Università
degli Studi
di Milano

In Partnership with KHALIFA UNIVERSITY

AI pipelines for Cyber Security
Ernesto Damiani

Cyber-Physical Systems Center (C2PS)

In Partnership with KHALIFA UNIVERSITY

SECURITY OF THE GLOBAL ICT INFRASTRUCTURE

Network and Communications Security

Business Process Security and Privacy

Security and Privacy of Big Data Platforms

SECURITY ASSURANCE

Security Risk Assessment and Metrics

Continuous Security Monitoring and Testing

DATA PROTECTION AND ENCRYPTION

High Performance Homomorphic Encryption

Lightweight Cryptography and Mutual Authentication

SESAR LAB

- Secure Software Architectures and Knowledge-based systems lab (SESAR) http://sesar.di.unimi.it

- Industry collaborations: SAP, British Telecom, ATOS, ENG, Cisco, TIM-Telecom Italia
- Part of the AI/Big-Data Community

Some Activities

KITE IT

enisa

Big Data Threat Landscape and Good Practice Guide

JANUARY 2016

www.enisa.europa.eu European Union Agency for Network and Information Security

BigData 2013

IEEE 2nd International Congress on Big Data
June 27-July 2, 2013, Santa Clara Marriott, CA, USA

TOREADOR

securescm

PRACTICE

International Journal of Big Data

Aristotele
Learning • Knowledge • Innovation

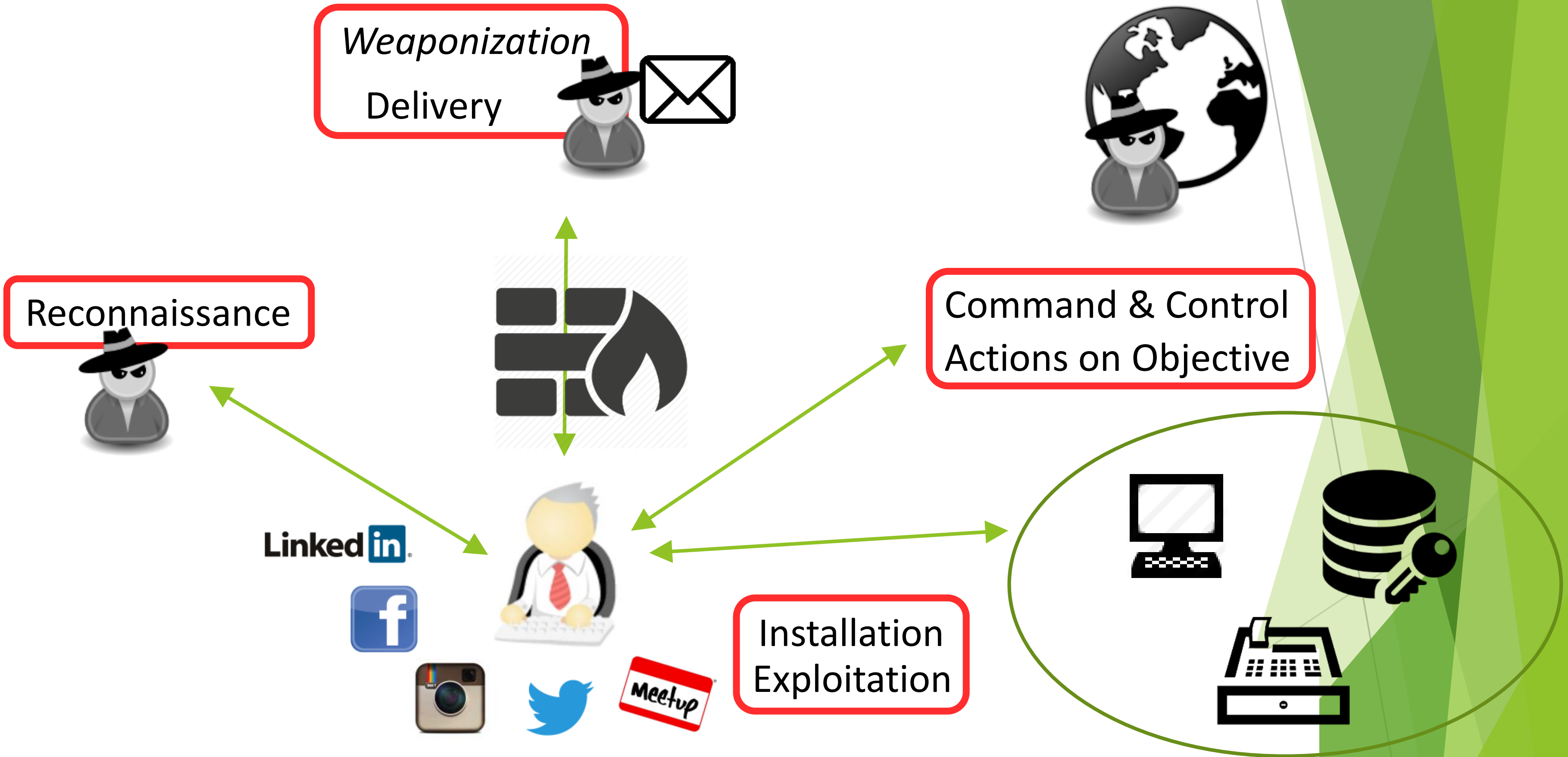
Cumulus
Certification infrastrUcture for MUlti-Layer cloUd Services

A4SOA

Today’s Agenda

- The danger outside (and inside)
- Data-driven cybersecurity
- AI pipelines for cybersecurity
- Wrap up

