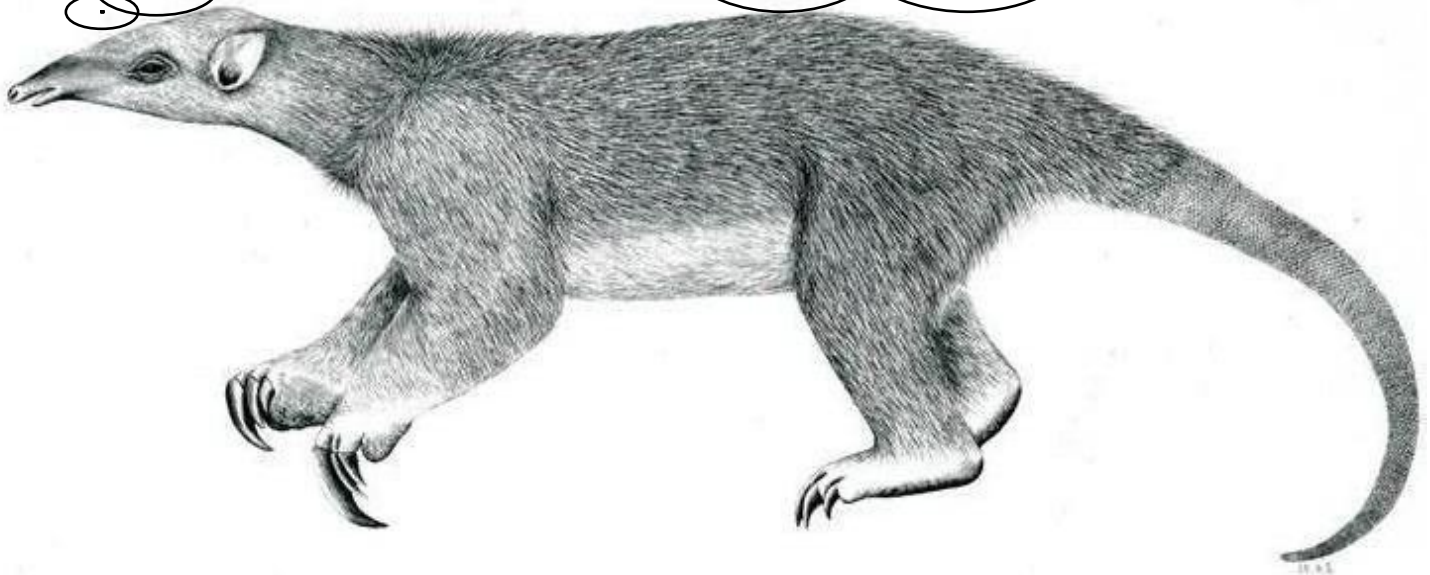
Anomaly Flow Graph: Centrality

Network / Host Classification



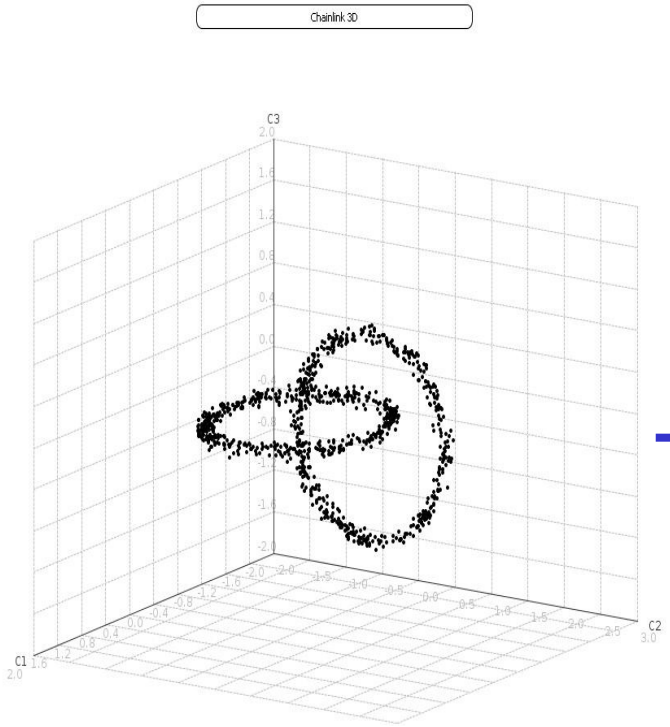
Together ahead. RUAG

Worst Case:
AI Brain Support

