MECT | Network Awareness Techniques

universidade de aveiro

QUIC DDoS Attack mitigation

Malicious actor detection

By Jodionísio Muachifi (97147) and Rúben Castelhano (97688)

Teacher:
Paulo Jorge Salvador Serra Ferreira

December 2023 | 2nd Presentation

A reminder - the problem

- DDoS attacks present serious problems for businesses:
  - Lost revenue
  - Decreased customer trust and reputation loss
  - Service unavailability is not permissible in certain fields (finance, health, military, etc.)
  - A pure brute-force defence is extremely expensive
  - Blocking access to the service is essentially letting the attacker win, since service unavailability is still achieved

- The average amount of downtime following a DDoS attack is 54 minutes and the average cost for each minute of downtime is $22,000[1]

1. Ponemon Institute, “Cyber security on the offense: A study of IT security experts”, November 2012

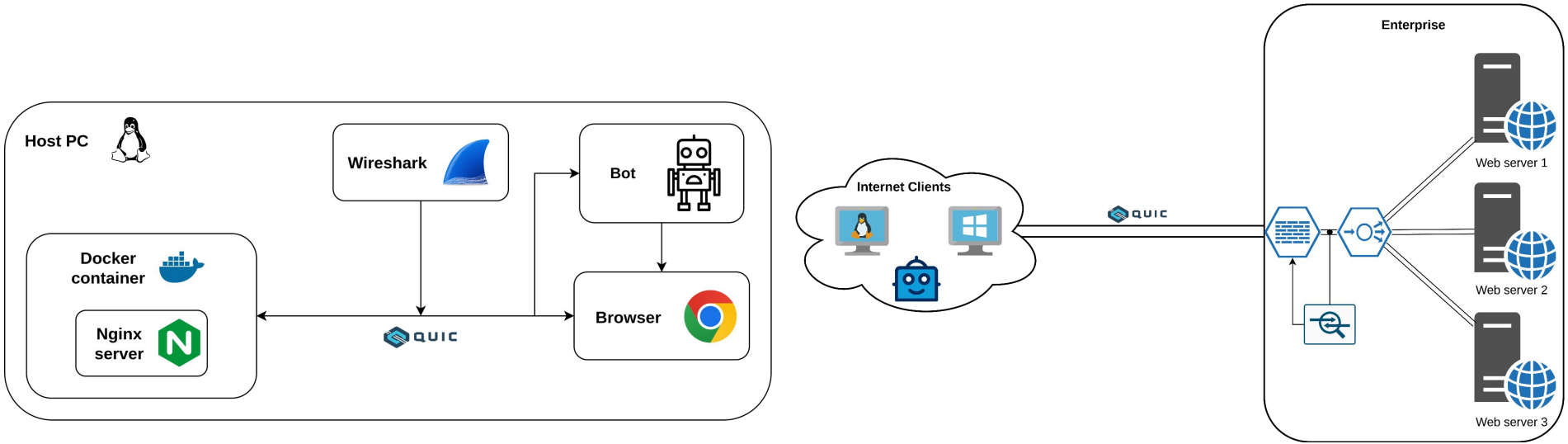
A reminder - our focus

- A proper DDoS mitigation solution is complex and based on several levels of detection and prevention:
  - Stateless firewalls blocking known bad actors
  - Load-balancers
  - Resource and network monitoring
    - **Distinguish between regular and malicious users <--- this will be our focus**

- We will monitor the network traffic patterns of known good actors to understand how they use a service to distinguish them from several levels of attackers

Data sources and real-world scenario

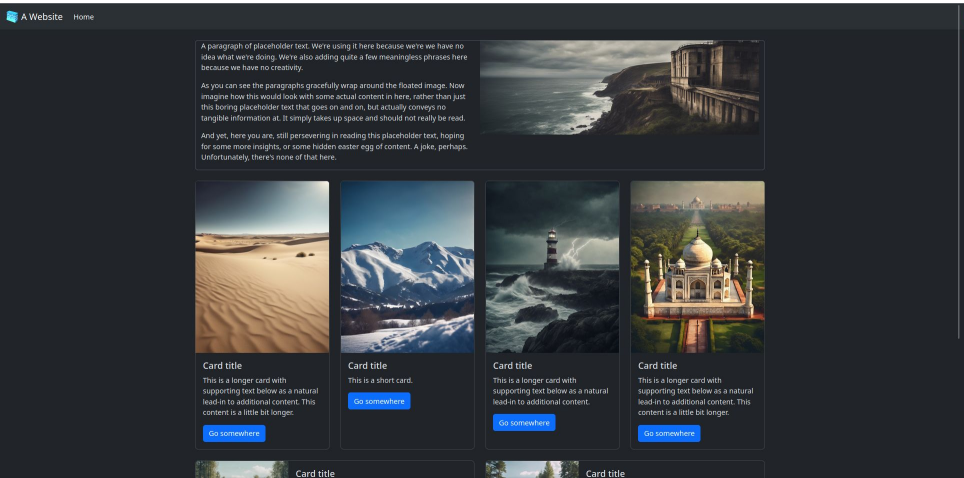
- We built a simple, mostly reading based website (with some images as well) and serve it in HTTP/3 via NGINX
- Captured packets using Wireshark

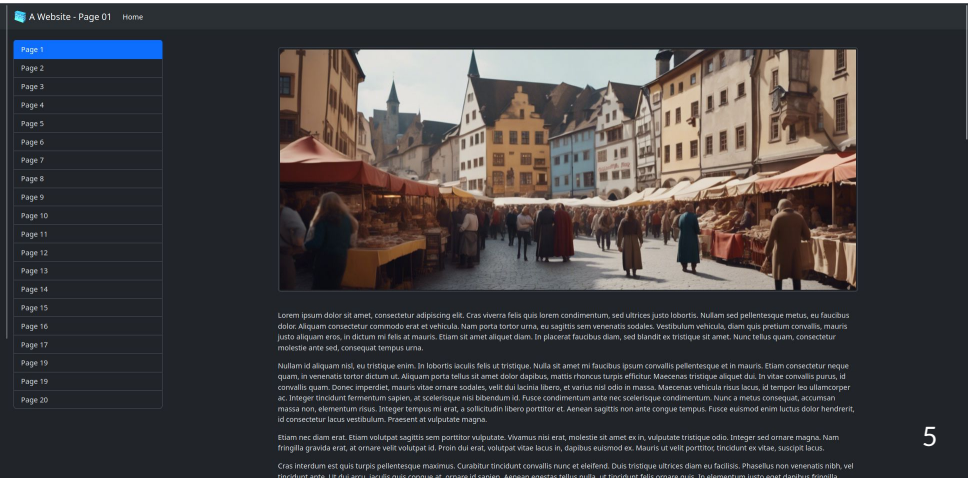


The website

We developed a website containing:

- A main page linking to 4 articles
- A total of 20 articles
  - all with a different amount of images and paragraphs
  - all images are unique across the entire website





Bots developed

- Human bot
  - **v1** - attempts to imitate human behaviour by navigating inside the browser through a normal distribution and taking into account the number of paragraphs and images inside a page
  - **v2** - attempts to imitate human behaviour by navigating inside the browser through a normal distribution
- Basic attacker (attacker 0)
  - spams curl requests of the main page
- Intermediate attacker (attacker 1)
  - navigates inside the browser by changing pages constantly
- Advanced attacker (attacker 2)
  - navigates inside the browser through a normal distribution
- Advanced attacker 2 (attacker 3)
  - navigates inside the browser through a normal distribution and taking into account the number of paragraphs and images inside a page

Data processing

- Collect raw packet data with a sampling period of 0.01 and 0.001 seconds
- Filter data to allow only QUIC packets between the clients and the web server
- Aggregate data by client (source IP address)
- Detect anomalous user behaviour

Observation process:

- Multiple sliding windows (planned)
  - 30 seconds and 8 minutes long
  - sliding every 1 and 10 seconds, respectively

Collected metrics

To extract these metrics for every sampling period a custom application was written using the C language and libpcap.

Collected metrics:

- Number of download packets
- Number of download bytes
- Number of upload packets
- Number of upload bytes

Extracted features - 1

To extract the features for sliding window a custom application was written using the C language and libgsl.