Big Data for Attackers: Scenario



Weaponization

Delivery

Reconnaissance

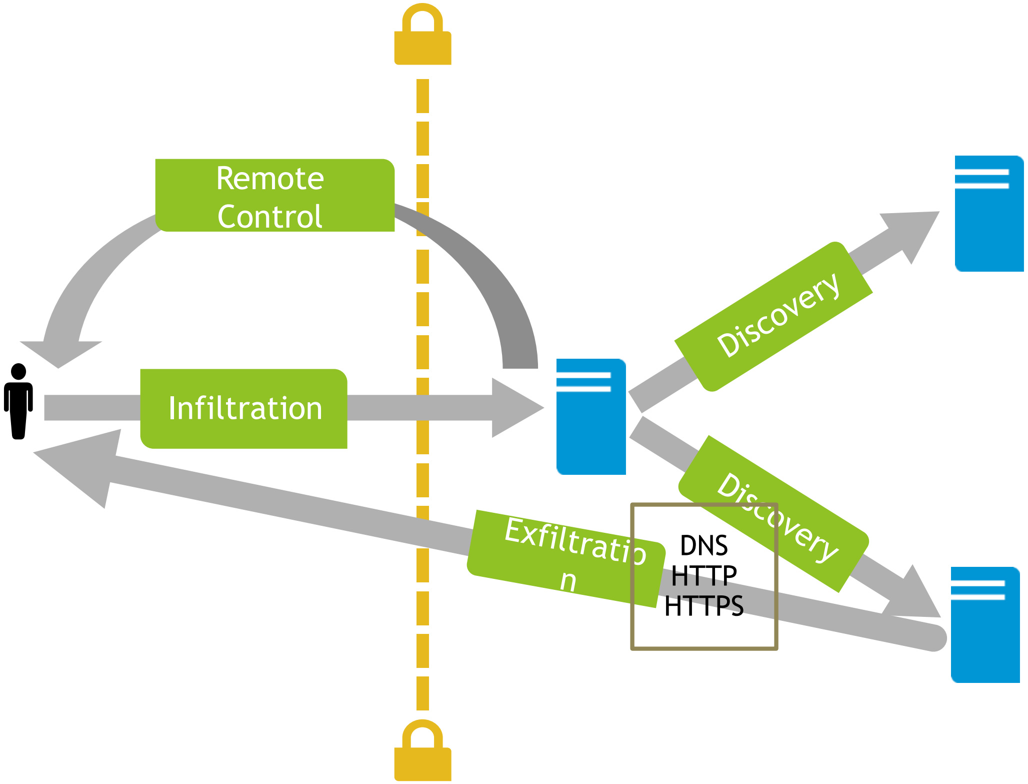
Command & Control

Actions on Objective

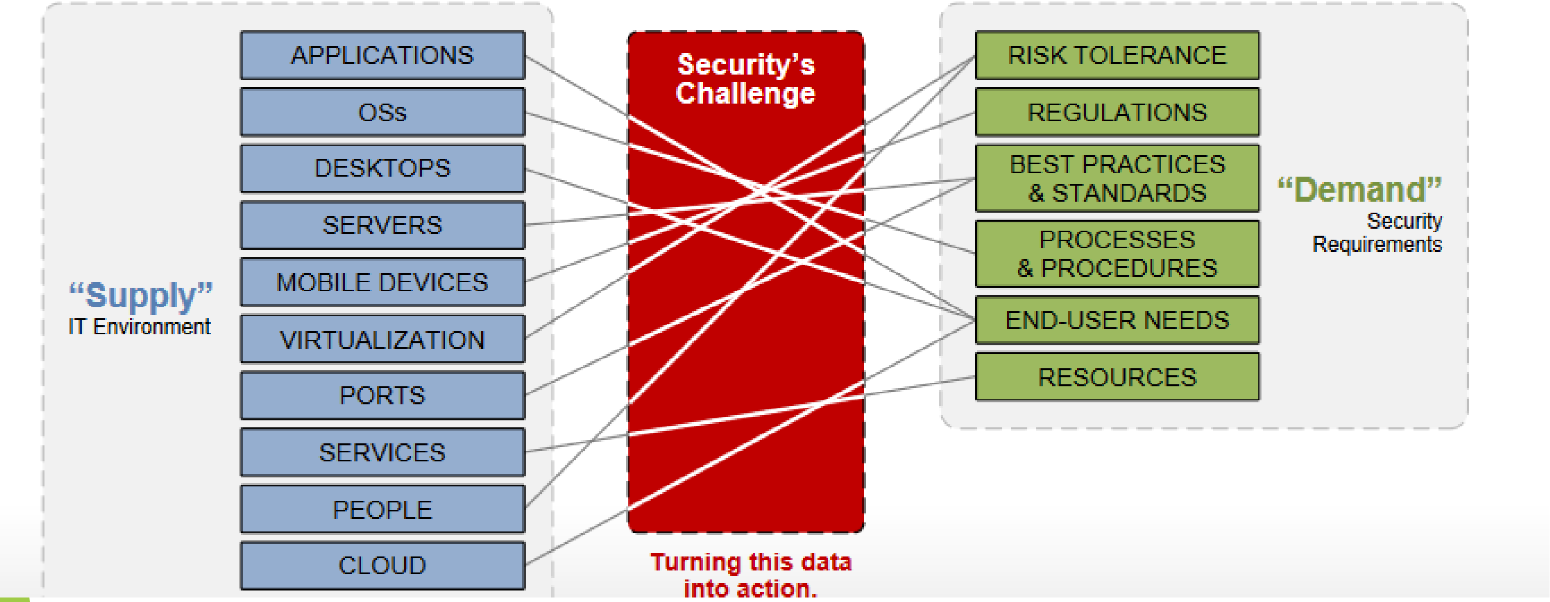
Installation

Exploitation

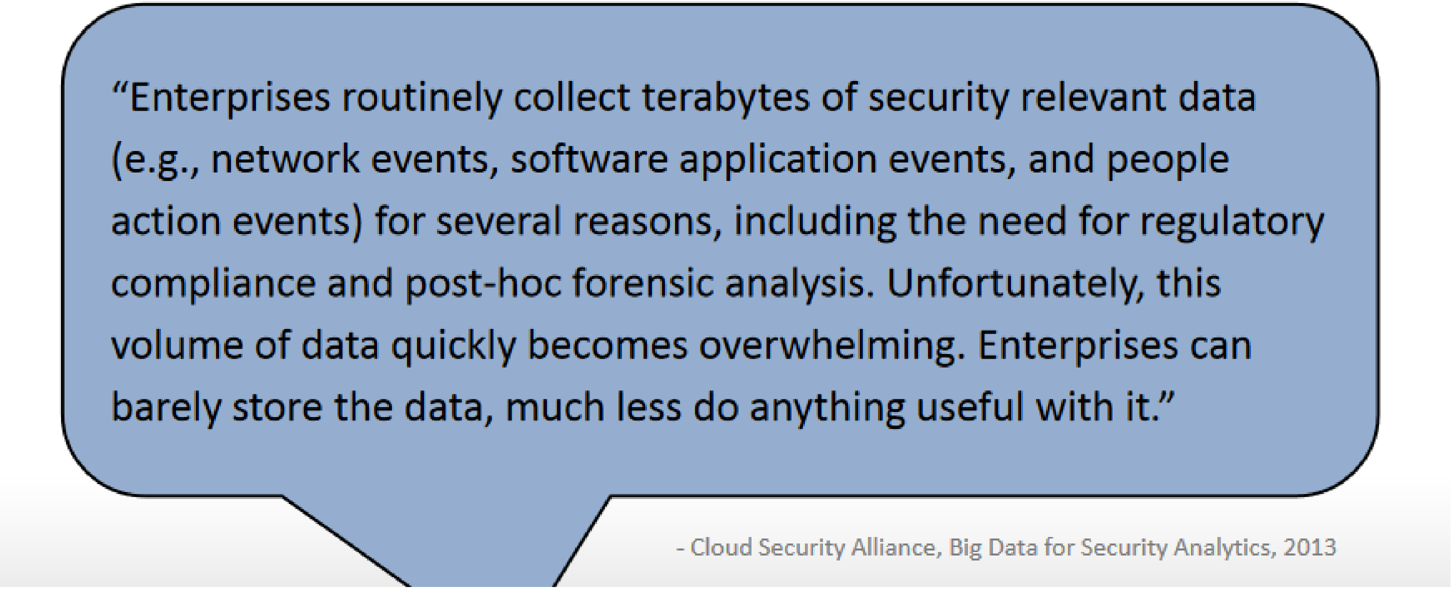
Advanced threats: The new landscape



Big Data For Defenders



But how to use them?



“Enterprises routinely collect terabytes of security relevant data (e.g., network events, software application events, and people action events) for several reasons, including the need for regulatory compliance and post-hoc forensic analysis. Unfortunately, this volume of data quickly becomes overwhelming. Enterprises can barely store the data, much less do anything useful with it.”

- Cloud Security Alliance, Big Data for Security Analytics, 2013

Sensitive data identification

- Regular expression
  - social security numbers, telephone numbers, addresses, and other data that has a significant amount of structure.
- Keywords
  - small number of known keywords can identify private data, e.g., medical or financial records
- Fingerprints
  - Hashes of substrings of unstructured data

Signatures

- **Atomic signatures**
  - A single packet, activity, or event is examined to determine if the signature should trigger a signature action.
  - The entire inspection can be accomplished in an atomic operation that does not require any knowledge of past activities.
- **Stateful signatures**
  - Stateful signatures trigger on a sequence of events
  - Require the analytics device to maintain state for a duration known as the *event horizon*.
  - Configuring the length of the event horizon is a tradeoff between consuming system resources and being able to detect an attack that occurs over a long period of time.

Atomic Signature Pros and Cons

- Atomic signatures advantages:
  - Consume minimal resources (e.g. memory) on the analytics device.
  - Easy to understand (search only for specific events).
  - Traffic analysis performed quickly and efficiently.
- Major drawbacks:
  - One has to know all the atomic events of interest and create the corresponding signatures. As the number of atomic signatures increases, managing them becomes overwhelming.

Example

- A simple string match triggers an alert action whenever the traffic that it is analyzing contains **/etc/passwd**.
- Knowing this string signature is sought in TCP traffic, an attacker can generate alerts by sending a flood of TCP packets with the **/etc/passwd** string in payload.
- The alerts are generated even if the connection is not part of any valid TCP connection
  - A large number of bogus alerts can impact the performance of your monitoring applications and devices.

Types of atomic signatures

- Host-Based Examples
  - Host-based IPSs examine many operations, including function calls, files accessed, and so on. The best method for detecting anomalous user behavior is to establish a baseline of the operations that a user normally performs on the system.
  - By monitoring deviations from the baseline, you can detect potentially malicious activity.
    - For example, if a function call is never invoked normally (except in connection with malicious activity), then triggering a signature action whenever it is called is a simple example of a host-based atomic signature..

Types of atomic signatures

- Network-Based Examples
  - LAND attack: a denial-of-service (DoS) attack in which the attacker sends TCP packet (with the SYN bit set) in which the source and destination IP address (along with source and destination port) are the same.
    - When it was first discovered, many IP stacks crashed the system when they received a LAND attack.
  - By inspecting a single packet, a Network-based IPS can identify LAND. Because everything is contained in a single packet, no state information is needed to identify this attack.

Stateful signatures and horizons

- Stateful signatures require several pieces of data to match an attack signature.
- The maximum amount of time over which an attack signature can successfully be detected (from the initial to the final data piece needed to complete the attack signature) is known as the *event horizon*.
- The analytics device must maintain state information for the entire event horizon.

Stateful Signatures Pros and Cons

- Major advantage: requiring a specific event to be detected in a known context increases the likelihood that the activity represents legitimate attack traffic. This minimizes the false positives.
- Main drawback: maintaining state consumes memory resources on the IDS device.
  - If the IDS does not efficiently manage resources when maintaining state, the large consumption of resources (such as memory and CPU) can lead to a slow response time, dropped packets, and missed signatures.
- *Slow attacks* exploit the fact that an IPS cannot maintain state information indefinitely without eventually running out of resources.

