Intelligent Thresholding

AMOSSYS
EXPERT IN CYBERSECURITY

Alban Siffer

November 20, 2018

Contents

Context

Motivations

- Massive usage of the Internet

Motivations

- Massive usage of the Internet
  - More and more vulnerabilities

Tesla Model S Hack Could Let Thieves Clone Key Fobs to Steal Cars

Motivations

Hackers Infect Over 200,000 MikroTik Routers With Crypto Mining Malware

- Massive usage of the Internet
  - More and more vulnerabilities
  - More and more threats

Tesla Model S Hack Could Let Thieves Clone Key Fobs to Steal Cars

Motivations

Hackers Infect Over 200,000 MikroTik Routers With Crypto Mining Malware

- Massive usage of the Internet
  - More and more vulnerabilities
  - More and more threats
- Awareness of the sensitive data and infrastructures

Tesla Model S Hack Could Let Thieves Clone Key Fobs to Steal Cars

Motivations

Hackers Infect Over 200,000 MikroTik Routers With Crypto Mining Malware

- Massive usage of the Internet
  - More and more vulnerabilities
  - More and more threats
- Awareness of the sensitive data and infrastructures
- Network security : a major concern

Tesla Model S Hack Could Let Thieves Clone Key Fobs to Steal Cars

- IDS (Intrusion Detection System)
  - Monitor traffic
  - Detect attacks

A Solution

- IDS (Intrusion Detection System)
  - Monitor traffic
  - Detect attacks
- Current methods : rule-based
  - Work fine on common and well-known attacks
  - Cannot detect new attacks

A Solution

- IDS (Intrusion Detection System)
  - Monitor traffic
  - Detect attacks
- Current methods : rule-based
  - Work fine on common and well-known attacks
  - Cannot detect new attacks
- Emerging methods : anomaly-based
  - Use the network data to estimate a normal behavior
  - Apply algorithms to detect abnormal events ($\rightarrow$ attacks)

Overview

- Overall design machine learning/data mining techniques for intrusion detection

data → ALGORITHM → alerts

Overview

- Overall design machine learning/data mining techniques for intrusion detection

$$\text{data} \longrightarrow \boxed{\text{ALGORITHM}} \longrightarrow \text{alerts}$$

- All ”standard” algorithms have been tested …

Overview

- Overall design machine learning/data mining techniques for intrusion detection

data → ALGORITHM → alerts

- All "standard" algorithms have been tested ...
- ... mostly on KDD99 dataset
  - not really representative
  - encourage supervised algorithms

Behind magic

- Algorithms are not magic
    - They give some information about data (scores)

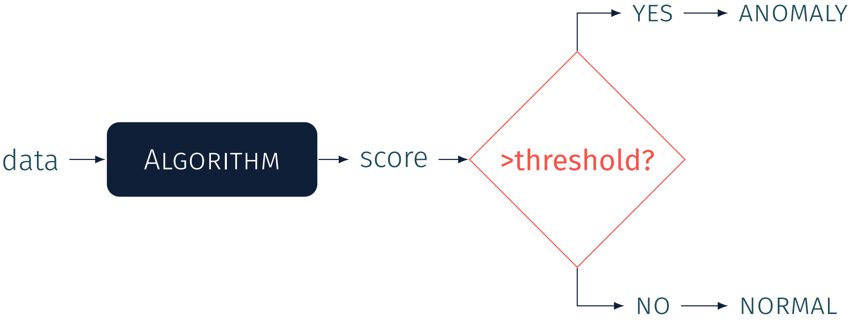
data → ALGORITHM → score

Behind magic

- Algorithms are not magic
  - They give some information about data (scores)
  - But the decision often rely on a human choice

Behind magic

- Algorithms are not magic
  - They give some information about data (scores)
  - But the decision often rely on a human choice



Inherent problem

- Two points are often not tackled

Inherent problem

- Two points are often not tackled
  - How to set the threshold?
  - What does this threshold mean?

Inherent problem

- Two points are often not tackled
  - How to set the threshold?
  - What does this threshold mean?
- Common approaches to set it

Hard-coded

Probabilistic

Inherent problem

- Two points are often not tackled
  - How to set the threshold?
  - What does this threshold mean?
- Common approaches to set it

Lack of adaptability

Lack of interpretability

Hard-coded

Probabilistic

Expertise or fine-tuning are required

- Two points are often not tackled
  - How to set the threshold?
  - What does this threshold mean?
- Common approaches to set it

Lack of adaptability

Lack of interpretability

Hard-coded

Expertise or fine-tuning are required

Probabilistic

Adaptable

Real meaning

Distribution assumption

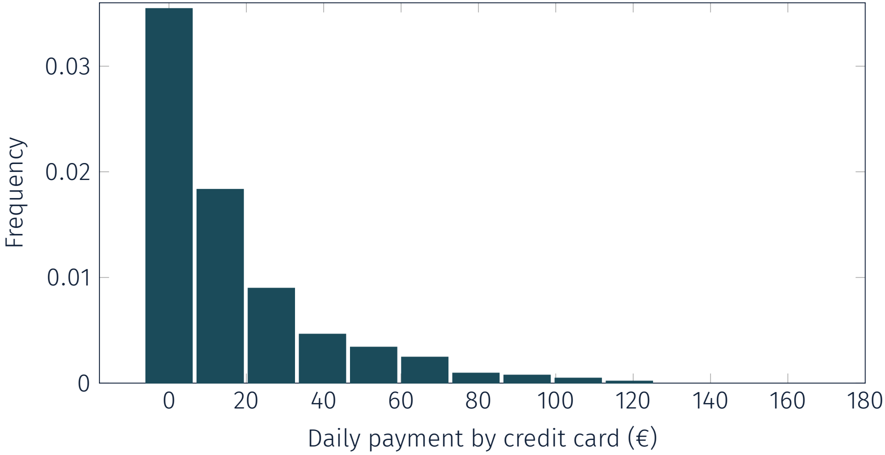
Recent examples

- *GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training* [1]
  → Hard-coded
- *Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications* [2] → Hard-coded
- *Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection* [3]
  → Distribution assumption (log-normal)

