Expert-in-the-Loop Supervised Learning for Computer Security Detection Systems

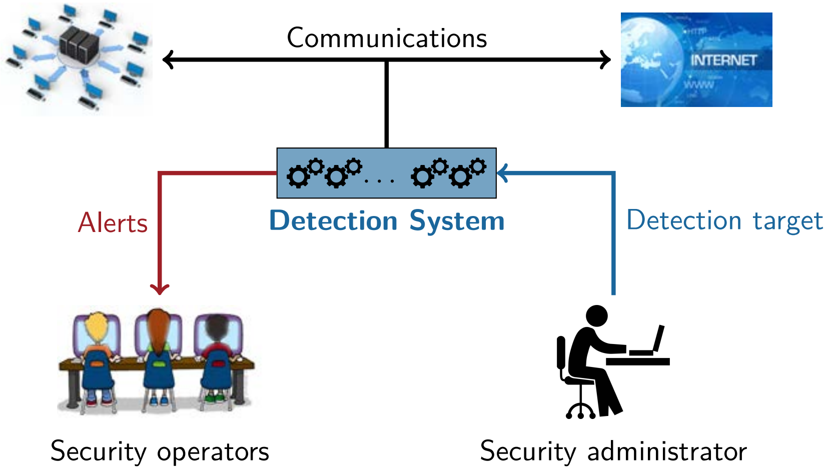
Anaël Beaugnon

anael.beaugnon@ssi.gouv.fr

Laboratoire Exploration et recherche en Détection (LED)

SoSySec 30/11/2018

Computer Security Detection Systems



Machine Learning for Detection Systems

Commercial Brochures

✓ 0-day

Academic Papers

✓ Outstanding results

Machine Learning for Detection Systems

Commercial Brochures

✓ 0-day

Academic Papers

✓ Outstanding results

Computer Security Experts

✗ Incomprehensible black box

✗ Too many false positives

✗ Unable to reproduce academic results in production

Machine Learning for Detection Systems

Commercial Brochures

✓ 0-day

Academic Papers

✓ Outstanding results

Computer Security Experts

✗ Incomprehensible black box

✗ Too many false positives

✗ Unable to reproduce academic results in production

Objective

How to make machine learning suit detection systems ?

Outline

1 Machine Learning Pipeline

2 ILAB: End-to-End Active Learning System

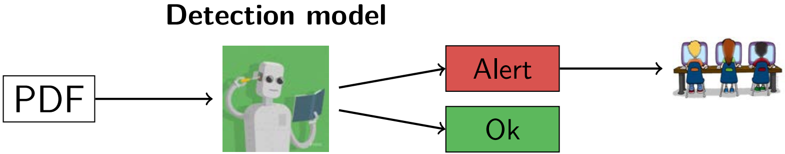
3 SecuML: Machine Learning for Computer Security

Outline

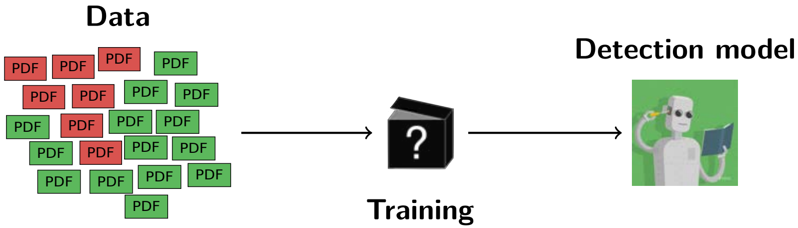
1. Machine Learning Pipeline
2. ILAB: End-to-End Active Learning System
3. SecuML: Machine Learning for Computer Security

Machine Learning Detection Models

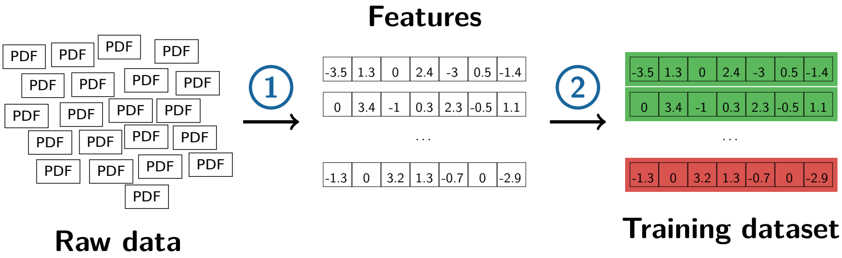
Binary Classifier



Building a Machine Learning Detection Model



Raw Data → Training Dataset



1. Feature extraction
2. Annotation

Feature Extraction

PDF Files

- Presence of JavaScript
- Presence of OpenActions
- Average size of the objects
- Num. images
- etc.

NetFlow Data

- Num. packets sent/received
- Num. bytes sent/received
- Num. contacted IP adresses
- Num. contacted ports
- etc.

Discriminating Features

- Expert knowledge
- Academic publications

Training and Evaluation

Training algorithm

-3.5	1.3	0	2.4	-3	0.5	-1.4
0	3.4	-1	0.3	2.3	-0.5	1.1

...

-1.3	0	3.2	1.3	-0.7	0	-2.9

3 →

→

4 → Production ?