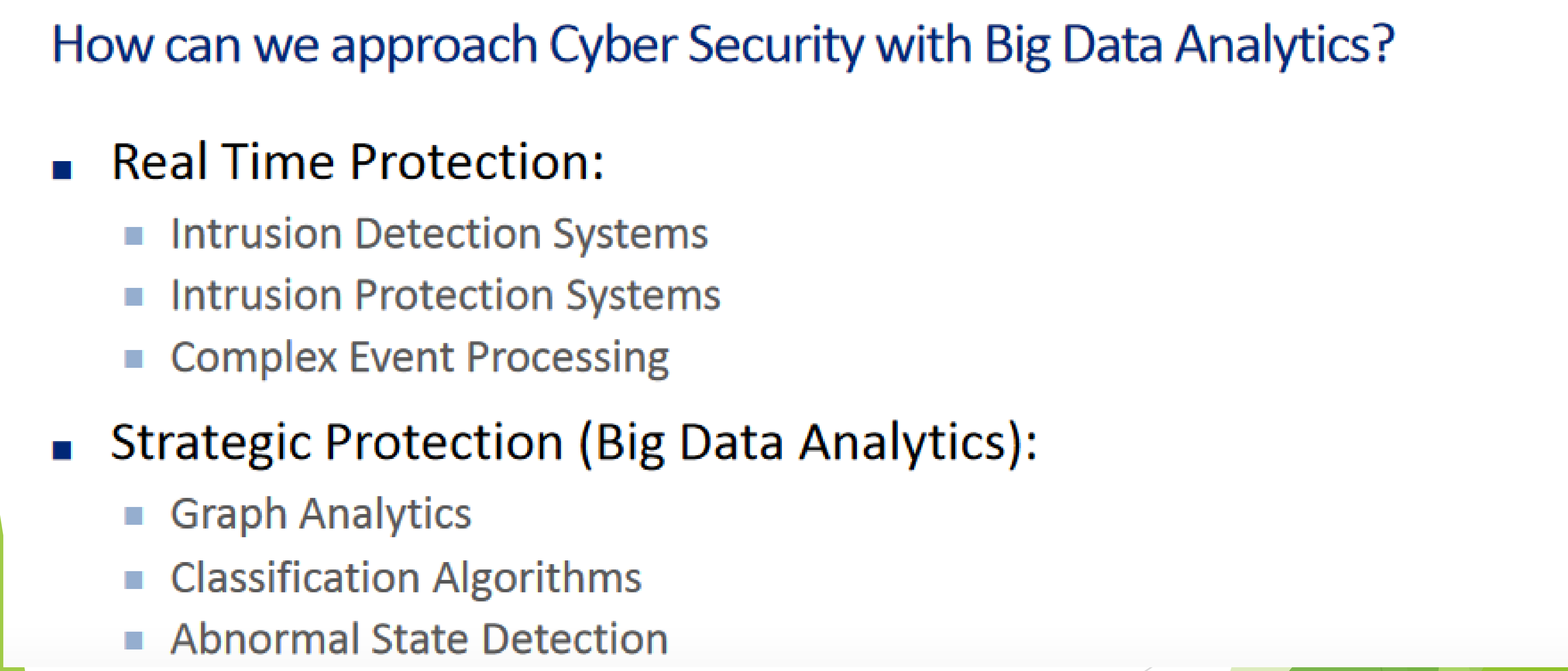
Examples of stateful signatures

- Host-Based
  - Many attacks invoke cmd.exe remotely. To remotely execute cmd.exe, the attacker must make a network connection to the host.
    - We do not want to trigger a signature action whenever cmd.exe is invoked (because our users use this program frequently).
  - A stateful signature triggers an action when cmd.exe is invoked, only if the application invoking cmd.exe first accepted a network connection.
- Network-Based
  - To minimize the ability of attackers to generate bogus alarms, most TCP attack signatures are triggered only if the signature is observed on a valid TCP connection.

From Signatures to Pattern Detection

- *Signature-based analytics* can only detect attacks for which a signature has previously been created
- Data Analytics Techniques use *behavioral patterns* to detect behavior that falls outside of normal system operation

Enter AI



How can we approach Cyber Security with Big Data Analytics?

- Real Time Protection:
  - Intrusion Detection Systems
  - Intrusion Protection Systems
  - Complex Event Processing
- Strategic Protection (Big Data Analytics):
  - Graph Analytics
  - Classification Algorithms
  - Abnormal State Detection

Analytics Challenges

- High Volume Data Streams
  - Threats Emerging at High Rate
  - Short Lived Patterns
  - Evasive Threats: hard to detect, harder to predict
  - Slow Threats
- Increasing Cost of Traditional Techniques
  - Signature generation takes time and money
  - Talent pool scarce and expensive
- Need for Real Time
  - Adaptive defense should respond in real-time (seconds to minutes) to changing attack vectors

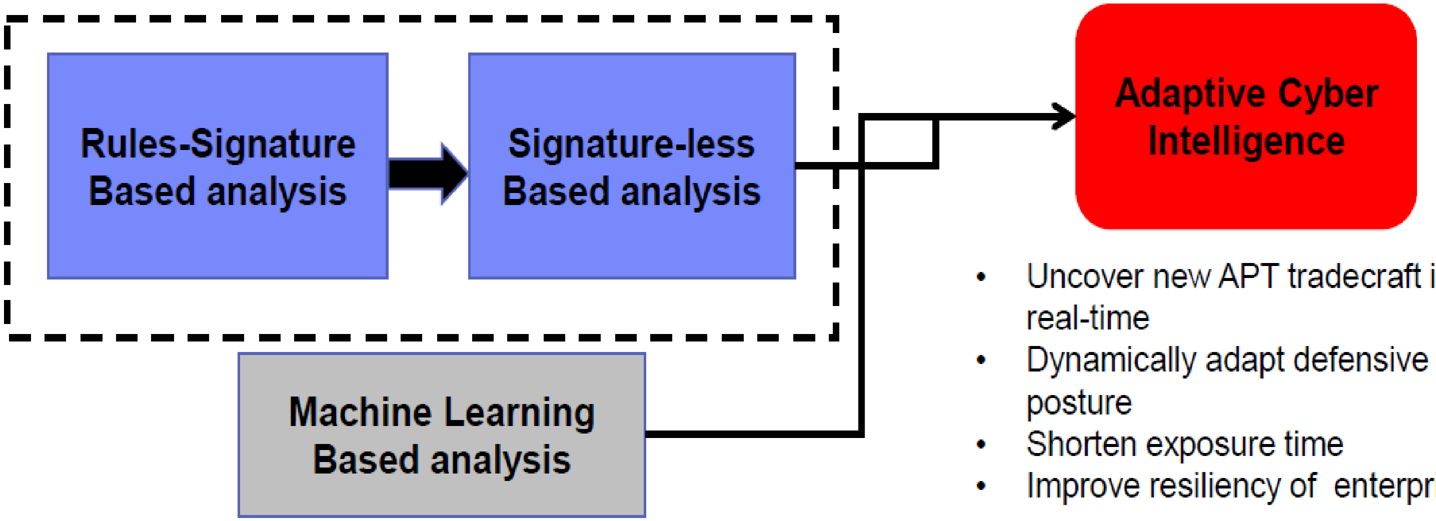
Use of Machine Learning

Machine Learning is best used for:

- Dynamically discovering new or subtle changes in attack signatures-tradecraft
- Behavior modeling to characterize normal versus anomalous activity
- Provide lower sensitivity of analysis to reduce false alarms by balancing bias-variance and precision-versus recall