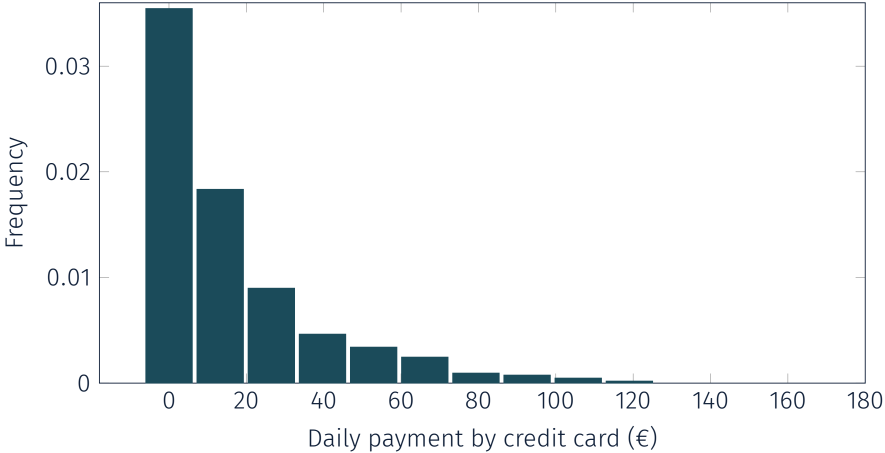
[1] Akcay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. arXiv preprint (2018)
[2] Xu, Haowen, et al. Proceedings of the 2018 World Wide Web Conference on World Wide Web
[3] Yisroel Mirsky, Tomer Doitshman, Yuval Elovici, and Asaf Shabtai (NDSS'18)

Providing better thresholds

My problem

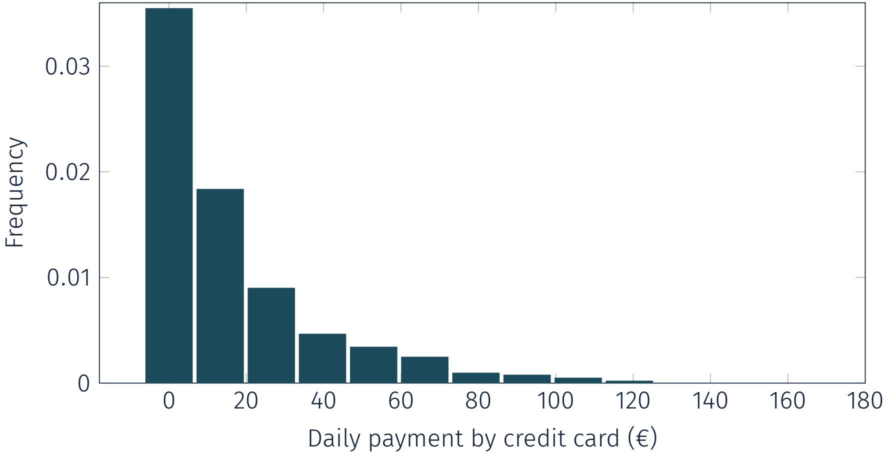

0.03
0.02
0.01
0
Frequency
0
20
40
60
80
100
120
140
160
180
Daily payment by credit card (€)

My problem

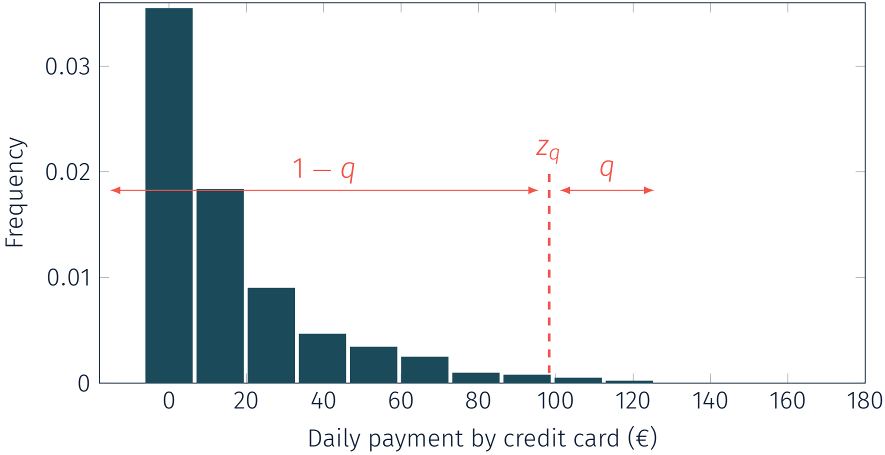


⊸ How to set z_q such that $\mathbb{P}(X€ > z_q) < q$?