- Number of download/upload packets
  - mean, median, variance, stdev
  - 99th, 98th, 95th, 1st, 2nd, 5th percentiles
- Download/upload packet size
  - mean, median, variance, stdev
  - 99th, 98th, 95th, 1st, 2nd, 5th percentiles
- Periods of silence down/up (0 threshold)
  - len, mean, median, variance, stdev
  - 99th, 98th, 1st, 2th percentiles

Extracted features - 2

- Covariance
  - upload packet size and packet count
  - download packet size and packet count
  - upload and download packet size
  - upload and download packet count
- Correlation
  - upload packet size and packet count
  - download packet size and packet count
  - upload and download packet size
  - upload and download packet count

Total of 66 features

Detection - data processing

- Data is split into 2 datasets
  - A training dataset containing 75% of all normal data points
  - A testing dataset containing the remaining 25 % of all normal data points and of all abnormal data points

- The data is scaled using a standard scaler
  - fitted on normal data
- PCA is performed to reduce the number of features from 66 to 50
  - fitted on normal data

Imitating human behaviour - human bot v1

- The human bot behaves as follows:
  - Opens the main page and waits for a few seconds (normal distribution) before deciding what to do (called stun period)
  - Picks one of 4 articles inside the main page
  - Reads the article
    - chooses the number of paragraphs and images to be read/looked-at (uniform distribution)
    - chooses the time spent reading based on the reading speed (normal distribution) and number of paragraphs and images to be read/looked-at
    - scrolls up and down the page
  - Picks a new article
  - Mouse movements and scrolls are also based on normal distributions

Imitating human behaviour - human bot v1

We used the data from the first version of the human bot and compared it with the behaviour of the 4 attacker bots next.

The time spent reading each paragraph normal distribution is as follows:

- **mean**: 20 seconds
- **stdev**: 10 seconds
- **min**: 2 seconds

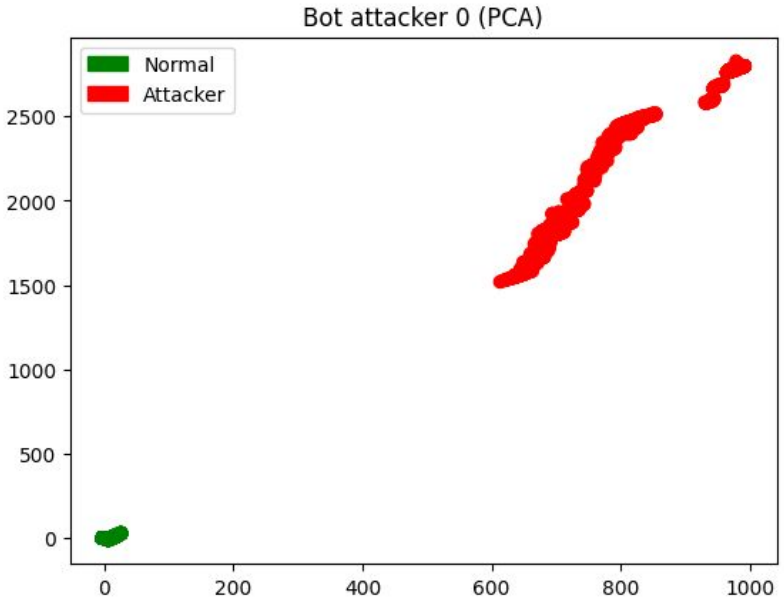
The time spent looking at each image normal distribution is as follows:

- **mean**: 1 seconds
- **stdev**: 1.5 seconds
- **min**: 0.5 seconds

Bot attacker 0

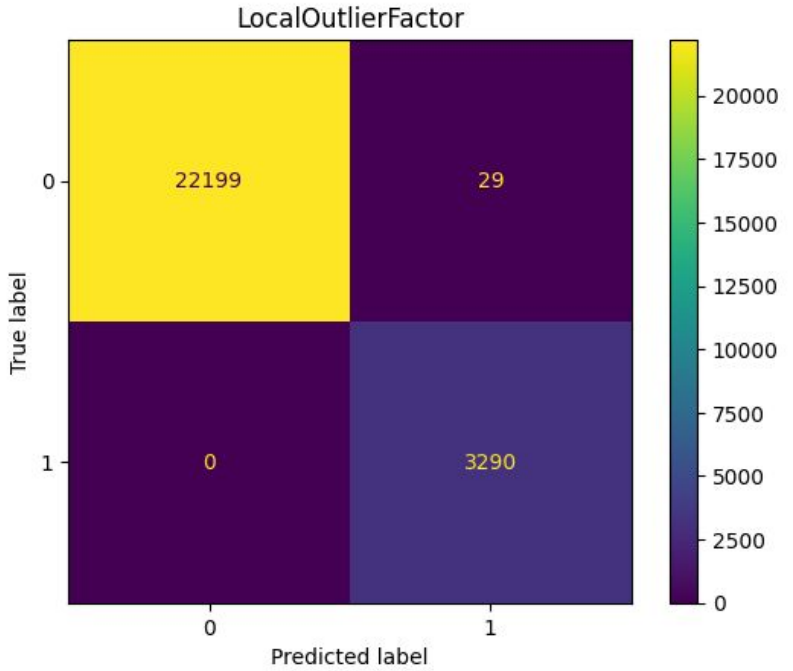
Using a custom version of curl with HTTP 3 support, this bot spams get requests to the main page in an infinite loop as fast as possible.

- Sampling period of 10 ms
- Sliding window of 30 seconds sliding every second



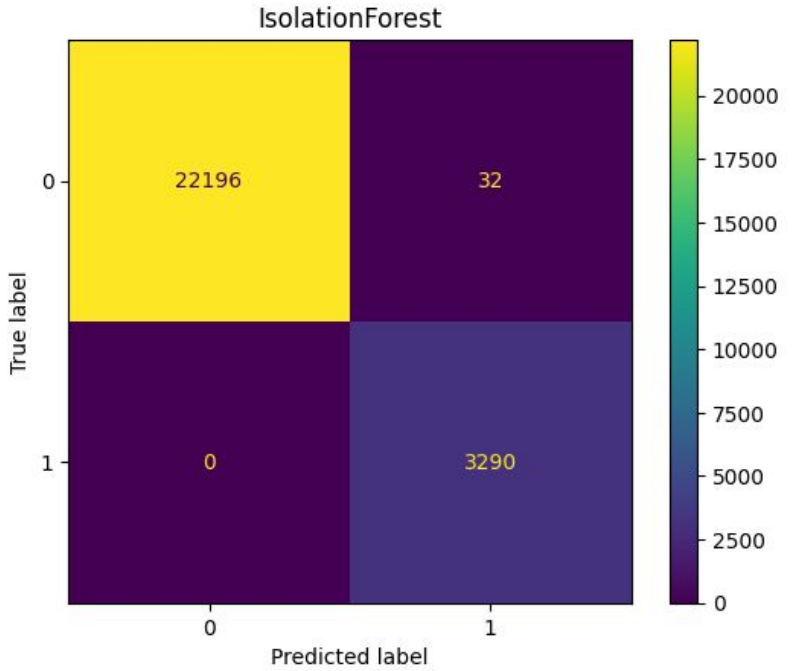
Bot attacker 0 - LocalOutlierFactor results

- **Accuracy**: 99.886 %
- **Precision**: 99.126 %
- **Recall**: 100%
- **F-1**: 99.561 %