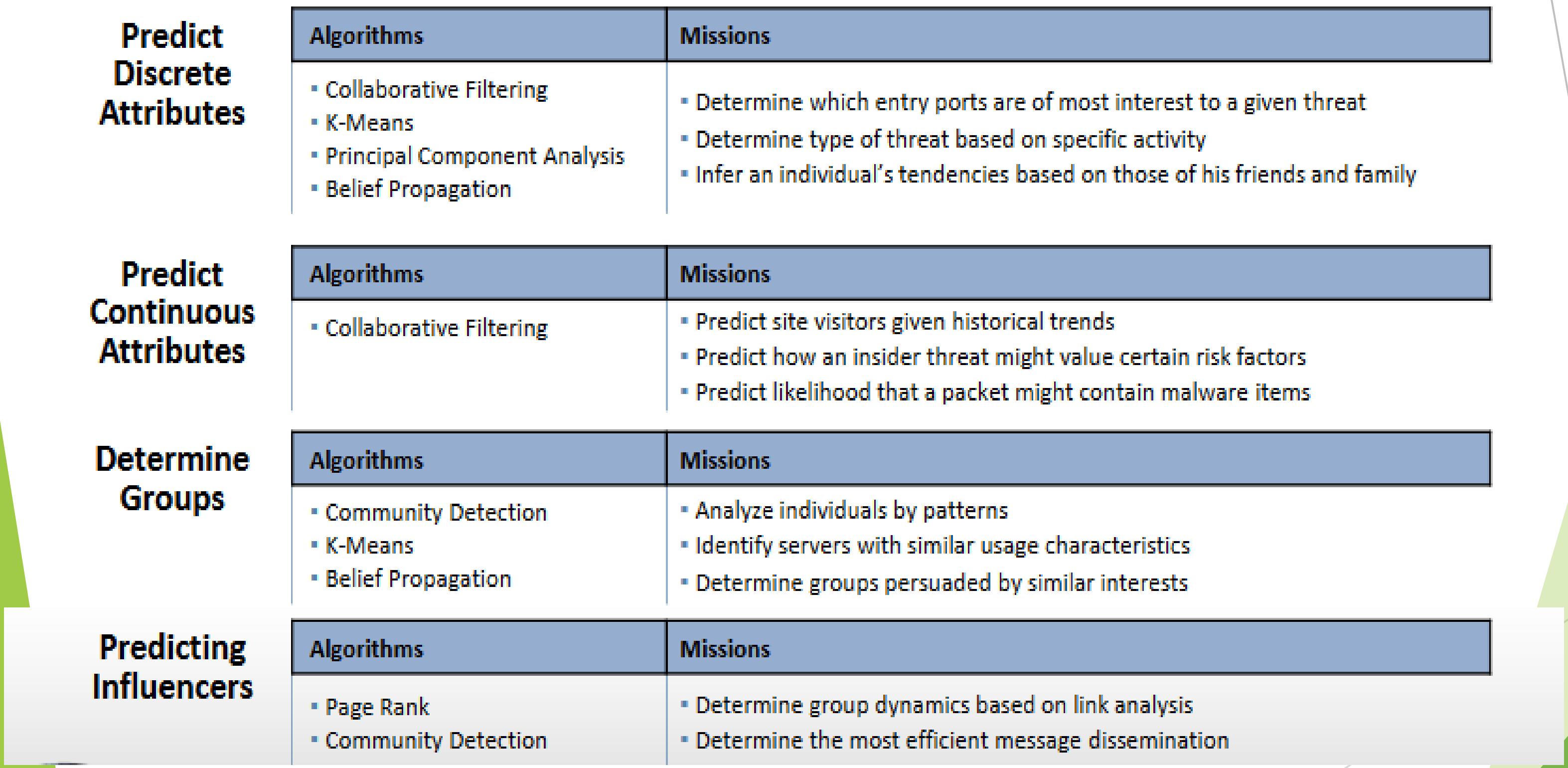
Complementary, not alternative

- ML improves signature based responsiveness and increases precision


Rules-Signature Based analysis
Signature-less Based analysis
Machine Learning Based analysis
Adaptive Cyber Intelligence

- Uncover new APT tradecraft in real-time
- Dynamically adapt defensive posture
- Shorten exposure time
- Improve resiliency of enterprise

What AI and Big Data Analytics can deliver today



Predict Discrete Attributes

Algorithms	Missions
Collaborative Filtering K-Means Principal Component Analysis Belief Propagation	Determine which entry ports are of most interest to a given threat Determine type of threat based on specific activity Infer an individual's tendencies based on those of his friends and family

Predict Continuous Attributes

Algorithms	Missions
Collaborative Filtering	Predict site visitors given historical trends Predict how an insider threat might value certain risk factors Predict likelihood that a packet might contain malware items

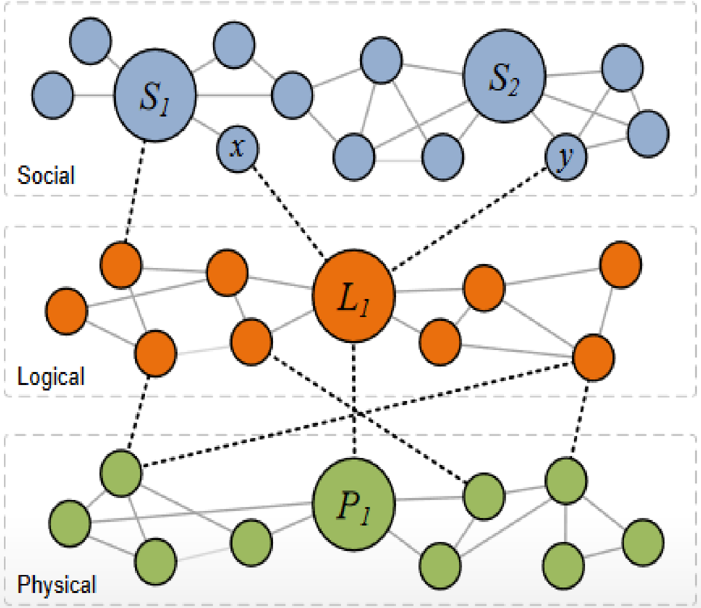
Determine Groups

Algorithms	Missions
Community Detection K-Means Belief Propagation	Analyze individuals by patterns Identify servers with similar usage characteristics Determine groups persuaded by similar interests

Predicting Influencers

Algorithms	Missions
Page Rank Community Detection	Determine group dynamics based on link analysis Determine the most efficient message dissemination

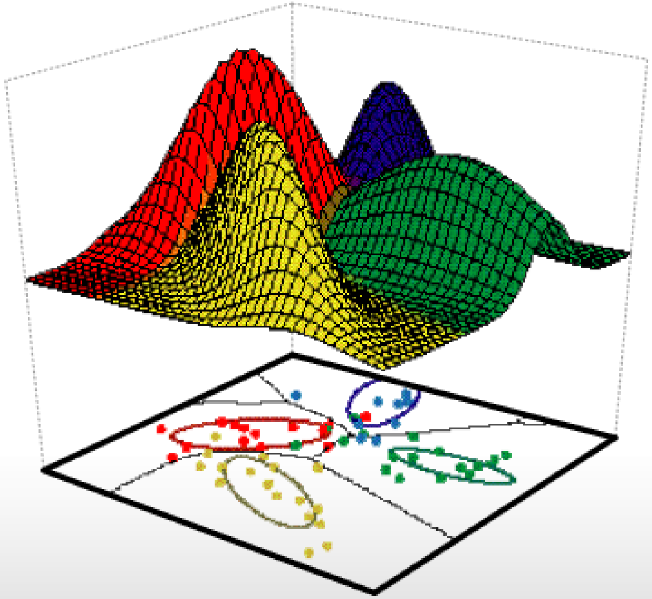
Graph Analytics



- PageRank:
  - Which nodes have the most connections?
- Belief Propagation:
  - Where might the next incident occur?
- Community Detection
  - What is the behavior of the connections?
  - Are they growing?

Classification

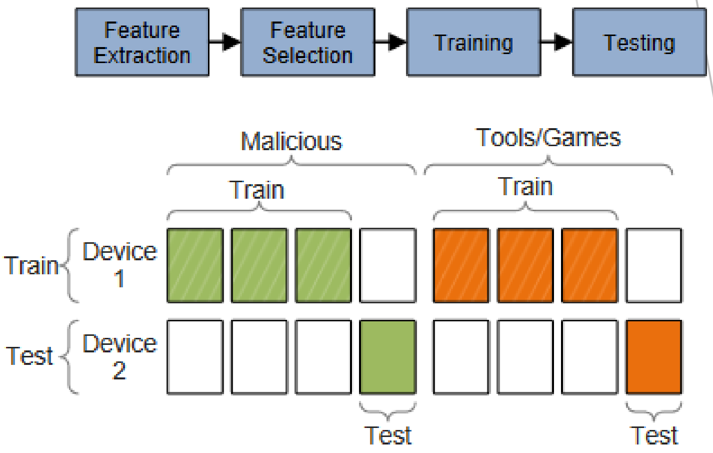
- Rules can be created from the raw data gathered using different Classification Algorithms:
  - Artificial Neural Networks (ANN)
  - Decision Trees (DT)
  - Naive Bayes (NB)
  - Support Vector Machines (SVM)
  - Boosted Decision Trees (BDT)
  - Boosted Naive Bayes (BNB)



Identifying Features

Abnormal State Detection

- Identify the most informative features to monitor.
- Evaluating various detection methods and algorithms.
- Understanding the feasibility of running these methods for detection.

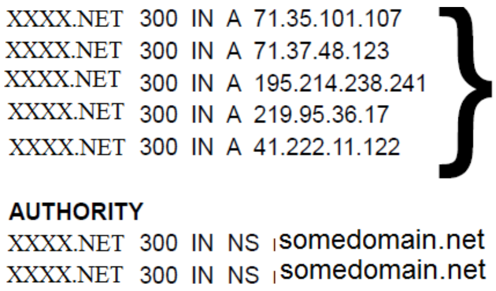


- Detection Algorithms: K-Means, Histograms, Logistic Regression, Decision Tree, Bayesian Net, Naive Bayes
- Feature Selection: InfoGain, Chi Square, Fisher Score

Example: Fast Flux

- Fast flux DNS is a technique that a cybercriminal can use to prevent identification of his key host server's IP address.
- By abusing the way the DNS works, the criminal can create a botnet with nodes that join and drop off the network faster than law enforcement officials can trace them.
- The basic idea behind Fast flux is to have numerous IP addresses associated with a single fully qualified domain name(e.g "xxx.yyy.com"), where the IP addresses are swapped in and out with extremely high frequency (after the low TTL has expired), through changing DNS records.

