Detecting Novel Variants of Application Layer (D)DoS Attacks using Supervised Learning

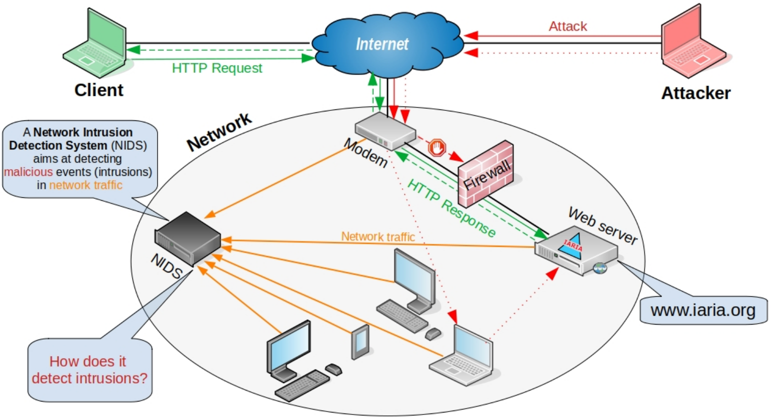
Etienne van de Bijl, Jan Klein, Joris Pries,
Rob van der Mei, Sandjai Bhulai

Stochastics, Centrum Wiskunde & Informatica. (Netherlands)
Faculty of Science, Vrije Universiteit Amsterdam. (Netherlands)
Contact email: evdb@cwi.nl

ABOUT ETIENNE VAN DE BIJL

- BSc (2017) & MSc (2020) in business analytics at the Vrije Universiteit Amsterdam, Netherlands
- Academy assistant at the Vrije Universiteit Amsterdam, Netherlands (2017 - 2018)
- PhD student (2019 - 2023) at the Centrum Wiskunde & Informatica, Netherlands
- Topics of interests: data mining, machine learning, cybersecurity, diffusion models, spread of misinformation

INTRODUCTION



INTRUSION DETECTION SYSTEMS

Signature-based	Anomaly-based
Compares observed network events against patterns that correspond to known threats	*Searches for malicious traffic by constructing a notion of normal behavior and flags activities which do not conform to this notion*
+ Effective against known attacks − Time consuming for experts − Only finds known attacks	+ Able to detect novel attacks − Suffers a high false-positive rate

Machine learning (ML) → ability to overcome this high false-positive rate

RESEARCH GOAL

Research on (novel) intrusion detection with ML:

1. Closed-world assumption → identical attacks
2. Open-world assumption → unrelated attacks

What about detecting related cyberattacks in an open-world setting?

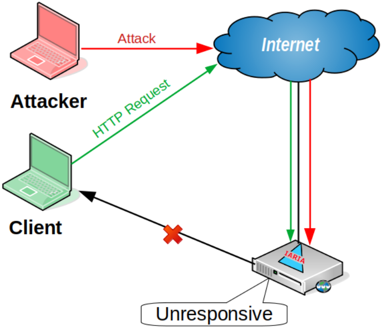
We study to what extent **ML models** are accurately able to detect **novel variants** of known cyberattacks

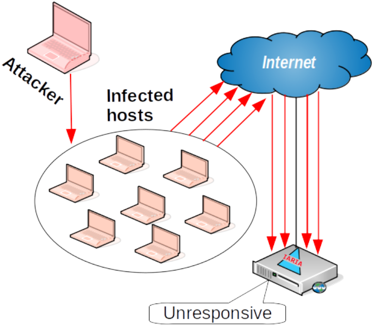
Scope: detecting application layer (distributed) denial-of-service attacks targeting the HTTP protocol of a web server

(D)DOS ATTACKS EXPLAINED

Denial-of-Service attack (DoS)

Distributed Denial-of-Service attack (DDoS)





DATASETS

Selected the **CIC-IDS-2017** & **CIC-IDS-2018** intrusion detection datasets from the Canadian Institute of Cybersecurity → contains a variety of DoS and DDoS attacks & publicly available

SlowHTTPTest

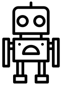
Goldeneye

Slowloris

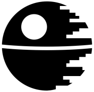
Hulk

LOIC

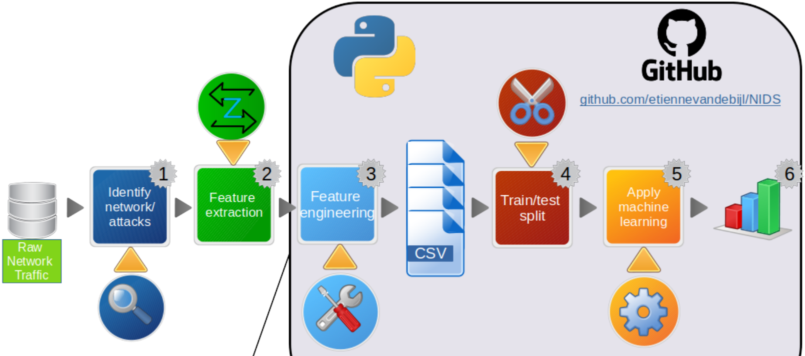
Botnet



HOIC



WORKFLOW - FROM DATA TO RESULTS



Contribution: we combined network, transport and application layer features into one dataset