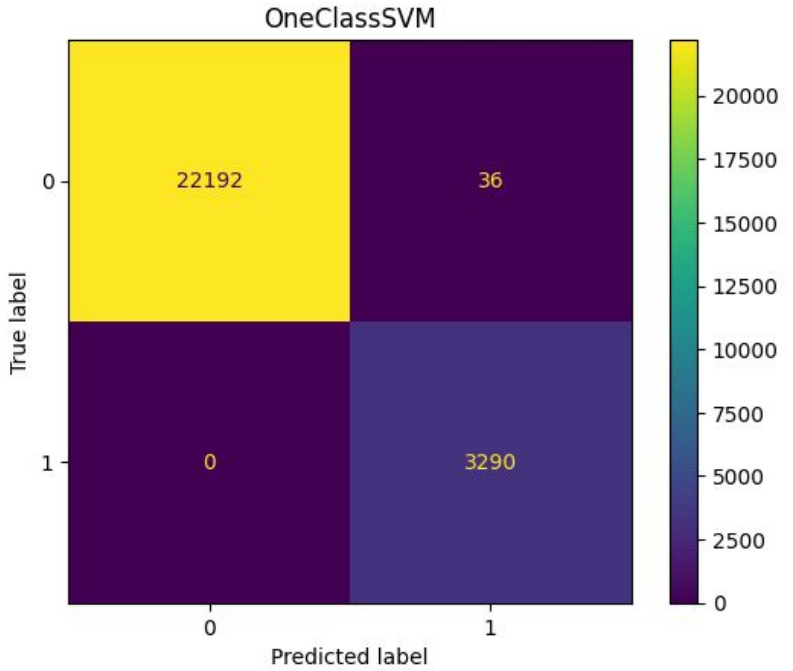
Bot attacker 0 - IsolationForest results

- **Accuracy**: 99.875 %
- **Precision**: 99.037 %
- **Recall**: 100 %
- **F-1**: 99.516 %



Bot attacker 0 - OneClassSVM (RBF) results

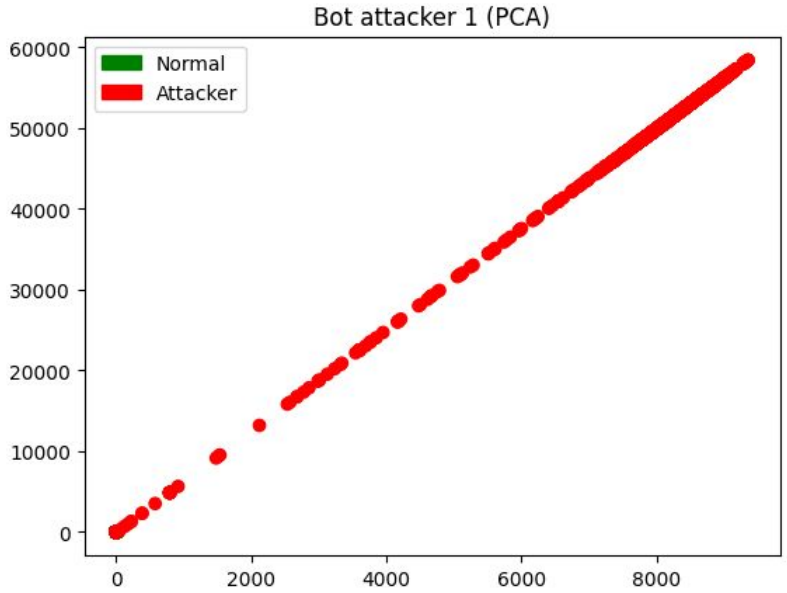
- **Accuracy**: 99.859 %
- **Precision**: 98.918 %
- **Recall**: 100 %
- **F-1**: 99.456 %



Bot attacker 1

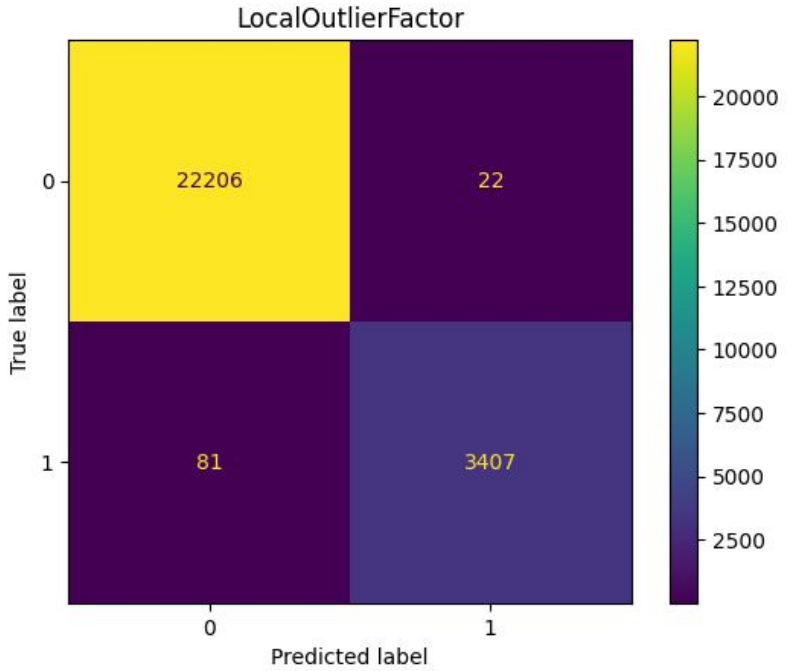
This bot opens the main page, immediately picks an article and then switches articles as fast as possible (or refreshes the page)

- Sampling period of 10 ms
- Sliding window of 30 seconds sliding every second



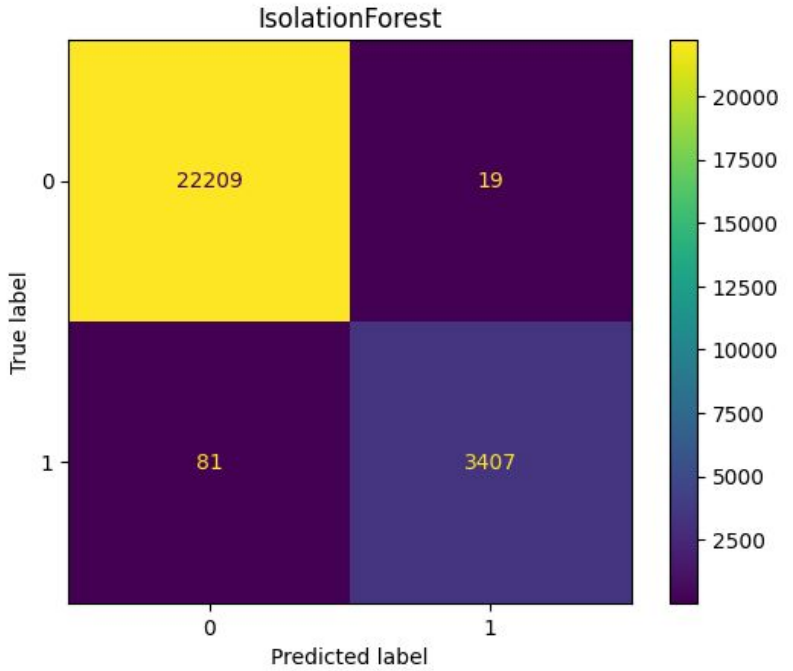
Bot attacker 1 - LocalOutlierFactor results

- **Accuracy**: 99.599 %
- **Precision**: 99.358 %
- **Recall**: 97.678 %
- **F-1**: 98.511 %



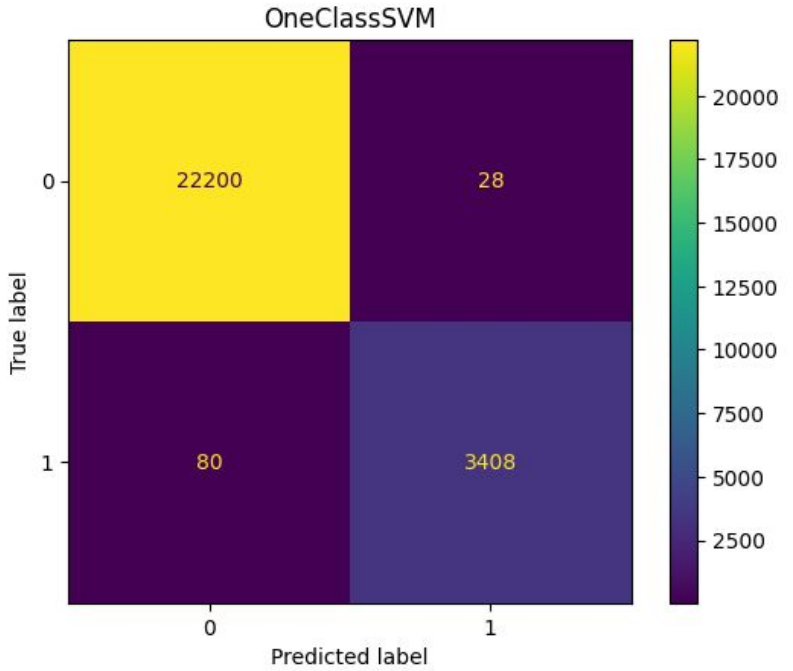
Bot attacker 1 - IsolationForest results