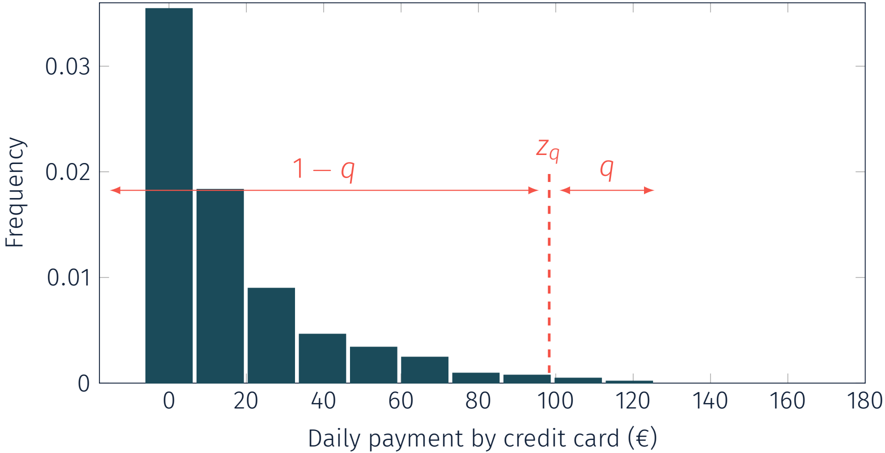
Solution 1: Empirical Approach



Solution 1: Empirical Approach

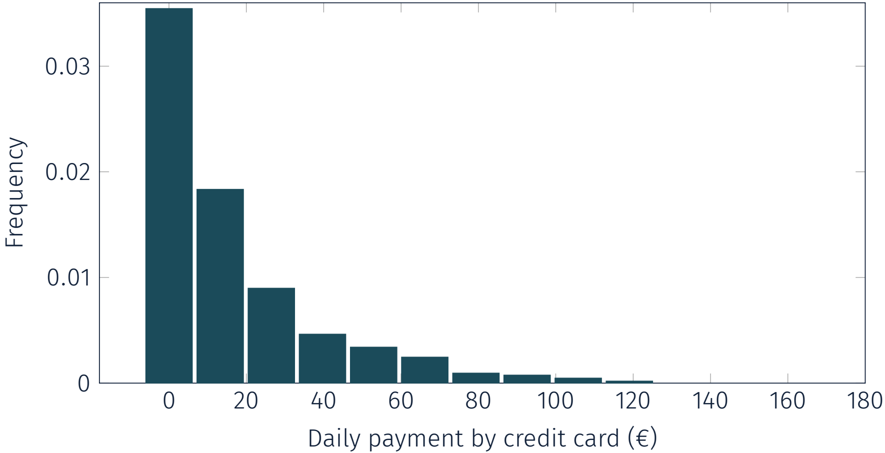


Solution 1: Empirical Approach

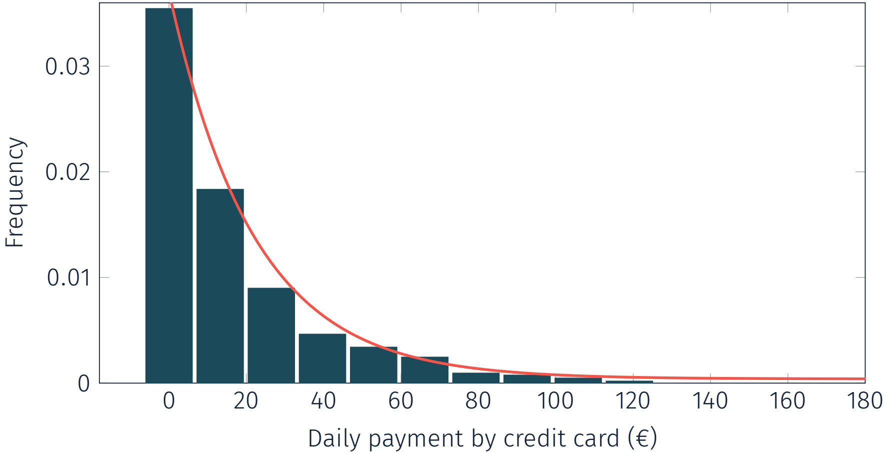


⟜ Drawbacks: stuck in the interval, poor resolution

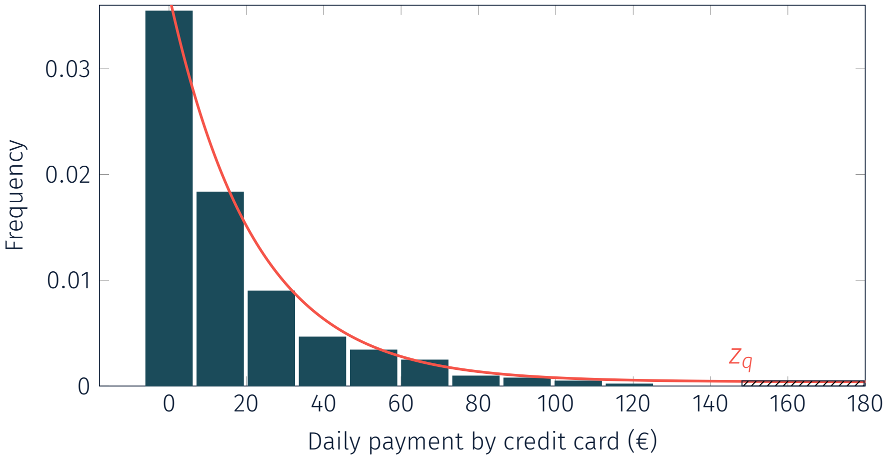
Solution 2: Standard Model



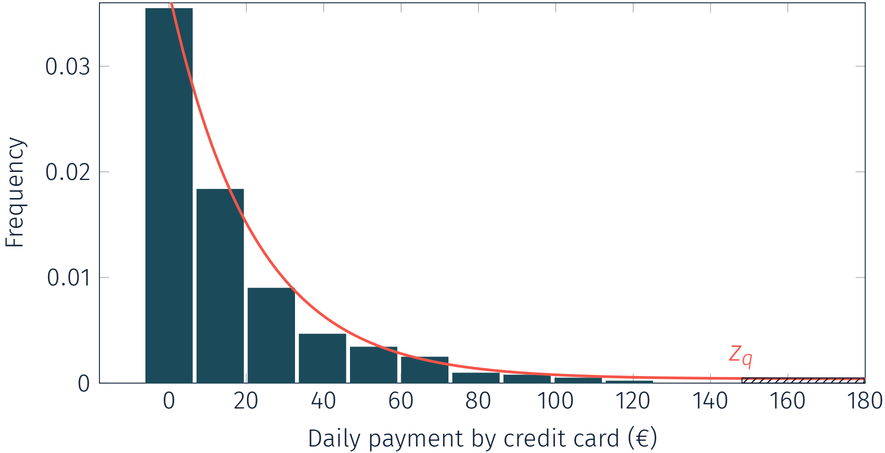