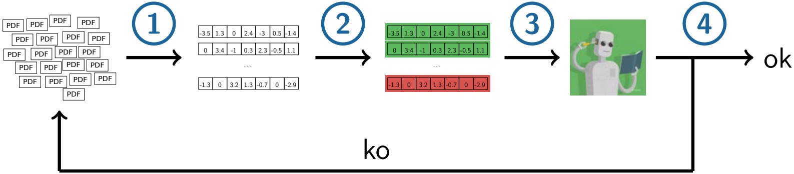
Training dataset

Detection model

3 Which model class ?

4 Evaluation

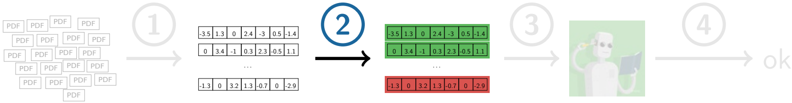
The Whole Machine Learning Pipeline



1. Feature extraction
2. Annotation
3. Which model class ?
4. Evaluation

PhD Thesis A. Beaugnon, *Expert-in-the-Loop Supervised Learning for Computer Security Detection Systems*

Annotating a Dataset with a Reduced Workload



References for Other Steps

PhD Thesis	A. Beaugnon, *Expert-in-the-Loop Supervised Learning for Computer Security Detection Systems*
SSTIC'17	A. Beaugnon, A. Husson, P. Chifflier, *Le Machine Learning confronté aux contraintes opérationnelles des systèmes de détection*
C&ESAR'18	A. Beaugnon, P.Chifflier, *Machine Learning for Computer Security Detection Systems: Practical Feedback and Solutions*

Outline

1 Machine Learning Pipeline

2 ILAB: End-to-End Active Learning System

3 SecuML: Machine Learning for Computer Security

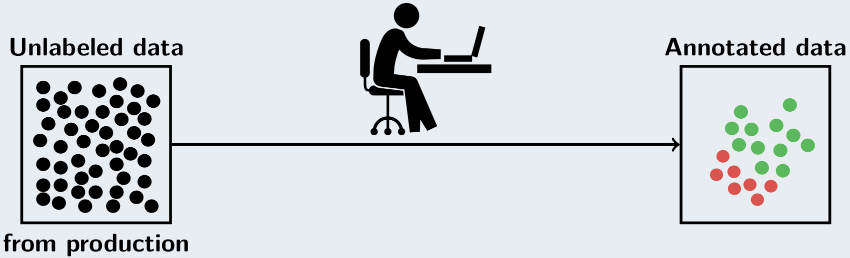
Lack of Good Training Data

- ✗ Public annotated datasets (often biased)
- ✗ Crowd-sourcing

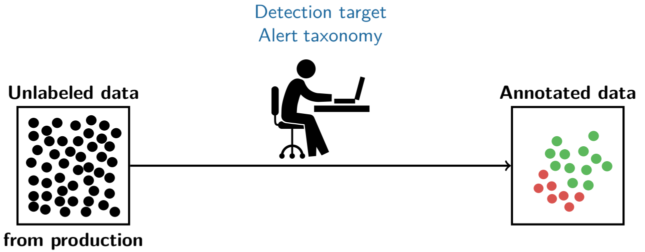
Lack of Good Training Data

- ✗ Public annotated datasets (often biased)
- ✗ Crowd-sourcing

Solution: *In-situ* Annotations



In-situ Annotations



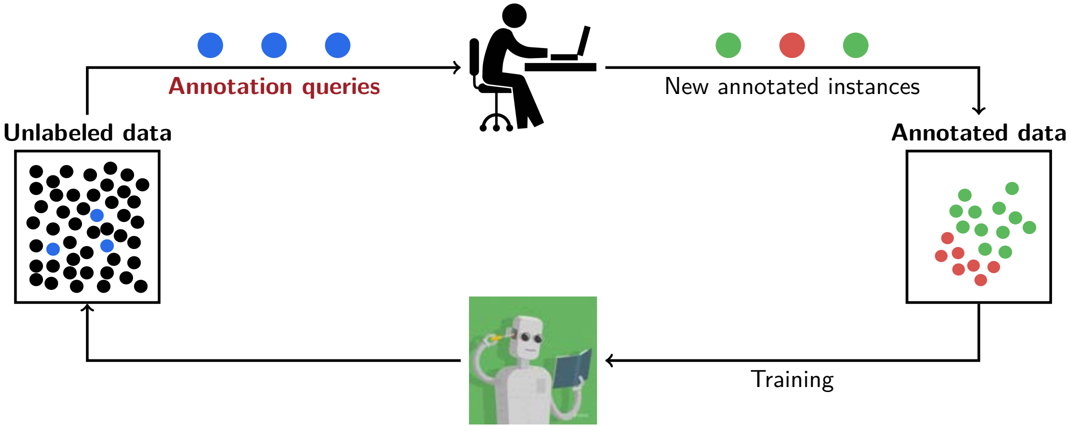
Annotation

binary label
+
family

Binary labels $\longleftrightarrow$ Detection target

Malicious families $\longleftrightarrow$ Alert taxonomy

Iterative Process



In-situ Annotations

Objectives

- Maximize the detection performance
- Minimize the human workload
  - Number of manual annotations
  - Global annotation time

In-situ Annotations

Objectives

- Maximize the detection performance
- Minimize the human workload
  - Number of manual annotations
  - Global annotation time

Challenges

1. Which instances should be annotated ?
   - ✗ Uniform random selection
2. How to design the user interface ?

In-situ Annotations with ILAB

End-to-End Active Learning System

Active Learning Strategy
$+$
Annotation System

Active Learning Strategy

Queries the instances to annotate cleverly.

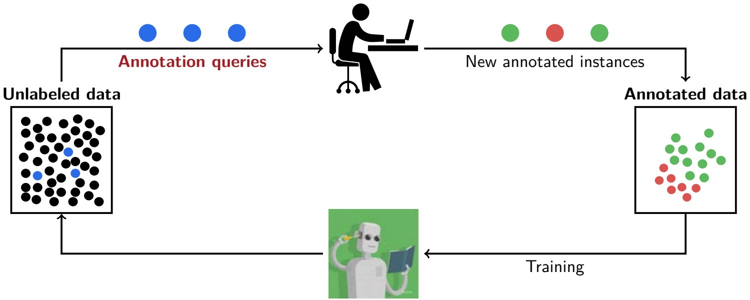
RAID'17 *ILAB: An Interactive Labelling Strategy for Intrusion Detection*

Annotation System

Suits computer security experts' needs.