- **Accuracy**: 99.611 %
- **Precision**: 99.445 %
- **Recall**: 97.678 %
- **F-1**: 98.554 %



Bot attacker 1 - OneClassSVM (RBF) results

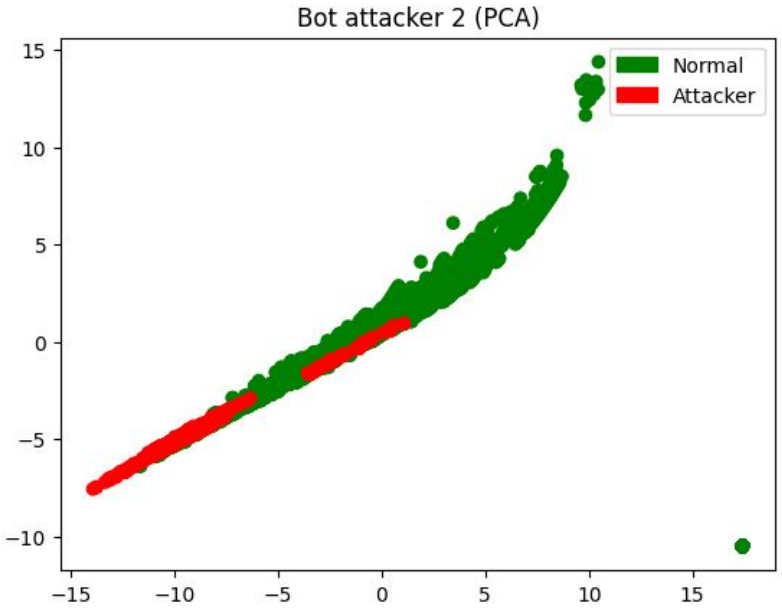
- **Accuracy**: 99.580 %
- **Precision**: 99.185 %
- **Recall**: 97.706 %
- **F-1**: 98.440 %



Bot attacker 2

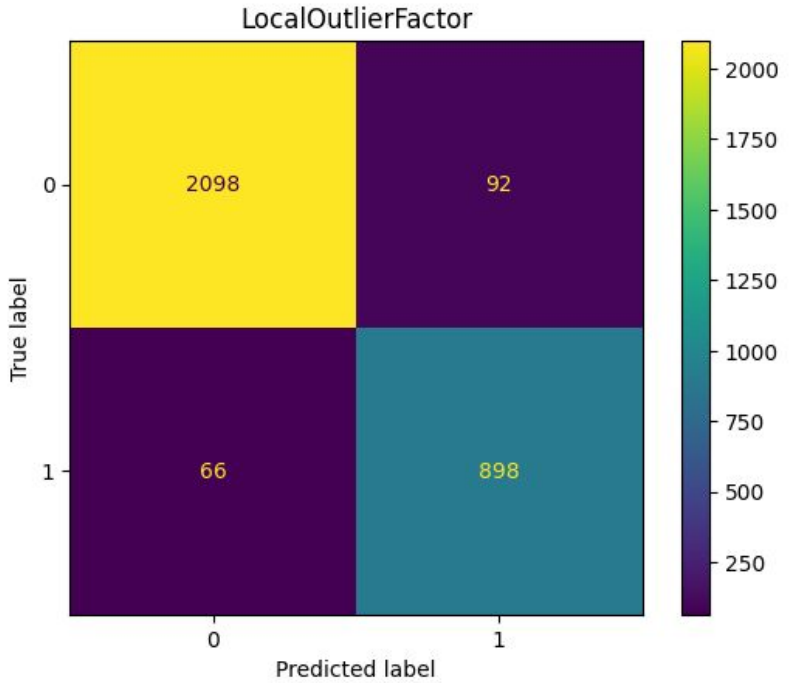
This bot opens the main page, picks an article and then switches articles in a normal distribution interval (the same used on the human bot) but doesn't take into account the number of paragraphs and images in the page

- Sampling period of 1 ms
- Sliding window of 8 minutes sliding every 10 seconds



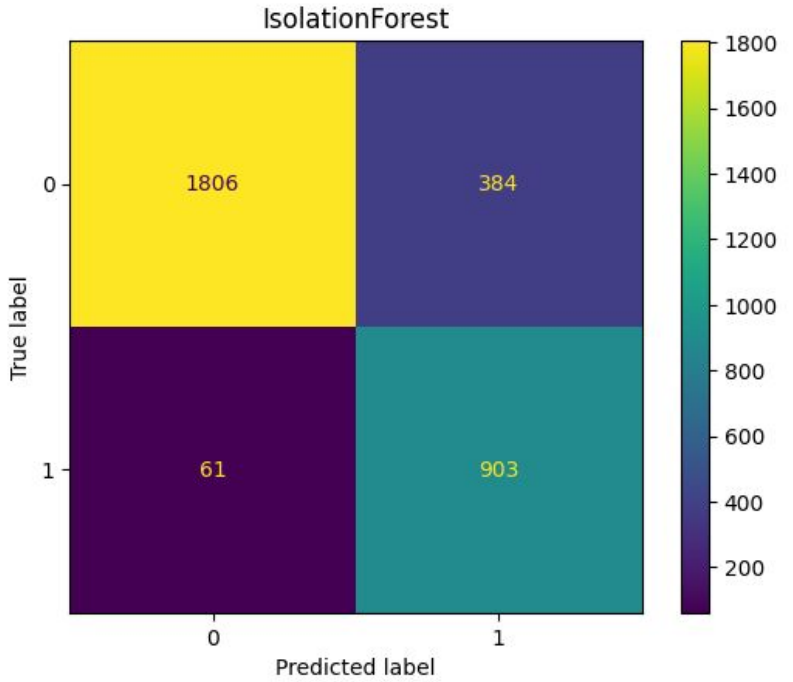
Bot attacker 2 - LocalOutlierFactor results

- **Accuracy**: 94.990 %
- **Precision**: 90.707 %
- **Recall**: 93.154 %
- **F-1**: 91.914 %



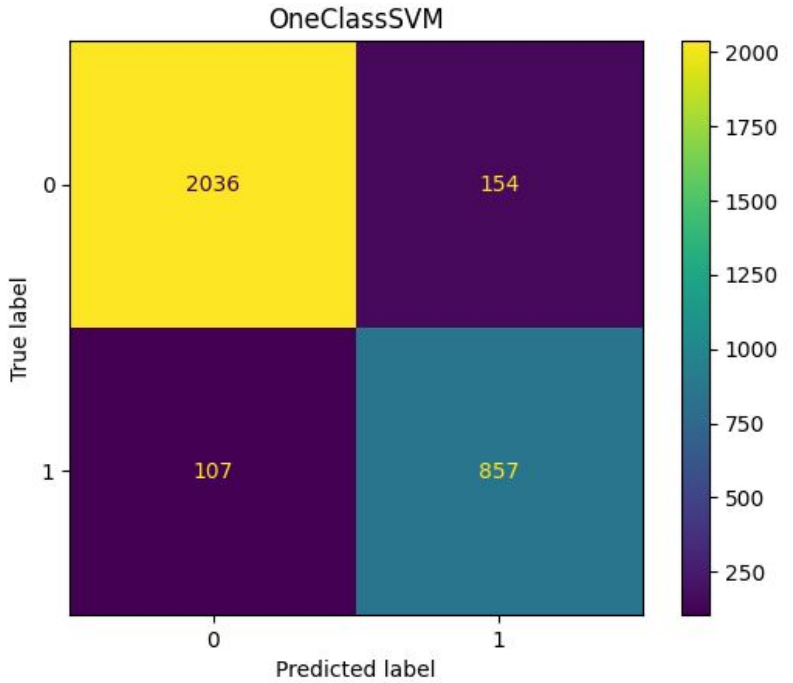
Bot attacker 2 - IsolationForest results

- **Accuracy**: 85.891 %
- **Precision**: 70.163 %
- **Recall**: 93.154 %
- **F-1**: 80.231 %