Solution 2: Standard Model



Solution 2: Standard Model

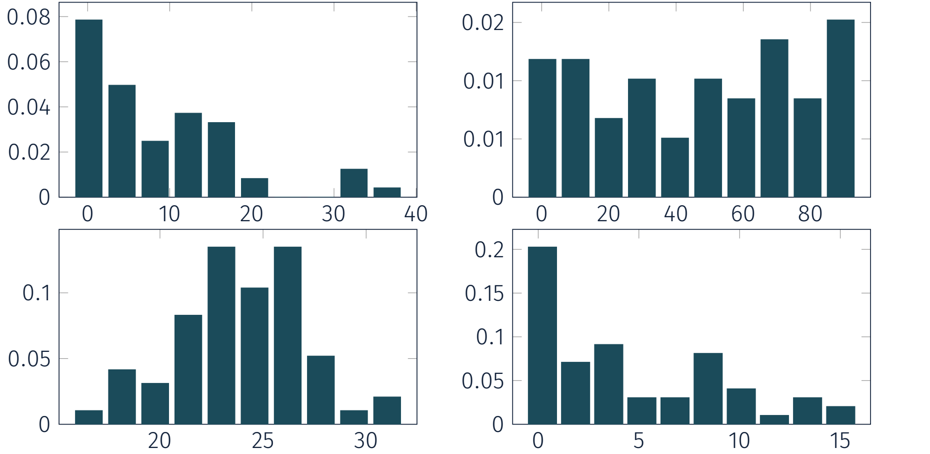


Solution 2: Standard Model

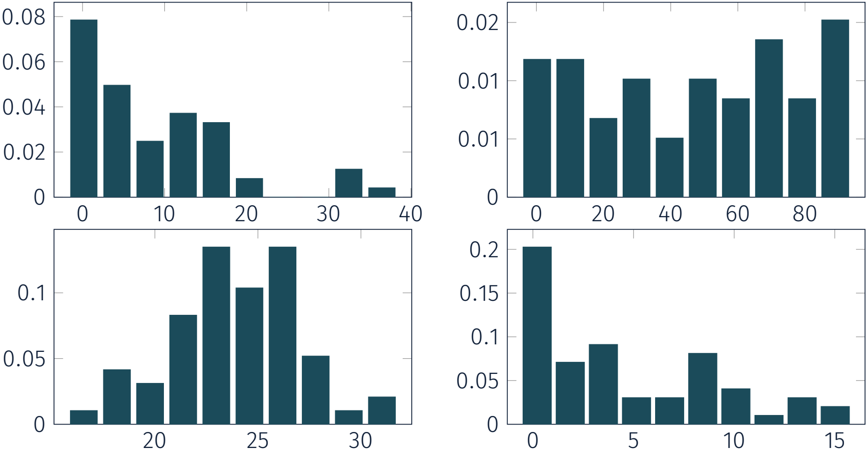


⟶ Drawbacks: manual step, distribution assumption

Realities


0.08
0.06
0.04
0.02
0
0
10
20
30
40
0.02
0.01
0.01
0
0
20
40
60
80
0.1
0.05
0
20
25
30
0.2
0.15
0.1
0.05
0
0
5
10
15

