
Anaël Beaugnon
anael.beaugnon@ssi.gouv.fr

Laboratoire Exploration et recherche en Détection (LED)

SoSySec 30/11/2018
Computer Security Detection Systems

Communications

Detection System

Alerts

Security operators

Security administrator

Detection target
Commercial Brochures
✓ 0-day

Academic Papers
✓ Outstanding results
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
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?
1 Machine Learning Pipeline

2 ILAB: End-to-End Active Learning System

3 SecuML: Machine Learning for Computer Security
1. Machine Learning Pipeline

2. ILAB: End-to-End Active Learning System

Machine Learning Detection Models

Detection model

PDF -> Alert
PDF -> Ok
Building a Machine Learning Detection Model

Data → Training → Detection model
Raw Data → Training Dataset

1. Feature extraction
2. Annotation

Features

Raw data

Training dataset
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

Training dataset

Detection model

③ Which model class?

④ Evaluation

Production?
The Whole Machine Learning Pipeline

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

Annotating a Dataset with a Reduced Workload

ILAB: and End-to-End Active Learning System

References for Other Steps


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

Outline

1. Machine Learning Pipeline
2. ILAB: End-to-End Active Learning System
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

Unlabeled data from production

Annotated data
In-situ Annotations

Unlabeled data from production

Detection target Alert taxonomy

Annotated data

Annotation

- Binary label
- Family

Binary labels $\leftrightarrow$ Detection target

Malicious families $\leftrightarrow$ Alert taxonomy
Iterative Process

Unlabeled data → Annotation queries → New annotated instances → Annotated data → Training → Unlabeled data
**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

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

Objectives

Which instances should be annotated?

For an annotation budget $B$:
- Maximize the detection performance
- Minimize the waiting-periods
Uncertainty Sampling

An Active Learning Strategy

Uncertainty Sampling

An Active Learning Strategy

Uncertainty Sampling

An Active Learning Strategy

Uncertainty Sampling

An Active Learning Strategy

Uncertainty Sampling

An Active Learning Strategy

Uncertainty Sampling

An Active Learning Strategy

Sampling Bias
A misclassified cluster is completely overlooked!

Schütz et al. Performance thresholding in practical text classification, CIKM’06.
Sampling biases degrade the detection performance.

<table>
<thead>
<tr>
<th>Uncertainty Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximize the detection performance</td>
</tr>
<tr>
<td>Minimize the waiting-periods</td>
</tr>
</tbody>
</table>
### Sampling biases degrade the detection performance.

<table>
<thead>
<tr>
<th></th>
<th>Uncertainty Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximize the detection performance</td>
<td>✗</td>
</tr>
<tr>
<td>Minimize the waiting-periods</td>
<td>✓</td>
</tr>
</tbody>
</table>

How to avoid sampling biases without lengthening the waiting-periods?
Annotation: binary label + family
ILAB: Active Learning Strategy

Avoiding Sampling Biases

Annotation: binary label + family

1. Binary logistic regression

\[ P(y = 1 \mid x) = \frac{1}{1 + \exp(- (w^T x + b))} \]
1 Binary logistic regression

\[ P(y = 1 \mid x) = \frac{1}{1 + \exp(- (w^T x + b))} \]
**Annotation:** binary label + family

1. **Binary logistic regression**
   
   \[ P(y = 1 \mid x) = \frac{1}{1 + \exp(- (w^T x + b))} \]

2. **Uncertainty sampling**
ILAB: Active Learning Strategy

Avoiding Sampling Biases

Annotation: binary label + family

1. Binary logistic regression
   \[ P(y = 1 \mid x) = \frac{1}{1+\exp(- (w^T x + b))} \]

2. Uncertainty sampling

3. Rare category detection

Clusters = User-defined families
ILAB: Active Learning Strategy

Avoiding Sampling Biases

Annotation: binary label + family

1. Binary logistic regression
   \[ P(y = 1 \mid x) = \frac{1}{1 + \exp(- (w^T x + b))} \]

2. Uncertainty sampling

3. Rare category detection

Clusters = User-defined families
Rare Category Detection

Avoiding Sampling Biases

Multi-class logistic regression

Gaussian mixture:
\[ p_{\mathcal{N}}(\mu_f, \Sigma_f)(x) \propto \exp\left(-\frac{1}{2} \text{\textbf{v}}^T \Sigma^{-1} \text{\textbf{v}}\right) \]