Bot attacker 2 - OneClassSVM (RBF) results

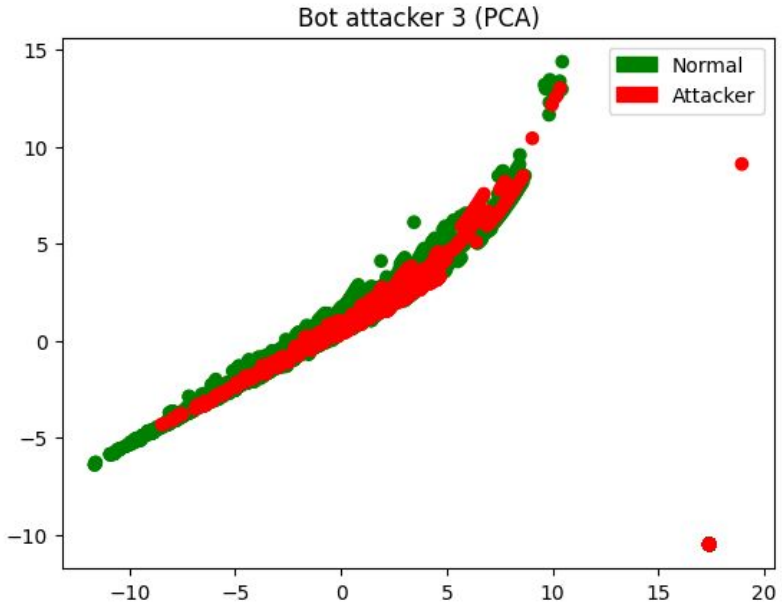
- **Accuracy**: 91.725 %
- **Precision**: 84.768 %
- **Recall**: 93.154 %
- **F-1**: 86.785 %



Bot attacker 3

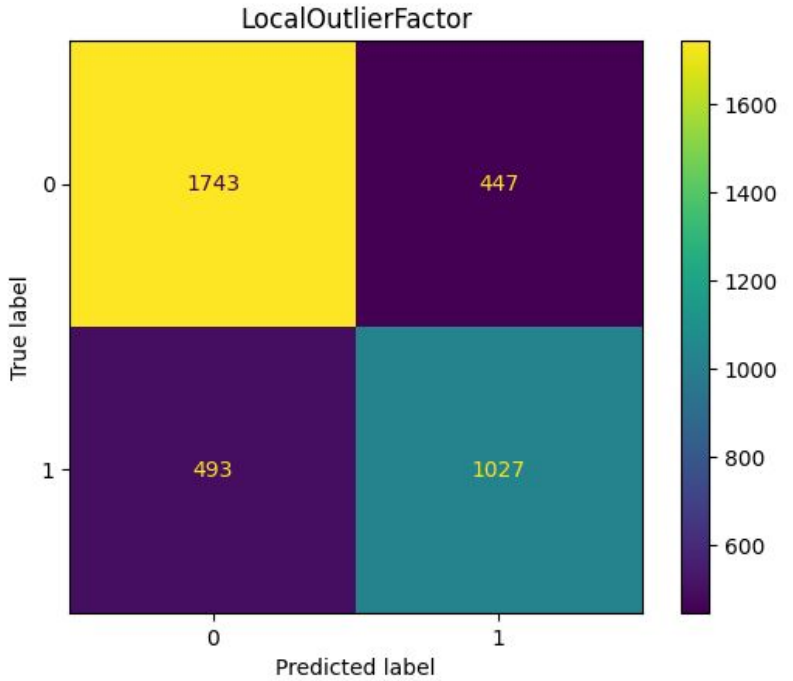
This bot opens the main page, picks an article and then switches articles in a normal distribution interval (the same used on the human bot) while taking into account the number of paragraphs and images on the page

- Sampling period of 1 ms
- Sliding window of 8 minutes sliding every 10 seconds



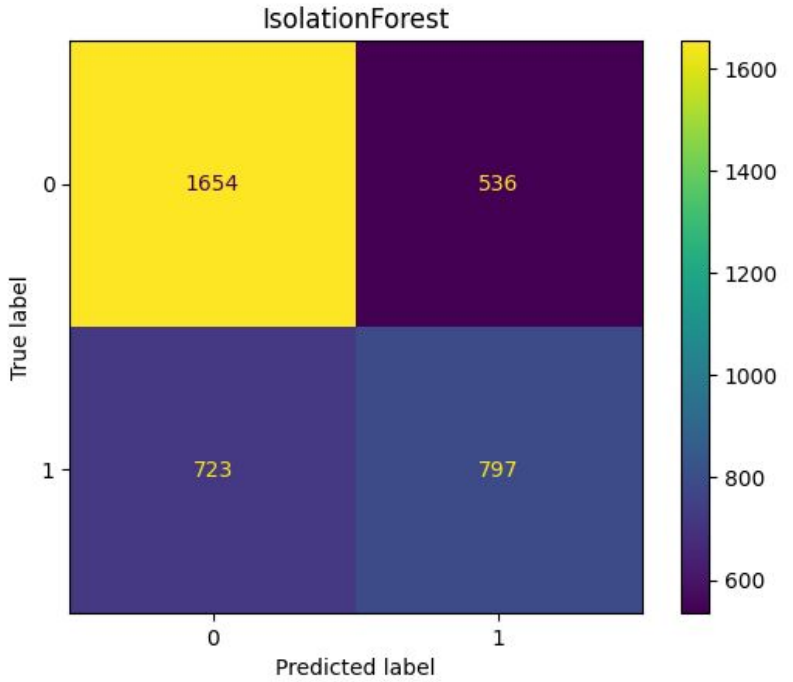
Bot attacker 3 - LocalOutlierFactor results

- **Accuracy**: 74.663 %
- **Precision**: 69.674 %
- **Recall**: 67.566 %
- **F-1**: 68.604 %



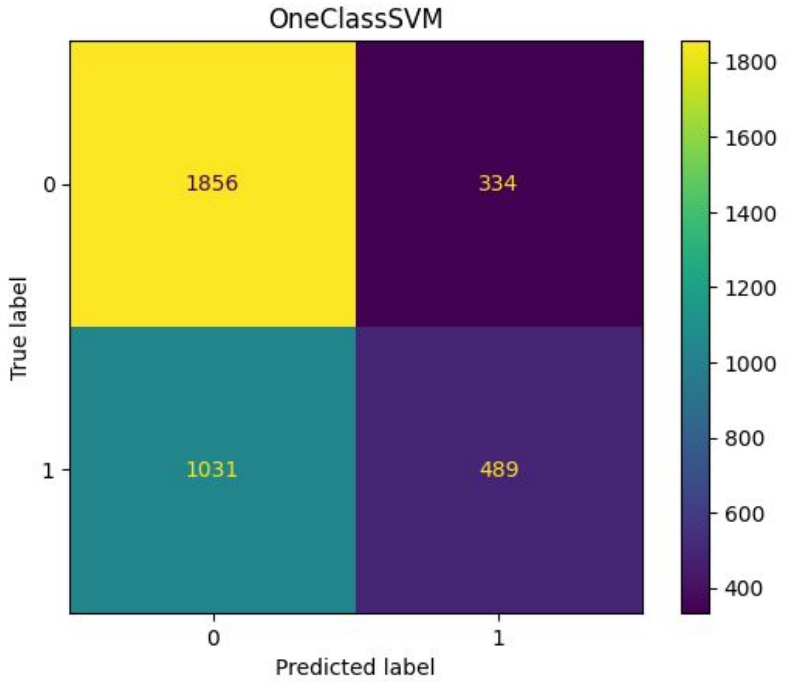
Bot attacker 3 - IsolationForest results

- **Accuracy**: 66.065 %
- **Precision**: 59.790 %
- **Recall**: 52.434 %
- **F-1**: 55.871 %



Bot attacker 3 - OneClassSVM (RBF) results

- **Accuracy**: 63.208 %
- **Precision**: 59.417 %
- **Recall**: 32.171 %
- **F-1**: 41.741 %



Imitating human behaviour - human bot v2

- The human bot behaves as follows:
  - Opens the main page and waits for a few seconds (normal distribution) before deciding what to do (called stun period)
  - Picks one of 4 articles inside the main page
  - Reads the article
    - chooses the number of paragraphs and images to be read/looked-at (uniform distribution)
    - chooses the time spent reading based on the reading speed (normal distribution)
  - Picks a new article
  - Mouse movements and scrolls are also based on normal distributions

Imitating human behaviour - human bot v2

- To try and improve the human bot behaviour, we decided to collect statistics from our usage of a real website. In this case https://www.abola.pt/. We measured the time we stayed in each article and extracted the mean and standard deviation of these values. This was then fed to a modified data of the human bot (which didn't take into account the amount of paragraphs and images in the page) by assuming each page is of roughly the same length like in **a, bola**.
- We then replaced the previous "human" data with this new data based on our behaviour and evaluated attacker 2 and attacker 3 again.