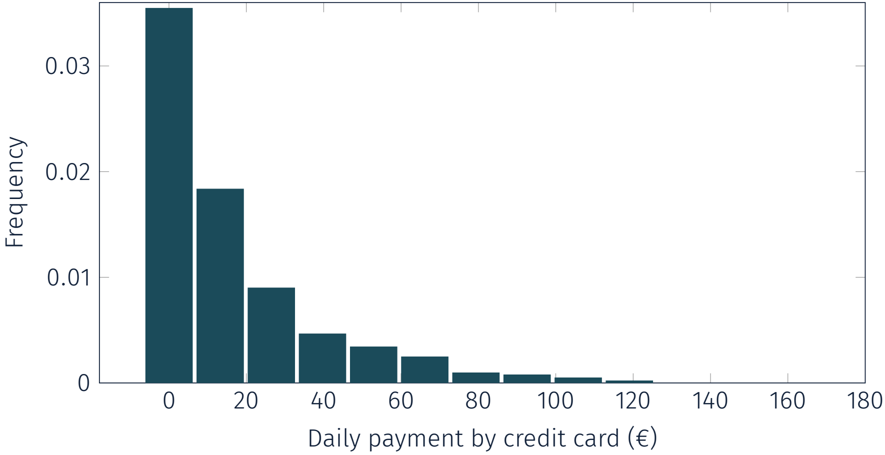
Realities

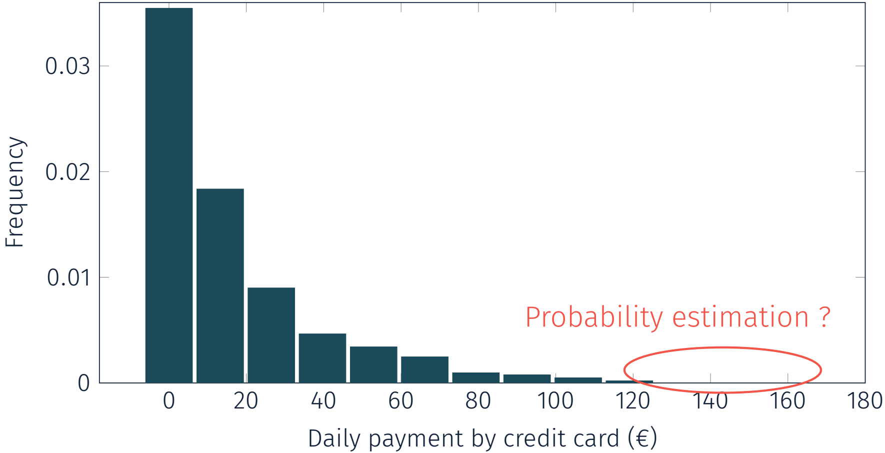


- Different behaviours, temporal drift

PROPERTIES	**Empirical quantile**	**Standard model**
statistical guarantees	Yes	Yes
easy to adapt	Yes	No
high resolution	No	Yes

Inspection of extreme events



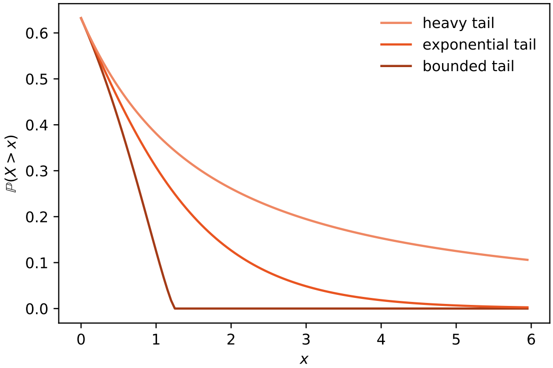

0.03
0.02
0.01
0
Frequency
0
20
40
60
80
100
120
140
160
180
Daily payment by credit card (€)
Probability estimation ?

- Main result (Fisher-Tippett-Gnedenko, 1928)

  The extreme values of any distribution have nearly the same distribution (called Extreme Value Distribution)

- Main result (Fisher-Tippett-Gnedenko, 1928)

  The extreme values of any distribution have nearly the same distribution (called Extreme Value Distribution)



How to use EVT

- Get some data $X_1, X_2 \ldots X_n$

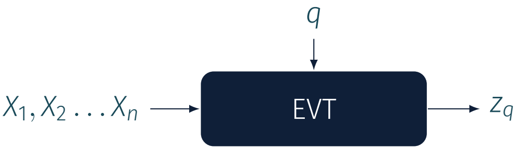
How to use EVT

- Get some data $X_1, X_2 \ldots X_n$
- Estimate the most appropriate tail

- Get some data $X_1, X_2 \ldots X_n$
- Estimate the most appropriate tail
- Compute z_q such as $\mathbb{P}(X > z_q) < q$

How to use EVT

- Get some data $X_1, X_2 \ldots X_n$
- Estimate the most appropriate tail
- Compute z_q such as $\mathbb{P}(X > z_q) < q$

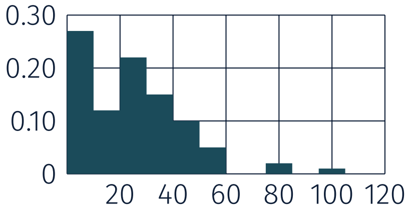


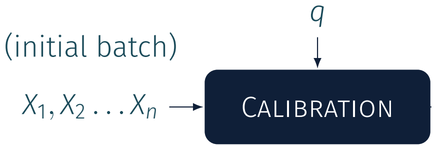
Finding anomalies in streams

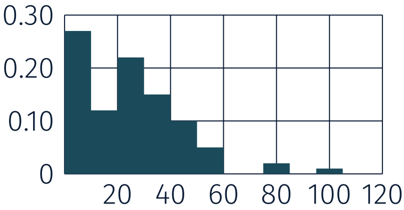
Streaming Peaks-Over-Threshold (SPOT) algorithm

(initial batch)