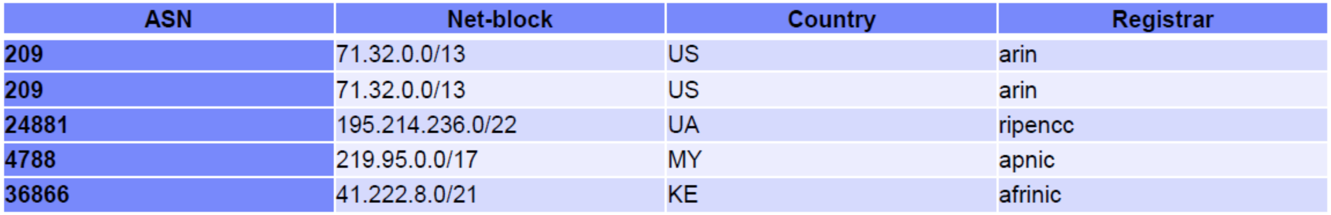
What can a ML model notice?



```
XXXX.NET  300  IN  A  71.35.101.107
XXXX.NET  300  IN  A  71.37.48.123
XXXX.NET  300  IN  A  195.214.238.241
XXXX.NET  300  IN  A  219.95.36.17
XXXX.NET  300  IN  A  41.222.11.122

AUTHORITY
XXXX.NET  300  IN  NS  somedomain.net
XXXX.NET  300  IN  NS  somedomain.net
```

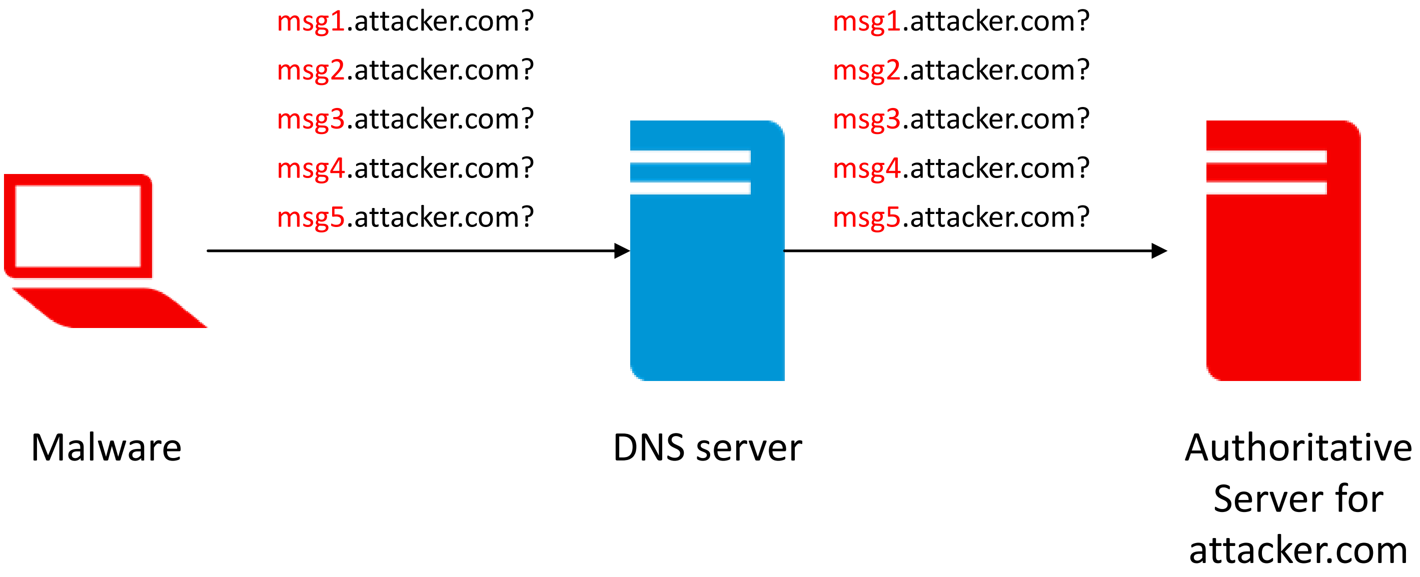
Notice non contiguous IP address TTL of 300 seconds for Domain somedomain.net



ASN	Net-block	Country	Registrar
209	71.32.0.0/13	US	arin
209	71.32.0.0/13	US	arin
24881	195.214.236.0/22	UA	ripencc
4788	219.95.0.0/17	MY	apnic
36866	41.222.8.0/21	KE	afrinic

Notice different ASN, Countries and Registrars

Example: DNS Exfiltration



msg1.attacker.com?
msg2.attacker.com?
msg3.attacker.com?
msg4.attacker.com?
msg5.attacker.com?

msg1.attacker.com?
msg2.attacker.com?
msg3.attacker.com?
msg4.attacker.com?
msg5.attacker.com?

Malware

DNS server

Authoritative
Server for
attacker.com

Vern Paxson, Mihai Christodorescu, Mobin Javed, Josyula Rao, Reiner Sailer, Douglas Schales, Marc Ph. Stoecklin, Kurt Thomas, Wietse Venema, and Nicholas Weaver.
Practical comprehensive bounds on surreptitious communication over DNS.
In *Proceedings of the 22nd USENIX conference on Security* (SEC'13). USENIX Association, Berkeley, CA, USA, 17-32.

A real world example

- Queries

BLGCOFDAGOOOESDULBOOBOOOOOOOOOOOOOOOOOOLDOSESKGKHHF.detacsufbo.ru

EUJSFLDAGOOOESDUDBOOBOOOOOOOOOOOOOOOOOOOSSJHGHFCLFOHCHLGHSAHAHU.CHLAAFHLSGHAFGFUOOEUGDKLCSHEKLJBOCOSECHFFUGBSKGDJGGGHOJHJCGJG.KCDOELDUOEGUCUOUHJUAKEGGGFGEKHLGFDFESJOEL.detacsufbo.ru

SHUDHFDAGOOOESDUGBOOBOOOOOOOOOOOOOOOOOOEDKDFBBHLEGGJLGUFABHCCU.DHDFFCHHKSHGHAOUBGEGEJLGFHUBDFGUGJDFFEAKFSBFFGSDACGHCSKBHLSCGHH.EHSHHJFHUAAOOGKKSDDAHAUBBJDCCKGSHKLGJGAS.detacsufbo.ru

OHDOBHDAGOOESDUGBOOHOOOAOOOOOOOOOOOOOOO.detacsufbo.ru

HBSGGCDAGOOESDUUSOOBOOOOOOOOOOOOOOOOOOO.detacsufbo.ru

- Responses (TXT records)

LLCDGHDABOOOSSUHOOOFOOOOOOOOOOOOOOOOOOO

KJGDUDABOOOSBSUHOOOFOOOOOOOOOOOOOOOOOOO

JJDHUDABOOOSBSUHOOOFOOOOOOOOOOOOOOOOOOO

HBEAGDABOOOSBSUHOOOUOOOOOOOOOOOOOOOOOOO

KALFCSDAOOOSBSUHOOOFOOOOOOOOOOOOOOOOOOO

Features for DNS exfiltration detection

- Lengths of DNS queries and responses
- Sizes of request and reply packets
- Entropy
- Total number/volume of DNS queries from a device
- Total number/volume of DNS queries to a domain

DESIGNING A PIPELINE

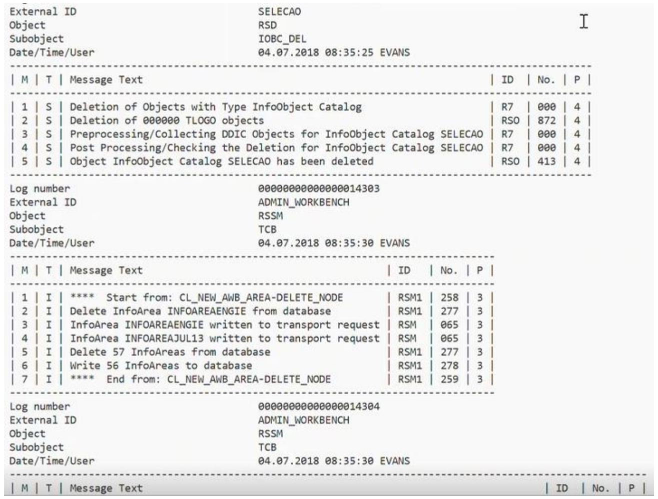
Scenario and Design of the pipeline

- 1 scenario
  - Automating security alerts on a flow of system log events
- 3 pipelines
  - Data provisioning: from Operation to Analyst
  - Log analysis: automating security alerts
  - Enhanced log analysis: automating security alerts

The problem

Security Incident?

