



Sandro Etalle
s.etalle@tue.nl



Bram C.M. Cappers
b.c.m.cappers@tue.nl



Jarke J. van Wijk
j.j.v.wijk@tue.nl

SPYSPOT - NETWORK TRAFFIC ANALYSIS USING DEEP PACKET INSPECTION AND DATA VISUALIZATION

Advanced Persistent Threats

2010

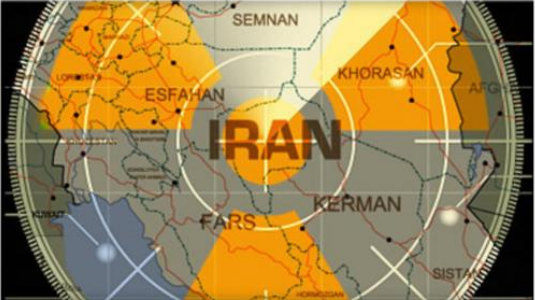
APT:

Targeted attack
designed to strike
once

CBSNEWS Video US World Politics Entertainment Health MoneyW

By CBSNEWS / CBS/AP November 29, 2010, 3:19 PM

Iran Confirms Stuxnet Worm Halted Centrifuges



generic Iran iranian map nuclear nukes target crosshair radar sonar missile bomb CBS/ISTOCKPHOTO

Iran's president has confirmed for the first time that a computer worm affected centrifuges in the country's uranium enrichment program.

Iran has previously denied the Stuxnet worm, which experts say is calibrated to destroy centrifuges, had caused any damage, saying they uncovered it before it could have any effect.

But President Mahmoud Ahmadinejad has said it "managed to create problems for a limited number of our centrifuges." Speaking to a press conference Monday, he said the problems were resolved.

Earlier in November, U.N. inspectors found Iran's enrichment program temporarily shut down, according to a recent report by the U.N. nuclear watchdog. The extent and cause of the shutdown were not known, but speculation fell on Stuxnet.

The finding was contained in a report from the International Atomic Energy Agency for the U.N. Security Council and the 35 IAEA board member nations.

Diplomats who spoke to the Associated Press that week said they did not know why the thousands of centrifuges stopped turning out material that Iran says it

“Worm
calibrated to
destroy
centrifuges”

Advanced Persistent Threats

2013

APT:

Advanced

Designed by highly organized criminal organizations

SECURELIST Kaspersky Lab

Video US World Politics Entertainment Health MoneyW

Menu

The Great Bank Robbery: the Carbanak APT

By GREAT on February 16, 2015

How the Carbanak cybergang stole \$1bn A targeted attack on a bank

- 1. Infection**
Carbanak backdoor sent as an attachment
Bank employee
Emails with exploits
Credentials stolen
100s of machines infected in search of the admin PC
- 2. Harvesting Intelligence**
Intercepting the clerks' screens
Cash transfer systems
- 3. Mimicking the staff**
How the money was stolen
 - Online-banking: Money was transferred to fraudulent accounts
 - E-payment systems: Money was transferred to banks in China and the US
 - Inflating account balances: The extra funds were pocketed via a fraudulent transaction
 - Controlling ATMs: Orders to dispense cash at a pre-determined time

© 2015 Kaspersky Lab

GREAT KASPERSKY

The story of Carbanak began when a bank from Ukraine asked us to help with a forensic investigation. Money was being mysteriously stolen from ATMs. Our initial thoughts tended towards the Tyupkin malware. However, upon investigating the hard disk of the ATM system we couldn't find anything except a rather odd VPN configuration (the netmask was set to 172.0.0.0).

At this time we regarded it as just another malware attack. Little did we know then that a few months later one of our colleagues would receive a call at 3 a.m. in the middle of the night. On the phone was an account manager, asking us to call a certain number as matter of urgency. The person at the end of the line was the CSO of a Russian bank. One of their systems was alerting that data was being sent from their Domain Controller to the People's Republic of China.

Each bank robbery took 2-4 months, from infecting the first computer to

“Attack took 2-4 months from infecting the first computer to cashing the money out”

Advanced Persistent Threats

2015

APT:
Persistent

Use of domain
knowledge to
stay under radar



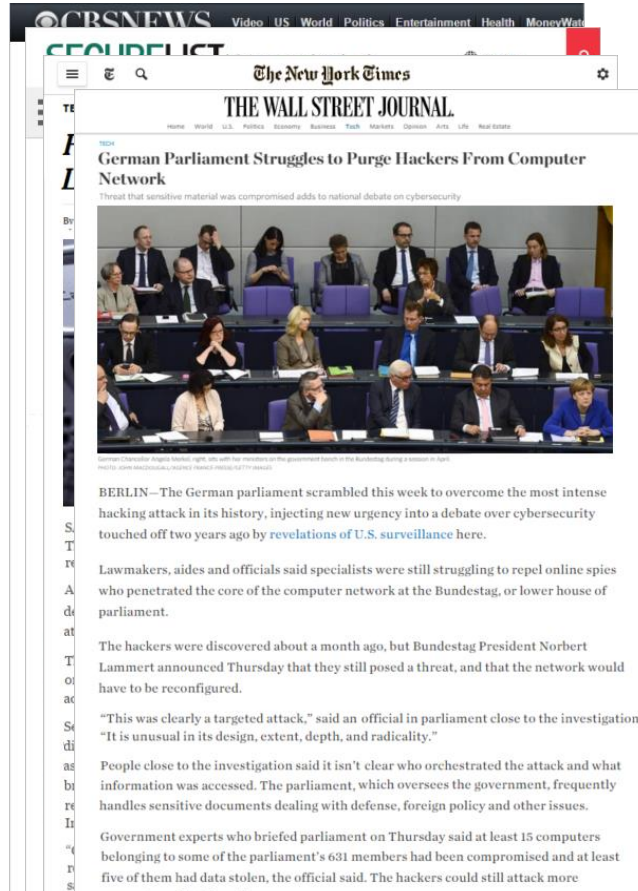
“Hackers
cloaked the
source of the
attacks by
routing
them via
universities”

Advanced Persistent Threats

2015

APT:
Threat hidden in
payload

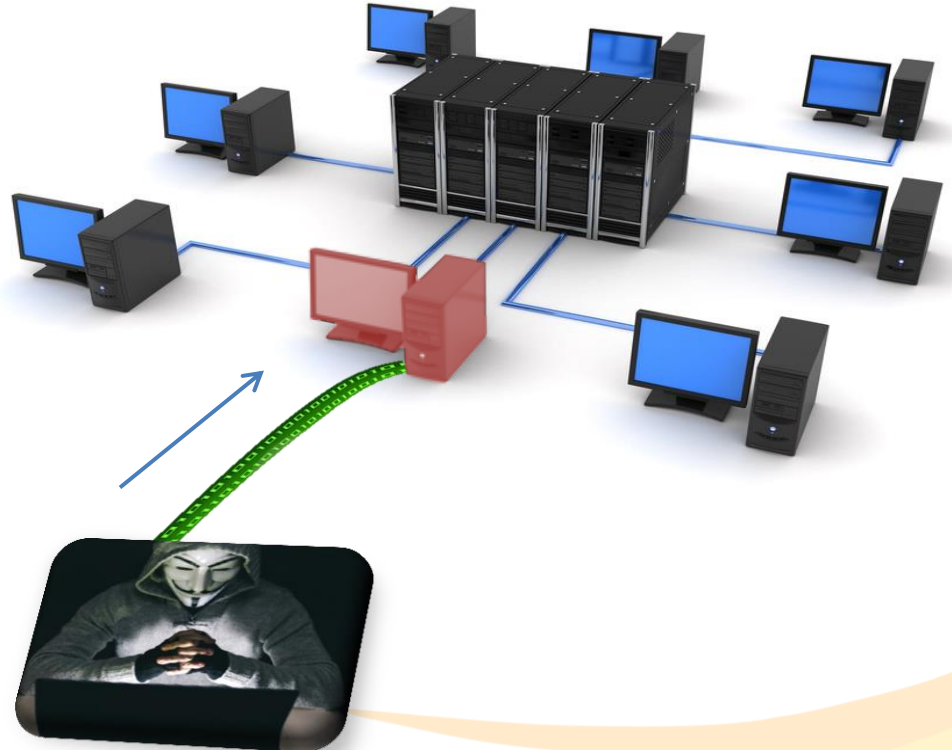
Flow analysis can't
detect them



“We haven’t
seen this
type of
attack
before.”