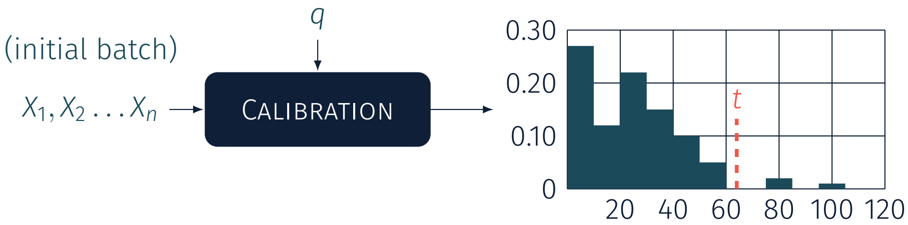
$X_1, X_2 \ldots X_n$



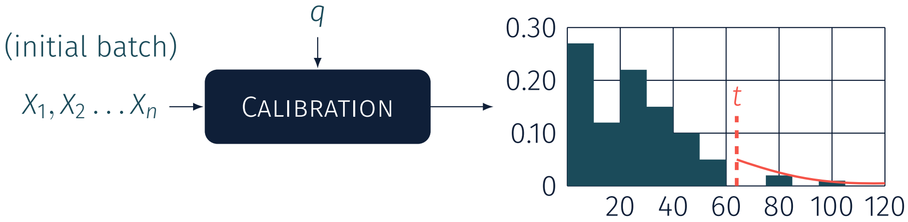

q
(initial batch)
$X_1, X_2 \ldots X_n$
Calibration


0.30
0.20
0.10
0
20
40
60
80
100
120

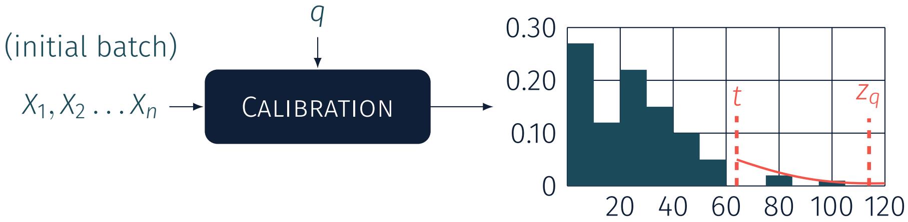
Streaming Peaks-Over-Threshold (SPOT) algorithm



Streaming Peaks-Over-Threshold (SPOT) algorithm



Streaming Peaks-Over-Threshold (SPOT) algorithm



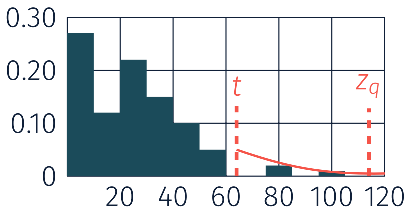
Streaming Peaks-Over-Threshold (SPOT) algorithm

q

(initial batch)

$X_1, X_2 \ldots X_n$

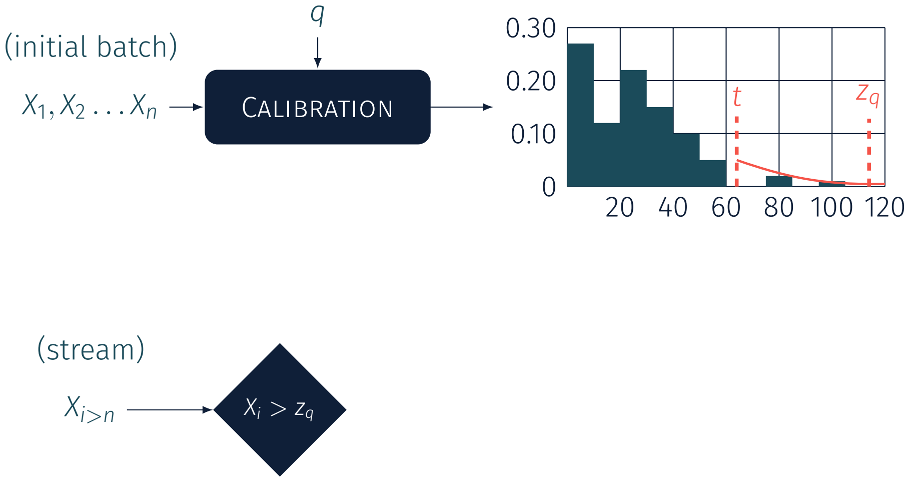
Calibration



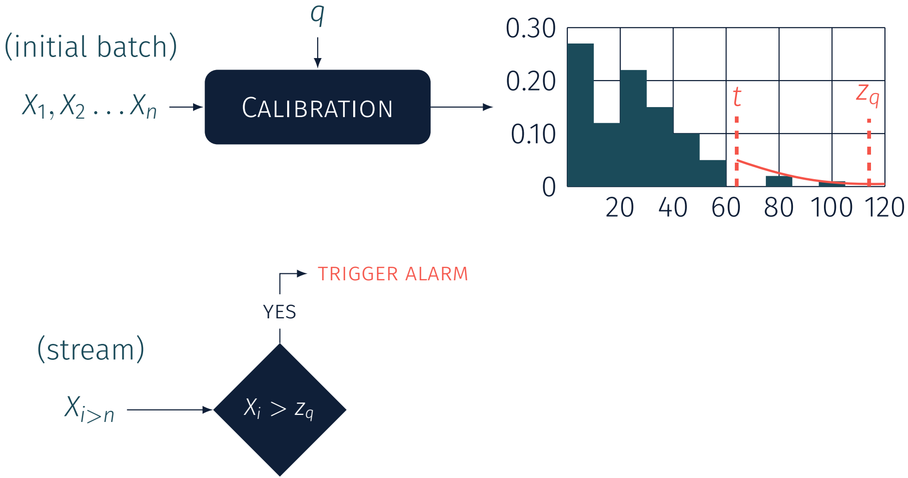
(stream)

