Statistical analysis of DDoS attack

Pranshu Kandoi (200101086)
Vatsal Gupta (200101105)

Under the guidance of Prof. Pinaki Mitra

Indian Institute of Technology, Guwahati

Problem Statement

The main problem statement is summarised as follows:

- Proposing statistical methods for IoT (Internet of Things) attack detection with **arbitrary traffic**.
- In particular, recognising DDoS (Distributed Denial of Service) attacks in a **timely fashion** with simple statistical implementation, integrable into a hardware probe.

Prior Work

- [1] Proposed one-parameter statistical methods for DDoS attack recognition focusing on simulated attack data and concluded with approaches suggesting possibilities of early detection statistical parameters.
- [2] Extended [1] to detect the "start" of a DDoS attack - attack time and used some statistical parameters to prove their efficiency on a custom dataset.
- Other papers *(as cited in the report)* also analyse DDoS attack detection.

[1] Hajtmanek et Al. One-Parameter Statistical Methods to Recognize DDoS Attacks. *Symmetry*. 2022
[2] Smiesko et Al. Machine Recognition of DDoS Attacks Using Statistical Parameters. *Mathematics*. 2024

Proposed Approach: Premise

- We re-envision attack detection in IoT with price detection in the finance world - we create a_t (Total packets forwarded in window ending at t) as the analogue of price.
- We hence adapt complex statistical financial strategies to IoT systems and analyse their performance.
- We have established a temporal metric μ to quantify performance of various statistical strategies.

Proposed Approach: Novel points

- The algorithms proposed in various strategies are **incremental** and hence are very useful for the **online task** nature of the problem.
- Using statistical strategies allows for **fast computation** (as compared to ML based strategies) and require less resources so can be used easily on **small scale compact IoT devices**.
- Multiple strategies can be used in conjunction allowing for a diverse, adaptable and **robust IDS** (Intrusion Detection System).

Proposed Approach: Methodology

- Select appropriate time quantum ($a_t =$ total fowarded packets in t^{th} window) to resample upon depending on compute-time and window-size trade-off.
- Create time-series data to test statistical strategies based on selected time quantum.
- Evaluate strategies based on proposed metric (μ).

Data creation: How time-sampling is done?

- $A(0,t) = \sum_{i=1}^{t} a(i), \quad a(t) = A(0,t) - A(0,t-1)$

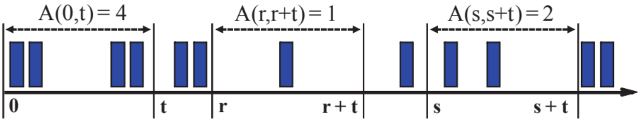


Figure: $A(s,t)$ is total packets flown between window s and t
$a(t)$ or a_t is the number of packets in window t

Data creation: Selecting the right time quantum

- 1s was chosen as the appropriate time quantum by running analysing running time of practical algorithms.

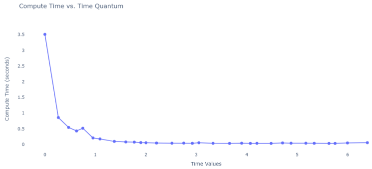


Figure: Compute time v/s Time Quantum.

- The trade-off gave 1s as the appropriate time-quantum, based on this the data is split into
  - **Regular data** Consisting of 1 value for each window.
  - **OHLC Data** Consisting of 4 values for each window.

Data Creation: Regular vs OHLC Data

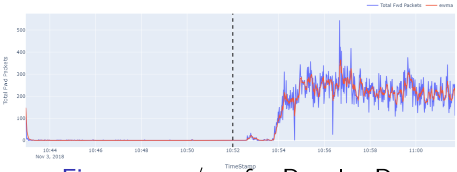


Figure: a_t v/s t for Regular Data

Figure: Zoomed-in OHLC candlestick Chart

Figure: The instant that attack begins (attack time) is shown by a vertical dashed black line.

Strategies: Studied

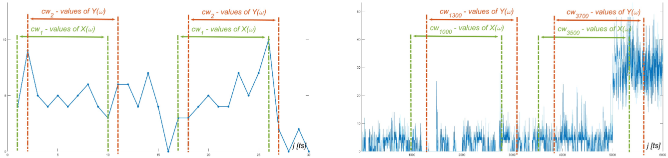
Trend & Momentum based Strategies
ATR
KCHB
AO
Stoch
AROON
UI

Multi-OHLC based Strategies
BB
RAPD
KST
MACD
RSI
TSI
TRIX
StochRSI

Non-Financial based Strategies
C/V
SKEW
KURT
R/S
D_{KL}
c

- Trend & momentum uses regular data where as Multi-OHLC uses the OHLC data. Both classes are financial strategies.