Annotation queries
▶ Detect new families
\[ \arg \min_{x \in C} p_{\mathcal{N}}(\mu_f, \Sigma_f)(x) \]
▶ Representative instances
\[ \arg \max_{x \in C} p_{\mathcal{N}}(\mu_f, \Sigma_f)(x) \]

Pelleg et Moore Active learning for anomaly and rare category detection, NIPS’05.
1. Multi-class logistic regression

\[ \text{Annotation queries} \]

\[ \text{Detect new families} \]

\[ \arg \min_{x \in C} f \]

\[ p_{N}(\mu_f, \Sigma_f)(x) \propto \exp \left( -\frac{1}{2} \| \Sigma^{-1/2} f (x - \mu_f) \|^2 \right) \]

\[ \text{Representative instances} \]

\[ \arg \max_{x \in C} f \]

\[ p_{N}(\mu_f, \Sigma_f)(x) \]

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

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

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

Pelleg et Moore  Active learning for anomaly and rare category detection, NIPS’05.
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**
     \[ \arg \min_{x \in C_f} p_{\mathcal{N}(\mu_f, \Sigma_f)}(x) \]

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

1. Multi-class logistic regression
2. Gaussian mixture:
   \[ p_{N(\mu_f,\Sigma_f)}(x) \propto \exp\left(-\frac{1}{2} \left\| \Sigma_f^{-\frac{1}{2}} (x - \mu_f) \right\|^2\right) \]
3. Annotation queries
   - Detect new families
     \[ \arg\min_{x \in C_f} p_{N(\mu_f,\Sigma_f)}(x) \]
   - Representative instances
     \[ \arg\max_{x \in C_f} p_{N(\mu_f,\Sigma_f)}(x) \]

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
     \[ \arg \min_{x \in C_f} p_{\mathcal{N}(\mu_f, \Sigma_f)}(x) \]
   - Representative instances
     \[ \arg \max_{x \in C_f} p_{\mathcal{N}(\mu_f, \Sigma_f)}(x) \]

Pelleg et Moore  Active learning for anomaly and rare category detection, NIPS'05.
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
     \[ \arg \min_{x \in C_f} p_{\mathcal{N}(\mu_f, \Sigma_f)}(x) \]
   - Representative instances
     \[ \arg \max_{x \in C_f} p_{\mathcal{N}(\mu_f, \Sigma_f)}(x) \]

Pelleg et Moore - Active learning for anomaly and rare category detection, NIPS'05.
ILAB: Active Learning Strategy

Reduce the temporal complexity
- Annotating while computing

Divide-and-Conquer Approach

- Uncertainty sampling
  - Uncertain queries
- Malicious analysis
  - Malicious queries
- Benign analysis
  - Benign queries
ILAB: Active Learning Strategy

Divide-and-Conquer Approach

- Reduce the temporal complexity
- Annotating while computing

Uncertainty sampling
Uncertain queries

Malicious analysis
Malicious queries

Benign analysis
Benign queries

Annotation
Computation ...

Computation ...
Avoiding Sampling Biases
Rare category detection

Reducing the Waiting-Periods
Divide-and-conquer
Comparison to State-of-the-Art Strategies

Simulations on Annotated Datasets

<table>
<thead>
<tr>
<th></th>
<th>#instances</th>
<th>#features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contagio</td>
<td>10,000</td>
<td>113</td>
</tr>
<tr>
<td>NSL-KDD</td>
<td>74,826</td>
<td>122</td>
</tr>
</tbody>
</table>

Active Learning Strategies

- **Uncertainty**

- **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**
Comparison to State-of-the-Art Strategies

Avoiding Sampling Biases

ILAB and Aladin detect properly the different families.
Comparison to State-of-the-Art Strategies

Reducing Waiting-Periods

Waiting-periods reduced thanks to ILAB.

<table>
<thead>
<tr>
<th>Num. Annotations</th>
<th>Execution Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>250</td>
</tr>
<tr>
<td>1,000</td>
<td>500</td>
</tr>
<tr>
<td>1,500</td>
<td>750</td>
</tr>
<tr>
<td>2,000</td>
<td>1,000</td>
</tr>
</tbody>
</table>

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

Anaël Beaugnon SoSySec 30/11/2018 - Machine Learning for Detection Systems
**Comparison to State-of-the-Art Strategies**

**ILAB** avoids sampling biases without increasing the waiting-periods.

<table>
<thead>
<tr>
<th></th>
<th>Uncertainty</th>
<th>Görnitz et al.</th>
<th>Aladin</th>
<th><strong>ILAB</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>No bias</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Quick</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
</tbody>
</table>

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**
Comparison to State-of-the-Art Strategies

ILAB avoids sampling biases without increasing the waiting-periods.

<table>
<thead>
<tr>
<th></th>
<th>Uncertainty</th>
<th>Görnitz et al.</th>
<th>Aladin</th>
<th>ILAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>No bias</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Quick</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

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

What do computer security experts think?

Don’t forget the expert!

Annotation queries → New annotated instances

Unlabeled data → Annotated data

Training

Mac Aodha et al. Putting the scientist in the loop: accelerating scientific progress with interactive machine learning, ICPR’14.
Don’t forget the expert!

<table>
<thead>
<tr>
<th>Simulations</th>
<th>GUI</th>
<th>User Exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncertainty</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Görnitz et al.</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td><strong>Aladin</strong></td>
<td>✓</td>
<td>~</td>
</tr>
<tr>
<td>Nissim et al.</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Moskovitch et al.</td>
<td>✓</td>
<td>×</td>
</tr>
</tbody>
</table>

**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/
Four security experts

<table>
<thead>
<tr>
<th></th>
<th>Jour 1</th>
<th>Jour 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num. flows</td>
<td>$1.2 \cdot 10^8$</td>
<td>$1.2 \cdot 10^8$</td>
</tr>
<tr>
<td>Num. IPs</td>
<td>463,913</td>
<td>507,258</td>
</tr>
<tr>
<td>Num. features</td>
<td>134</td>
<td>134</td>
</tr>
</tbody>
</table>

Initial Annotations
- Anomalous instances
  - obvious scans
- Normal instances
  - uniform selection
ILAB: Annotation Systems

Annotation Interface

**Network Flow Example**

- **Suggestion**: slow_scan
- **Malicious Families**:
  - ICMP_scan
  - TCP_Syn_flooding
  - misconfiguration
  - obvious_scan
  - slow_scan

- **Benign Families**:
  - DNS
  - SMTP
  - web

**Instance 374335**

**Description**

- **NetFlows**
  - **Start**: 08:22:23:341
  - **Duration**: 8.835
  - **Proto**: TCP
  - **Src IP**: 43805
  - **Src port**: 23
  - **Det IP**: ...
  - **Det port**: ...
  - **Flags**: ...
  - **Num bytes**: 168
  - **Num packets**: 3
### Problem-Specific Visualizations

Display any Data Type

**NetFlow**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetFlows</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Start</th>
<th>Duration</th>
<th>Proto</th>
<th>Src IP</th>
<th>Src port</th>
<th>Dst IP</th>
<th>Dst port</th>
<th>Flags</th>
<th>Num bytes</th>
<th>Num packets</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:22:23.341</td>
<td>8.835</td>
<td>TCP</td>
<td>43.805</td>
<td>23</td>
<td></td>
<td></td>
<td>S</td>
<td>168</td>
<td>3</td>
</tr>
</tbody>
</table>
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

Raw data

Features

ILAB

Annotated data

Features

ILAB

Annotated data
Strong Link between Features and Annotations

Features

Raw data

ILAB

Information Loss

Annotated data
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

Solutions
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.
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
In-situ annotations with ILAB (Interactive LABeling)

Reducing the annotation workload

- **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*
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/