$X_{i>n}$

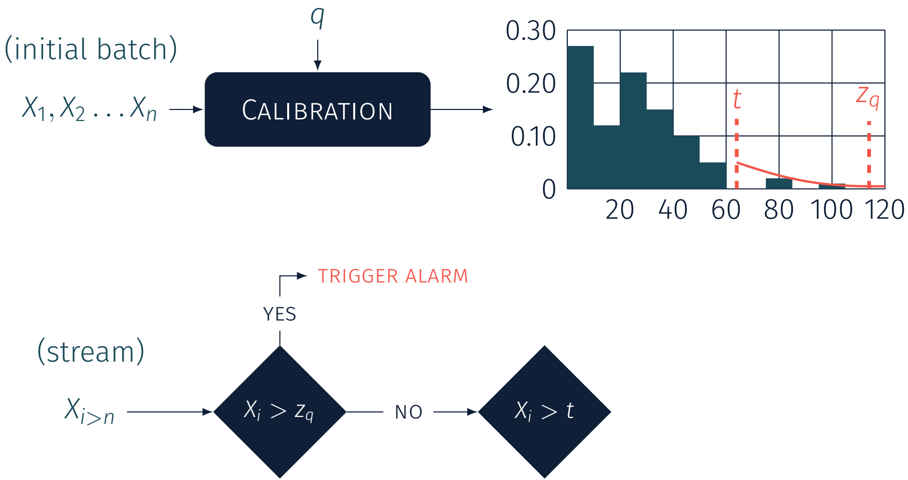
Streaming Peaks-Over-Threshold (SPOT) algorithm



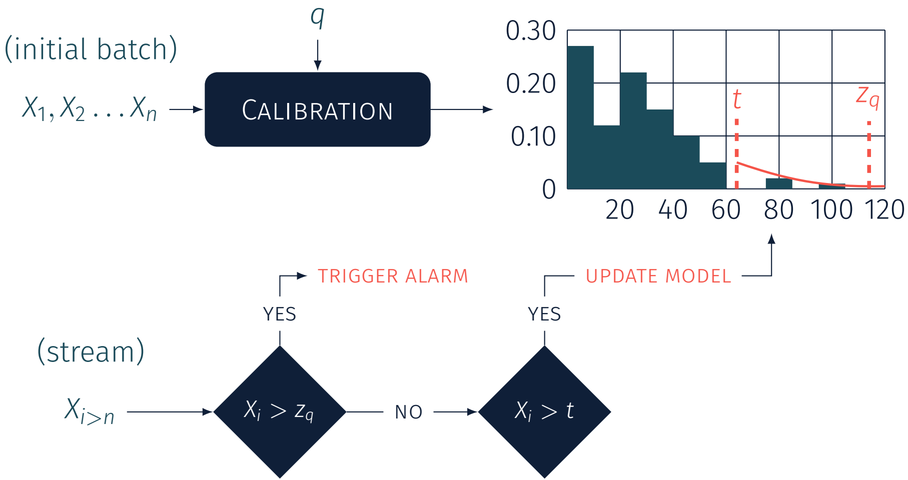
Streaming Peaks-Over-Threshold (SPOT) algorithm



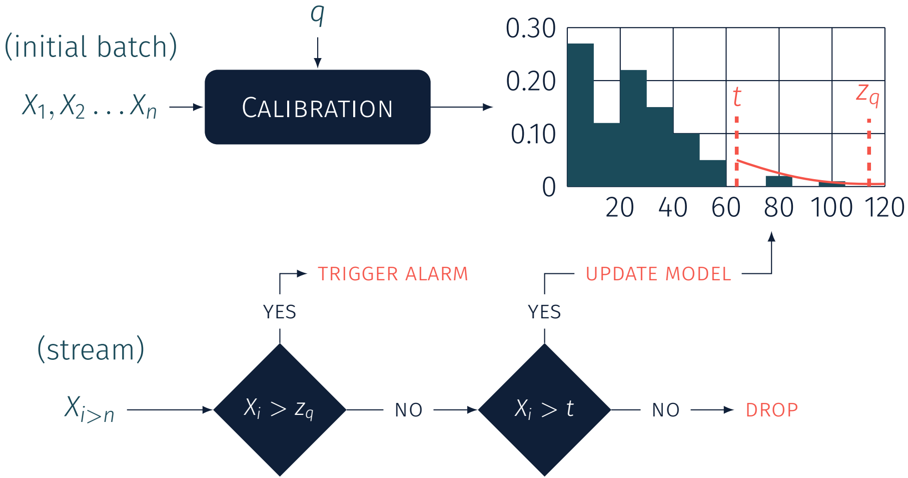
Streaming Peaks-Over-Threshold (SPOT) algorithm



Streaming Peaks-Over-Threshold (SPOT) algorithm



Streaming Peaks-Over-Threshold (SPOT) algorithm



Application to intrusion detection

About the data

- Lack of relevant public datasets to test the algorithms ...

About the data

- Lack of relevant public datasets to test the algorithms ...
- KDD99 ? See [McHugh 2000] and [Mahoney & Chan 2003]

About the data

- Lack of relevant public datasets to test the algorithms ...
- KDD99 ? See [McHugh 2000] and [Mahoney & Chan 2003]
- We rather use MAWI[1]
  - 15 min a day of real traffic (.pcap file)
  - Anomaly patterns given by the MAWILab [Fontugne *et al.* 2010] with taxonomy [Mazel et al. 2014]

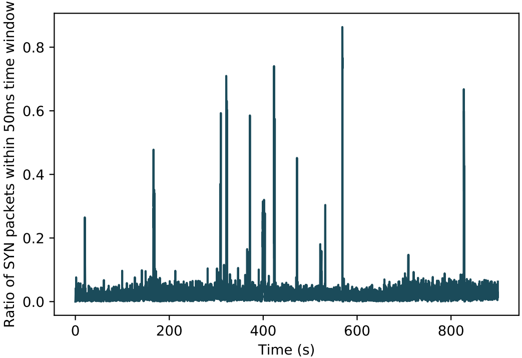
[1] http://www.fukuda-lab.org/mawilab/

An example to detect network SYN scan

- The ratio of SYN packets : relevant feature to detect network scan [Fernandes & Owezarski 2009]