AICS'18, IDEA'18 *End-to-End Active Learning for Computer Security Experts*

Active Learning Strategy



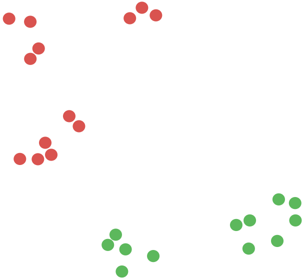
Which instances should be annotated ?

For an annotation budget B:

- Maximize the detection performance
- Minimize the waiting-periods

Uncertainty Sampling

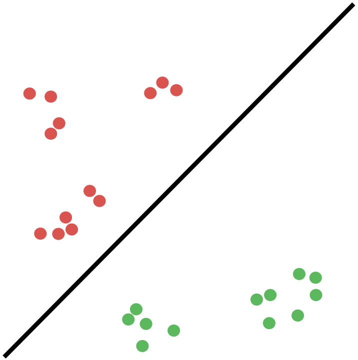
An Active Learning Strategy



Lewis et al. A sequential algorithm for training text classifiers, 1994.

Uncertainty Sampling

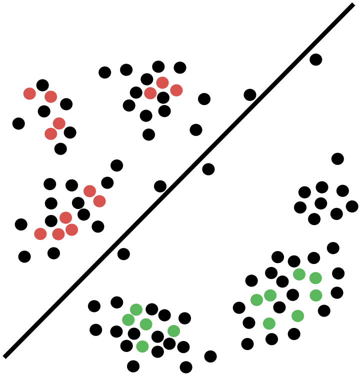
An Active Learning Strategy



Lewis et al. A sequential algorithm for training text classifiers, 1994.

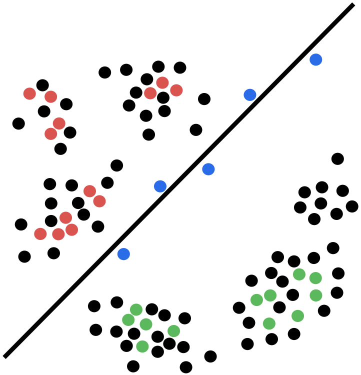
Uncertainty Sampling

An Active Learning Strategy



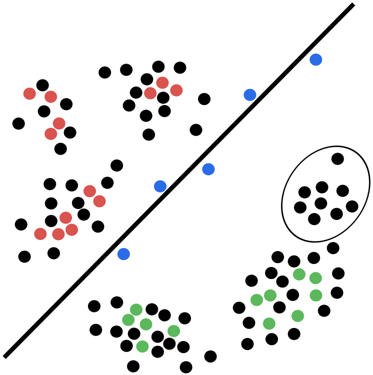
Lewis et al. A sequential algorithm for training text classifiers, 1994.

Uncertainty Sampling



Lewis et al. A sequential algorithm for training text classifiers, 1994.

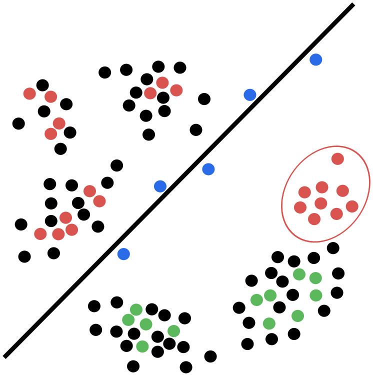
Uncertainty Sampling



Lewis et al. A sequential algorithm for training text classifiers, 1994.

Uncertainty Sampling

An Active Learning Strategy



Sampling Bias

A misclassified cluster is completely overlooked !

Lewis et al. A sequential algorithm for training text classifiers, 1994.

Schütz et al. Performance thresholding in practical text classification, CIKM'06.

Uncertainty Sampling

Sampling biases degrade the detection performance.

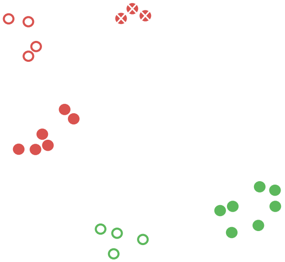
	Uncertainty Sampling
Maximize the detection performance	✗
Minimize the waiting-periods	✓

Sampling biases degrade the detection performance.

	Uncertainty Sampling
Maximize the detection performance	✗
Minimize the waiting-periods	✓

How to avoid sampling biases without lengthening the waiting-periods ?

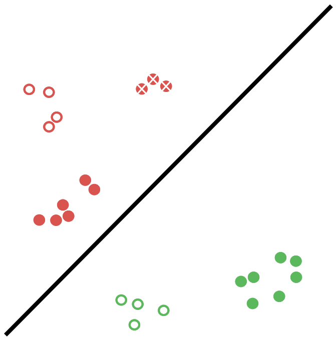
Annotation: binary label + family



Annotation: binary label + family

1. Binary logistic regression

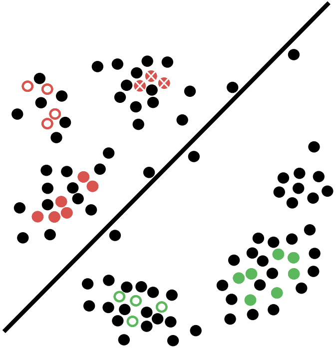
$P(y = 1 \mid x) = \frac{1}{1+\exp\left(-\left(\mathbf{w}^T x + b\right)\right)}$



Annotation: binary label + family

1. Binary logistic regression

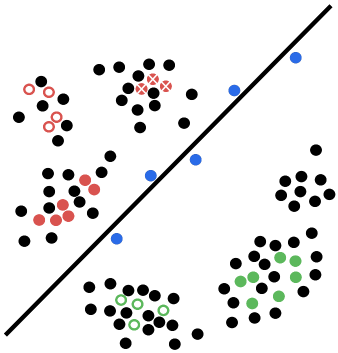
$$P(y = 1 \mid x) = \frac{1}{1+\exp\left(-\left(\mathbf{w}^T x+b\right)\right)}$$