Strategies: How are they calculated?



- As we gradually analyse obtain new datapoints, we obtain a time series of the estimated values of the given statistical parameter.
- This statistical parameter "looks back" i.e. uses k previous values to compute its new current value which we term as "lag".

Strategies: Bollinger band

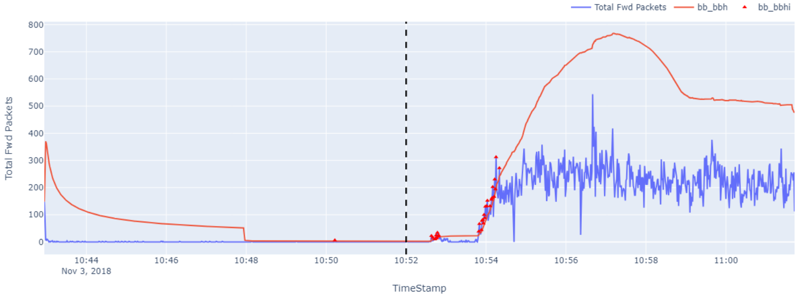


Figure: Understanding signal creation with Bollinger Bands

Results and Analysis: Metric proposed

- Previous metrics overlook temporal aspects.
- We propose a metric based on the exponential decay function which gives exponentially lower weights to signals further away from attack time (λ).

$$\boxed{\mu_\alpha(i) = e^{-\frac{x_i - \lambda}{\alpha}}} \tag{1}$$

$$\boxed{\mu_\alpha = \textstyle\sum_i \mu_\alpha(i), i \text{ ranges over all 1 in the signal array where } x > \lambda.} \tag{2}$$

- **Proximity** vs **Frequency** Trade-off is established by this metric

Results and Analysis: Analysis of strategies based on μ

- We compare Non-Financial and Financial strategies based on the μ metric.

Strategies	**FP**	**TP**	$\mu_{0.01}$	$\mu_{0.1}$	$\mu_{0.5}$	μ_1	μ_2	μ_{10}
Financial Strategies								
BB	1	29	3.24×10^{-28}	6.51×10^{-3}	2.40	6.90	13.41	24.65
RAPD	**2**	41	3.85×10^{-28}	$\mathbf{1.08 \times 10^{-2}}$	3.95	10.65	19.80	35.13
KST	0	31	4.10×10^{-91}	5.80×10^{-9}	0.30	3.04	9.68	24.55
MACD	0	57	1.10×10^{-123}	3.23×10^{-12}	0.09	2.18	11.05	40.99
RSI	0	33	3.12×10^{-28}	2.73×10^{-3}	1.16	5.05	12.41	27.00
TSI	0	**142**	1.15×10^{-89}	8.14×10^{-9}	0.50	7.06	29.99	**103.10**
TRIX	0	21	2.59×10^{-30}	6.81×10^{-3}	3.91	9.02	13.75	19.29
StochRSI	1	7	$\mathbf{5.54 \times 10^{-26}}$	7.18×10^{-3}	0.95	2.13	3.61	6.06
AO	0	3	1.11×10^{-29}	1.27×10^{-3}	0.29	0.67	1.13	2.22
Stoch	1	32	6.11×10^{-29}	3.31×10^{-3}	1.03	3.44	9.16	24.50
KCHB	0	127	7.27×10^{-29}	9.30×10^{-3}	**4.48**	**16.15**	**40.62**	99.52
Non-Financial Strategies								
SKEW	1	2	$\mathbf{3.71 \times 10^{-28}}$	$\mathbf{3.12 \times 10^{-3}}$	0.55	1.05	1.45	1.87
KURT	1	2	3.61×10^{-28}	2.80×10^{-3}	0.55	1.06	1.84	2.11
R/S	1	**46**	3.12×10^{-28}	1.80×10^{-3}	**0.88**	**5.60**	**15.74**	**37.03**

Results and Analysis: Analysis of Strategies based on compute time

- Compute-Times are analysed for Non-Financial vs Financial strategies.