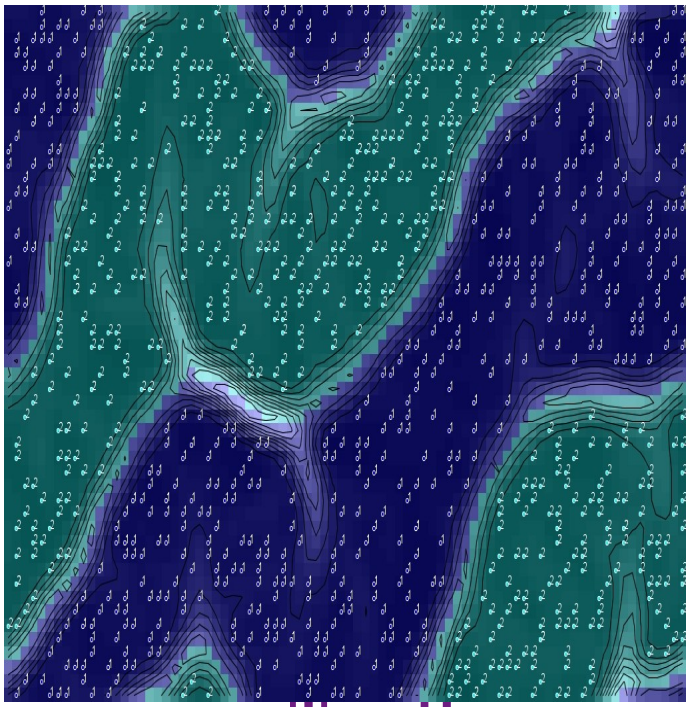


Together ahead. RUAG

Clustering of Multidimensional Data
ESOM: Nonlinear Mapping



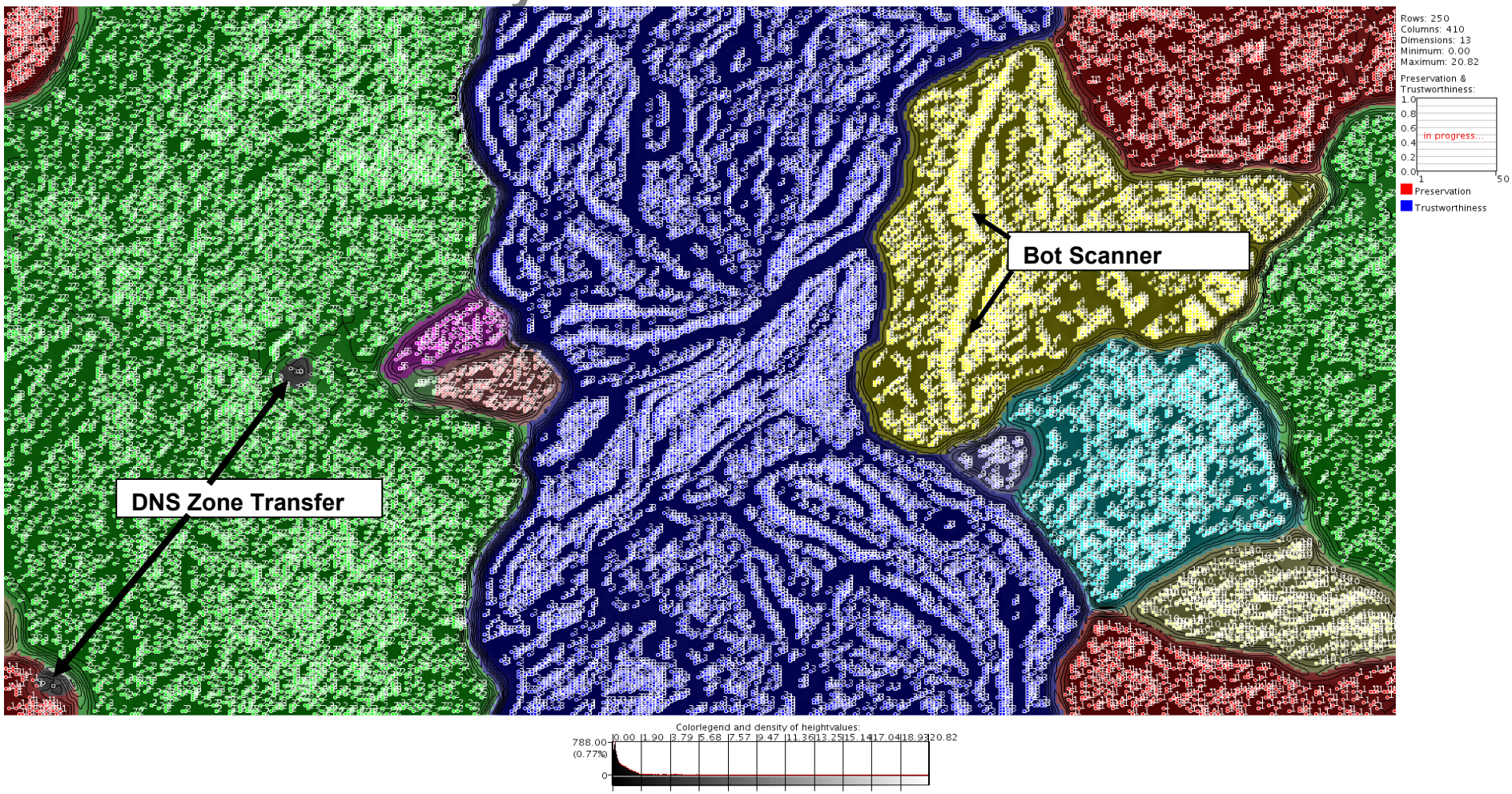
Retina



EsomViewer, Operational Picture:

13 Dim statistical T2 Flow parameters

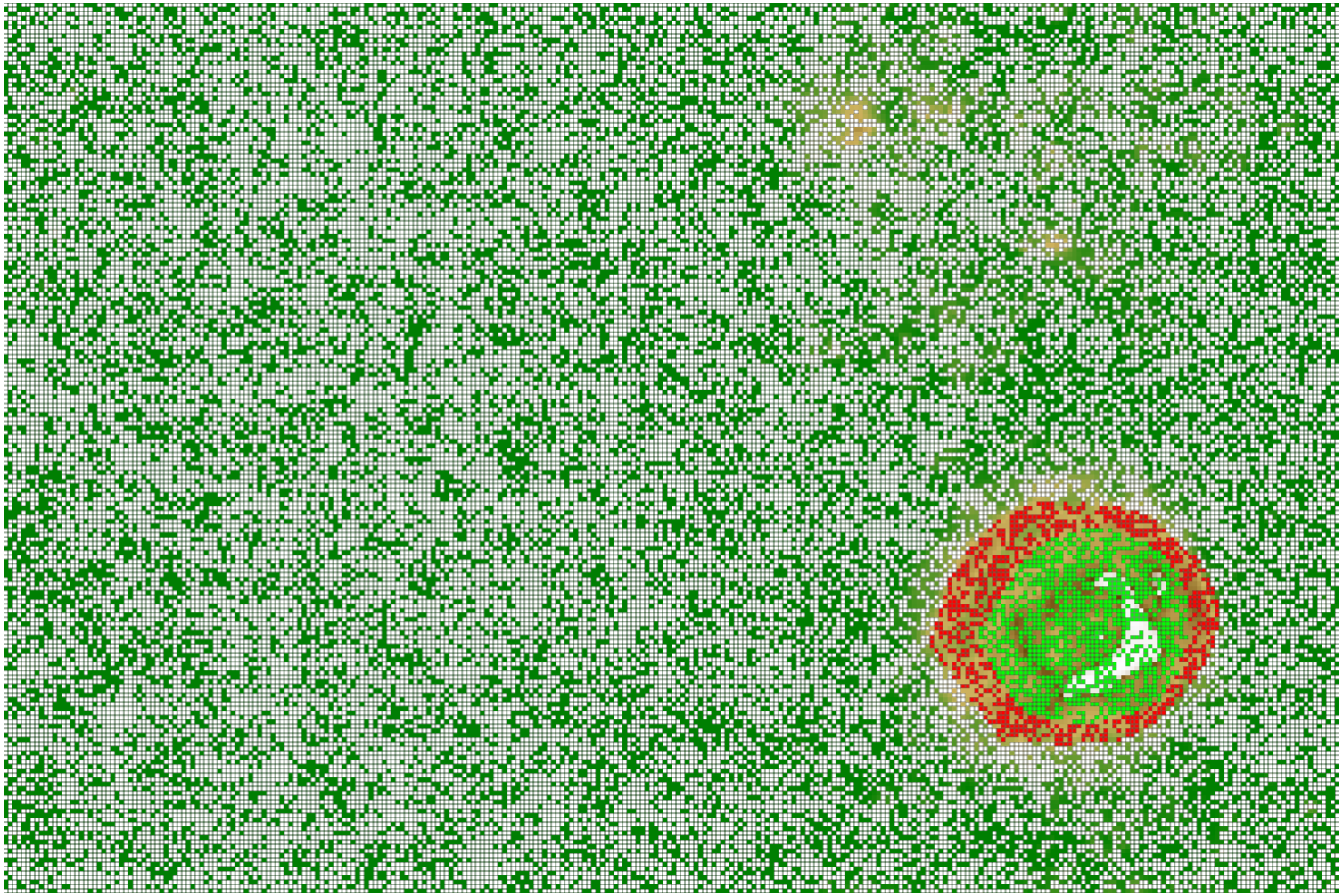
Now conceivable by human brain