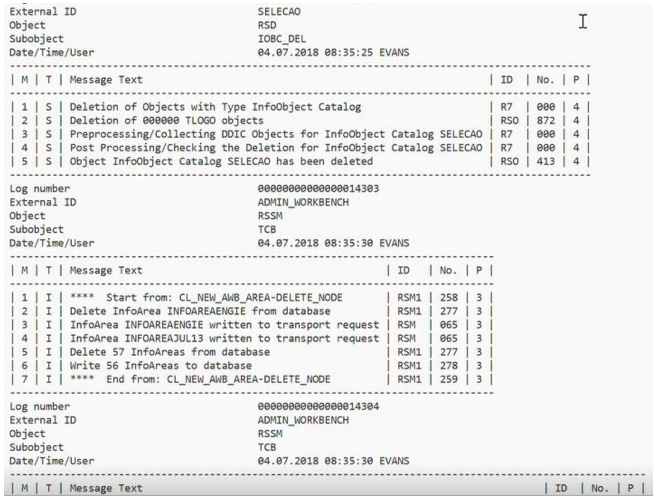
→ Automate w/ML



```
External ID                              SELECAO
Object                                   RSD
Subobject                                IOBC_DEL
Date/Time/User                           04.07.2018 08:35:25 EVANS
------------------------------------------------------------------------------------------------
| M | T | Message Text                                                         | ID  | No. | P |
------------------------------------------------------------------------------------------------
| 1 | S | Deletion of Objects with Type InfoObject Catalog                     | R7  | 000 | 4 |
| 2 | S | Deletion of 000000 TLOGO objects                                     | RSO | 872 | 4 |
| 3 | S | Preprocessing/Collecting DDIC Objects for InfoObject Catalog SELECAO | R7  | 000 | 4 |
| 4 | S | Post Processing/Checking the Deletion for InfoObject Catalog SELECAO | R7  | 000 | 4 |
| 5 | S | Object InfoObject Catalog SELECAO has been deleted                   | RSO | 413 | 4 |
------------------------------------------------------------------------------------------------
Log number                               00000000000000014303
External ID                              ADMIN_WORKBENCH
Object                                   RSSM
Subobject                                TCB
Date/Time/User                           04.07.2018 08:35:30 EVANS
--------------------------------------------------------------------------------
| M | T | Message Text                                        | ID   | No. | P |
--------------------------------------------------------------------------------
| 1 | I | ****  Start from: CL_NEW_AWB_AREA-DELETE_NODE       | RSM1 | 258 | 3 |
| 2 | I | Delete InfoArea INFOAREAENGIE from database         | RSM1 | 277 | 3 |
| 3 | I | InfoArea INFOAREAENGIE written to transport request | RSM  | 065 | 3 |
| 4 | I | InfoArea INFOAREAJUL13 written to transport request | RSM  | 065 | 3 |
| 5 | I | Delete 57 InfoAreas from database                   | RSM1 | 277 | 3 |
| 6 | I | Write 56 InfoAreas to database                      | RSM1 | 278 | 3 |
| 7 | I | ****  End from: CL_NEW_AWB_AREA-DELETE_NODE         | RSM1 | 259 | 3 |
--------------------------------------------------------------------------------
Log number                               00000000000000014304
External ID                              ADMIN_WORKBENCH
Object                                   RSSM
Subobject                                TCB
Date/Time/User                           04.07.2018 08:35:30 EVANS
--------------------------------------------------------------------------------------------------------
| M | T | Message Text                                                                | ID   | No. | P |
```

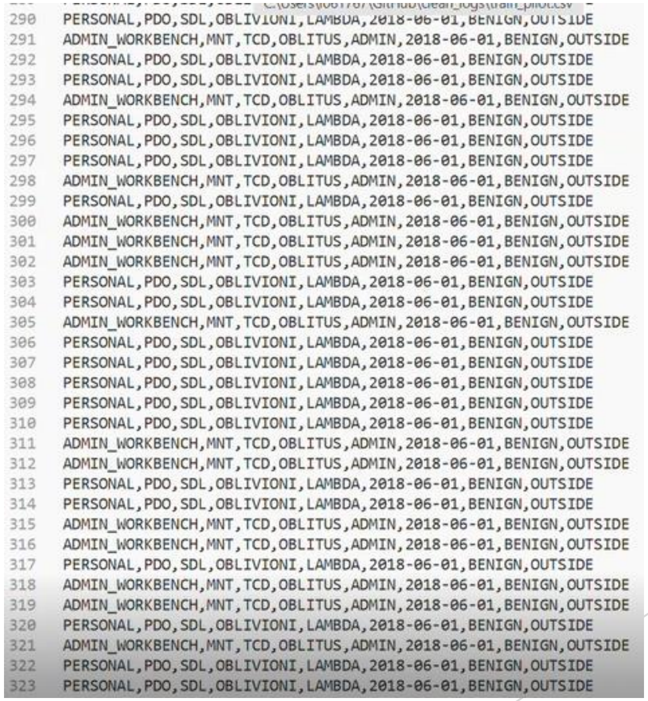
Step 1: Data preparation

Data minimization



```
External ID                          SELECAO
Object                               RSD
Subobject                            IOBC_DEL
Date/Time/User                       04.07.2018 08:35:25 EVANS
--------------------------------------------------------------------------------
| M | T | Message Text                                                | ID  | No. | P |
--------------------------------------------------------------------------------
| 1 | S | Deletion of Objects with Type InfoObject Catalog            | R7  | 000 | 4 |
| 2 | S | Deletion of 000000 TLOGO objects                            | RSO | 872 | 4 |
| 3 | S | Preprocessing/Collecting DDIC Objects for InfoObject Catalog SELECAO | R7  | 000 | 4 |
| 4 | S | Post Processing/Checking the Deletion for InfoObject Catalog SELECAO | R7  | 000 | 4 |
| 5 | S | Object InfoObject Catalog SELECAO has been deleted          | RSO | 413 | 4 |
--------------------------------------------------------------------------------
Log number                           00000000000000014303
External ID                          ADMIN_WORKBENCH
Object                               RSSM
Subobject                            TCB
Date/Time/User                       04.07.2018 08:35:30 EVANS
-------------------------------------------------------------------
| M | T | Message Text                                       | ID   | No. | P |
-------------------------------------------------------------------
| 1 | I | ****  Start from: CL_NEW_AWB_AREA-DELETE_NODE      | RSM1 | 258 | 3 |
| 2 | I | Delete InfoArea INFOAREAENGIE from database        | RSM1 | 277 | 3 |
| 3 | I | InfoArea INFOAREAENGIE written to transport request | RSM  | 065 | 3 |
| 4 | I | InfoArea INFOAREAJUL13 written to transport request | RSM  | 065 | 3 |
| 5 | I | Delete 57 InfoAreas from database                  | RSM1 | 277 | 3 |
| 6 | I | Write 56 InfoAreas to database                     | RSM1 | 278 | 3 |
| 7 | I | ****  End from: CL_NEW_AWB_AREA-DELETE_NODE        | RSM1 | 259 | 3 |
-------------------------------------------------------------------
Log number                           00000000000000014304
External ID                          ADMIN_WORKBENCH
Object                               RSSM
Subobject                            TCB
Date/Time/User                       04.07.2018 08:35:30 EVANS
--------------------------------------------------------------------------------------
| M | T | Message Text                                                    | ID  | No. | P |
```

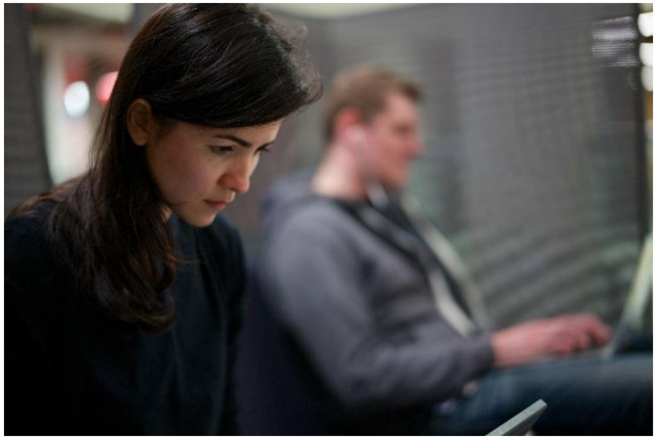
Send sample



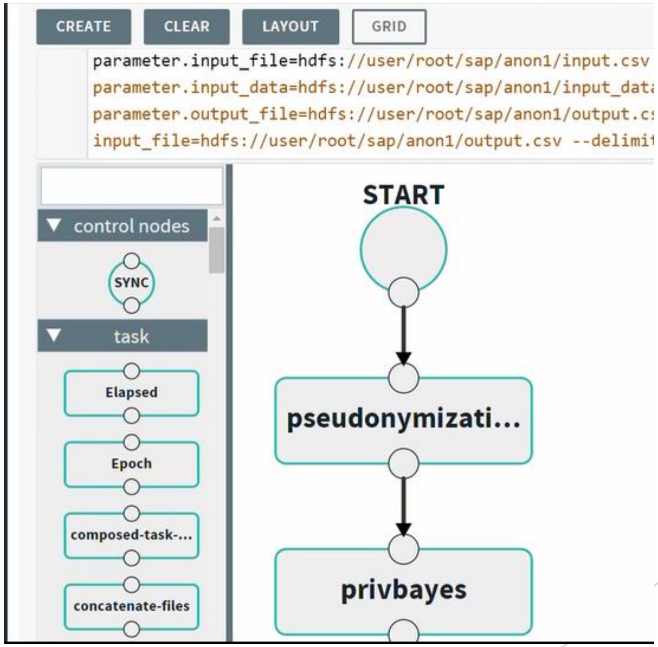
```
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
ADMIN_WORKBENCH,MNT,TCD,OBLITUS,ADMIN,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
PERSONAL,PDO,SDL,OBLIVIONI,LAMBDA,2018-06-01,BENIGN,OUTSIDE
```