Measured interval between articles (Rúben):

- **mean**: 38.015 seconds
- **stdev**: 20.514 seconds
- **min**: 11.44 seconds

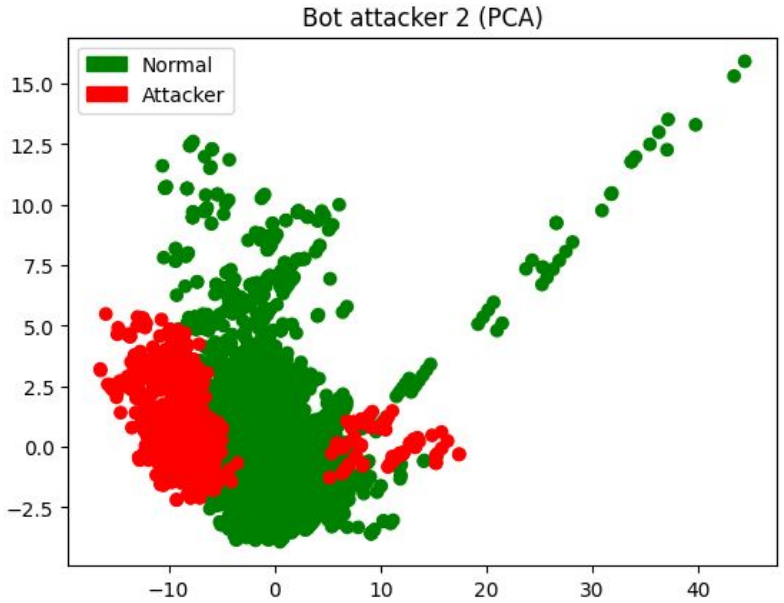
Measured interval between articles (Jodionísio):

- **mean**: 41.595 seconds
- **stdev**: 17.884 seconds
- **min**: 18.11 seconds

Bot attacker 2 - new human data

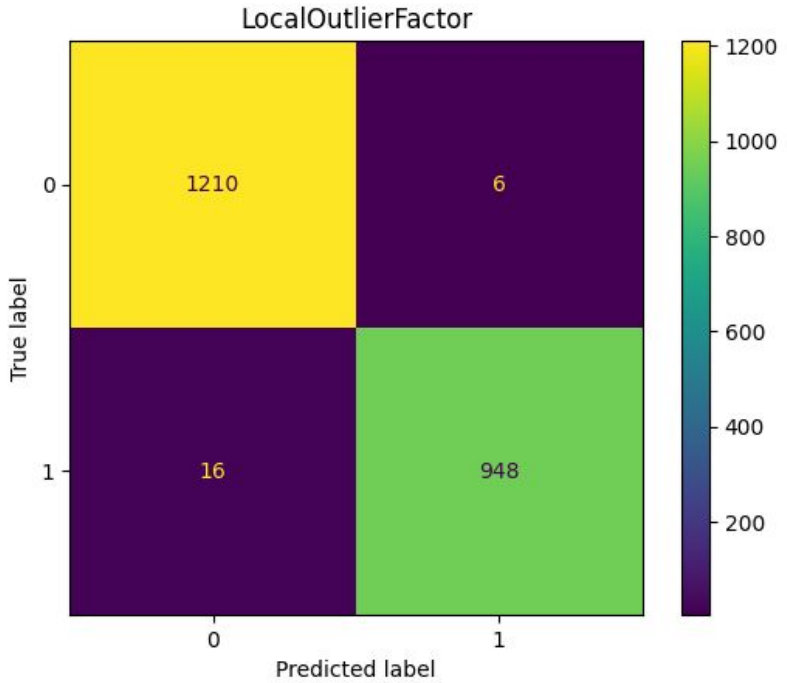
This bot opens the main page, picks an article and then switches articles in a normal distribution interval (the same used on the human bot v1)

- Sampling period of 1 ms
- Sliding window of 8 minutes sliding every 10 seconds



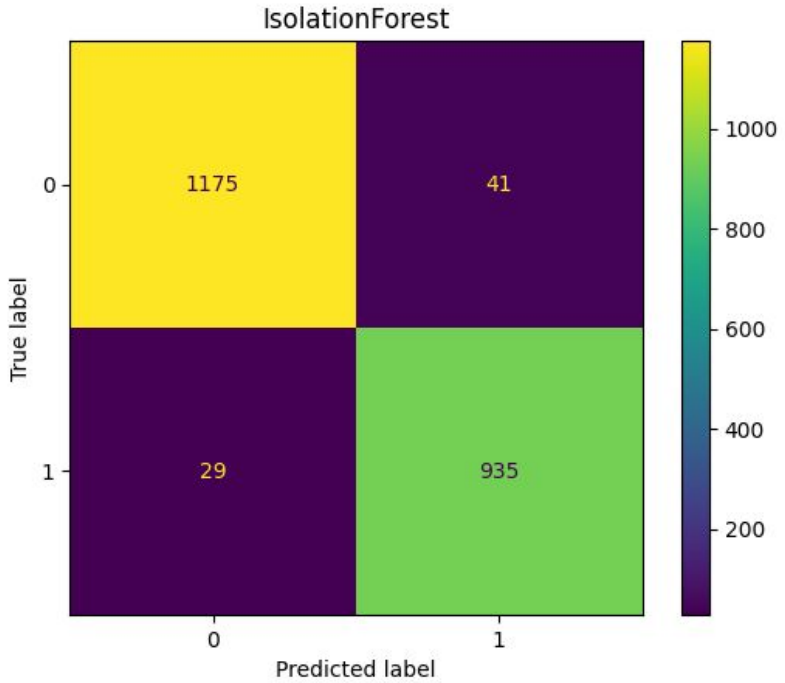
Bot attacker 2 - LocalOutlierFactor results

- **Accuracy**: 98.991 %
- **Precision**: 99.371 %
- **Recall**: 98.340 %
- **F-1**: 98.853 %



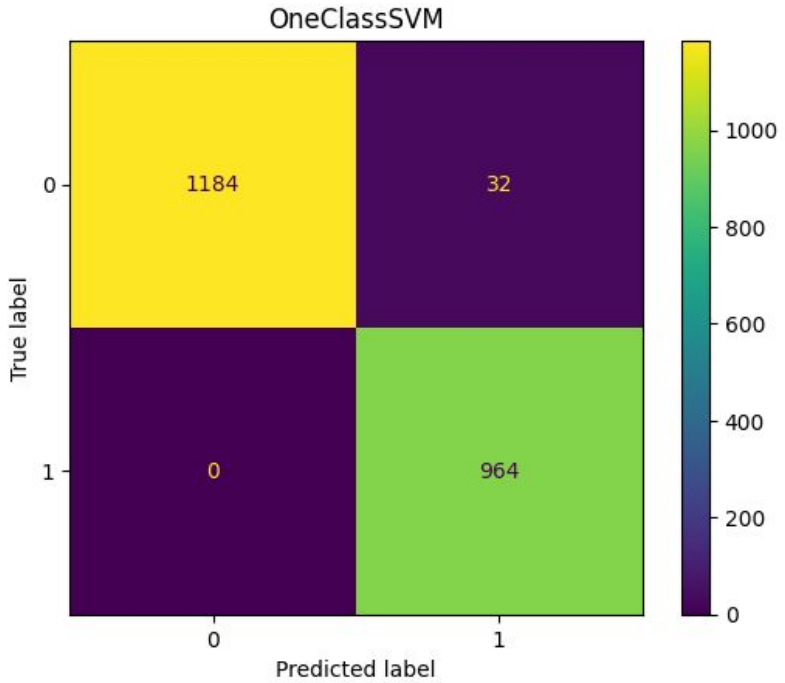
Bot attacker 2 - IsolationForest results