Together ahead. RUAG

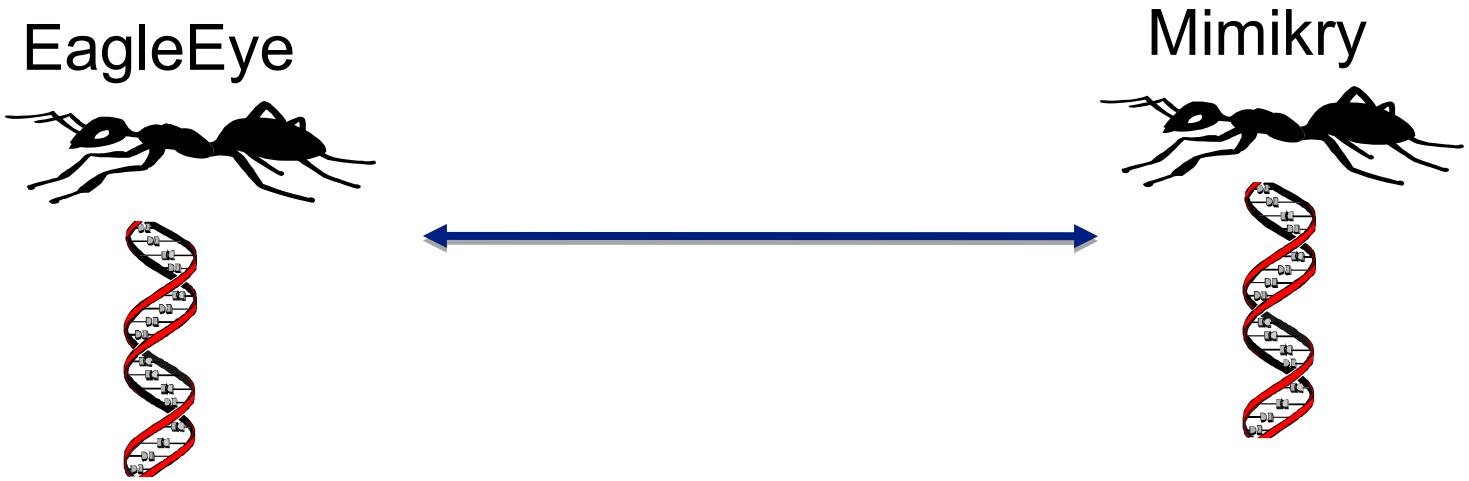
ESOM Pet Scan Analysis:

5 Dim Time dependent parameters



Together ahead. RUAG

Arms Race: Genetic Programming

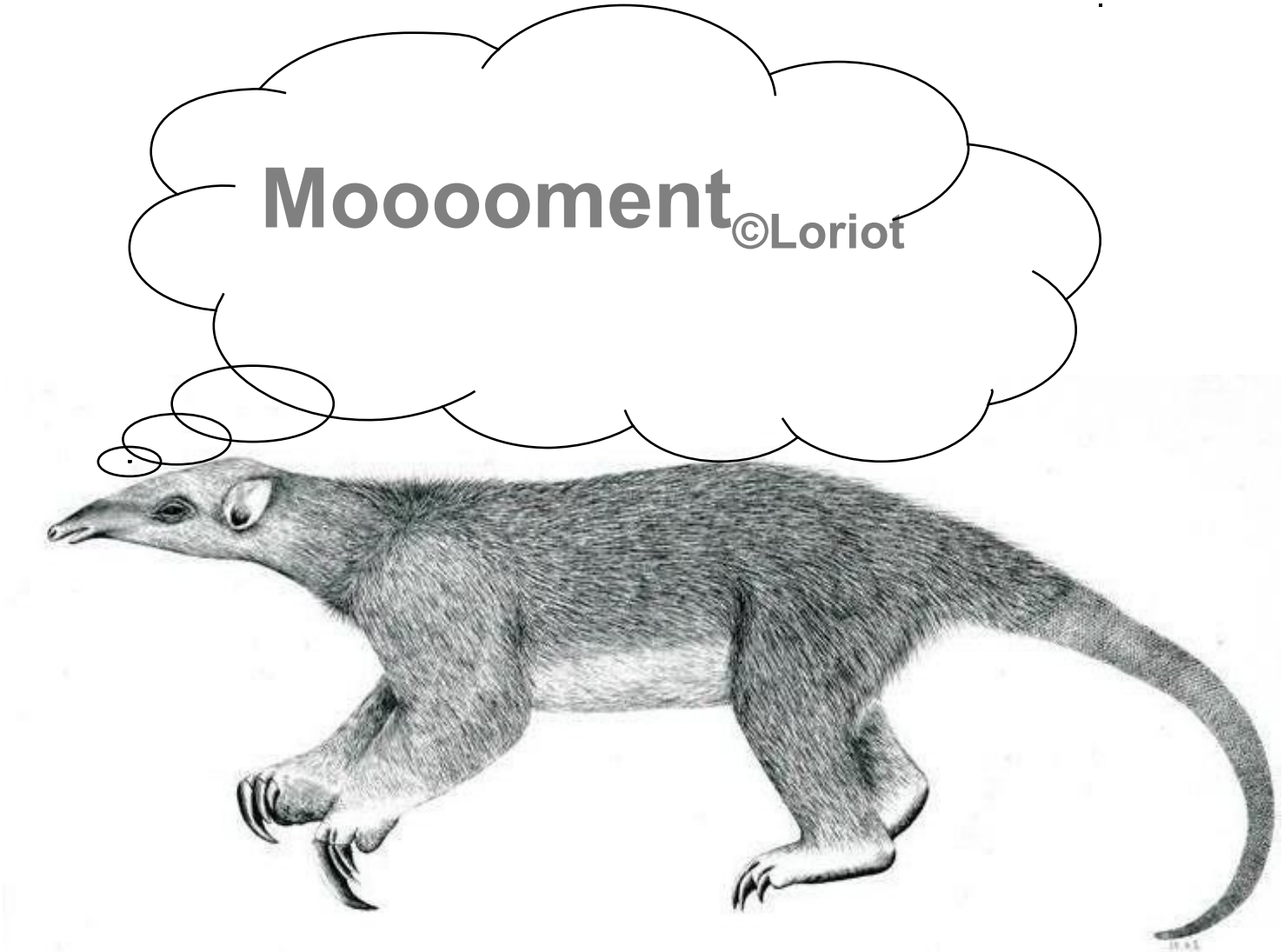


Defender: Detect the Undetecable

Attacker: Fly under Radar

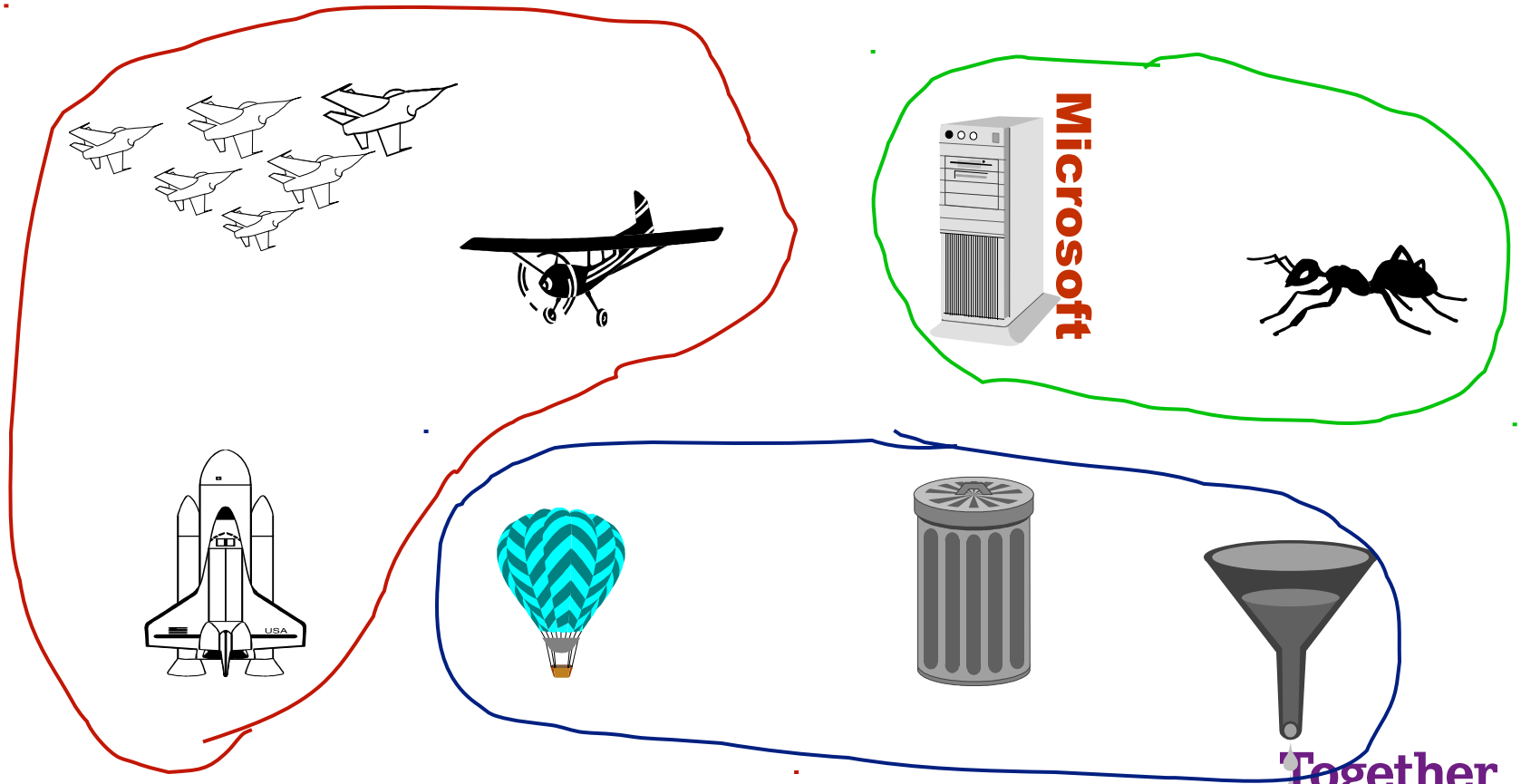
The problem of the missing performance function: bzip?!

Together ahead. RUAG



Together ahead. RUAG

Exercise: AI yourself



Exercise: Knowledge Extraction

Data: Bad Weather

- Low Pressure Rain Coat
- Clouds Storm
- Clouds Rain Noodle
- Sun Rain Clouds
- Spring Autumn
- Pressure sinking Heating

Data: Good Weather

- Sun Sea Bomb
- High Pressure Isobar Sun
- Pressure climb Grill Steak
- Sun Friends Beer
- Sun Beach Clouds
- Summer Sun

Good Weather = Sun & / Rain

Words or Word Chains, which separate most of data sets correctly, have highest Information gain.