FINAL DATASET

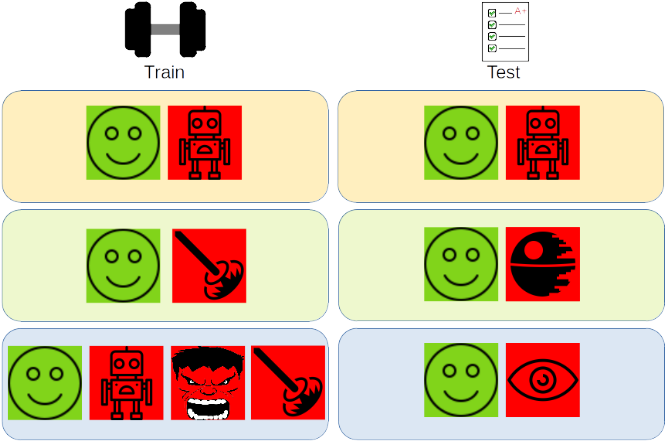
Meta-data

103 IP, TCP, and HTTP features
CIC-IDS-2017: 524,698 interactions (instances)
CIC-IDS-2018: 9,595,037 interactions

CIC-IDS-2017	49.2%	0.2%	1.5%	0.0%	30.2%	18.2%	0.3%	0.4%
CIC-IDS-2018	65.1%	1.3%	0.4%	11.2%	18.8%	3.1%	0.0%	0.1%

Observation 1: CIC-IDS-2018 is larger and more imbalanced
Observation 2: malicious class sizes differ a lot → each attack has its own characteristics

EXPERIMENTAL SETUP 1

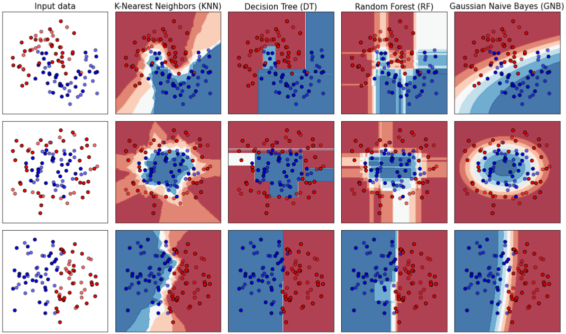


Experiment 1: Detecting known attacks

Experiment 2: Detecting novel variants

Experiment 3: Detecting novel variants using multiple attacks

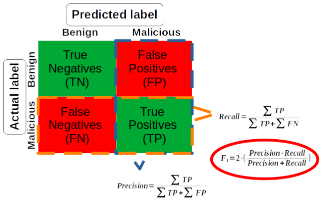
MACHINE LEARNING MODELS



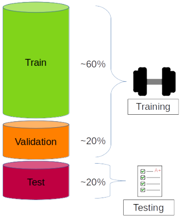
Key insight: models learn differently from data $\rightarrow$ different predictions for the instances

EXPERIMENTAL SETUP 2

Evaluation metric: F_1



Data split: holdout CV



Note 1: $F_1 \in [0, 1]$ where 1 is optimal

Note 2: stratified split respecting class distributions (slide 9)

Note 3: CIC-IDS-2017 $\rightarrow$ 20 random splits, CIC-IDS-2018 $\rightarrow$ 10

RESULTS EXPERIMENTS 1 & 2

Setup:
CIC-IDS-2017 average F_1 scores

Observation 1:
Diagonal showing detecting known attacks; almost perfect in all situations