Step 2: Privacy-preserving data transfer

Sample analysis



Data transfer



```
parameter.input_file=hdfs://user/root/sap/anon1/input.csv
parameter.input_data=hdfs://user/root/sap/anon1/input_data
parameter.output_file=hdfs://user/root/sap/anon1/output.cs
input_file=hdfs://user/root/sap/anon1/output.csv --delimit
```

Data preparation services

Pseudonymization

- Can be reverted by Bob
- Hides usernames
- Does not protect against background knowledge re-identification

Priv-Bayes

- Privacy-preserving data publishing
- Resistant to background knowledge attack
- Let Alice write, test and deploy her classifier without seeing the real data

The Dalenius requirement

Before contributing an entry *e* of *the training set f*, one could be tempted to require that ***computing F in production (i.e., the inference) should reveal absolutely nothing about f*****.**

- This is a re-phrasing of the classic Dalenius requirement for statistical databases
- Three decades of research in privacy have shown that <u>it cannot be achieved if side information about S is available</u>.

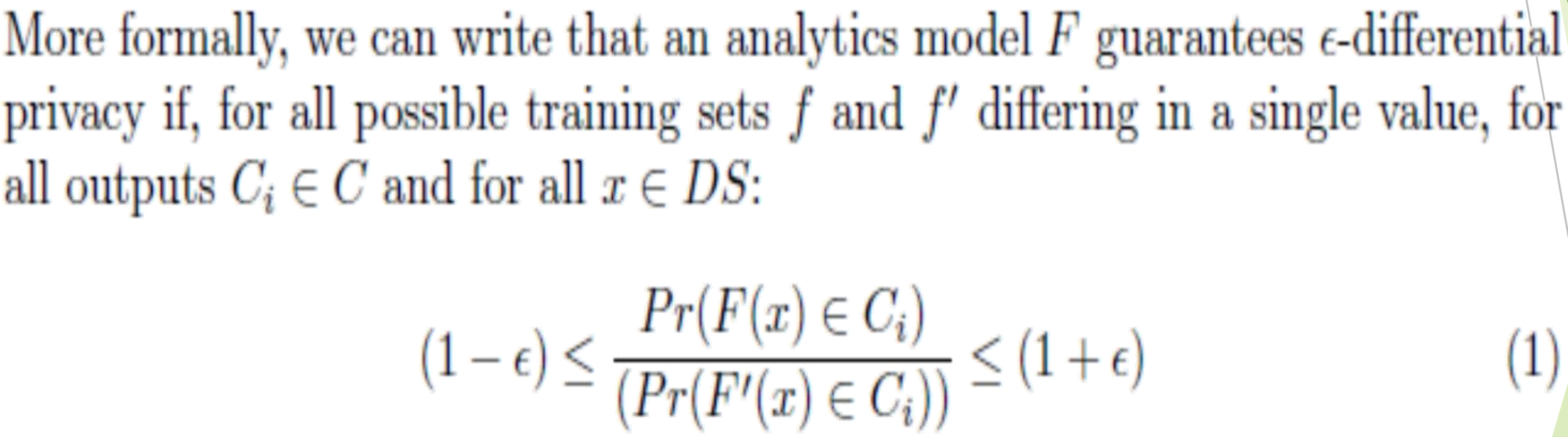
Differential privacy

Cynthia Dwork's seminal work has turned the "impossible" Dalenius requirement into an achievable goal:

Observing the execution of *F*, one should be able to infer the same information about any entry *e* of *f* as by observing *F'*, obtained using the training set *f' = f - {e} + {r}*, where *r* is a random entry.

This will provide the owner of *e* - assuming she has something to gain by knowing the result of *F* - with some rational motivation for contributing *e* to the training set, as she will be able to deny any specific claim on the value of e that anyone could put forward based on *F* **(plausible deniability)**.

Dwork’s formula



More formally, we can write that an analytics model F guarantees ϵ-differential privacy if, for all possible training sets f and f' differing in a single value, for all outputs $C_i \in C$ and for all $x \in DS$:

$$(1-\epsilon) \leq \frac{Pr(F(x) \in C_i)}{(Pr(F'(x) \in C_i))} \leq (1+\epsilon) \qquad (1)$$

Basic ideas

The basic approach to achieving differential privacy: Introducing a degree of randomization in the computation of F , making [F(x)] a *random variable* over DS.

Questions: How, where and when to inject randomization, depending on the nature of F ?

A sample randomization (1)

- A loan agency wishes to compute an estimate F of the *average amount of its loan requests*, and display it in a kiosk.
- There is however a privacy problem: *anyone who can guess the number n of borrowers, observing the average amount before and after a customer has applied for a loan, will be able to guess of the amount that the customer has borrowed*.
- The loan agency may protect its customers' privacy by adding to the loan requests some type of random noise. But which one ?

A sample randomization (2)

A popular probability density for such noise is the Laplace distribution

$$p(z) = e^{\frac{-|z|}{\sigma}} = e^{-|z|\epsilon}$$

- The distribution of this random variable is “concentrated around the truth": the probability that *[F]* is *z* units from F drops exponentially with *z*.
- This randomization surely introduces some uncertainty, as the screens no longer show *F* but the value of a random variable *[F]* with Laplace distribution whose average coincides with *F.*

A sample randomization (3)

Let’s see if this randomization guarantees that the overhead screen content will be (epsilon) *differentially private*.

Yes!

Replacing the last (actually, any) loan request by an arbitrary value in the range [0,MAX], one can shift the amount of the average loan by less than MAX/n; so, the probability value w $e^{\epsilon} \approx 1 + \epsilon$ by an amount smaller than as requested by Eq. (1)

Plausible deniability

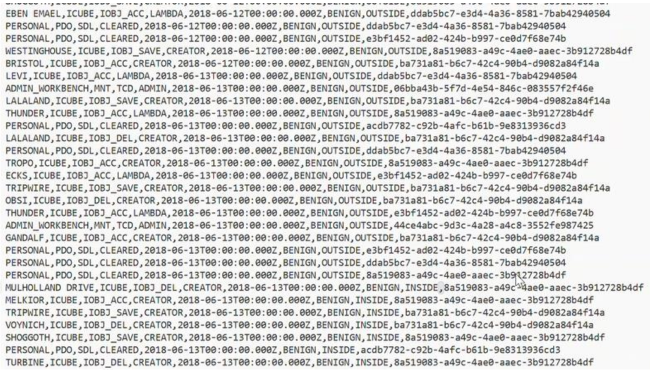
I deny it!
you borrowed 100$!

Bob is guessing the loan amount based on the new value in the kiosk
Alice is *plausibly denying* it, as any other value in the range could have caused the same change (with a probability depending on epsilon)
Epsilon measures the privacy achieved – and the accuracy lost.
Regulations (e.g., **EU GDPR**) dictate its value or at least its presence!

Step 3: Data analysis

Random Forest AI classifier

- Nosy admin: reads data it should not
- Escalation of privileges: non-admin user performs admin operation