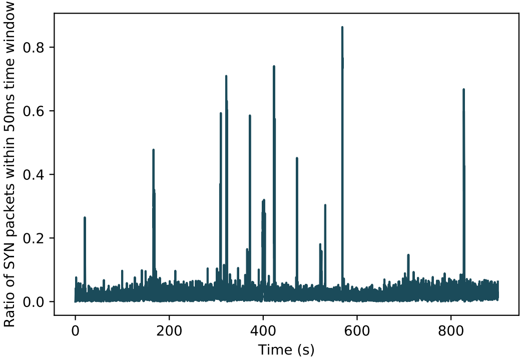
An example to detect network SYN scan

- The ratio of SYN packets : relevant feature to detect network scan [Fernandes & Owezarski 2009]



An example to detect network SYN scan

- The ratio of SYN packets : relevant feature to detect network scan [Fernandes & Owezarski 2009]

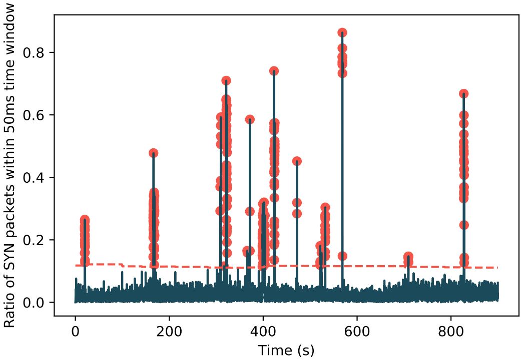


- Goal: find peaks

SPOT results

- Parameters : $q = 10^{-4}, n = 2000$ (from the previous day record)

- Parameters : $q = 10^{-4}, n = 2000$ (from the previous day record)

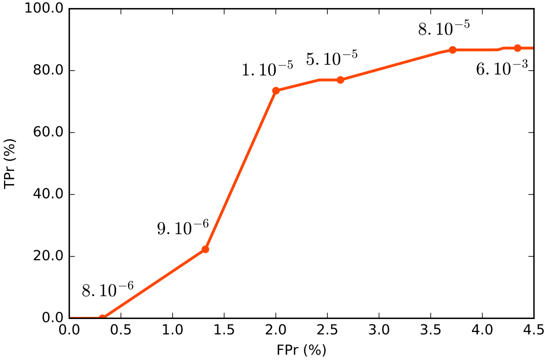


Do we really flag scan attacks ?

- The main parameter q: a False Positive regulator

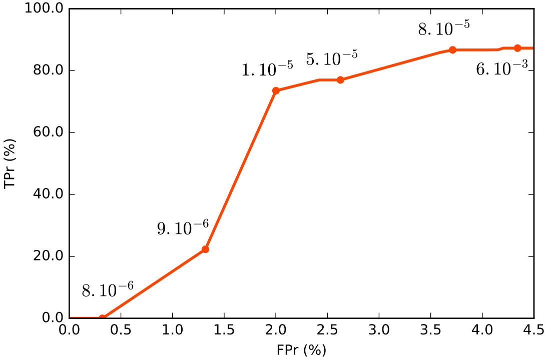
Do we really flag scan attacks ?

- The main parameter q: a False Positive regulator



Do we really flag scan attacks ?

- The main parameter q: a False Positive regulator



- 86% of scan flows detected with less than 4% of FP

In a nutshell

SPOT Specifications for automatic thresholding

- A single main parameter q
  - With a probabilistic meaning $\rightarrow \mathbb{P}(X > z_q) < q$
  - False Positive regulator

SPOT Specifications for automatic thresholding

- A single main parameter q
  - With a probabilistic meaning $\rightarrow \mathbb{P}(X > z_q) < q$
  - False Positive regulator
- Stream capable
  - Incremental learning
  - Online detection
  - Fast (current `C++` library: `libspot`, >100000 values/s)
  - Low memory usage (only the excesses)

SPOT Specifications for automatic thresholding

- A single main parameter q
  - With a probabilistic meaning $\rightarrow \mathbb{P}(X > z_q) < q$
  - False Positive regulator
- Stream capable
  - Incremental learning
  - Online detection
  - Fast (current `C++` library: `libspot`, >100000 values/s)
  - Low memory usage (only the excesses)
- Wide number of applications
  - Back-end of scoring methods
  - drifting contexts (with an additional parameter) $\rightarrow$ DSPOT

Conclusion

- Context: A great deal of work has been done to develop anomaly detection algorithms

Conclusion

- Context: A great deal of work has been done to develop anomaly detection algorithms
- Problem: Decision thresholds rely on either distribution assumption or expertise

Conclusion

- Context: A great deal of work has been done to develop anomaly detection algorithms
- Problem: Decision thresholds rely on either distribution assumption or expertise
- Our solution: Building dynamic threshold with a probabilistic meaning

Conclusion

- Context: A great deal of work has been done to develop anomaly detection algorithms
- Problem: Decision thresholds rely on either distribution assumption or expertise
- Our solution: Building dynamic threshold with a probabilistic meaning
    - Application to detect network anomalies

Conclusion

- Context: A great deal of work has been done to develop anomaly detection algorithms
- Problem: Decision thresholds rely on either distribution assumption or expertise
- Our solution: Building dynamic threshold with a probabilistic meaning
    - Application to detect network anomalies
    - But a general tool to monitor online time series in a blind way