ILAB: Active Learning Strategy

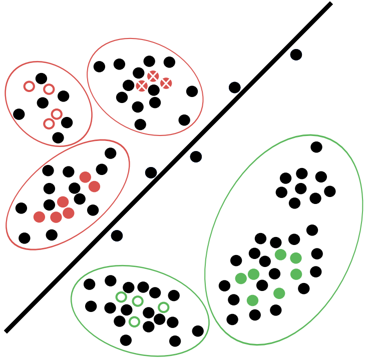
Annotation: binary label + family

1. Binary logistic regression
   $P(y = 1 \mid x) = \frac{1}{1+\exp\left(-\left(\mathbf{w}^T x + b\right)\right)}$
2. Uncertainty sampling



Annotation: binary label + family

1. Binary logistic regression
   $P(y = 1 \mid x) = \frac{1}{1+\exp\left(-\left(\mathbf{w}^T x + b\right)\right)}$
2. Uncertainty sampling
3. Rare category detection

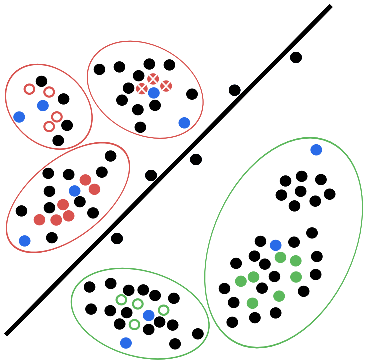


Clusters = User-defined families

ILAB: Active Learning Strategy

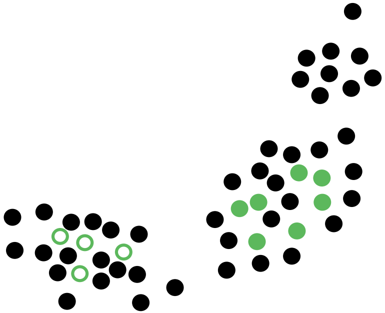
Annotation: binary label + family

1. Binary logistic regression
   $P(y = 1 \mid x) = \frac{1}{1+\exp\left(-\left(\mathbf{w}^T x + b\right)\right)}$
2. Uncertainty sampling
3. Rare category detection



Clusters = User-defined families

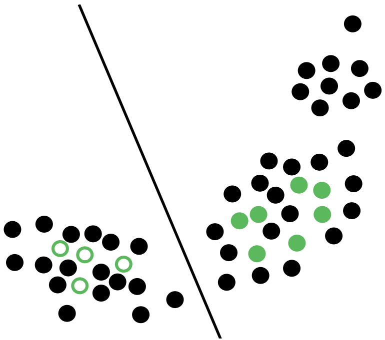
Rare Category Detection



Pelleg et Moore Active learning for anomaly and rare category detection, NIPS'05.

Rare Category Detection

1. Multi-class logistic regression

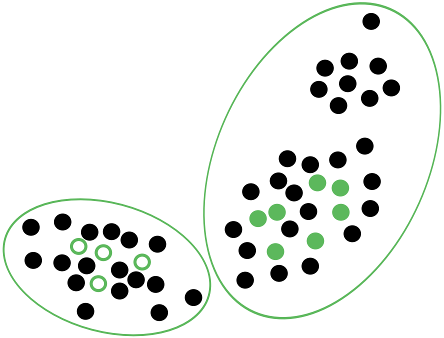


Pelleg et Moore Active learning for anomaly and rare category detection, NIPS'05.

Rare Category Detection

1. Multi-class logistic regression
2. Gaussian mixture:

$$p_{\mathcal{N}(\mu_f,\Sigma_f)}(x) \propto \exp\left(-\frac{1}{2}\left\|\Sigma_f^{-\frac{1}{2}}(x-\mu_f)\right\|^2\right)$$



Pelleg et Moore Active learning for anomaly and rare category detection, NIPS'05.

Rare Category Detection

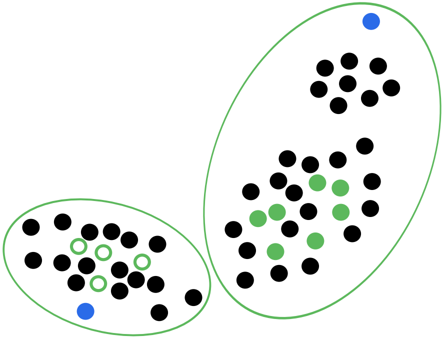
Avoiding Sampling Biases

1. Multi-class logistic regression
2. Gaussian mixture:

$$p_{\mathcal{N}(\mu_f,\Sigma_f)}(x) \propto \exp\left(-\frac{1}{2}\left\|\Sigma_f^{-\frac{1}{2}}(x-\mu_f)\right\|^2\right)$$

3. Annotation queries
   - **Detect new families**

$$\underset{x\in\mathcal{C}_f}{\arg\min}\; p_{\mathcal{N}(\mu_f,\Sigma_f)}(x)$$



Pelleg et Moore Active learning for anomaly and rare category detection, NIPS'05.

Rare Category Detection

1. Multi-class logistic regression
2. Gaussian mixture:

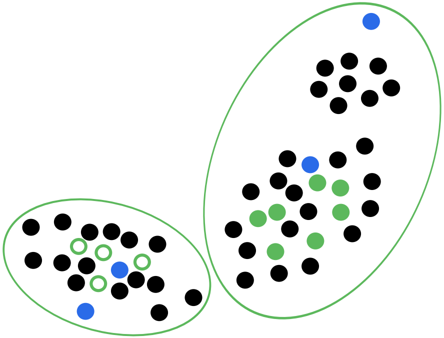
$$p_{\mathcal{N}(\mu_f,\Sigma_f)}(x) \propto \exp\left(-\frac{1}{2}\left\|\Sigma_f^{-\frac{1}{2}}(x-\mu_f)\right\|^2\right)$$

3. Annotation queries
   - **Detect new families**

$$\underset{x\in\mathcal{C}_f}{\arg\min}\; p_{\mathcal{N}(\mu_f,\Sigma_f)}(x)$$

   - **Representative instances**

$$\underset{x\in\mathcal{C}_f}{\arg\max}\; p_{\mathcal{N}(\mu_f,\Sigma_f)}(x)$$



Pelleg et Moore Active learning for anomaly and rare category detection, NIPS'05.