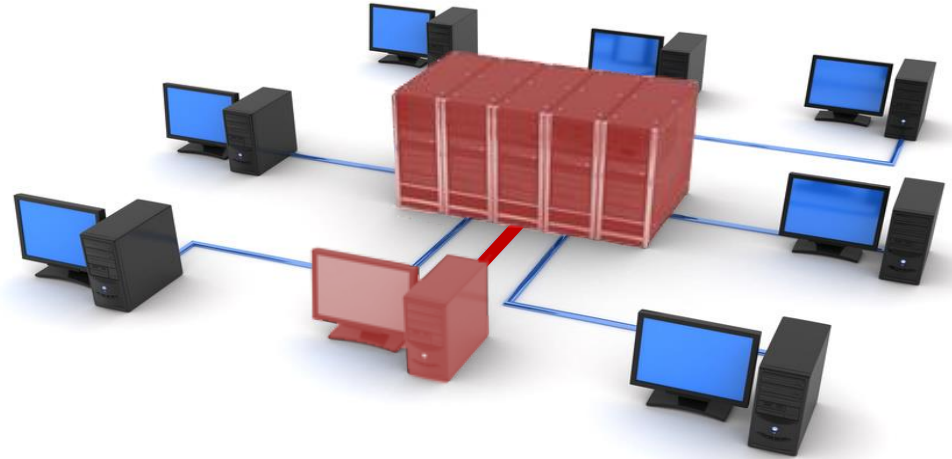
Advanced Persistent Threats

- Infiltration
- Expansion
- Sabotage



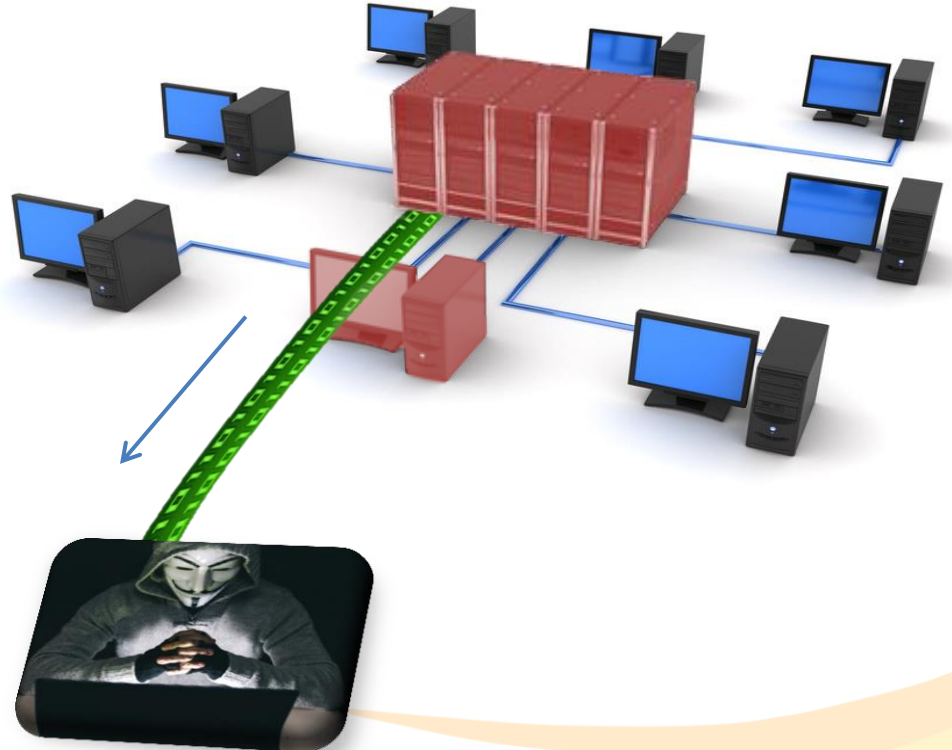
Advanced Persistent Threats

- Infiltration
- **Expansion**
- Sabotage



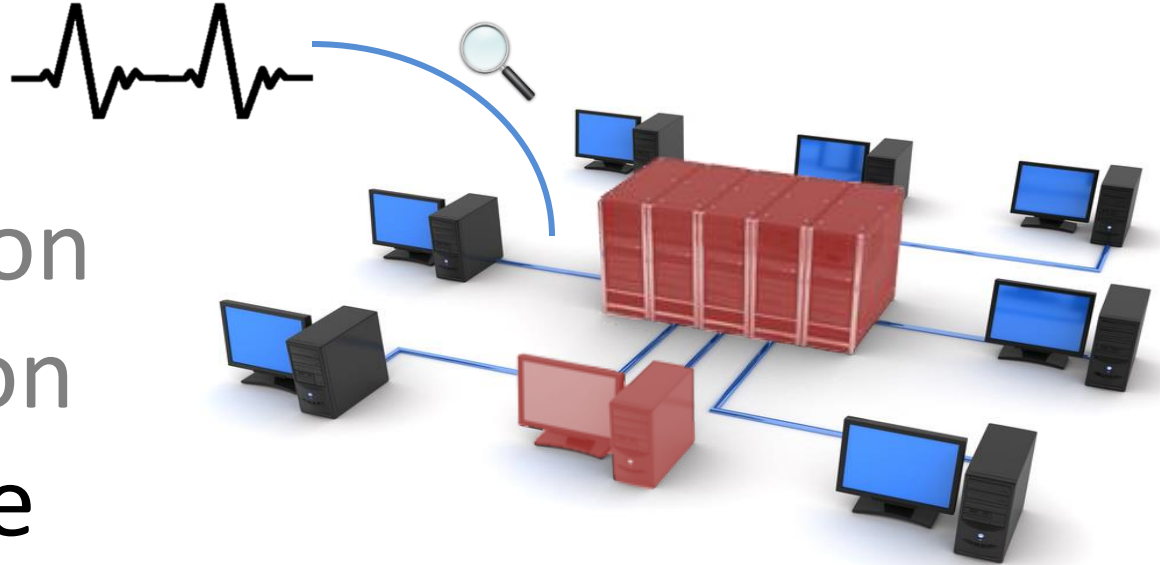
Advanced Persistent Threats

- Infiltration
- Expansion
- Sabotage
 - Espionage

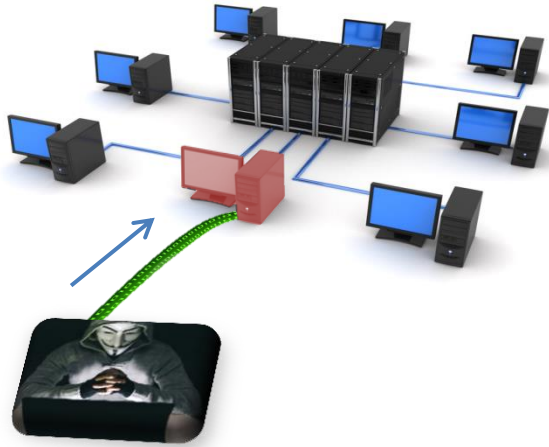


Advanced Persistent Threats

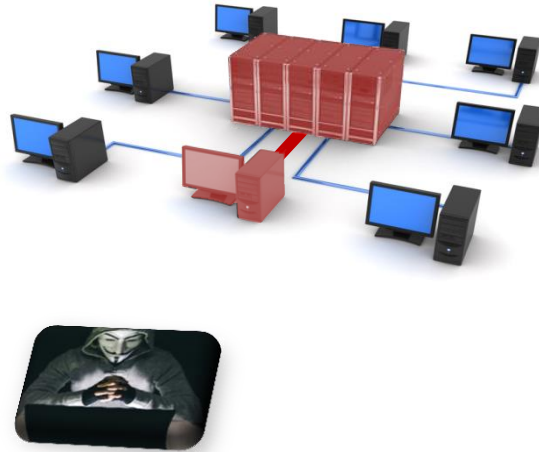
- Infiltration
- Expansion
- Sabotage
 - Espionage
 - Disrupting services



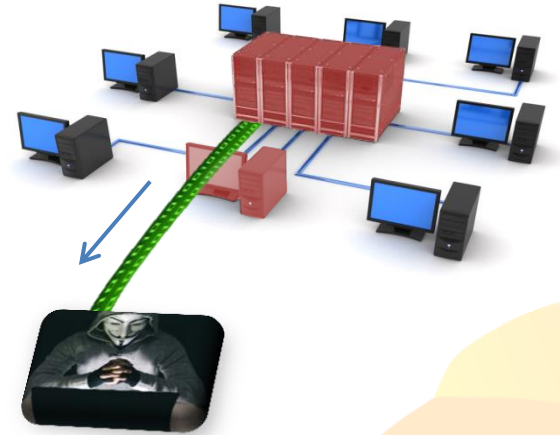
Advanced Persistent Threats



Infiltration



Expansion



Sabotage

How to detect APTs?

Analyze traffic content

Byte Analysis



Flow Analysis



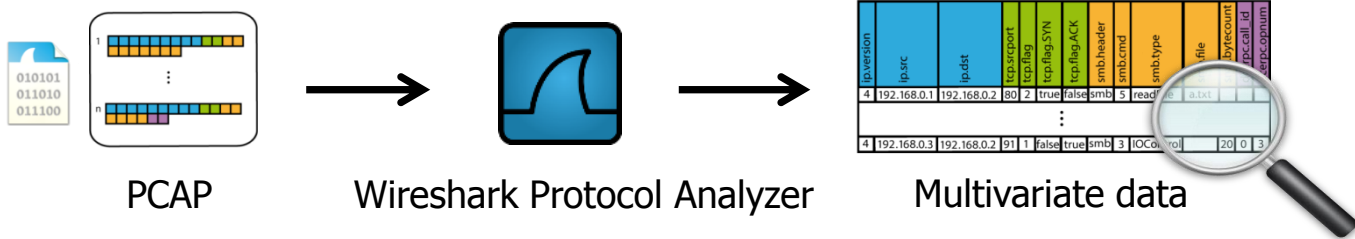
Semantic Analysis



How to obtain this data?

#Attributes	#Protocols
1	none
5-20	2 (TCP/IP)
50-200/ protocol	≥ 3

Data model

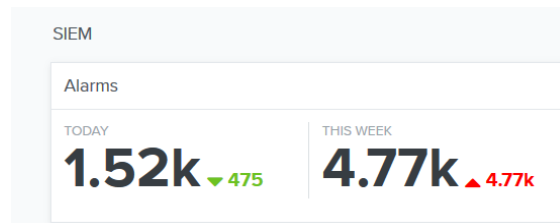


ip.version	ip.src	ip.dst	tcp.srcport	tcp.flag	tcp.flag.SYN	tcp.flag.ACK	smb.header	smb.cmd	smb.type	smb.file	smb.bytecount	dcerpc.call_id	dcerpc.opnum
4	192.168.0.1	192.168.0.2	80	2	true	false	smb	5	readFile	a.txt			
⋮													
4	192.168.0.3	192.168.0.2	91	1	false	true	smb	3	IOControl		20	0	3

WHY DO WE NEED VISUALIZATION?

Too many alerts!

- Explain anomalies



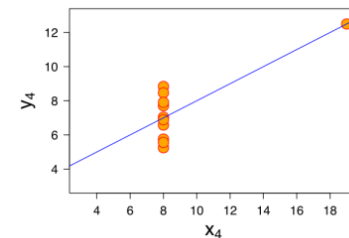
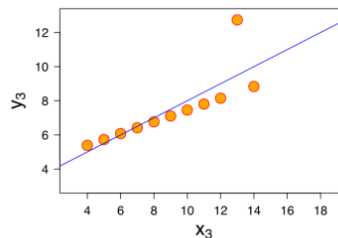
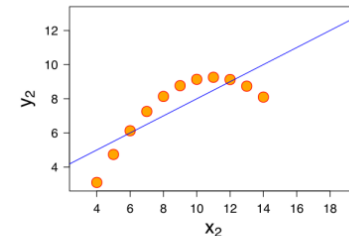
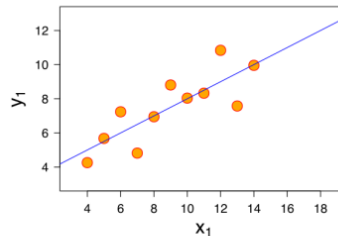
WHY DO WE NEED VISUALIZATION?

Too many alerts!

- Explain anomalies

Abuse

- Human cognition
- Domain knowledge



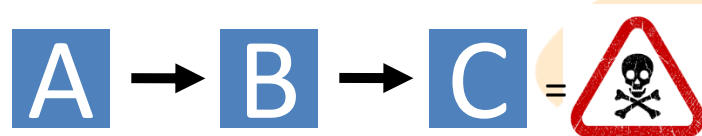
WHY DO WE NEED VISUALIZATION?

Too many alerts!

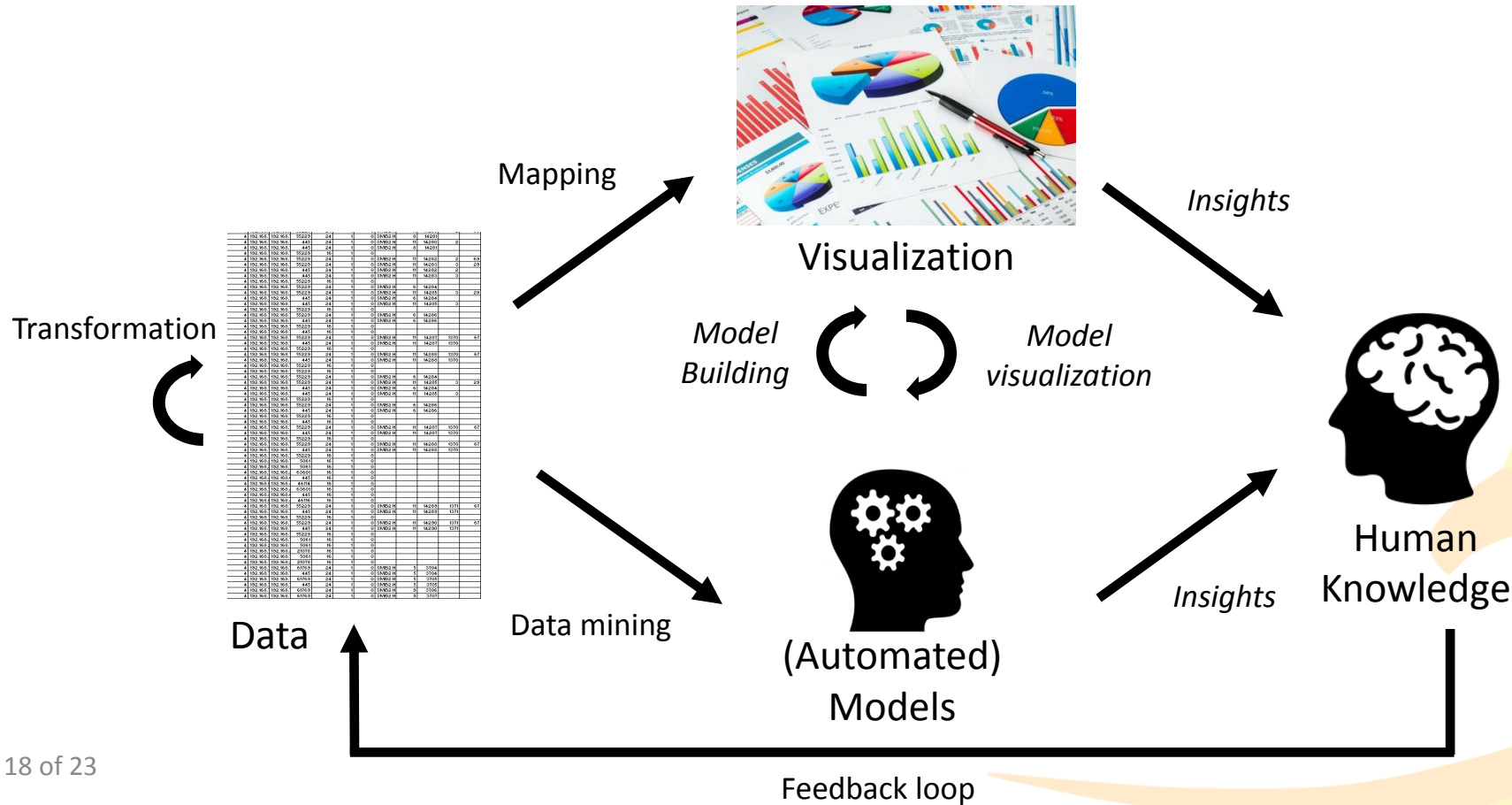
- Explain anomalies

Abuse

- Human cognition
- Domain knowledge



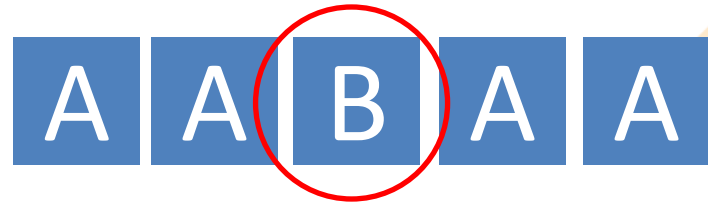
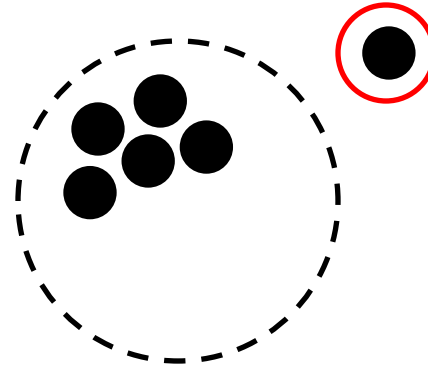
Visual Analytics



ANOMALY DETECTION

Anomalies:

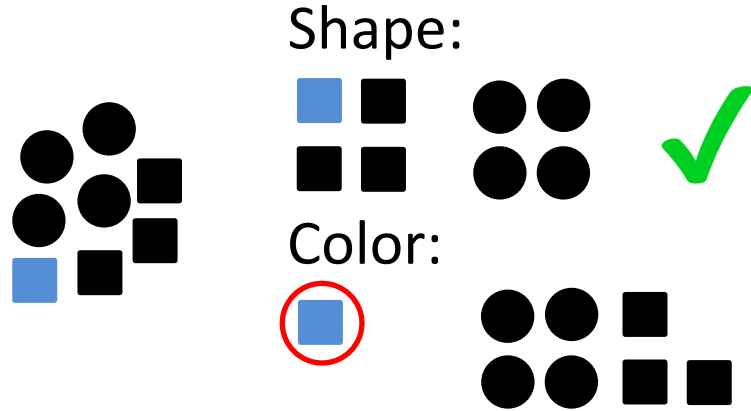
- Point
- Contextual
- Collective



ANOMALY DETECTION

Anomalies:

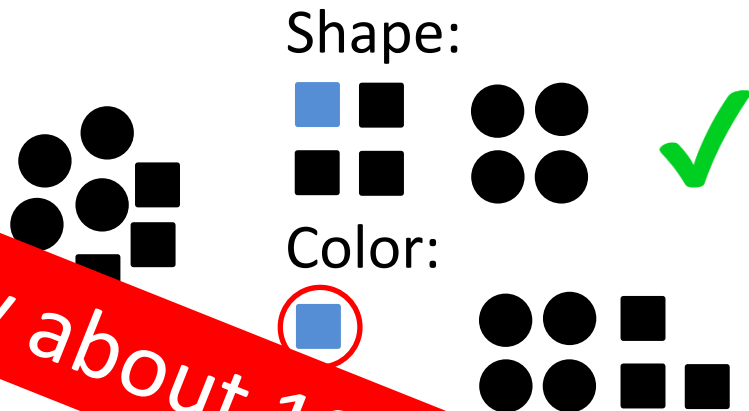
- Point
- Contextual
- Collective



ANOMALY DETECTION

Anomalies:

- Point
- Contextual
- Collective



How about 100 attributes?