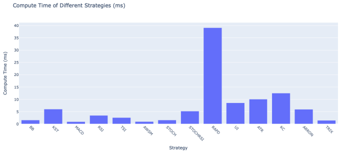


Figure: Compute time for financial strategies

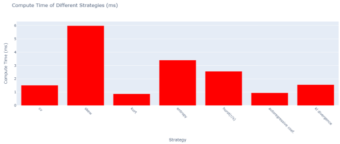


Figure: Compute time for non-financial strategies

Results and Analysis: Discussion on RAPD and StochRSI

- We choose to discuss 2 strategies showing great potential for creation of a well-balanced robust IDS and an early detection IDS i.e. RAPD and StochRSI respectively[1].
- The algorithm for RAPD was is discussed on the following slide. StochRSI has the following formula:

$$\boxed{\text{StochRSI} = \frac{\text{RSI}-\text{Min RSI}_K}{\text{Max RSI}_K-\text{Min RSI}_K}}$$

$$\boxed{\begin{aligned} \text{RSI} &= 100 - \frac{100}{1+\frac{\text{AG}}{\text{AL}}} \\ \text{AG} &= \frac{\sum_{i=1}^{n} \max(\text{C}_i - \text{C}_{i-1}, 0)}{n} \\ \text{AL} &= \frac{\sum_{i=1}^{n} \max(\text{C}_{i-1} - \text{C}_i, 0)}{n} \end{aligned}}$$

[1]Other strategies and their performance is discussed in detail in the report.

Results and Analysis: RAPD

for $i = lag + 1$ to t **do**
  if $|y(i) - \text{avgFilter}(i-1)| > \text{threshold} \times \text{stdFilter}(i-1)$ **then**
    if $y(i) > \text{avgFilter}(i-1)$ **then**
      set signals(i) to $+1$; ▷ Positive signal
    else
      set signals(i) to -1; ▷ Negative signal
    end if
    set filteredY(i) to influence $\times\ y(i) + (1 - \text{influence}) \times \text{filteredY}(i-1)$;
  else
    set signals(i) to 0; ▷ No signal
    set filteredY(i) to $y(i)$;
  end if
  set avgFilter(i) to mean(filteredY(i-lag+1),...,filteredY(i));
  set stdFilter(i) to std(filteredY(i-lag+1),...,filteredY(i));
end for

for i = lag + 1 to t **do**
  Compute *avgFilter*(i) : μ_n and *stdFilter*(i) : s_n incrementally by computing mean and standard deviation incrementally.
  $s_n^2 = \frac{n-2}{n-1} s_{n-1}^2 + \frac{1}{n}\left(x_n - \mu_{n-1}\right)^2$
  $\mu_n = \mu_{n-1} + \frac{x_n - \mu_{n-1}}{k}$
  if $|y(i) - \text{avgFilter}(i-1)| > \text{threshold} \times \text{stdFilter}(i-1)$ **then**
    $\mathit{signals}(i) = \text{sign}(y(i) - \text{avgFilter}(i-1))$;
    $\mathit{filteredY}(i) = \text{influence} \times y(i) + (1 - \text{influence}) \times \text{filteredY}(i-1)$;
  else
    $\mathit{signals}(i) = 0$;
    $\mathit{filteredY}(i) = y(i)$;
  end if
end for

Figure: Non-Incremental vs Incremental RAPD

Contd.

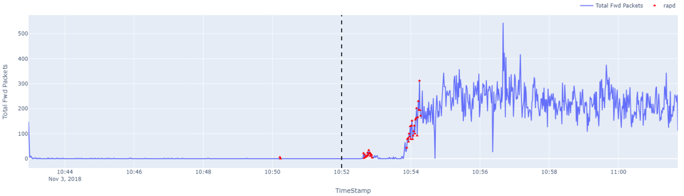


Figure: RAPD on a_t

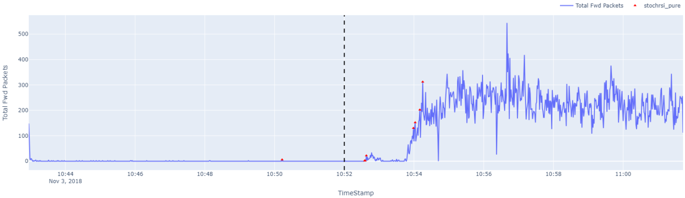


Figure: StochRSI on a_t

Non-Financial V/S Financial Strategies

- Financial strategies are **out-performing on time**
- Financial strategies **are more diverse** than non-financial strategies.
- Financial strategies **outperform on proximity** and can serve early detection systems better than Non-Financial strategies.
- Non-financial strategies like C/V which give a generalised broad peak **can work well in conjunction** with financial strategies.

Conclusion

- **Enhanced Performance with Financial Strategies** proving that adaption of IoT systems into finance world is an appreciable approach.
- **Evaluation Metric** The μ metric proves effective in assessing strategy performance.
- **Optimal Strategy Selection** We provide computational and frequency/proximity insights guiding creation of IDS systems using these strategies.
- **Robustness through Strategy Combination** Considering an attack only when the ensemble of multiple strategies are in agreement, thereby reducing false alarms and improving overall IDS reliability and resilience.

Questions and open discussion

- Any Questions?
- Our report can be found on this t.ly/BQAdC link.
- It can be downloaded directly from this t.ly/CKT39 link.

Thank you

Thank you