Rare Category Detection

1. Multi-class logistic regression
2. Gaussian mixture:

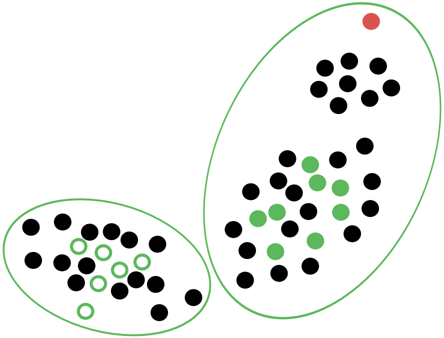
$$p_{\mathcal{N}(\mu_f,\Sigma_f)}(x) \propto \exp\left(-\frac{1}{2}\left\|\Sigma_f^{-\frac{1}{2}}(x-\mu_f)\right\|^2\right)$$

3. Annotation queries
   - **Detect new families**

$$\arg\min_{x \in \mathcal{C}_f} \; p_{\mathcal{N}(\mu_f,\Sigma_f)}(x)$$

   - **Representative instances**

$$\arg\max_{x \in \mathcal{C}_f} \; p_{\mathcal{N}(\mu_f,\Sigma_f)}(x)$$



Pelleg et Moore Active learning for anomaly and rare category detection, NIPS'05.

Rare Category Detection

1. Multi-class logistic regression
2. Gaussian mixture:

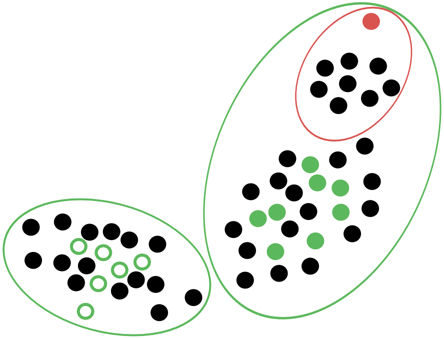
$$p_{\mathcal{N}(\mu_f,\Sigma_f)}(x) \propto \exp\left(-\frac{1}{2}\left\|\Sigma_f^{-\frac{1}{2}}(x-\mu_f)\right\|^2\right)$$

3. Annotation queries
   - **Detect new families**

$$\arg\min_{x\in\mathcal{C}_f} \; p_{\mathcal{N}(\mu_f,\Sigma_f)}(x)$$

   - **Representative instances**

$$\arg\max_{x\in\mathcal{C}_f} \; p_{\mathcal{N}(\mu_f,\Sigma_f)}(x)$$



Pelleg et Moore   Active learning for anomaly and rare category detection, NIPS'05.

ILAB: Active Learning Strategy

Divide-and-Conquer Approach

- Reduce the temporal complexity
- Annotating while computing

Uncertainty sampling	Malicious analysis	Benign analysis
↓	↓	↓
Uncertain queries	Malicious queries	Benign queries

ILAB: Active Learning Strategy

Divide-and-Conquer Approach

- Reduce the temporal complexity
- Annotating while computing

Uncertainty sampling	Malicious analysis	Benign analysis
Uncertain queries	Malicious queries	Benign queries
Annotation	**Computation ...**	**Computation ...**

ILAB: Active Learning Strategy

Avoiding Sampling Biases

Rare category detection

Reducing the Waiting-Periods

Divide-and-conquer

Comparison to State-of-the-Art Strategies

Simulations on Annotated Datasets

	#instances	#features
Contagio	10,000	113
NSL-KDD	74,826	122

Active Learning Strategies

Uncertainty Almgren et al., Using Active Learning in Intrusion Detection, CSFW 2004.

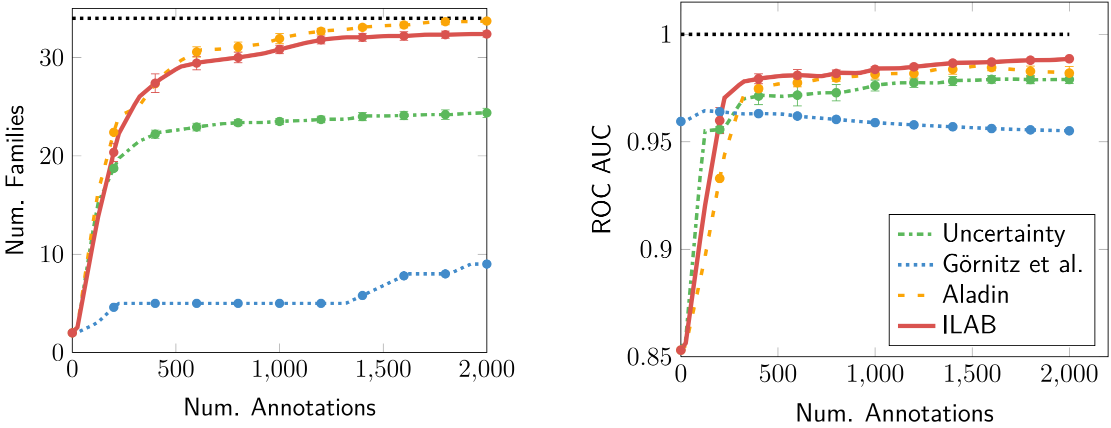
Görnitz et al. Görnitz et al., Toward Supervised Anomaly Detection, JAIR 2013.

Aladin Stokes et al., Aladin: Active Learning of Anomalies to Detect Intrusions, 2008.