A B A B A B

User 1: A A A

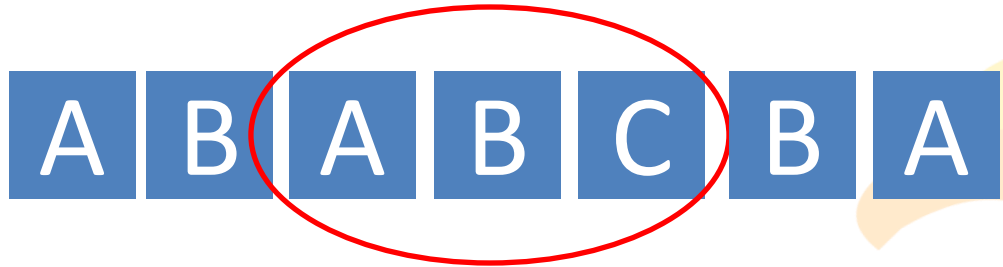
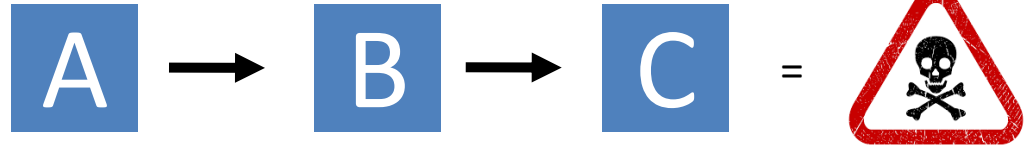
User 2: B B A B B

ANOMALY DETECTION

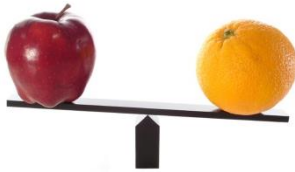
Anomalies:

- Point
- Contextual
- Collective

Knowledge:



DIFFERENT STRATEGIES



- Data-driven
 - What does the data want to be?
- Alert-driven
 - What does machine learning say?
- Knowledge-driven
 - Define what you know – Discover the unknown

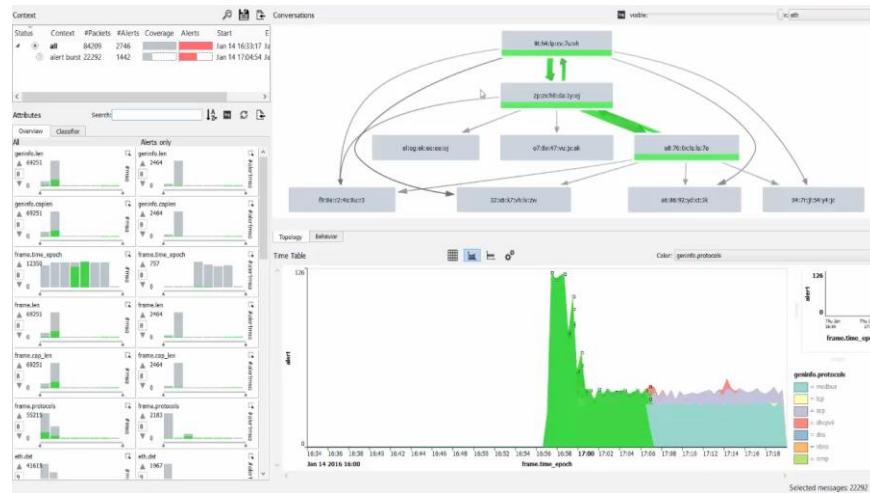
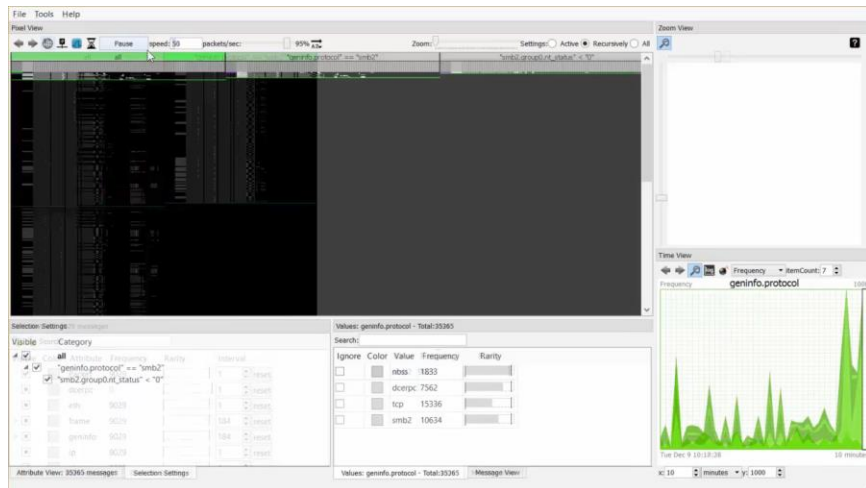
DIFFERENT STRATEGIES



Data-driven



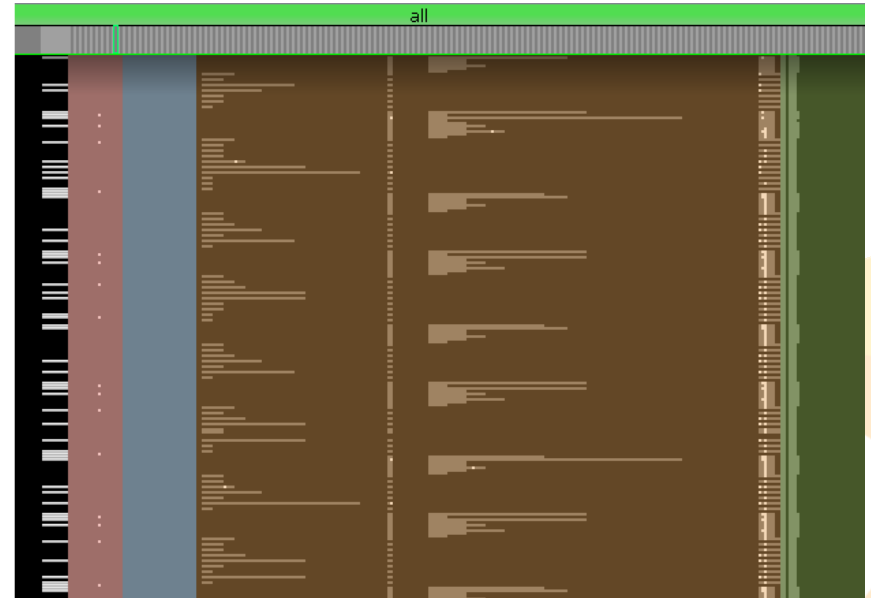
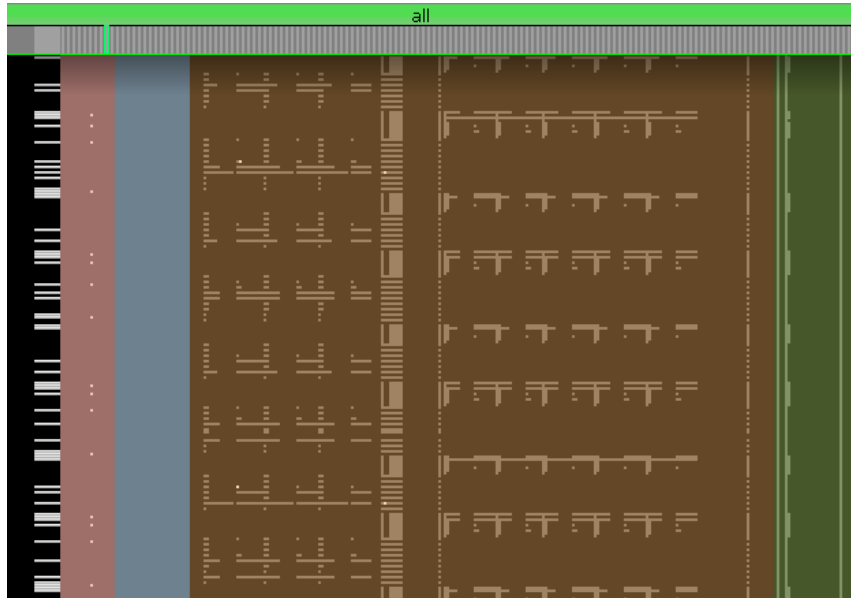
Alert-driven



Attribute Ordering



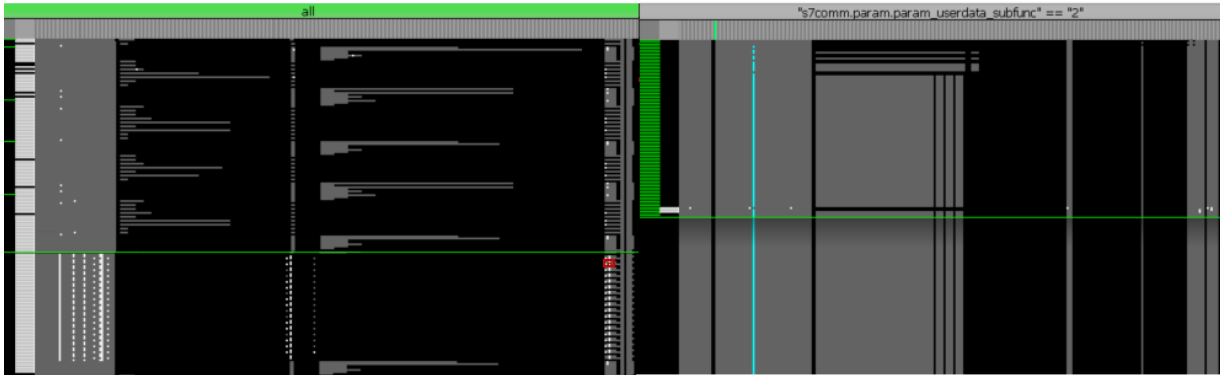
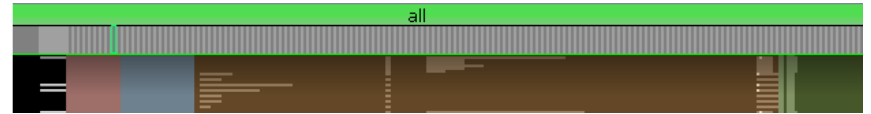
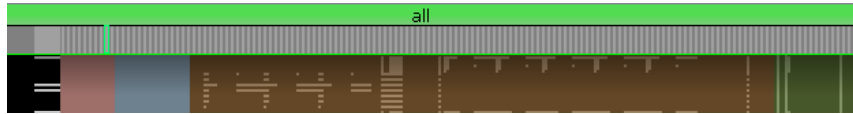
Data-driven



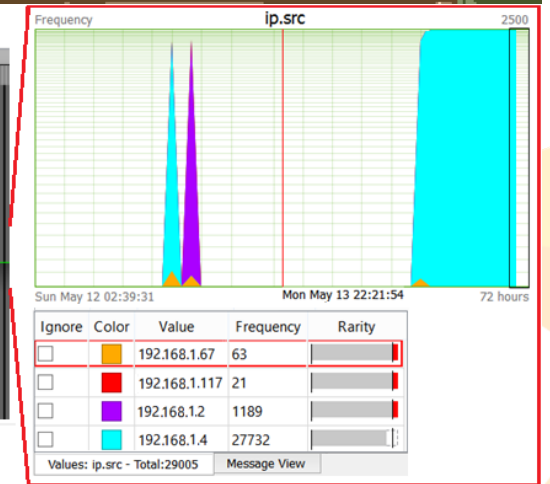
Attribute Ordering



Data-driven



(a)

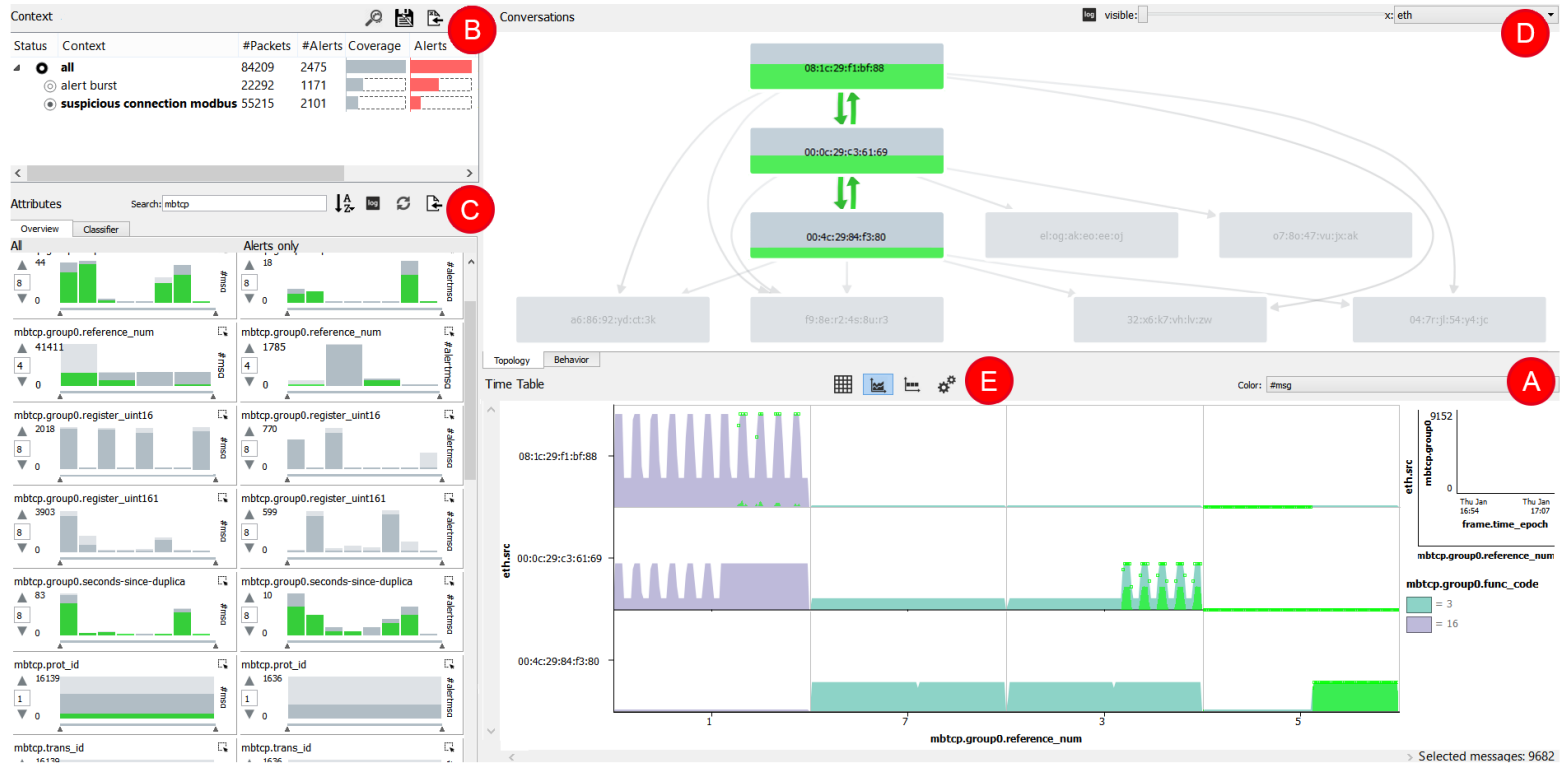


(b)