Observation 2.1:
Botnet attacks cannot help detecting other attacks

Observation 2.2:
Botnet attacks cannot be detected using other attacks

Observation 3:
The matrix is not symmetrical in the diagonal

	Botnet	GoldenEye	Hulk	LOIC	SlowHTTPTest	Slowloris
Botnet-DT	0.997	0.000	0.001	0.000	0.000	0.043
Botnet-GNB	1.000	0.000	0.000	0.000	0.000	0.000
Botnet-KNN	0.991	0.000	0.000	0.000	0.000	0.000
Botnet-RF	1.000	0.000	0.000	0.000	0.000	0.000
GoldenEye-DT	0.000	1.000	0.871	0.949	0.018	0.569
GoldenEye-GNB	0.000	0.997	0.771	0.997	0.004	0.860
GoldenEye-KNN	0.000	0.998	0.997	0.997	0.354	0.001
GoldenEye-RF	0.000	1.000	0.807	0.844	0.004	0.498
Hulk-DT	0.012	0.483	1.000	0.253	0.147	0.436
Hulk-GNB	0.000	0.476	0.999	0.998	0.004	0.003
Hulk-KNN	0.000	0.821	1.000	1.000	0.371	0.001
Hulk-RF	0.000	0.540	1.000	0.804	0.002	0.000
LOIC-DT	0.000	0.000	0.002	1.000	0.000	0.000
LOIC-GNB	0.000	0.000	0.000	1.000	0.000	0.000
LOIC-KNN	0.000	0.666	0.923	1.000	0.000	0.000
LOIC-RF	0.000	0.000	0.000	1.000	0.000	0.000
SlowHTTPTest-DT	0.002	0.073	0.066	0.006	0.995	0.618
SlowHTTPTest-GNB	0.000	0.001	0.029	0.807	0.234	0.297
SlowHTTPTest-KNN	0.000	0.002	0.000	0.000	0.987	0.602
SlowHTTPTest-RF	0.000	0.005	0.012	0.001	0.996	0.829
Slowloris-DT	0.003	0.087	0.146	0.000	0.159	0.998
Slowloris-GNB	0.000	0.000	0.000	0.000	0.181	0.901
Slowloris-KNN	0.000	0.001	0.000	0.000	0.208	0.993
Slowloris-RF	0.000	0.001	0.000	0.000	0.018	0.997

(D)DoS variant in training data – ML model

(D)DoS variant in test data

Conclusion 1:
ML models can detect known (D)DoS attacks
See (1)

Conclusion 2:
If an ML algorithm learned to detect attack A and is able to detect B, vice versa is not necessarily the case
See (3)

Remarkable: GNB cannot learn to detect SlowHTTPTest, but can find LOIC

Similar results for the CIC-IDS-2018

RESULTS EXPERIMENT 3

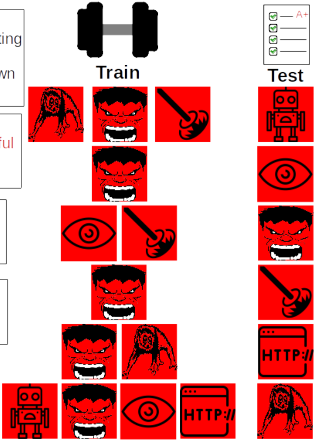
Setup: CIC-IDS-2017 detecting novel attack using combinations of known other attacks

Observation 1: Hulk dominantly useful to achieve highest score

Observation 2.1: KNN and DT obtain highest scores

Observation 2.2: Hulk and LOIC best detected with ML

Observation 3: Highest score obtained with small set of known attacks



f(x)=y

Model	Score
DT	0.460
KNN	0.821
KNN	**0.997**
KNN	**0.999**
KNN	0.399
DT	0.878

Highest average F_1 score

Conclusion 1: ML models using more known attacks do not achieve a higher novel detection rate See (3)

What about the CIC-IDS-2018?

RESULTS EXPERIMENT 3

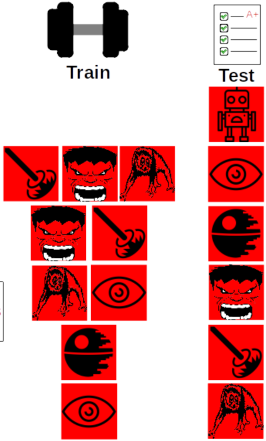
Setup:
CIC-IDS-2018 detecting novel attack using combinations of known other attacks

Observation 1.1:
GNB more robust against strong class imbalance

Observation 2.1:
Botnet attacks could not be detected

Observation 2.2:
Hulk, LOIC and Slowloris best detected with ML

Observation 3:
Highest score obtained with small set of known attacks



f(x)=y **Model**	**Score**
	0.0
GNB	0.862
GNB	0.853
GNB	**0.999**
KNN	**0.985**
GNB	**0.922**

Highest average F_1 score

Conclusion 1:
ML models using more known attacks do not achieve a higher novel detection rate
See (3)

Conclusion 2:
Higher imbalance of classes affects performance DT and RF considerably
See (1.1)

SUMMARY

We observed that:

- **ML models** are to a great extent able to detect known (D)DoS attacks in a closed world setting
- There are situations where these models are able to detect a novel variant when they are trained to detect a different variant
- Training on imbalanced data has an adverse effect on the evaluation performance of some ML classifiers
- It is not necessary to use many (D)DoS variants to detect a novel attack → sometimes a few known attacks can already lead to the highest novel detection rate

CONCLUSION AND FUTURE WORK

ML models can detect (D)DoS cyberattacks almost as well as signature-based approaches, but also have the capability to detect novel variants

Future directions:

1. Study a different type of cyberattacks (e.g. web-attacks)
2. Use different combinations of protocols (e.g. TCP and FTP)

THANKS FOR LISTENING

Do you have any questions?
Or email me: evdb@cwi.nl