ILAB **Beaugnon et al., ILAB: An Interactive Labelling Strategy for Intrusion Detection, RAID 2017.**

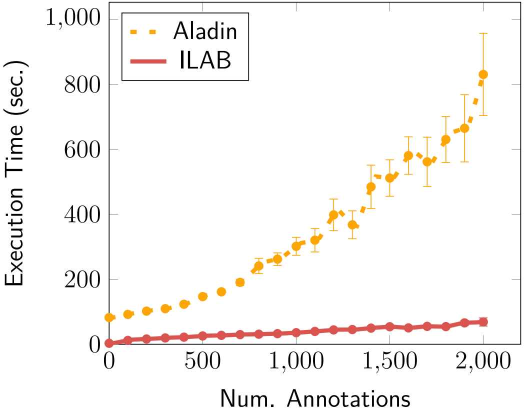
Comparison to State-of-the-Art Strategies

ILAB and Aladin detect properly the different families.



Comparison to State-of-the-Art Strategies

Waiting-periods reduced thanks to ILAB.



Comparison to State-of-the-Art Strategies

ILAB avoids sampling biases without increasing the waiting-periods.

	Uncertainty	Görnitz et al.	Aladin	**ILAB**
No bias	✗	✗	✓	✓
Quick	✓	✗	✗	✓

https://github.com/ANSSI-FR/SecuML

Uncertainty Almgren et al., Using Active Learning in Intrusion Detection, CSFW 2004.
Görnitz et al. Görnitz et al., Toward Supervised Anomaly Detection, JAIR 2013.
Aladin Stokes et al., Aladin: Active Learning of Anomalies to Detect Intrusions, 2008.
ILAB **Beaugnon et al., ILAB: An Interactive Labelling Strategy for Intrusion Detection, RAID 2017.**

Comparison to State-of-the-Art Strategies

ILAB avoids sampling biases without increasing the waiting-periods.

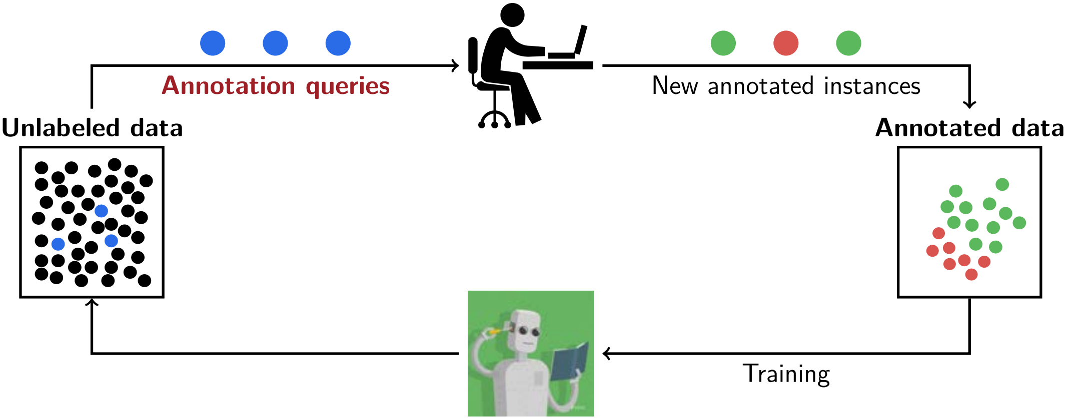
	Uncertainty	Görnitz et al.	Aladin	**ILAB**
No bias	✗	✗	✓	✓
Quick	✓	✗	✗	✓

`https://github.com/ANSSI-FR/SecuML`

Uncertainty Almgren et al., Using Active Learning in Intrusion Detection, CSFW 2004.
Görnitz et al. Görnitz et al., Toward Supervised Anomaly Detection, JAIR 2013.
Aladin Stokes et al., Aladin: Active Learning of Anomalies to Detect Intrusions, 2008.
ILAB **Beaugnon et al., ILAB: An Interactive Labelling Strategy for Intrusion Detection, RAID 2017.**

What do computer security experts think ?

Don’t forget the expert !



Don't forget the expert !

	Simulations	GUI	User Exp.
Uncertainty	✓	✗	✗
Görnitz et al.	✓	✗	✗
Aladin	✓	∼	∼
Nissim et al.	✓	✗	✗
Moskovitch et al.	✓	✗	✗

Aladin

- No information about the user interface
- 1000 annotations a day without any feedback !

Uncertainty Almgren et al., Using Active Learning in Intrusion Detection, CSFW 2004

Görnitz et al. Toward Supervised Anomaly Detection, JAIR 2013

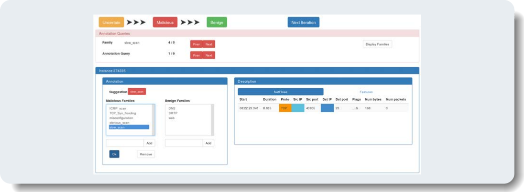
Aladin Stokes et al., Aladin: Active Learning of Anomalies to Detect Intrusions, 2008

Nissim et al. ALPD: Active learning framework for enhancing the detection of malicious PDF files, 2014.

Moskovitch et al. Malicious code detection using active learning, 2009.

ILAB: Annotation System

A user interface that suits security experts' needs.