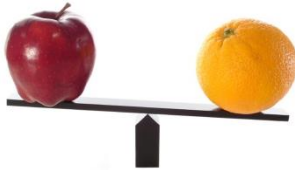
Man-in-the-middle



Alert-driven



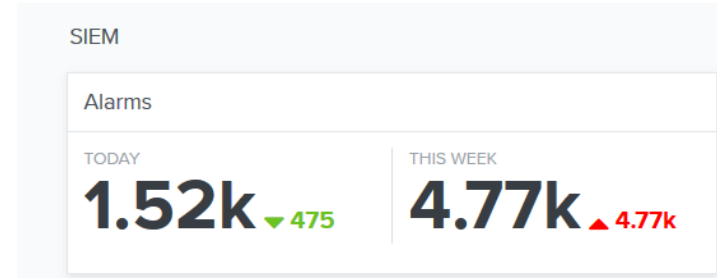
DIFFERENT STRATEGIES



- Data-driven
 - What does the data want to be?
- Alert-driven
 - What does machine learning say?
- Knowledge-driven
 - Define what you know – Discover the unknown

Motivation

- Machine learning is **difficult**
 - **Time-consuming** to setup
 - **Complex** to tune
 - **Horrible** to explain
- We need **faster** results!
 - No configuration, immediate results



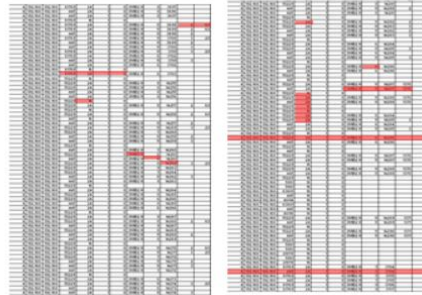
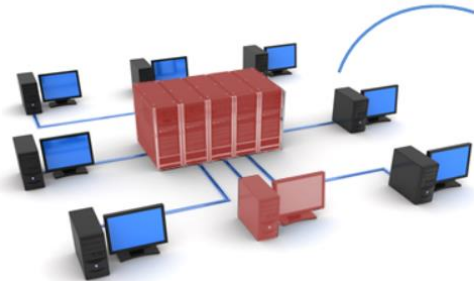
IEEE Visual Analytics Challenge 2017



Elegant Support
For Hypothesis Generation
and Testing

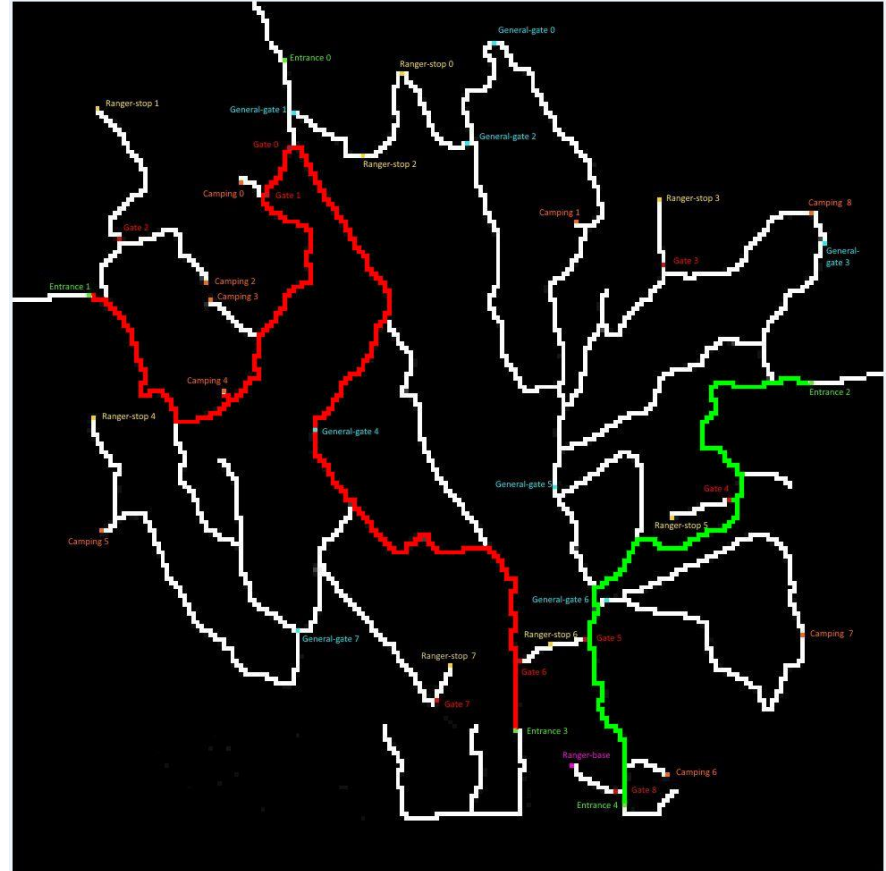
On average Industry and Academia teams worked **2 months** on the data to solve the challenge. We did it in **2 hours**...

“Great demonstration of a flexible existing tool EventPad to organize sequences of events in an intuitive way. Very powerful!”



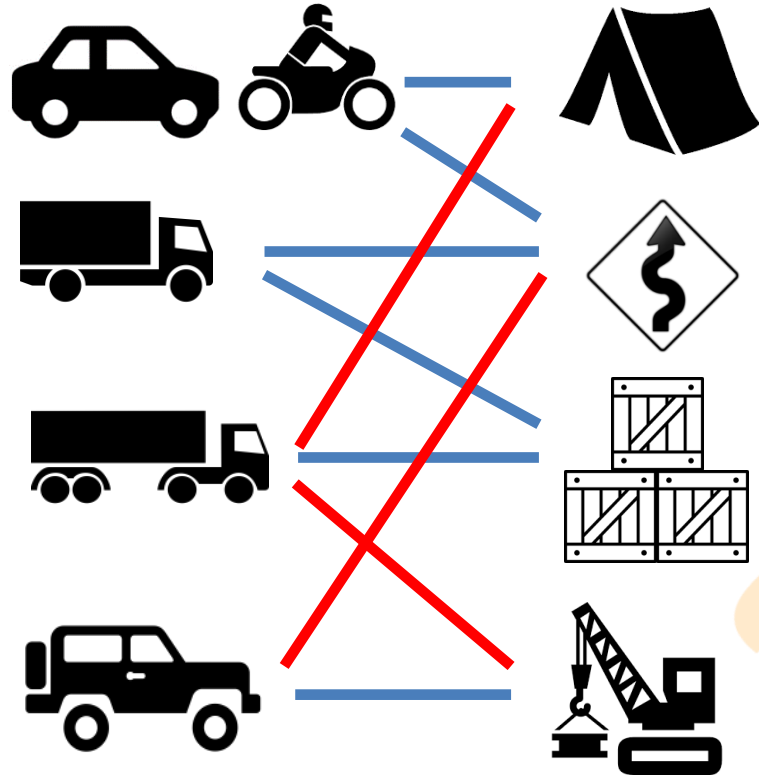
Event data

Time	Car-id	Type	Gate-name
00:43	262	4axle	Entrance1
01:03	262	4axle	General-gate1
01:06	262	4axle	Ranger-stop2
01:12	937	4axle	General-gate2
01:31	262	Car	Entrance3
01:53	937	4axle	Entrance2
01:56	937	Car	General-gate1
02:03	937	Car	Ranger-stop2
02:05	937	Car	General-gate2



Event data

Time	Car-id	Type	Gate-name
00:43	262	4axle	Entrance1
01:03	262	4axle	General-gate1
01:06	262	4axle	Ranger-stop2
01:12	937	4axle	General-gate2
01:31	262	Car	Entrance3
01:53	937	4axle	Entrance2
01:56	937	Car	General-gate1
02:03	937	Car	Ranger-stop2
02:05	937	Car	General-gate2



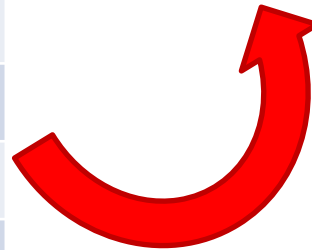
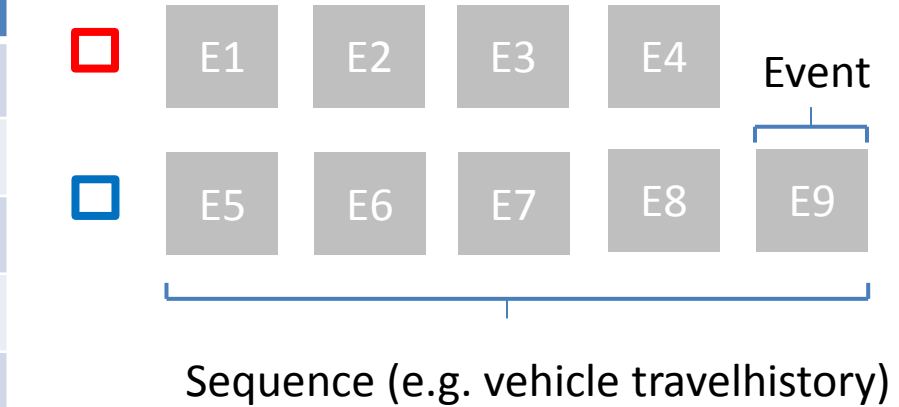
Event data

Time	Car-id	Type	Gate-name
00:43	262	4axle	Entrance1
01:03	262	4axle	General-gate1
01:06	262	4axle	Ranger-stop2
01:12	937	4axle	General-gate2
01:31	262	Car	Entrance3
01:53	937	4axle	Entrance2
01:56	937	Car	General-gate1
02:03	937	Car	Ranger-stop2
02:05	937	Car	General-gate2



Sequence analysis

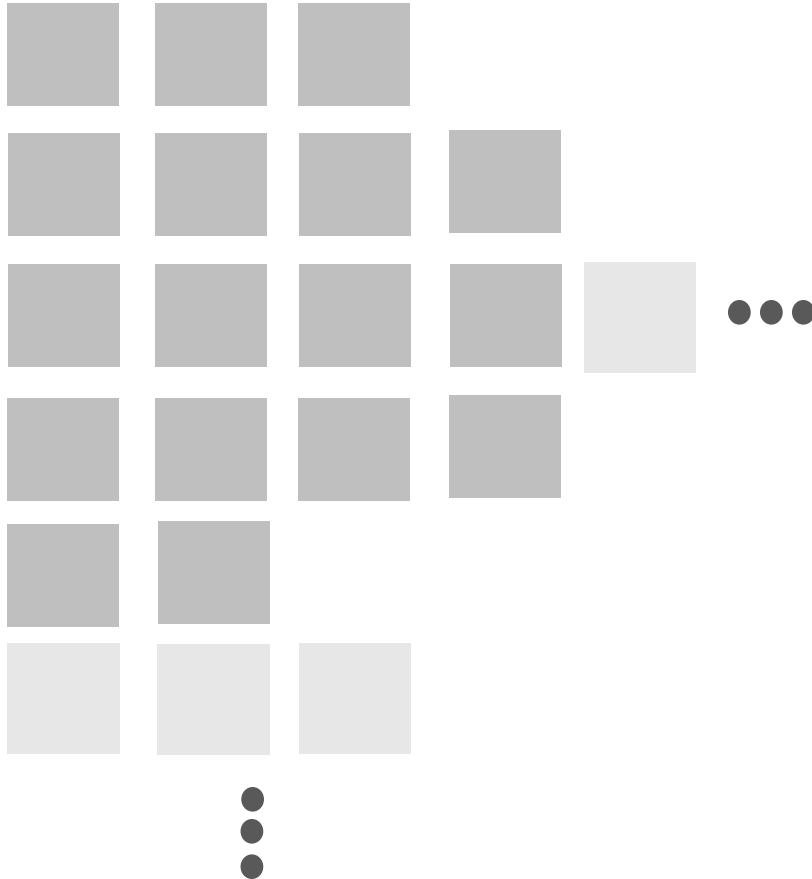
Time	Car-id	Type	Gate-name
00:43	262	4axle	Entrance1
01:03	262	4axle	General-gate1
01:06	262	4axle	Ranger-stop2
01:12	937	4axle	General-gate2
01:31	262	Car	Entrance3
01:53	937	4axle	Entrance2
01:56	937	Car	General-gate1
02:03	937	Car	Ranger-stop2
02:05	937	Car	General-gate2