Questions / Comments

Try me !
Who wants Bootcamp?

stefan.burschka@ruag.com

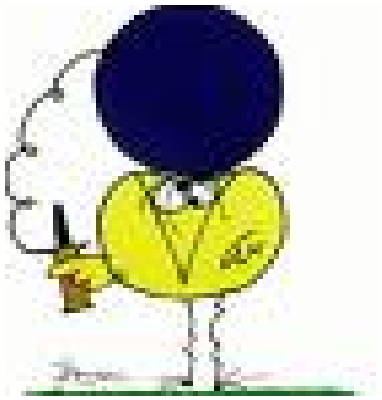
https://tranalyzer.com
Google: Dataming for Hackers

Together ahead. RUAG

Prescription for Desaster

"We've arranged a civilization in which the most crucial elements profoundly depend on science and technology. We have also arranged things so that almost no one understands science and technology. This is a prescription for disaster. We might get away with it for a while, but sooner or later this combustible mixture of ignorance and power is going to blow up in our faces."

Carl Sagan

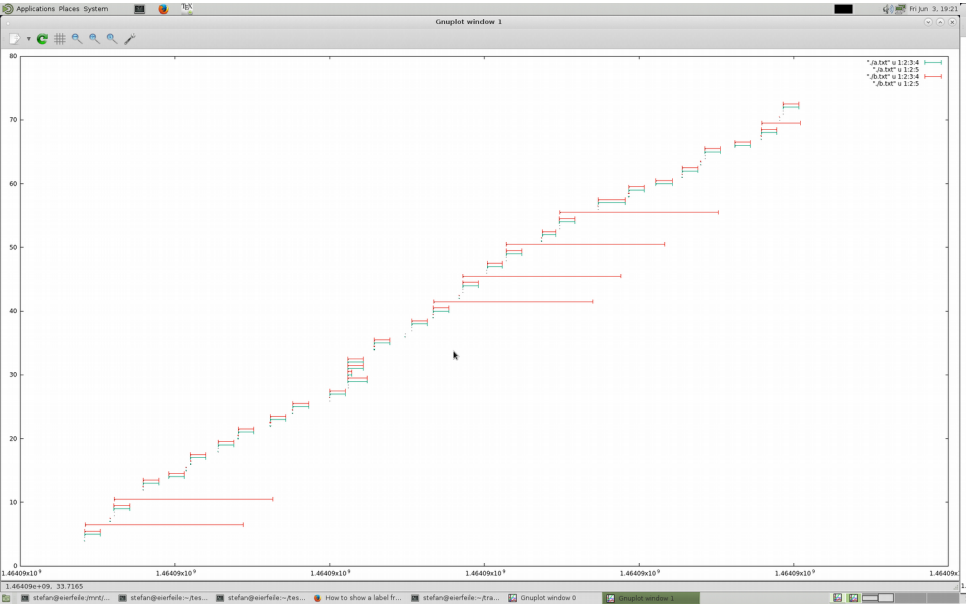


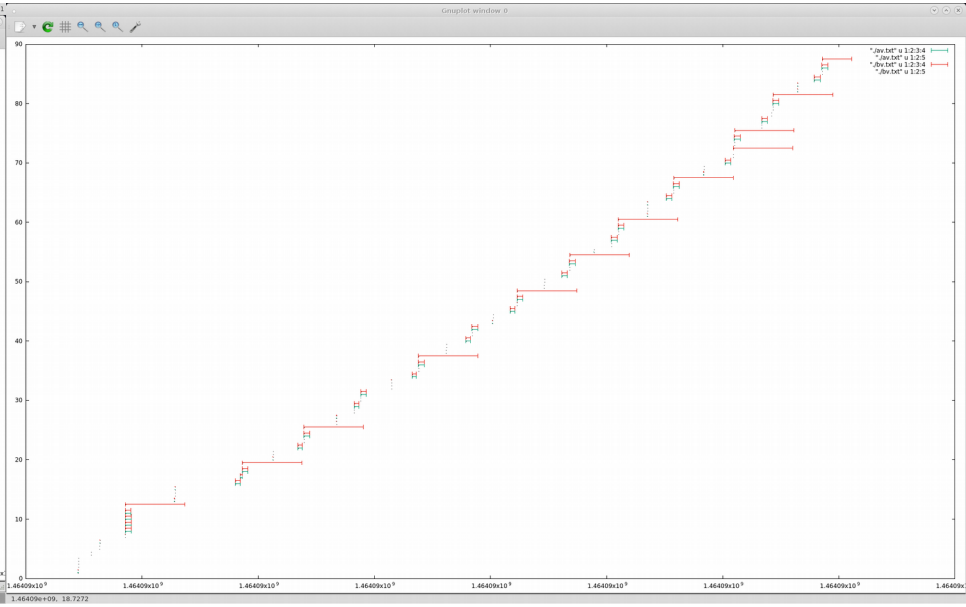
Flow Timeline, as Training Data Polygraph