Online documentation: https://anssi-fr.github.io/SecuML/

User Experience

Four security experts

Datasets

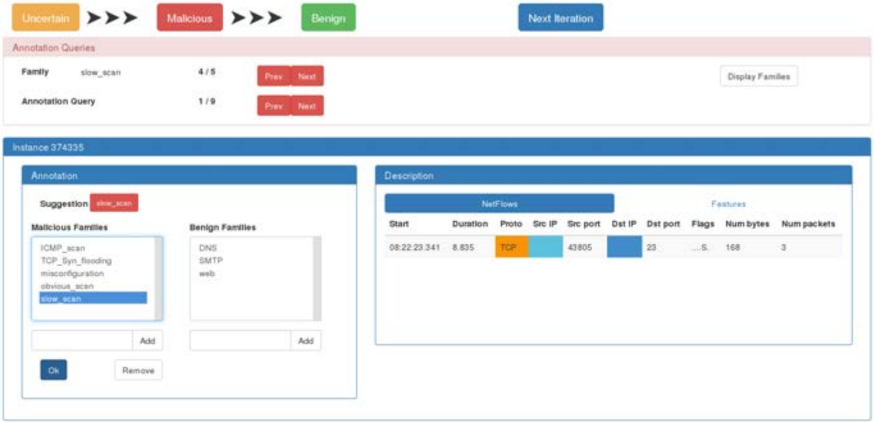
	Jour 1	Jour 2
Num. flows	$1.2 \cdot 10^8$	$1.2 \cdot 10^8$
Num. IPs	$463,913$	$507,258$
Num. features	134	134

Initial Annotations

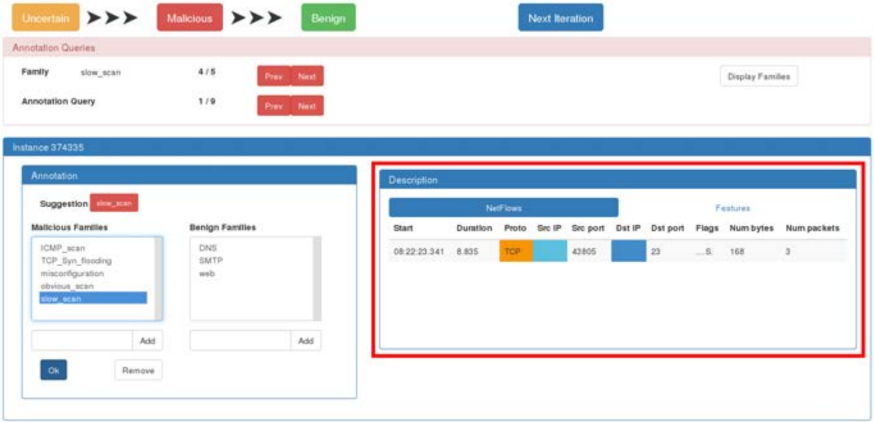
- Anomalous instances
  obvious scans
- Normal instances
  uniform selection

ILAB: Annotation Systems

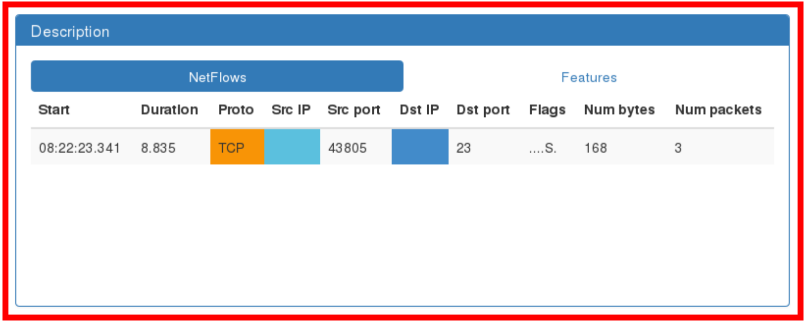
Annotation Interface



ILAB: Annotation Systems



NetFlow


Description

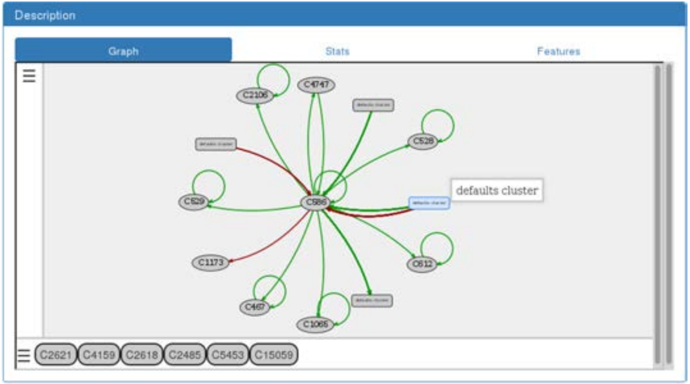
NetFlows | Features

Start	Duration	Proto	Src IP	Src port	Dst IP	Dst port	Flags	Num bytes	Num packets
08:22:23.341	8.835	TCP		43805		23	....S.	168	3

Problem-Specific Visualizations

Display any Data Type

Windows Event Logs

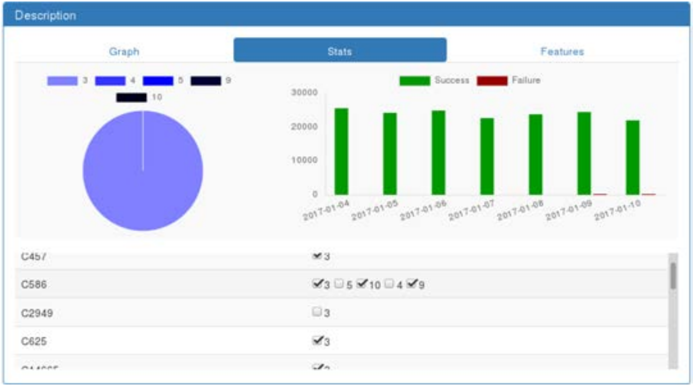


Courtesy C. Larroche

Problem-Specific Visualizations

Display any Data Type

Windows Event Logs



Courtesy C. Larroche

ILAB: Annotation System

Help Annotators to Remain Consistent

Family Editor

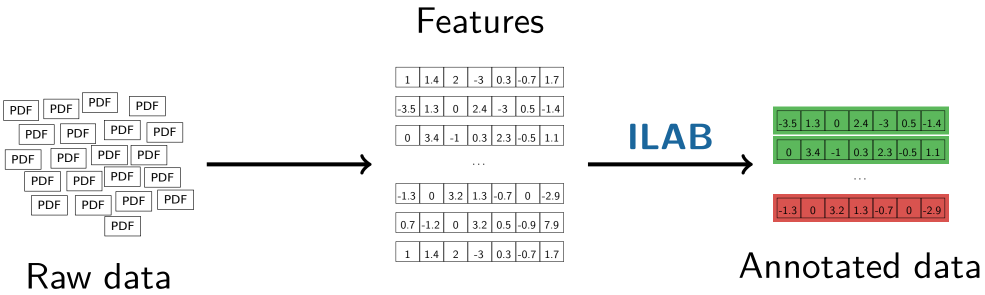
- Change a family name
- Change the label associated to a family
- Merge families

Kulesza et al. Structured labeling for facilitating concept evolution in machine learning, CHI 2014.