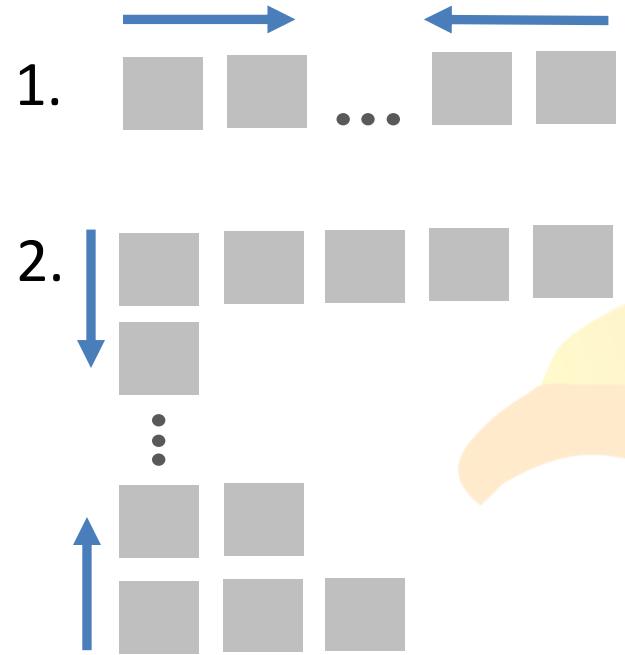
Sequence analysis



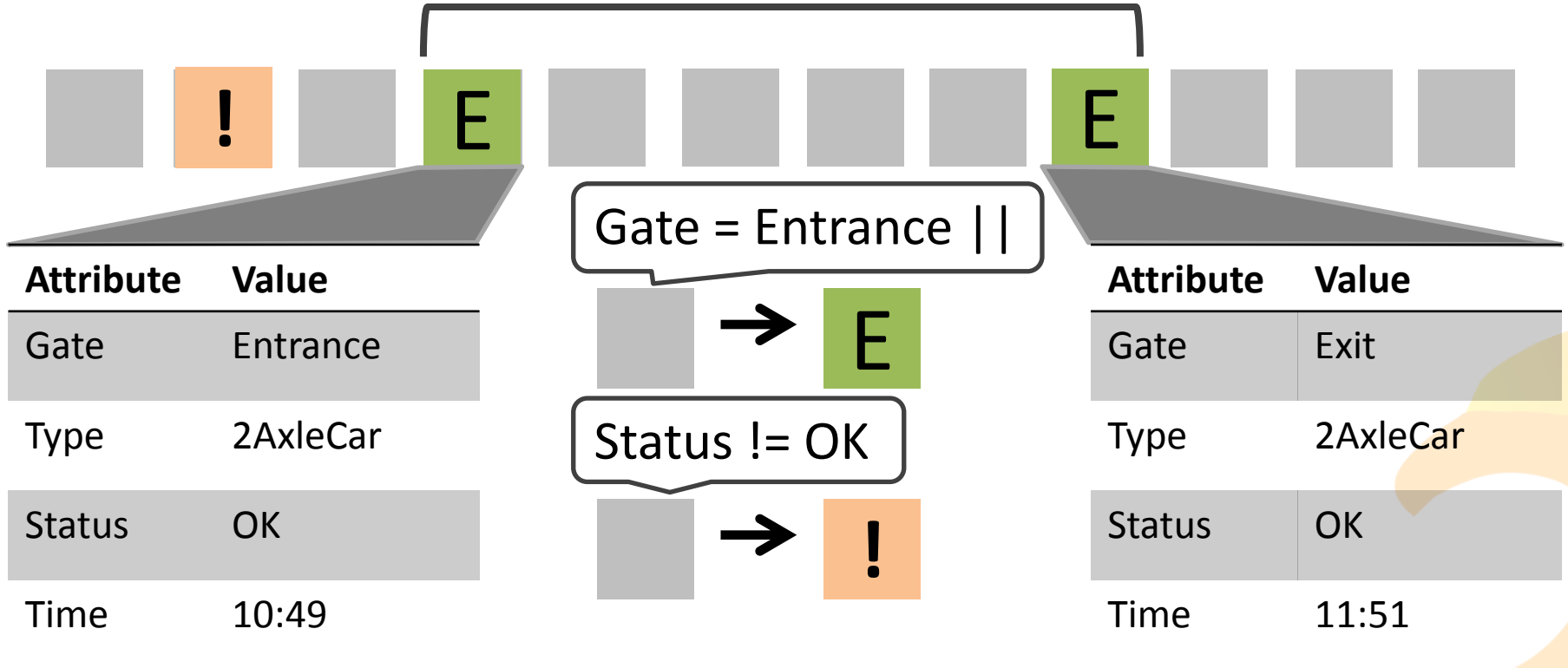
Result:



Simplify:



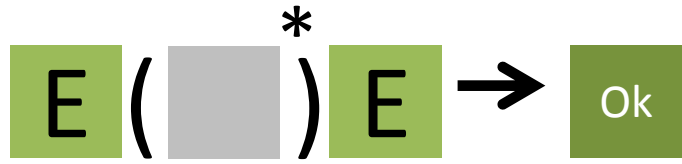
Conditional Formatting



Regular Expressions

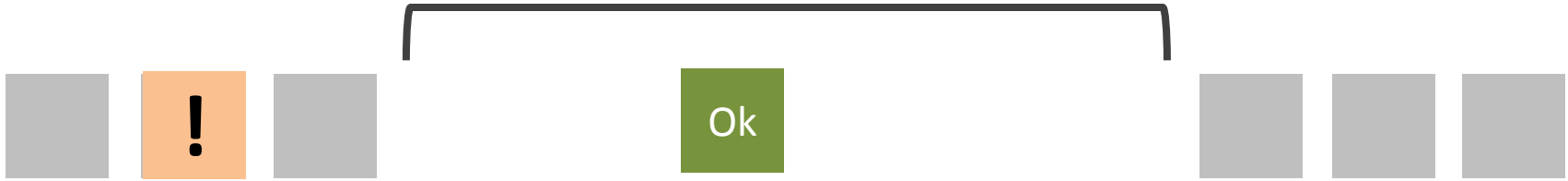


Attribute	Value
Gate	Entrance
Type	2AxleCar
Status	OK
Time	10:49



Attribute	Value
Gate	Exit
Type	2AxleCar
Status	OK
Time	11:51

Regular Expressions



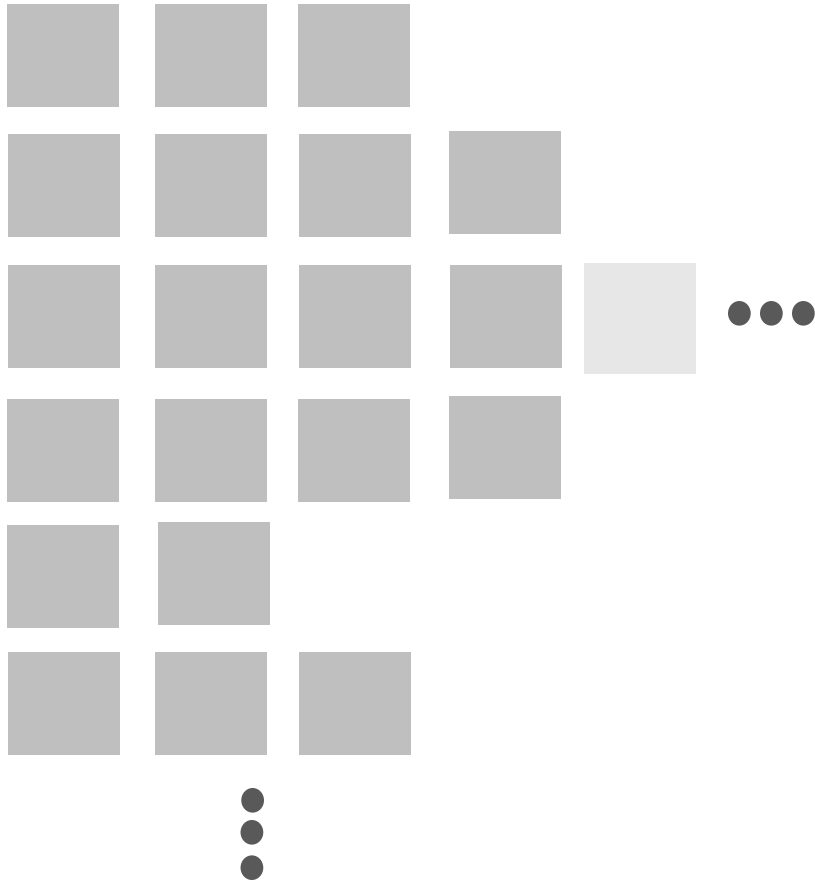
Attribute	Value
Gate	Entrance
Type	2AxleCar
Status	OK
Time	10:49



Attribute	Value
Gate	Exit
Type	2AxleCar
Status	OK
Time	11:51

Sequence analysis

Result:



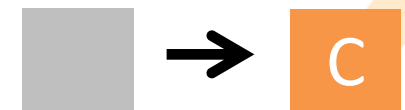
Simplify:



Type = Enter

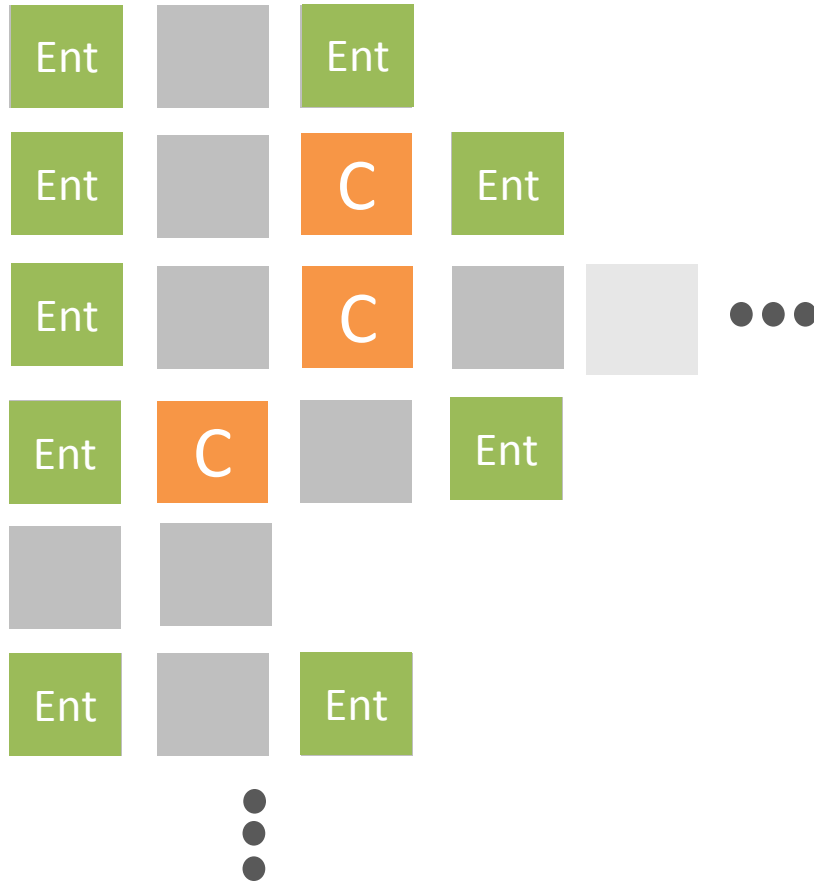


Type = Camping



Sequence analysis

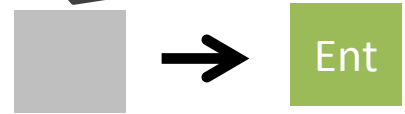
Result:



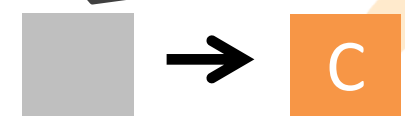
Simplify:



Type = Enter

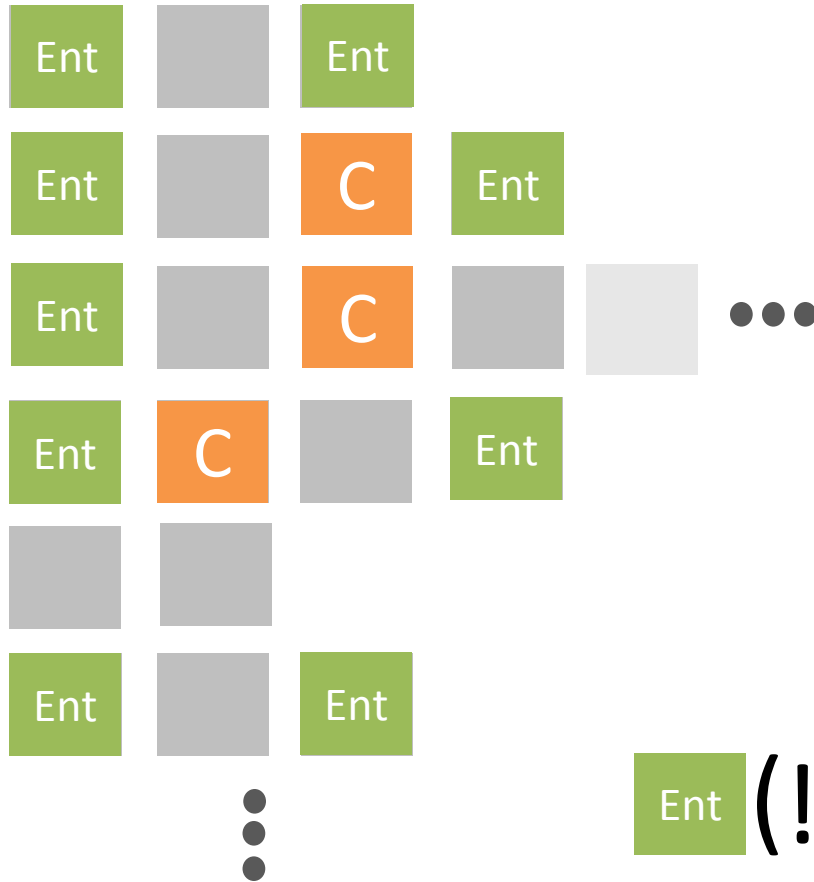


Type = Camping



Sequence analysis

Result:



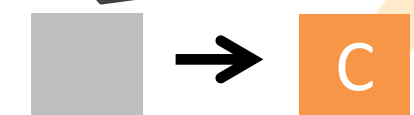
Simplify:



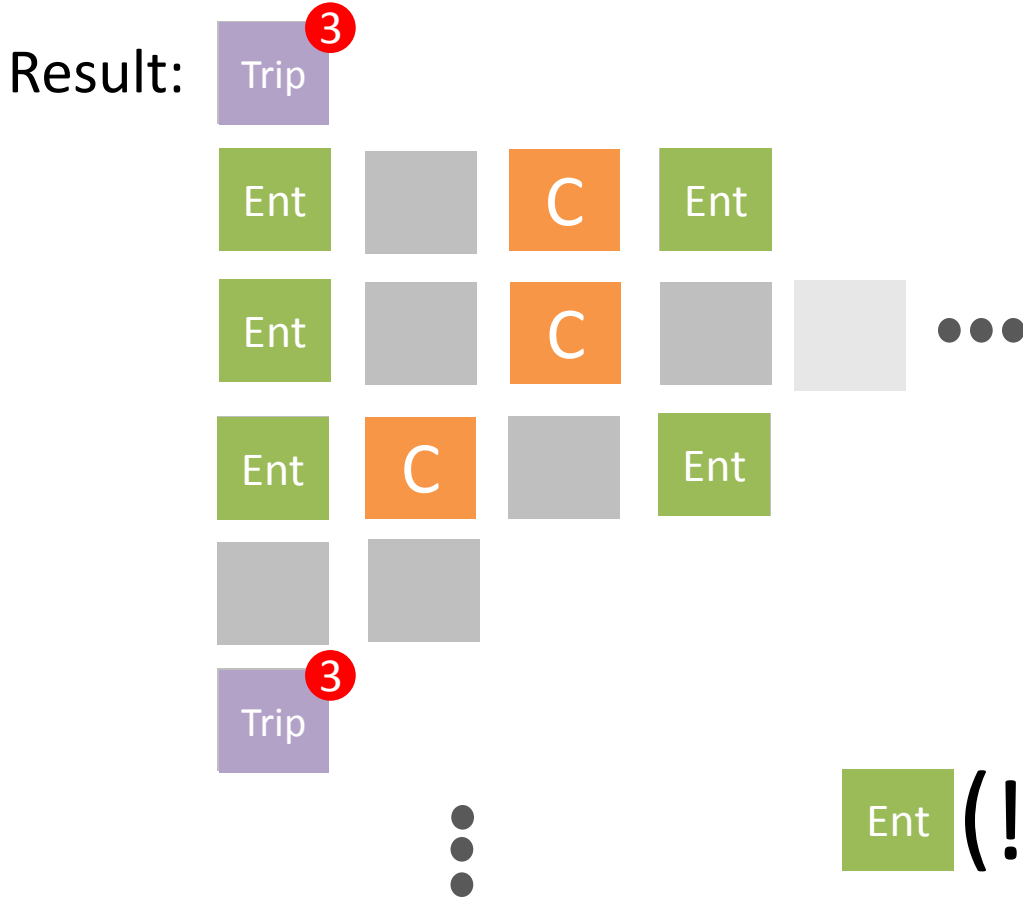
Type = Enter



Type = Camping



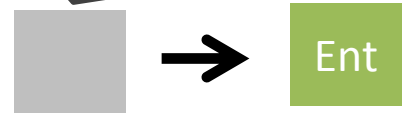
Sequence analysis



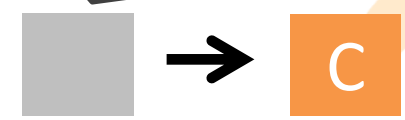
Simplify:

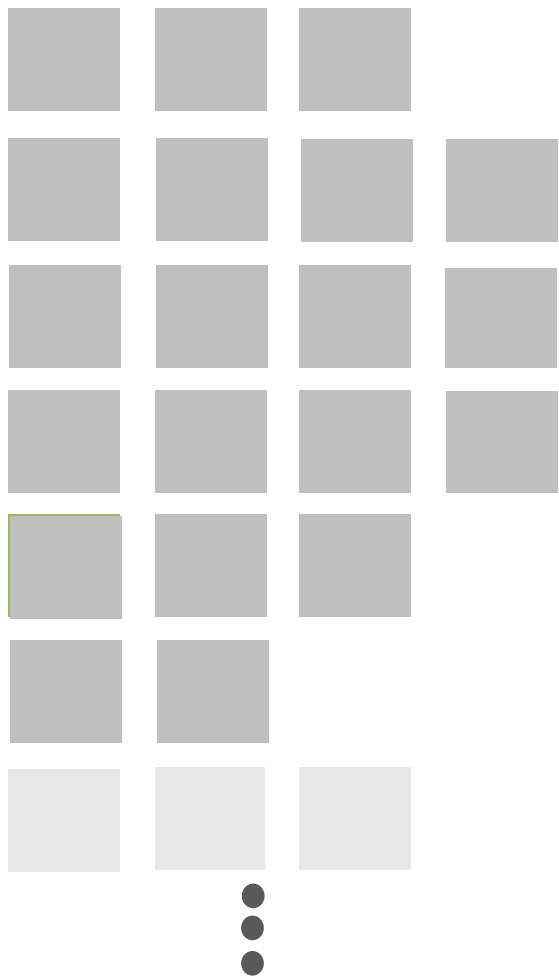


Type = Enter

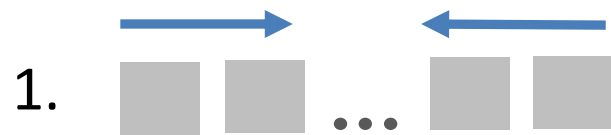


Type = Camping

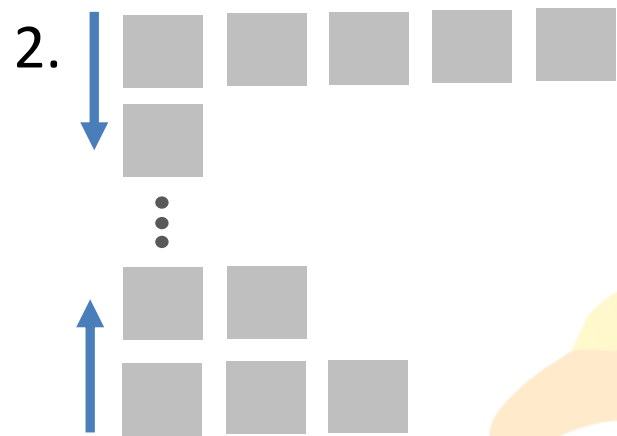


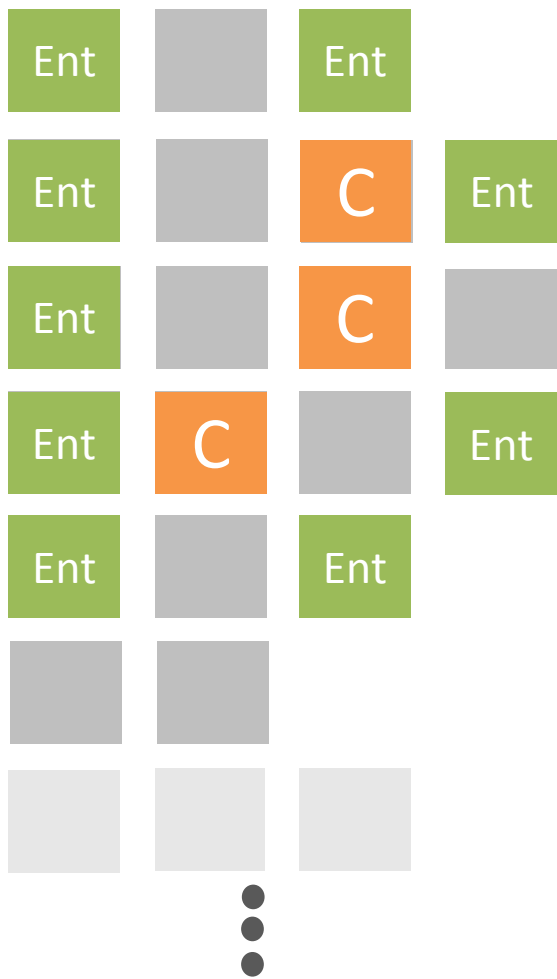


Rules:

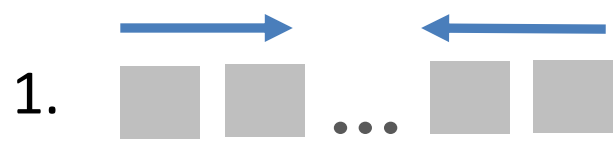


Aggregations:

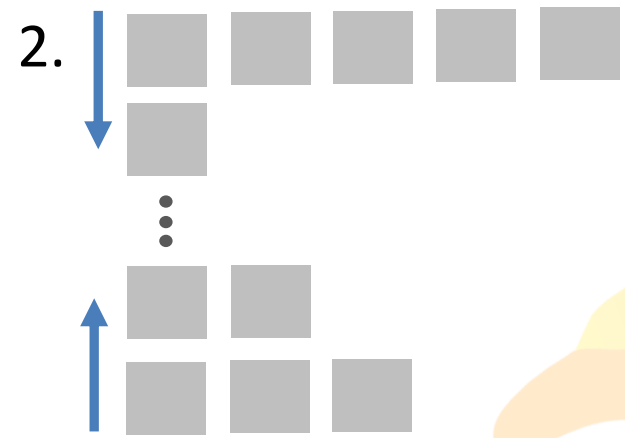


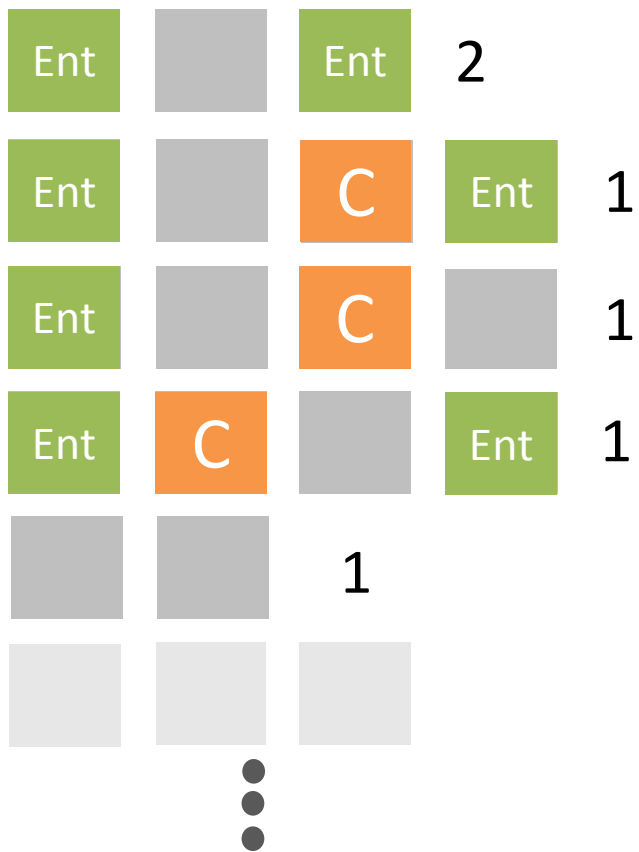


Rules:



Aggregations:

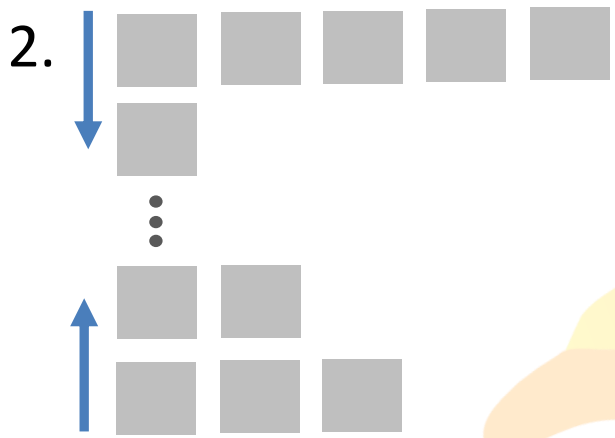


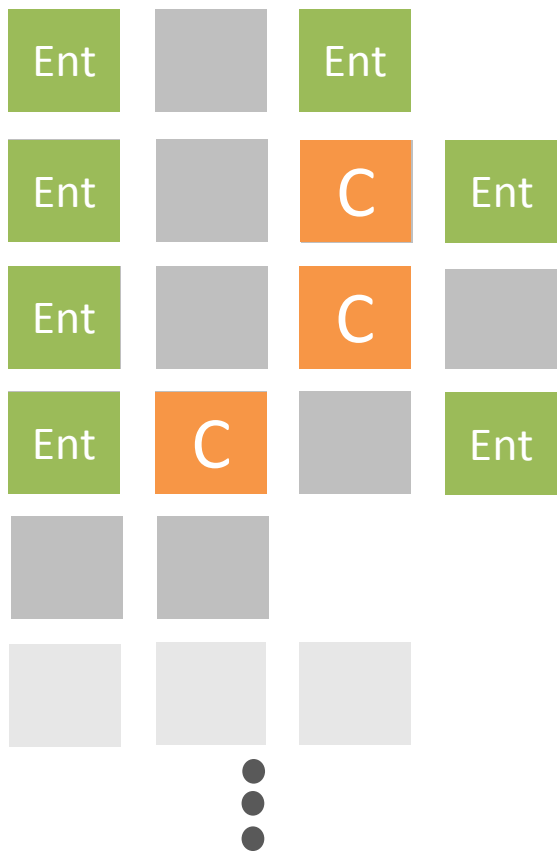


Rules:

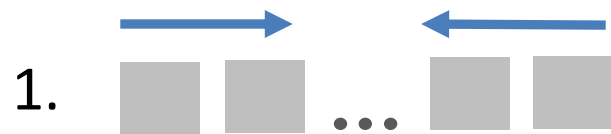


Aggregations:

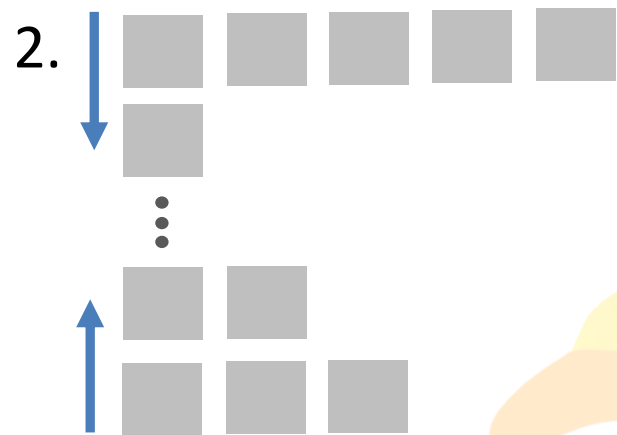


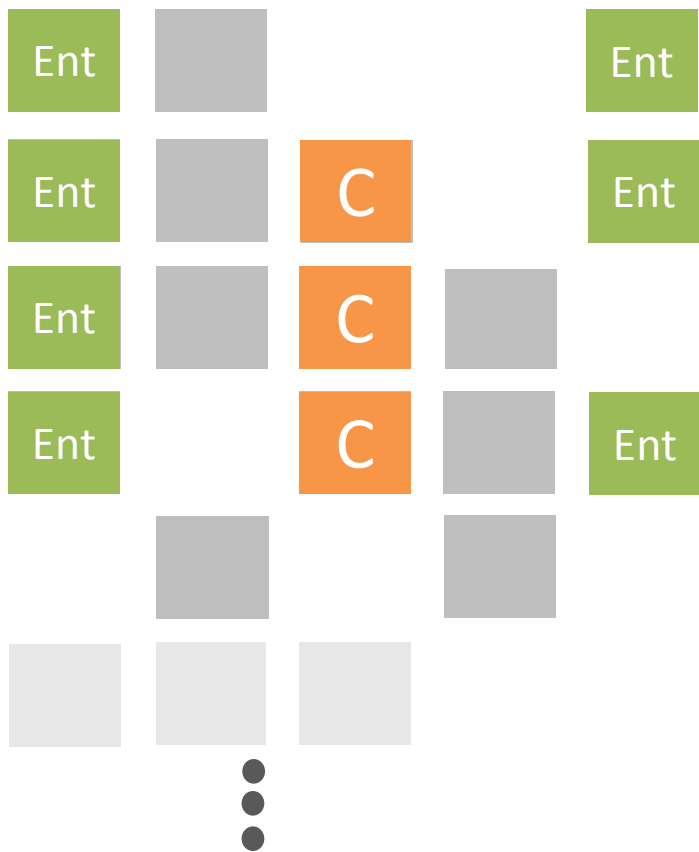


Rules:

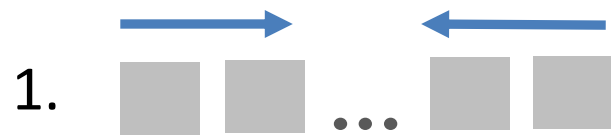


Aggregations:

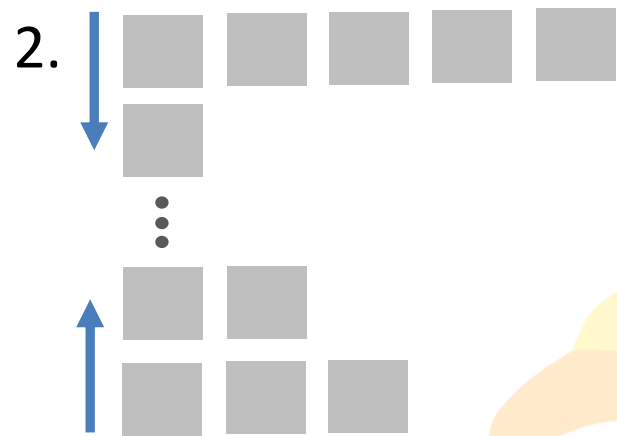




Rules:

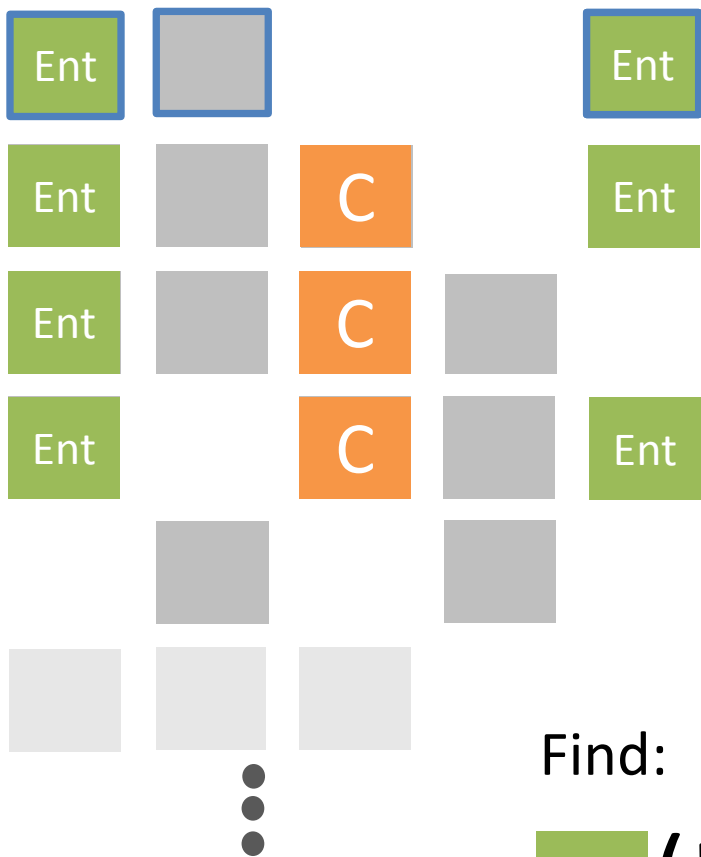


Aggregations:

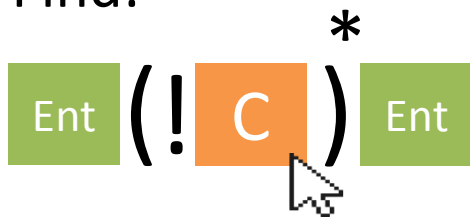


Selections:

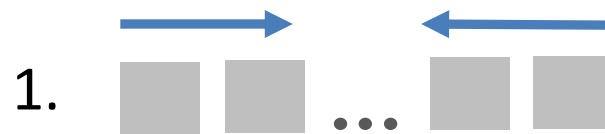




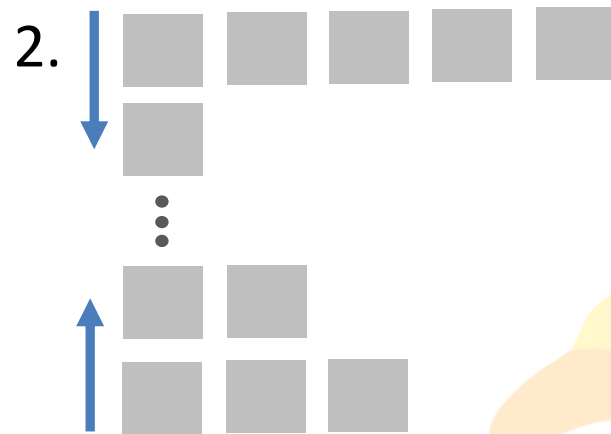
Find:



Rules:

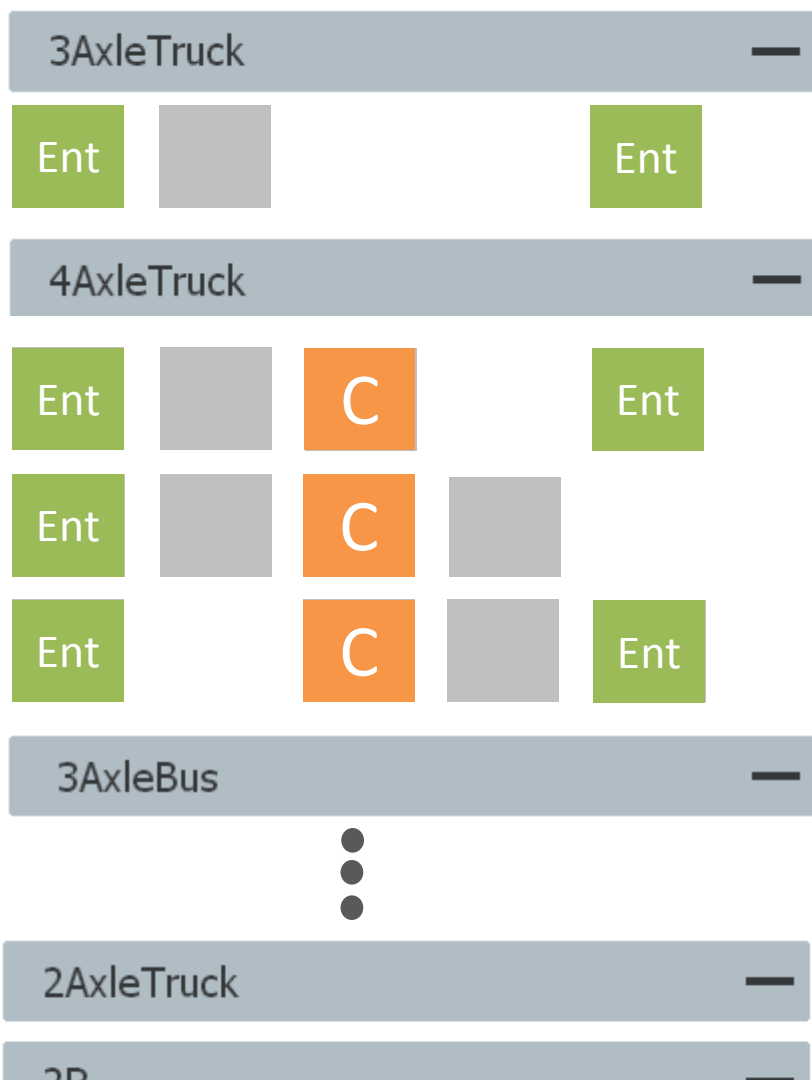


Aggregations:

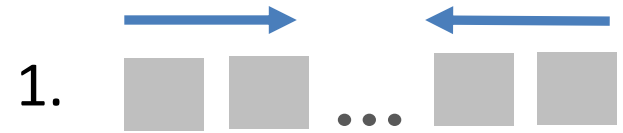


Selections:

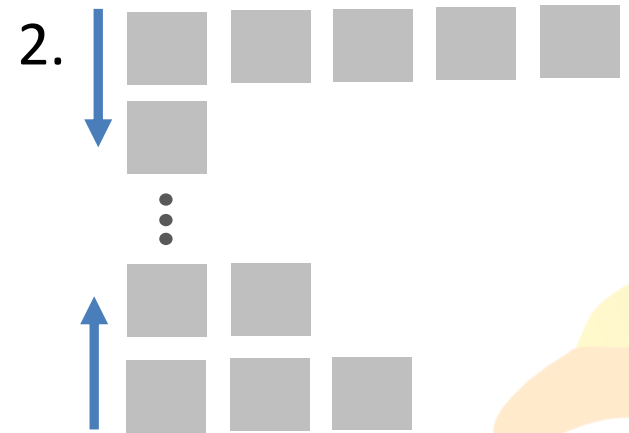




Rules:



Aggregations:



Selections:



Demo

File View Help

Attributes Traces

Search: ↓ ↕ ↕ ↕

frame.time_second

frame.time_strday

frame.time_stryea

frame.time_week

frame.time_year

Selections:

Context	#Events	#Traces	Coverag
All	171477	25523	
fromto	68334	9645	
Selection 0	34419	6527	
davtrio	76562	10973	

Rules:

Find Glyphs Rules

Multi-match Show labels Show popups

Group: Alphabetical ASC Sort: Frequency DESC Stack glyph size: none

gap penalty: 1 substitute penalty: 1

All

All

C G E	2293
E G C	1816
E G R R G E	1599
E E	1569
E G R R G G G C	1145
C G G G R R G E	1136
E G G C	910
E G G R R G G G C	878
C G G G R R G G E	831
E G G E	829
E C	758
E G E	746
E G R R G G E	741
E G G R R G E	729
C G G E	682
C E	554

SequenceView Multiple Sequence Alignment

Why this is cool

Efficient & Fast

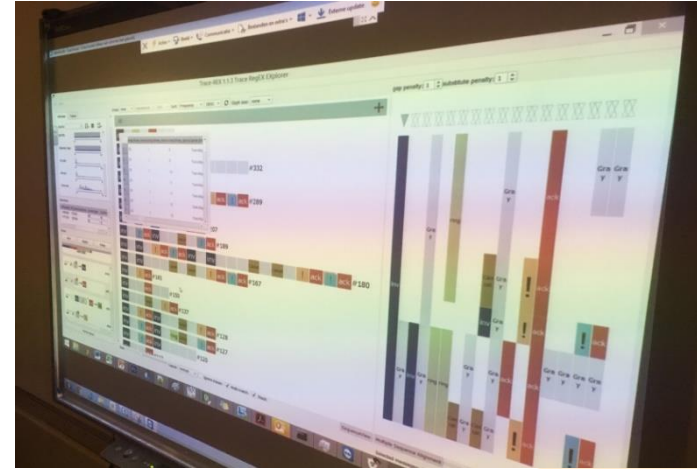
- Instant results (spot outliers)
- Suitable for Realtime stream analytics

Plug and play

- No complex configuration required, load the data and start playing!