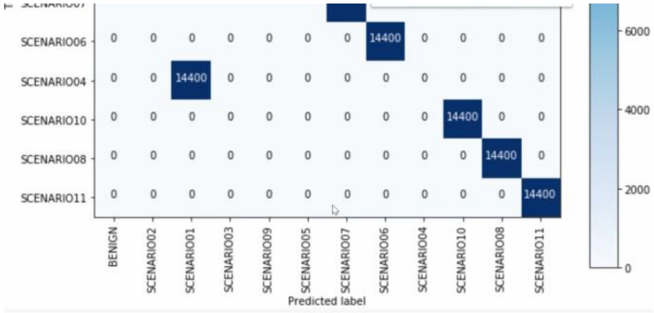


```
EBEN EMAEL,ICUBE,IOBJ_ACC,LAMBDA,2018-06-12T00:00:00.000Z,BENIGN,OUTSIDE,ddab5bc7-e3d4-4a36-8581-7bab42940504
PERSONAL,PDO,SDL,CLEARED,2018-06-12T00:00:00.000Z,BENIGN,OUTSIDE,ddab5bc7-e3d4-4a36-8581-7bab42940504
PERSONAL,PDO,SDL,CLEARED,2018-06-12T00:00:00.000Z,BENIGN,OUTSIDE,e3bf1452-ad02-424b-b997-ce0d7f68e74b
WESTINGHOUSE,ICUBE,IOBJ_SAVE,CREATOR,2018-06-12T00:00:00.000Z,BENIGN,OUTSIDE,8a519083-a49c-4ae0-aaec-3b912728b4df
BRISTOL,ICUBE,IOBJ_ACC,CREATOR,2018-06-12T00:00:00.000Z,BENIGN,OUTSIDE,ba731a81-b6c7-42c4-90b4-d9082a84f14a
LEVI,ICUBE,IOBJ_ACC,LAMBDA,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,ddab5bc7-e3d4-4a36-8581-7bab42940504
ADMIN_WORKBENCH,MNT,TCD,ADMIN,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,06bba43b-5f7d-4e54-846c-083557f2f46e
LALALAND,ICUBE,IOBJ_SAVE,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,ba731a81-b6c7-42c4-90b4-d9082a84f14a
THUNDER,ICUBE,IOBJ_ACC,LAMBDA,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,8a519083-a49c-4ae0-aaec-3b912728b4df
PERSONAL,PDO,SDL,CLEARED,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,acdb7782-c92b-4afc-b61b-9e8313936cd3
LALALAND,ICUBE,IOBJ_DEL,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,ba731a81-b6c7-42c4-90b4-d9082a84f14a
PERSONAL,PDO,SDL,CLEARED,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,ddab5bc7-e3d4-4a36-8581-7bab42940504
TROPO,ICUBE,IOBJ_ACC,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,8a519083-a49c-4ae0-aaec-3b912728b4df
ECKS,ICUBE,IOBJ_ACC,LAMBDA,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,e3bf1452-ad02-424b-b997-ce0d7f68e74b
TRIPWIRE,ICUBE,IOBJ_SAVE,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,ba731a81-b6c7-42c4-90b4-d9082a84f14a
OBSI,ICUBE,IOBJ_DEL,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,ba731a81-b6c7-42c4-90b4-d9082a84f14a
THUNDER,ICUBE,IOBJ_ACC,LAMBDA,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,e3bf1452-ad02-424b-b997-ce0d7f68e74b
ADMIN_WORKBENCH,MNT,TCD,ADMIN,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,44ce4abc-9d3c-4a28-a4c8-3552fe987425
GANDALF,ICUBE,IOBJ_ACC,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,ba731a81-b6c7-42c4-90b4-d9082a84f14a
PERSONAL,PDO,SDL,CLEARED,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,e3bf1452-ad02-424b-b997-ce0d7f68e74b
PERSONAL,PDO,SDL,CLEARED,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,ddab5bc7-e3d4-4a36-8581-7bab42940504
PERSONAL,PDO,SDL,CLEARED,2018-06-13T00:00:00.000Z,BENIGN,OUTSIDE,8a519083-a49c-4ae0-aaec-3b912728b4df
MULHOLLAND DRIVE,ICUBE,IOBJ_DEL,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,INSIDE,8a519083-a49c-4ae0-aaec-3b912728b4df
MELKIOR,ICUBE,IOBJ_ACC,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,INSIDE,8a519083-a49c-4ae0-aaec-3b912728b4df
TRIPWIRE,ICUBE,IOBJ_SAVE,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,INSIDE,ba731a81-b6c7-42c4-90b4-d9082a84f14a
VOYNICH,ICUBE,IOBJ_DEL,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,INSIDE,ba731a81-b6c7-42c4-90b4-d9082a84f14a
SHOGGOTH,ICUBE,IOBJ_SAVE,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,INSIDE,8a519083-a49c-4ae0-aaec-3b912728b4df
PERSONAL,PDO,SDL,CLEARED,2018-06-13T00:00:00.000Z,BENIGN,INSIDE,acdb7782-c92b-4afc-b61b-9e8313936cd3
TURBINE,ICUBE,IOBJ_DEL,CREATOR,2018-06-13T00:00:00.000Z,BENIGN,INSIDE,8a519083-a49c-4ae0-aaec-3b912728b4df
```

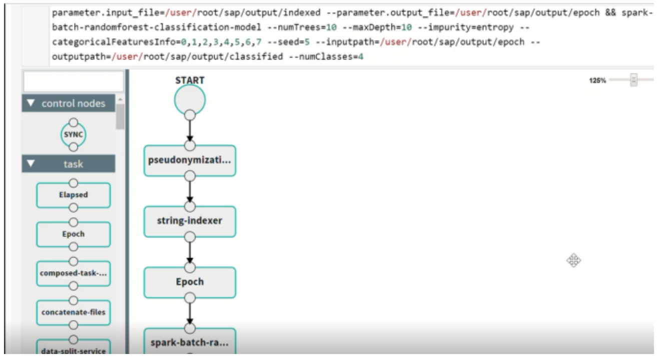
Step 3: Data analysis

Tuning accuracy…

Analytics pipeline



	BENIGN	SCENARIO02	SCENARIO01	SCENARIO03	SCENARIO09	SCENARIO05	SCENARIO07	SCENARIO06	SCENARIO04	SCENARIO10	SCENARIO08	SCENARIO11
SCENARIO06	0	0	0	0	0	0	0	14400	0	0	0	0
SCENARIO04	0	0	14400	0	0	0	0	0	0	0	0	0
SCENARIO10	0	0	0	0	0	0	0	0	0	14400	0	0
SCENARIO08	0	0	0	0	0	0	0	0	0	0	14400	0
SCENARIO11	0	0	0	0	0	0	0	0	0	0	0	14400



```
parameter.input_file=/user/root/sap/output/indexed --parameter.output_file=/user/root/sap/output/epoch && spark-
batch-randomforest-classification-model --numTrees=10 --maxDepth=10 --impurity=entropy --
categoricalFeaturesInfo=0,1,2,3,4,5,6,7 --seed=5 --inputpath=/user/root/sap/output/epoch --
outputpath=/user/root/sap/output/classified --numClasses=4
```

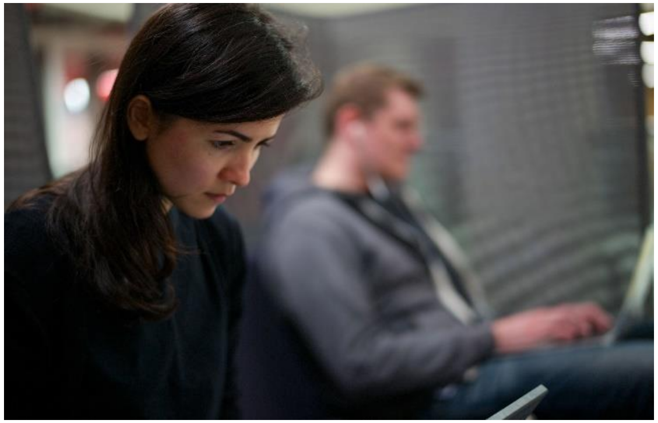
Step 3: Data analysis

- ERP → Log → Classifier
  - Categories
    - Benign
    - Nosy admin
    - Escalation of Privilege

Step 4: Additional analysis

Suspicious activity

“forgotten user”

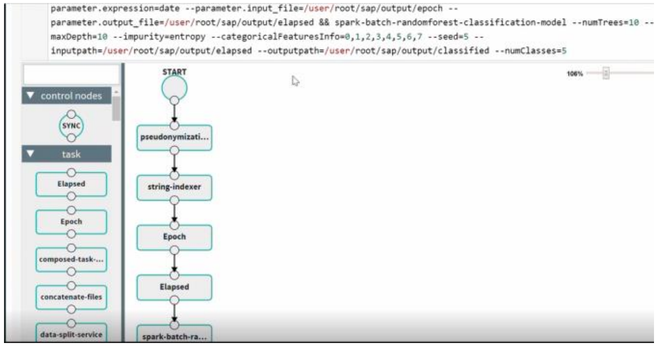


Pipeline update

Alice’s new pipeline

- Extra feature required: ‘Elapsed time’ (since last action)
- Re-trained model
  - Nosy admin
  - Elevation of privileges
  - Forgotten user

Pipeline update



```
parameter.expression=date --parameter.input_file=/user/root/sap/output/epoch --
parameter.output_file=/user/root/sap/output/elapsed && spark-batch-randomforest-classification-model --numTrees=10 --
maxDepth=10 --impurity=entropy --categoricalFeaturesInfo=0,1,2,3,4,5,6,7 --seed=5 --
inputpath=/user/root/sap/output/elapsed --outputpath=/user/root/sap/output/classified --numClasses=5
```

Problem solved. Again ☺

- ERP → Log → Classifier
  - Benign
  - Nosy admin
  - Escalation of Privilege
  - Forgotten user

Results

- Classifier metrics:
  - F1 score = 0.973141
  - Accuracy = 0.971297
  - weightedPrecision = 0.978695
  - weightedRecall = 0.971297
- Effort to switch from pipeline 2 to pipeline 3: 5 person-days
- The dataset has been released as opensource on openAIRE