- **Accuracy**: 96.789 %
- **Precision**: 95.799 %
- **Recall**: 98.340 %
- **F-1**: 96.392 %



Bot attacker 2 - OneClassSVM (RBF) results

- **Accuracy**: 98.532 %
- **Precision**: 96.787 %
- **Recall**: 98.340 %
- **F-1**: 98.367 %



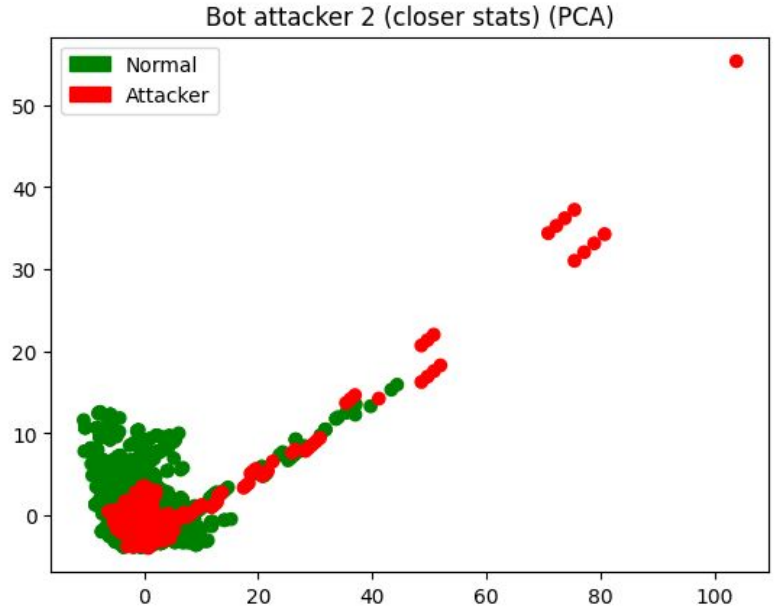
Bot attacker 2 (more accurate) - new human data

This bot opens the main page, picks an article and then switches articles in a normal distribution interval (close to the ones in the human bot v2)

- Sampling period of 1 ms
- Sliding window of 8 minutes sliding every 10 seconds

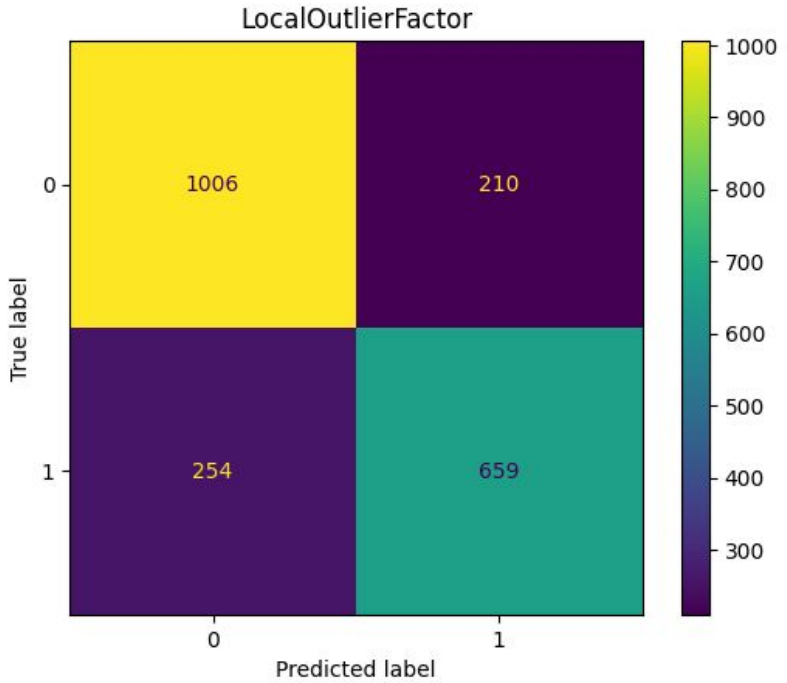
Time spent per page distribution:

- **mean**: 32 seconds
- **stdev**: 15 seconds
- **min**: 5 seconds



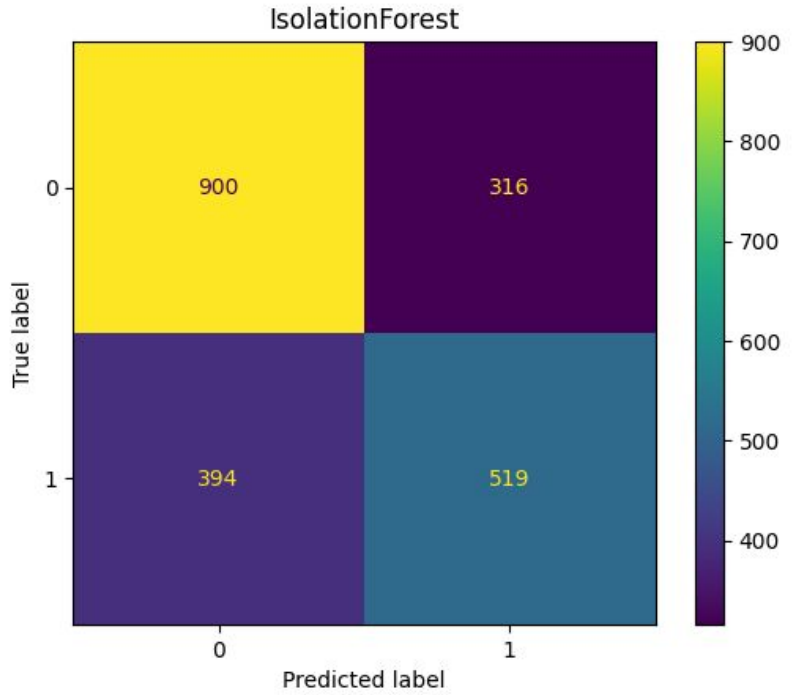
Bot attacker 2 - LocalOutlierFactor results

- **Accuracy**: 78.206 %
- **Precision**: 75.834 %
- **Recall**: 72.180 %
- **F-1**: 73.962 %



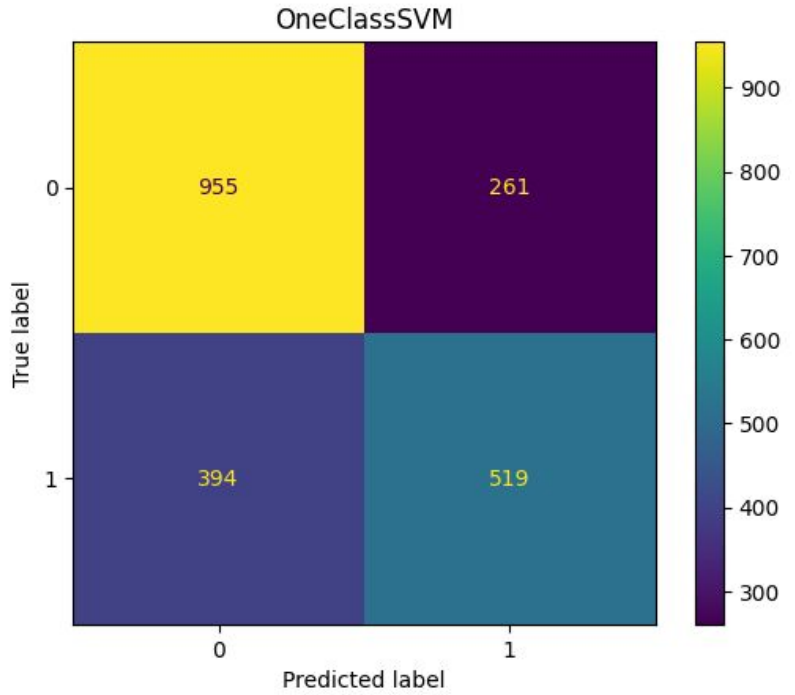
Bot attacker 2 - IsolationForest results