Malware & NoMalware should look the same

No Malware

Malware





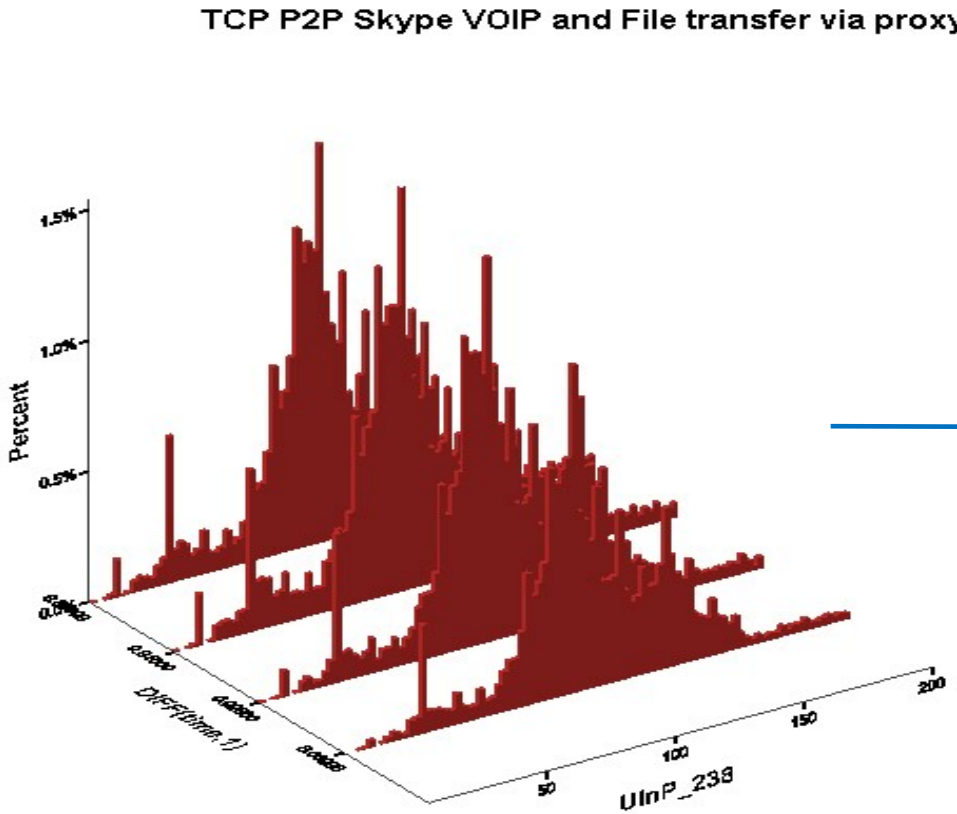
Somebody is cheating here !!!!!!!!!!

Costs you 95% of your prep time!

Together ahead. RUAG

3D Statistical Application / User profiling

Packet length-Interdistance Statistics: Fingerprinting



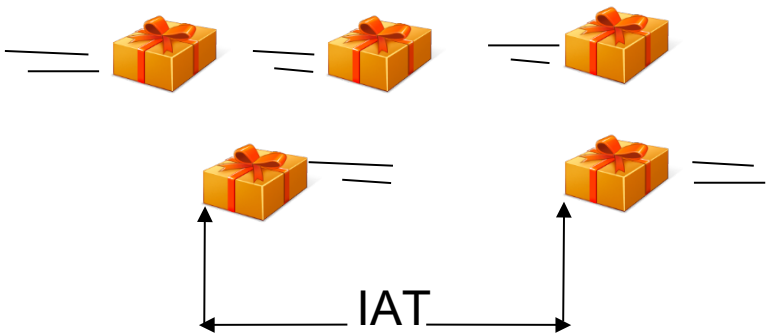
Vulnerable against TM

Chance for Covert Channels

Together ahead. RUAG

Covert Channel Detection

- Conventional Storage Channels
  - Steganography: Header, Content
  - Timing Channels (Pkt Length, IAT)
  - IAT Statistical Undetectable: Anael Fiaux Thesis EPFL



Together ahead. RUAG