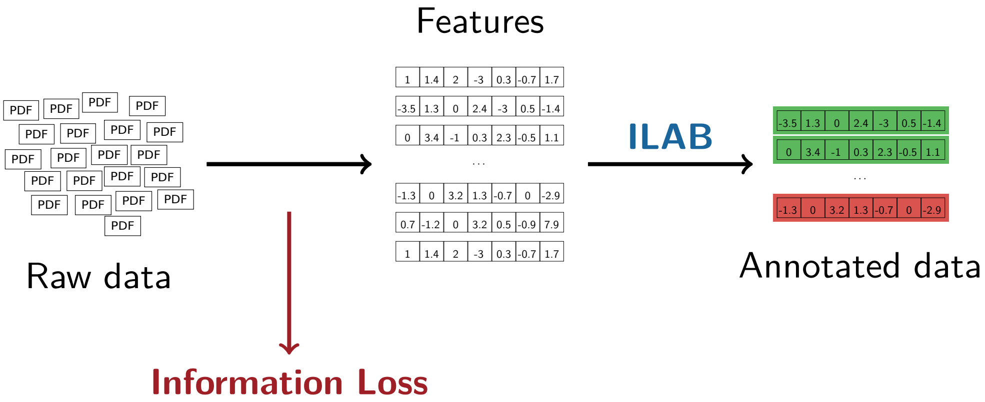
Widely used during the user experiments

- Delineate the detection target
- Define the alert taxonomy

Strong Link between Features and Annotations



Strong Link between Features and Annotations



Strong Link between Features and Annotations

The extracted features may not be expressive enough.

Features

Num. bytes sent/received :

- globally
- on port 80
- on port 53
- on port 25

Annotation

- full TCP connection
- on **port 22**
- **normal**

Annotation

- full TCP connection
- on **port 1258**
- **anomalous**

Strong Link between Features and Annotations

Knowledge about the Features

- Expressiveness

Strong Link between Features and Annotations

Knowledge about the Features

- Expressiveness

Make Features Evolve

- manually
- or even better, manually

Khiops	Boulle, Towards automatic feature construction for supervised classification, ECML'14.
Featuretools	Kanter et al., Deep feature synthesis: towards automating data science endeavors, DSAA' 15.
Hidost	Šrndić et al., Hidost: a static machine learning based detector of malicious files, EURASIP'16.

ILAB: and End-to-End Active Learning System

Active Learning Strategy

- Avoid sampling biases
- Maintain low waiting-periods

RAID'17 *ILAB: An Interactive Labelling Strategy for Intrusion Detection*

Annotation System

- Generic annotation interface
- Family editor

IDEA'18 *End-to-End Active Learning for Computer Security Experts*

User experiments !

Outline

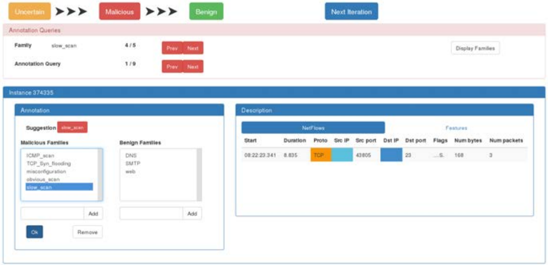
1 Machine Learning Pipeline

2 ILAB: End-to-End Active Learning System

3 SecuML: Machine Learning for Computer Security

In-situ annotations with ILAB (Interactive LABeling)

- **Active Learning Strategy**
  Selects cleverly the instances to be annotated.
- **Annotation System**
  GUI that suits security experts' needs.



RAID'17 A. Beaugnon, P.Chifflier, F. Bach, *ILAB: An Interactive Labelling Strategy for Intrusion Detection*

AICS'18, IDEA'18 A. Beaugnon, P.Chifflier, F. Bach, *End-to-End Active Learning for Computer Security Experts*

DIADEM (DIAgnosis of DEtection Models)

- Hide the machine learning machinery
- Diagnosis GUI
  - Performance indicators
  - Model behavior
  - Individual predictions

SSTIC'17 A. Beaugnon, A. Husson, P. Chifflier, *Le Machine Learning confronté aux contraintes opérationnelles des systèmes de détection*

C&ESAR'18 A. Beaugnon, P.Chifflier, *Machine Learning for Computer Security Detection Systems: Practical Feedback and Solutions*

Clustering

- K-means
- Gaussian mixtures
- DBSCAN
- etc.

Data Visualization with Projections

- **Unsupervised**
  - Principal Component Analysis (PCA)
- **Semi-supervised**
  - Relative Components Analysis (RCA)
  - Large Margin Nearest Neighbor (LMNN)
  - Neighborhood Components Analysis (NCA)
  - etc.

SecuML: Machine Learning for Computer Security Experts

Analysis Modules

- Data annotation with ILAB
- Diagnosis of detection models with DIADEM
- Clustering
- Data visualization with projections

Generic Solution

- Problem-specific visualizations
- Features as input

- Open source implementation: `https://github.com/ANSSI-FR/SecuML`
- Online documentation: `https://anssi-fr.github.io/SecuML/`