- **Accuracy**: 66.651 %
- **Precision**: 62.156 %
- **Recall**: 56.846 %
- **F-1**: 59.382 %



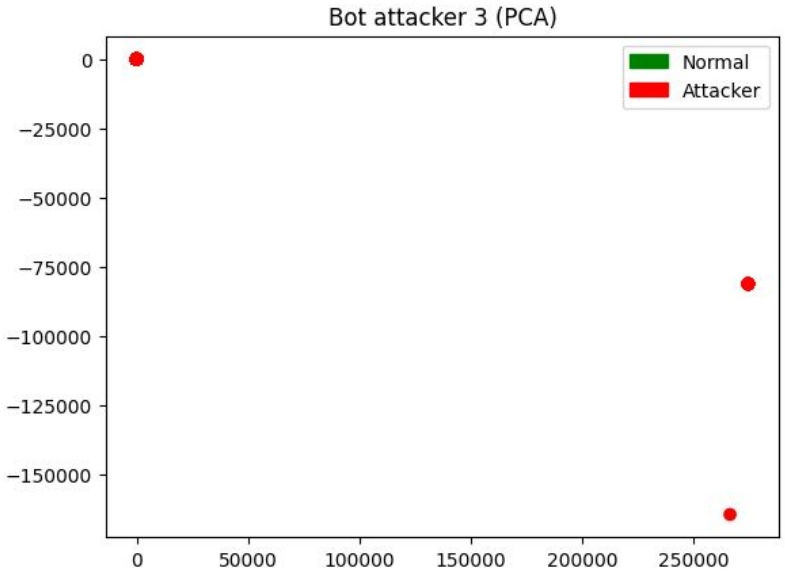
Bot attacker 2 - OneClassSVM (RBF) results

- **Accuracy**: 69.234 %
- **Precision**: 66.538 %
- **Recall**: 56.846 %
- **F-1**: 61.311 %



Bot attacker 3 - new human data

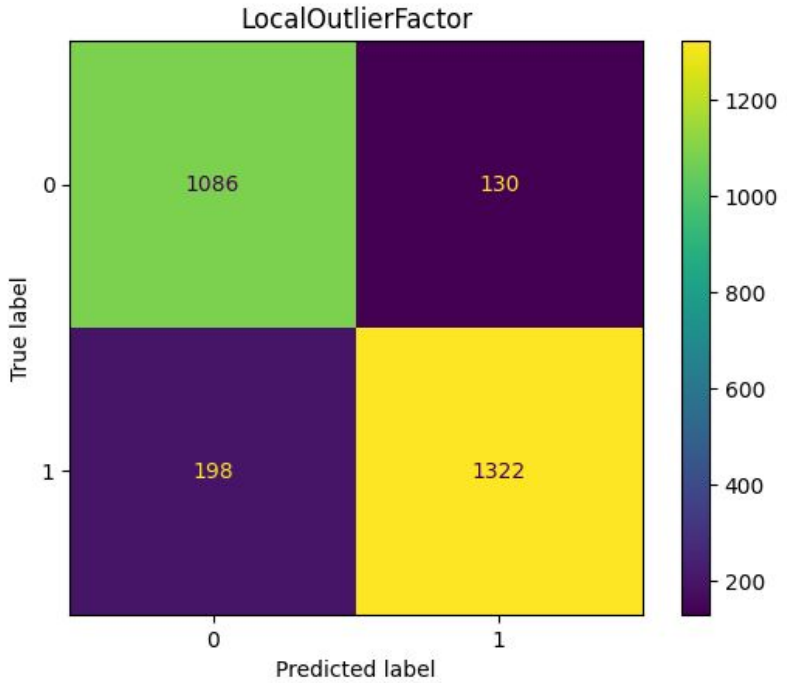
This bot opens the main page, picks an article and then switches articles in a normal distribution interval (the same used on the human bot v1) while taking into account the number of paragraphs and images on the page



- Sampling period of 1 ms
- Sliding window of 8 minutes sliding every 10 seconds

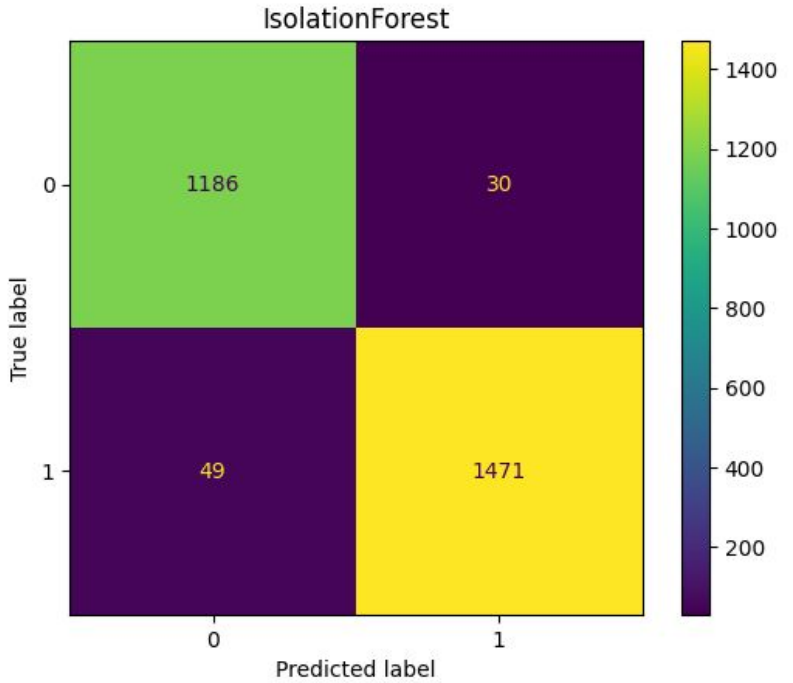
Bot attacker 3 - LocalOutlierFactor results

- **Accuracy**: 88.012 %
- **Precision**: 91.047 %
- **Recall**: 86.974 %
- **F-1**: 88.964 %



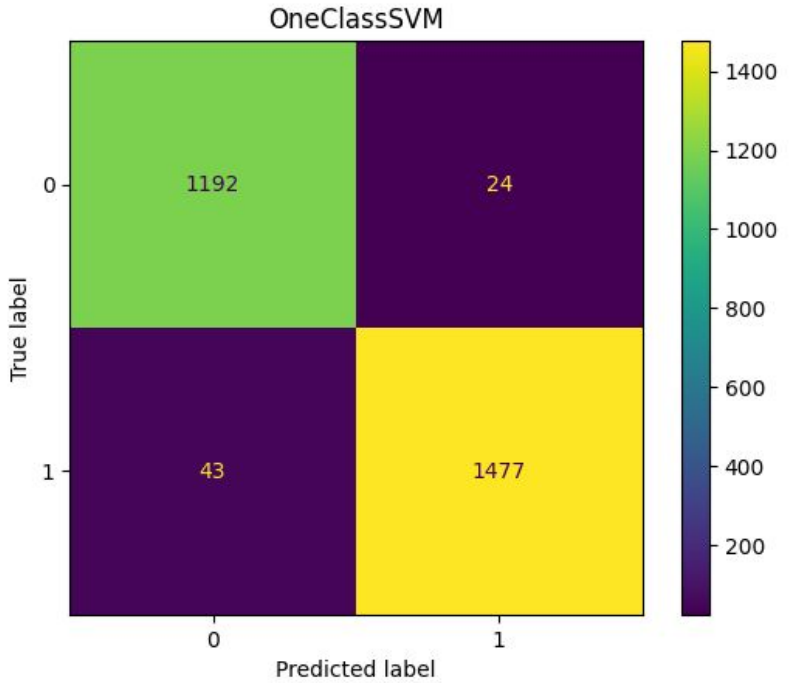
Bot attacker 3 - IsolationForest results

- **Accuracy**: 97.113 %
- **Precision**: 98.001 %
- **Recall**: 96.776 %
- **F-1**: 97.385 %



Bot attacker 3 - OneClassSVM (RBF) results

- **Accuracy**: 97.551 %
- **Precision**: 98.401 %
- **Recall**: 97.171 %
- **F-1**: 97.782 %



Thank you for your attention!
Any questions ?

References

- https://www.gnu.org/software/gsl/
- https://www.tcpdump.org/