Feedback loop

- Learn from automated methods, automated methods learn from you



Illegal Traffic Movement

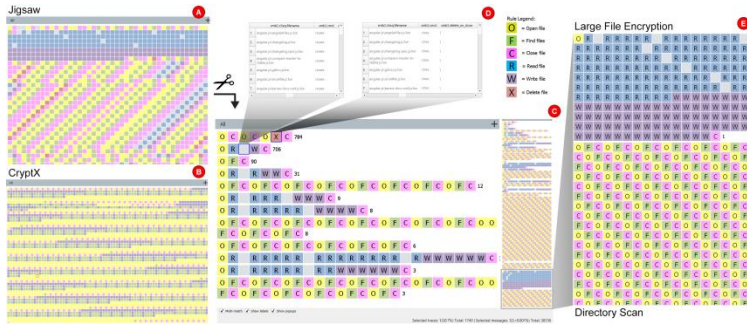


Elegant Support
For Hypothesis Generation
and Testing

Misuse detection in Wildlife Preserve



Applications



Ransomware Detection (94GB)



Voice Over IP fraud (40.000.000 events)



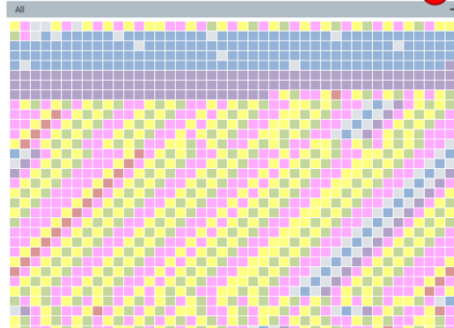
Healthcare Record Analysis (≈700.000 events)



Illegal traffic movement (≈200.000 events)

Ransomware Reverse Engineering

Jigsaw



CryptX

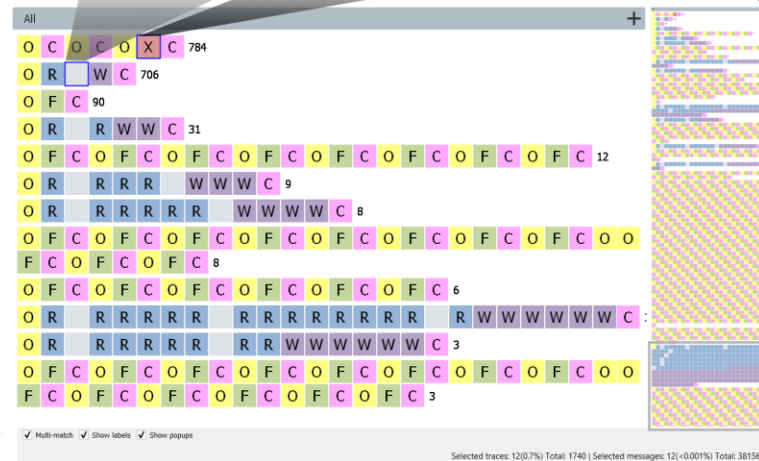


Clipboard content showing SMB command logs:

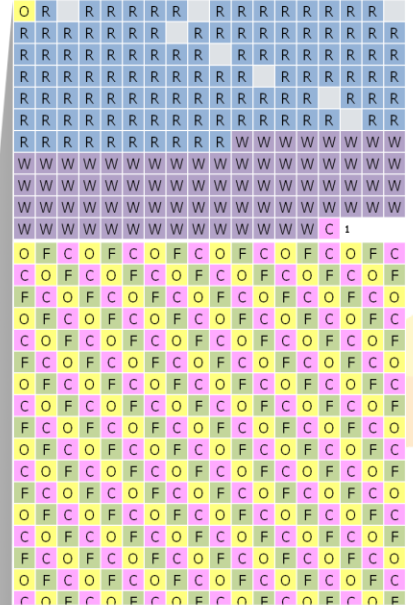
smb2_cseq.filename	smb2_cmd
angular.js\angularfiles.js.fun	create
angular.js\changelog.js.fun	create
angular.js\changelog.spec.js.fun	create
angular.js\compare-master-to-stable.js.fun	create
angular.js\gitdocs.js.fun	create
angular.js\gruntfile.js.fun	create
angular.js\karma-docs.conf.js.fun	create

smb2_cseq.filename	smb2_cmd	smb2_deletes_on_close
angular.js\angularfiles.js.fun	close	1
angular.js\changelog.js.fun	close	1
angular.js\changelog.spec.js.fun	close	1
angular.js\compare-master-to-stable.js.fun	close	1
angular.js\gitdocs.js.fun	close	1
angular.js\gruntfile.js.fun	close	1
angular.js\karma-docs.conf.js.fun	close	1

- Rule Legend:
- O = Open file
 - F = Find files
 - C = Close file
 - R = Read file
 - W = Write file
 - X = Delete file

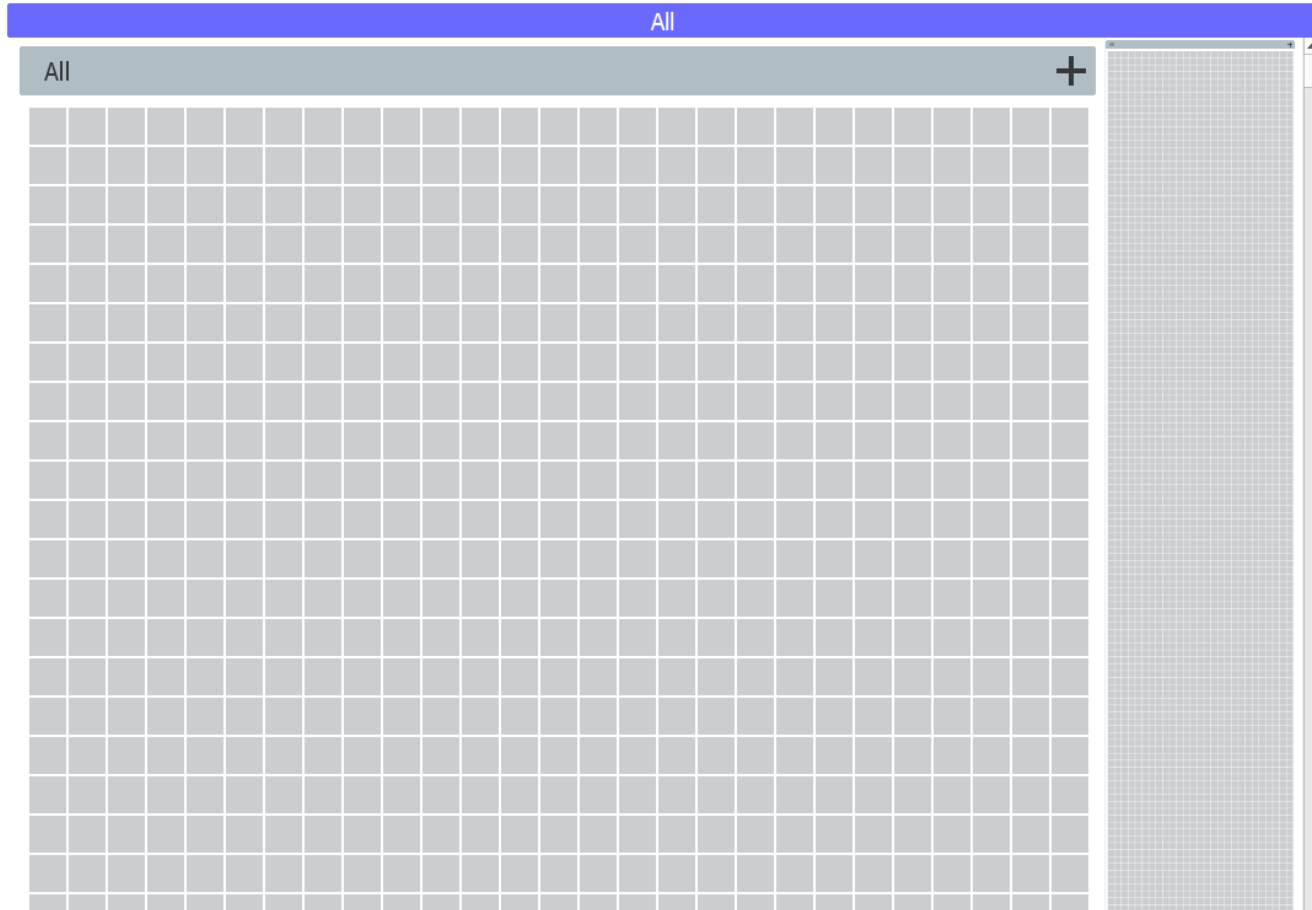


Large File Encryption

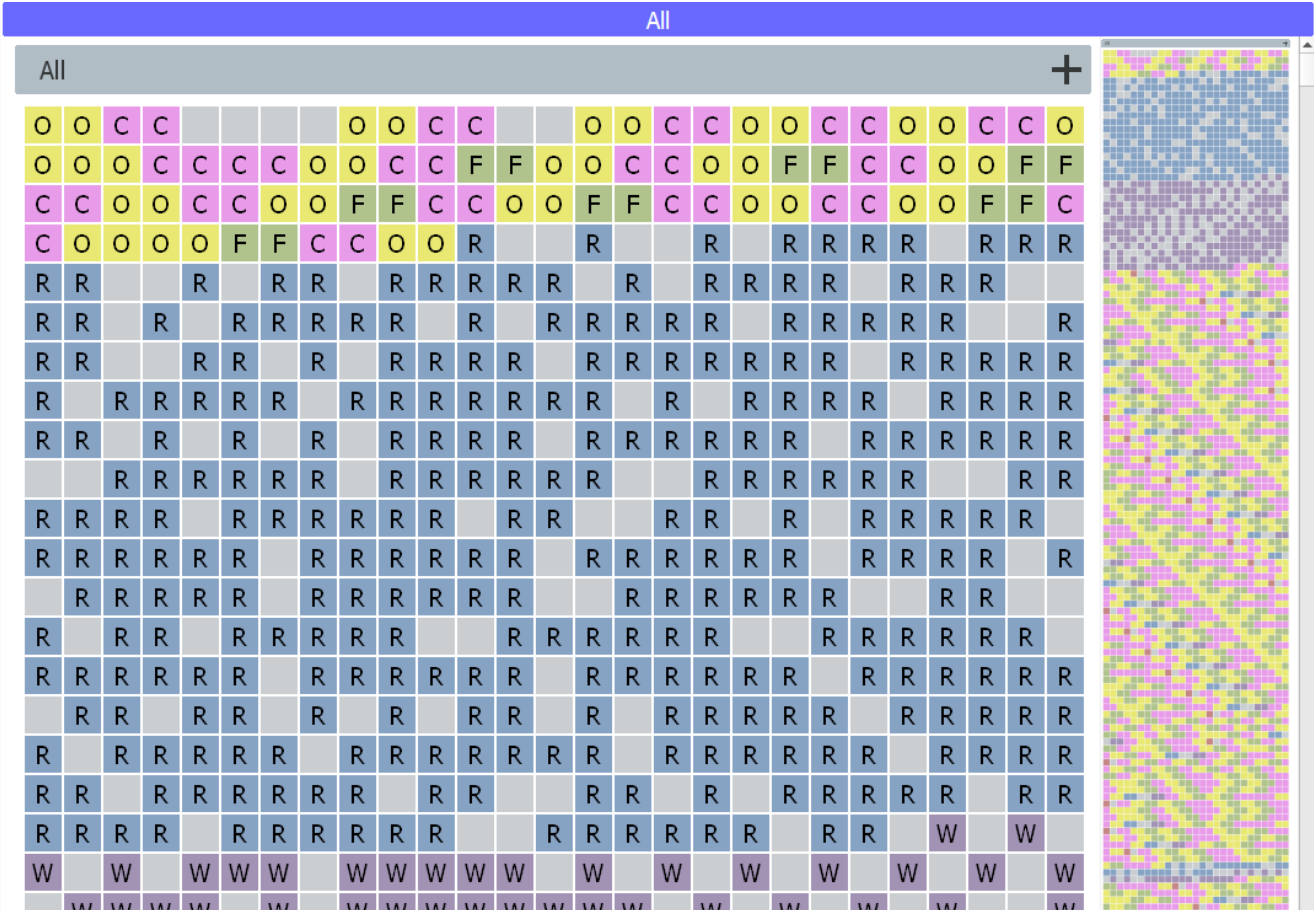


Directory Scan

See patterns instantly



See patterns instantly



Hospital Records



Possibilities are endless

Network Traffic Analysis



Workflow analysis



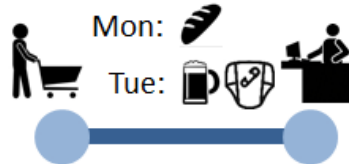
Hospital Treatments



Web analytics



Buy history analysis



Eventpad



Explore

- Instant detection of (un)desired patterns
- Discover unknown patterns with artificial intelligence



Understand

- Compare patterns to normal data
- Create rules to detect them



Prevent

- Monitor your data with specialized rule sets



Summary

- As long as AI is not perfect, we cannot go the beach
- Humans are still vital in data exploration&analysis
- Visualization can assist in
 - Data exploration
 - Making process mining techniques practical
 - Bridging the gap between machine learning and domain knowledge

Thanks!



The Note**pad** editor for **Event** Data

<http://event-pad.com>

Contact:



Bram Cappers

b.c.m.cappers@tue.nl

www.bramcappers.nl

Project:



Industrial Partners:

