SSH Attacks

Técnicas de Percepção de Redes

MIECT

Ana Luísa Ferreira, Nº93301
João Gameiro, Nº93097

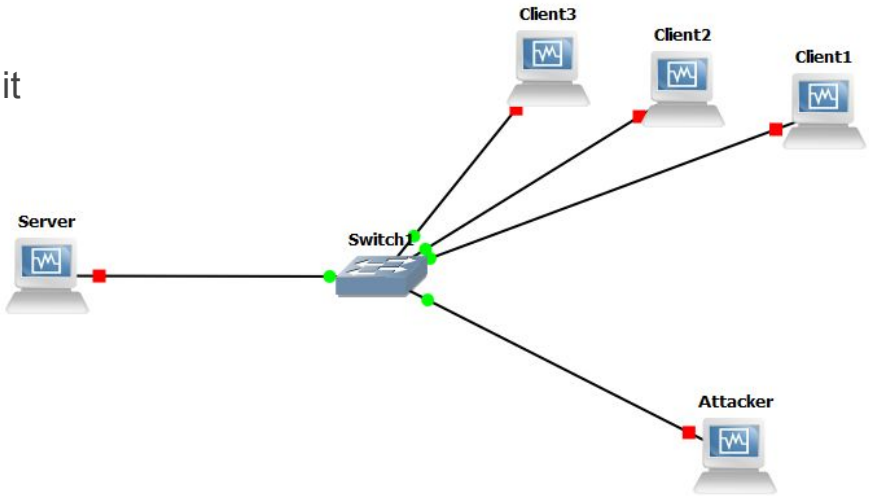
January 4th 2022

Our Scenario

We assumed:

- Using SSH connections inside the company is a normal behaviour.
- Users use the ssh connection to open, close, edit files, travel through directories, etc…
- The attacker got access to a normal machine.

Scenario: begins to copy files (scp command) which incites abnormal behaviour.



- A normal user behaviour will be used to build a profile
- Deviations from the profile will be identified as abnormalities (Anomaly Detection)

Good Behavior

- We developed three behavior patterns:
  - **Pattern 1:** Travels through directories in order to list and check their content. Creates a helloworld.py file, adds python code to it, runs it and watches the results.
  - **Pattern 2:** For a set of files that contains a set of algorithms, test and analyse the results. The clients traverses through the several directories and executes the programs.
  - **Pattern 3:** The user runs a “decrypt md5” algorithm to decrypt a large random.txt and analyses the obtained result.

Good Behavior

- The overall good client behavior pattern consists in:
  - Connecting to the server through SSH
    - Randomly choosing one of the three behaviors defined
    - Randomly deciding to repeat previous step or not
  - Disconnecting from the session
  - Wait a random gaussian delay time
  - Repeat the process

Bad Behavior

- Considering an important sensitive file:
  - The attacker logs into the SSH server
  - Splits the file into smaller chunks (of variable sizes)
  - Copies the chunks by a random order
    - Attacker exits and enters the SSH session every time a copy happens
    - There is a random gaussian delay between the copy of each file
  - Deletes the chunk of files afterwards
  - Repeat process

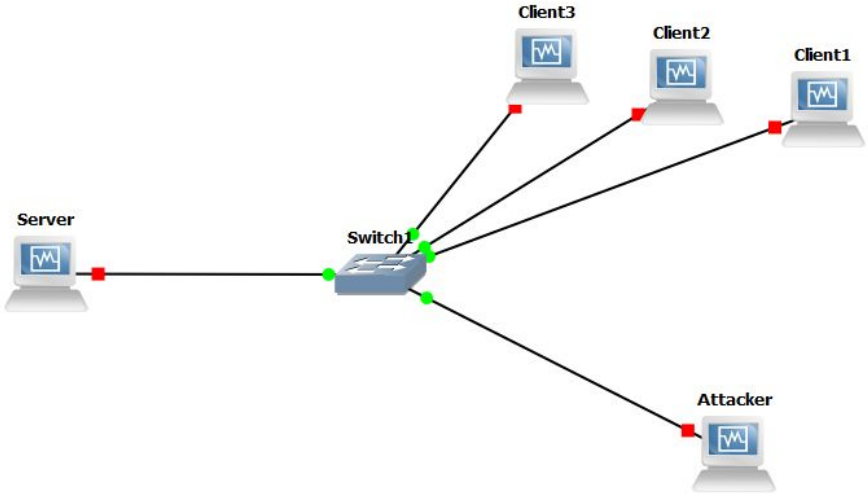
Random Gaussian Delays

We used several random gaussian delays to better simulate the behavior of a person:

- Time between cd or ls commands:
  - Mean 1 and deviation 1
  - Mean 2 and deviation 1
- Time between testing multiple algorithms:
  - Mean 5 and deviation 3
- Time between testing different implementations of the same algorithm:
  - Mean 10 and deviation 5
- For the attacker, time between each copy:
  - Mean 90 and deviation 75

Data sources

- Captures of the SSH Traffic
  - 4 Captures
    - 3 Clients
    - 1 Attacker



Observation Window

- Sliding window
  - So abnormalities can be detected as fast as possible
- Size of 2 minutes in order to detect abnormalities
  - Decision period of 20 seconds
- Considering a sampling period of 1 second to gather data

Metrics Extracted

- For each capture:
  - SSH Packets
    - Number of Uploaded Packets
    - Number of Downloaded Packets
  - Number of Bytes
    - Number of Uploaded Bytes
    - Number of Downloaded Bytes

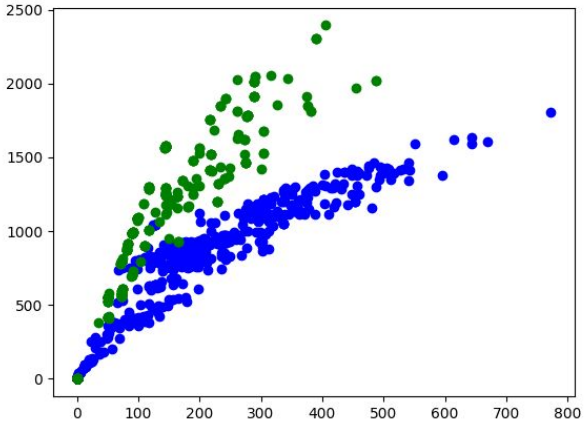
Features Extracted

- For each metric defined previously, the following features were extracted:
  - Mean
  - Median
  - Standard Deviation
  - Percentiles of 75%, 80%, 90%, 98%
- The silence/activity features extracted from each metric were the following:
  - Size of the silence and activity periods
  - Mean of the silence and activity periods
  - Standard deviation of the silence and activity periods

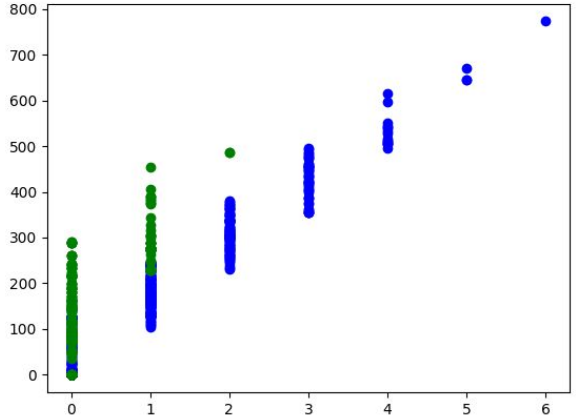
Classes Defined

- Two classes
  - Client
    - Since all three clients are doing the same behavior patterns only one client class was created
  - Attacker
    - Only one class to detect anomalies

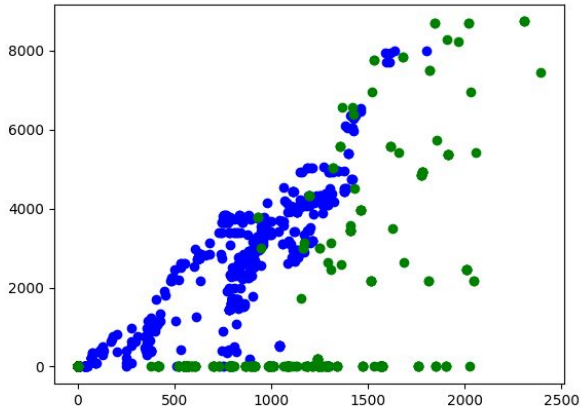
Plots - Features



(Mean Download Bytes, Std Download Bytes)

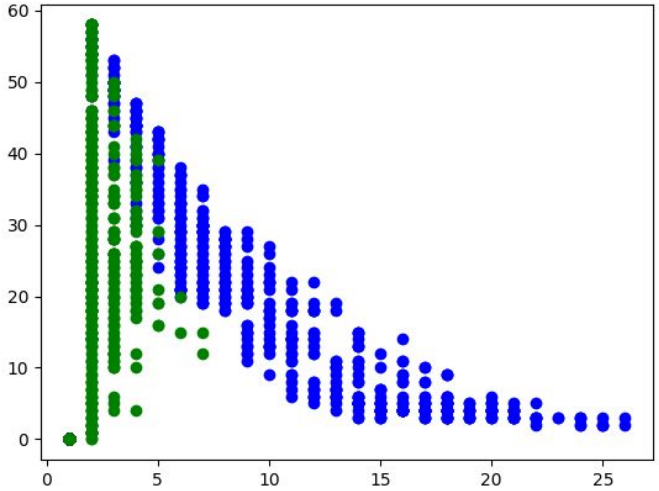


(Mean Uploaded Packets, Mean Downloaded Bytes)

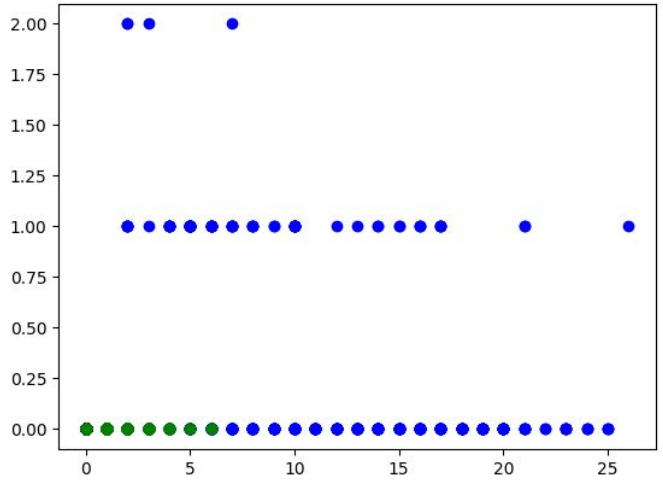


(Std Download Bytes, Percentil 98 Download Bytes)

Plots - Silence & Activity Features



(Size Silence Upload Packets,
Std Silence Upload Packets)



(Size Activity Download Packets,
Std Activity Download Packets)

Training & Test Features

- Two approaches were followed
  - First Approach
    - One train set with 50% of the features of each client
    - One test set with the attacker features
    - One test set with the rest of the client features
  - Second Approach
    - One train set with the whole client features of two of the three sets
    - One test set with the attacker features
    - One test set with the set not used of the client features
- The approach that presented better results was the first one

Anomaly Detection

- Statistical Analysis based on:
  - Centroids distances without PCA features
  - Centroids distances with PCA features
  - Multivariate PDF with PCA features
- Machine Learning based on:
  - One Class Support Vector Machines without PCA features
    - Linear, RBF and Poly Kernels
  - One Class Support Vector Machines with PCA features
    - Linear, RBF and Poly Kernels

Anomaly Detection

```
-- Anomaly Detection based on Centroids Distances --
True Positives: 347, True Negatives: 36
False Positives: 477, False Negatives: 157
Accuracy: 37.65978367748279%
Precision: 42.11165048543689%
Recall: 68.84920634920636%
F1-Score: 0.5225903614457831

-- Anomaly Detection based on Centroids Distances (PCA Features)--
True Positives: 339, True Negatives: 36
False Positives: 477, False Negatives: 165
Accuracy: 36.87315634218289%
Precision: 41.544117647058826%
Recall: 67.26190476190477%
F1-Score: 0.5136363636363637
```

```
-- Anomaly Detection based Multivariate PDF (PCA Features) --
True Positives: 357, True Negatives: 397
False Positives: 116, False Negatives: 147
Accuracy: 74.13962635201572%
Precision: 75.47568710359408%
Recall: 70.83333333333334%
F1-Score: 0.7308085977482088
```

Anomaly Detection

```
-- Anomaly Detection based on One Class Support Vector Machines (PCA Features) --

Kernel Linear Statistics
True Positives: 181, True Negatives: 305
False Positives: 208, False Negatives: 323
Accuracy: 47.78761061946903%
Precision: 46.52956298200514%
Recall: 35.91269841269841%
F1-Score: 0.40537513997760355

Kernel RBF Statistics
True Positives: 359, True Negatives: 337
False Positives: 176, False Negatives: 145
Accuracy: 68.43657817109144%
Precision: 67.10280373831776%
Recall: 71.23015873015873%
F1-Score: 0.6910490856592878

Kernel Poly Statistics
True Positives: 493, True Negatives: 91
False Positives: 422, False Negatives: 11
Accuracy: 57.42379547689283%
Precision: 53.87978142076503%
Recall: 97.81746031746032%
F1-Score: 0.6948555320648344
```

```
-- Anomaly Detection based on One Class Support Vector Machines --

Kernel Linear Statistics
True Positives: 189, True Negatives: 350
False Positives: 163, False Negatives: 315
Accuracy: 52.99901671583087%
Precision: 53.69318181818182%
Recall: 37.5%
F1-Score: 0.44158878504672894

Kernel RBF Statistics
True Positives: 328, True Negatives: 344
False Positives: 169, False Negatives: 176
Accuracy: 66.07669616519173%
Precision: 65.99597585513078%
Recall: 65.07936507936508%
F1-Score: 0.6553446553446554

Kernel Poly Statistics
True Positives: 442, True Negatives: 85
False Positives: 428, False Negatives: 62
Accuracy: 51.81907571288102%
Precision: 50.804597701149426%
Recall: 87.6984126984127%
F1-Score: 0.6433770014556041
```

Anomaly Detection - Results

- PCA features do not improve results in centroids distances technique
- PCA features improve results in Support Vector Machines techniques in all cases
- Anomaly Detection based on Multivariate PDF with PCA features showcased the overall best results
- Results improve when silence features are also used

Best number of components - PCA

```
--------------------PCA Stats Centroid Distances --------------------
With 9 components the maximum number of true positives is 61
With 2 components the minimum number of false positives is 87
With 9 components the best accuracy is 58.65834633385335
With 9 components the best precision is 41.21621621621622
With 9 components the best recall is 12.103174603174603
With 9 components the best f1-score is 0.18711656441717792

--------------------PCA Stats Centroid Distances w/Silence--------------------
With 14 components the maximum number of true positives is 350
With 2 components the minimum number of false positives is 686
With 14 components the best accuracy is 30.733229329173167
With 14 components the best precision is 32.28782287822878
With 14 components the best recall is 69.44444444444444
With 14 components the best f1-score is 0.4408060453400504
```

```
--------------------PCA Stats Multivariate --------------------
With 2 components the maximum number of true positives is 504
With 12 components the minimum number of false positives is 22
With 9 components the best accuracy is 82.7613104524181
With 12 components the best precision is 93.25153374233128
With 2 components the best recall is 100.0
With 9 components the best f1-score is 0.7474285714285714

--------------------PCA Stats Multivariate w/Silence--------------------
With 2 components the maximum number of true positives is 504
With 18 components the minimum number of false positives is 35
With 18 components the best accuracy is 83.22932917316692
With 18 components the best precision is 90.25069637883009
With 2 components the best recall is 100.0
With 17 components the best f1-score is 0.7583892617449665
```

Best number of components - PCA

```
--------------------PCA Stats SVM Linear--------------------
With 4 components the maximum number of true positives is 488
With 9 components the minimum number of false positives is 65
With 5 components the best accuracy is 65.600624024961
With 5 components the best precision is 54.75113122171946
With 4 components the best recall is 96.82539682539682
With 5 components the best f1-score is 0.6221079691516711

--------------------PCA Stats SVM RBF--------------------
With 4 components the maximum number of true positives is 495
With 2 components the minimum number of false positives is 248
With 7 components the best accuracy is 67.39469578783151
With 7 components the best precision is 56.86900958466453
With 4 components the best recall is 98.21428571428571
With 7 components the best f1-score is 0.6300884955752212

--------------------PCA Stats SVM Poly--------------------
With 3 components the maximum number of true positives is 441
With 7 components the minimum number of false positives is 239
With 2 components the best accuracy is 65.600624024961
With 7 components the best precision is 55.822550831792974
With 3 components the best recall is 87.5
With 2 components the best f1-score is 0.5827814569536424
```

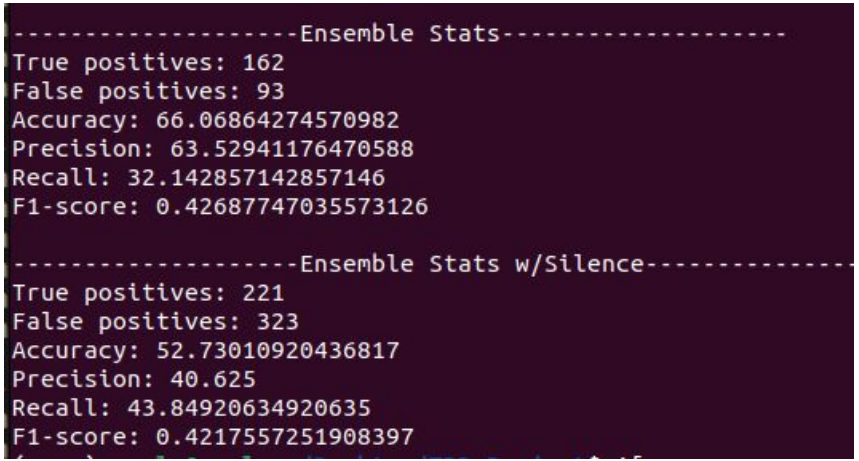
```
--------------------PCA Stats SVM Linear w/Silence--------------------
With 3 components the maximum number of true positives is 326
With 2 components the minimum number of false positives is 161
With 9 components the best accuracy is 63.026521060842434
With 9 components the best precision is 52.62237762237763
With 3 components the best recall is 64.68253968253968
With 9 components the best f1-score is 0.5594795539033457

--------------------PCA Stats SVM RBF w/Silence--------------------
With 14 components the maximum number of true positives is 499
With 4 components the minimum number of false positives is 250
With 10 components the best accuracy is 68.48673946957878
With 10 components the best precision is 58.090614886731395
With 14 components the best recall is 99.0079365079365
With 15 components the best f1-score is 0.64

--------------------PCA Stats SVM Poly w/Silence--------------------
With 4 components the maximum number of true positives is 483
With 5 components the minimum number of false positives is 211
With 5 components the best accuracy is 71.06084243369735
With 5 components the best precision is 61.98198198198198
With 4 components the best recall is 95.83333333333334
With 3 components the best f1-score is 0.650375939849624
```

Best Threshold

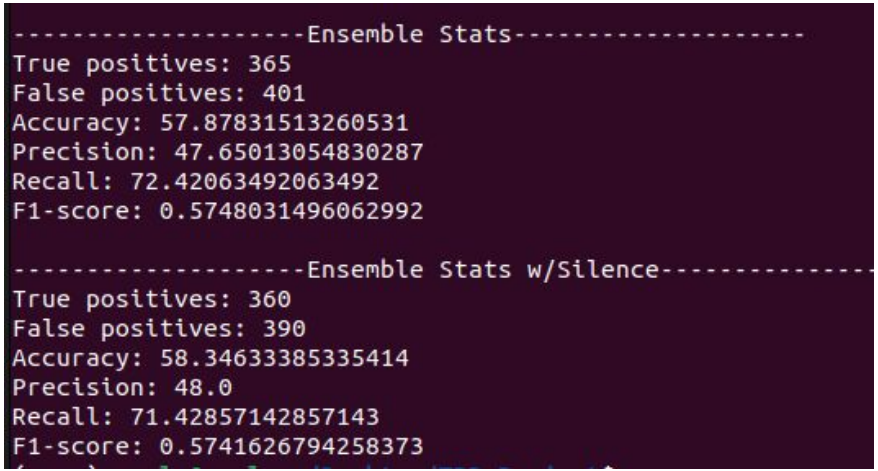
```
--------------------Threshold Stats Centroid Distances with PCA--------------------
For threshold 0.1 the maximum number of true positives is 504
For threshold 2.1  the minimum number of false positives is 88
For threshold 1.9 the best accuracy is 55.148205928237125
For threshold 0.1 the best precision is 39.31357254290172
For threshold 0.2 the best recall is 100.0
For threshold 0.2 the best f1-score is 0.5643896976483763

--------------------Threshold Stats Centroid Distances without PCA--------------------
For threshold 0.1 the maximum number of true positives is 350
For threshold 0.1  the minimum number of false positives is 734
For threshold 0.1 the best accuracy is 30.733229329173167
For threshold 0.1 the best precision is 32.28782287822878
For threshold 0.2 the best recall is 69.44444444444444
For threshold 0.2 the best f1-score is 0.4408060453400504

--------------------Threshold Stats Multivariate --------------------
For threshold 1.9 the maximum number of true positives is 328
For threshold 0.1  the minimum number of false positives is 32
For threshold 1.6 the best accuracy is 83.46333853354135
For threshold 0.1 the best precision is 90.64327485380117
For threshold 2.0 the best recall is 65.07936507936508
For threshold 2.0 the best f1-score is 0.7557603686635945
```

Ensemble

If 50% or more of the methodologies used say it is an “Anomaly” the final result is “Anomaly”.

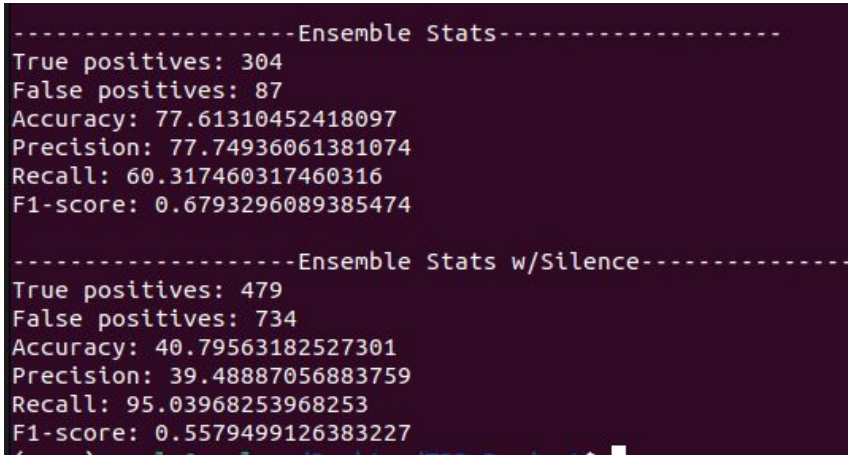


All 6 methodologies

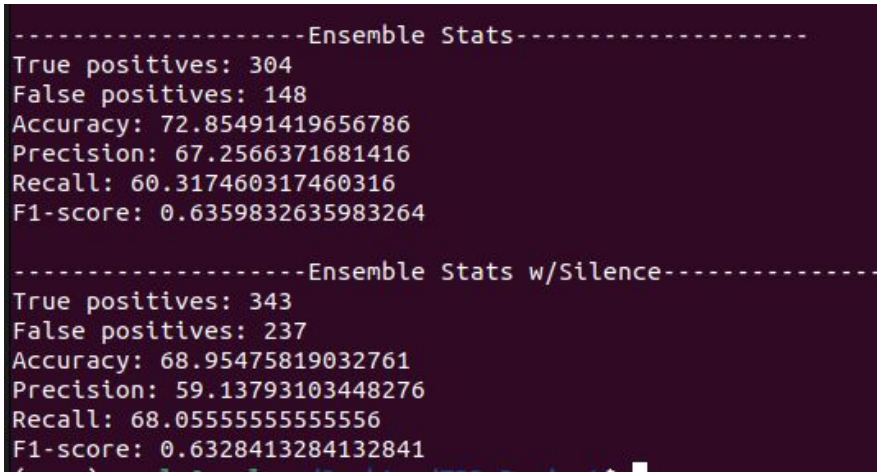


One Class Support Vector Machines with PCA features with all the 3 kernels

Ensemble



3 methodologies of Statistical Analysis



One Class Support Vector Machines (PCA features) with all the 3 kernels + Multivariate (PCA features)

Code

- Github Repository
  - https://github.com/JPCGameiro/TPR_Project