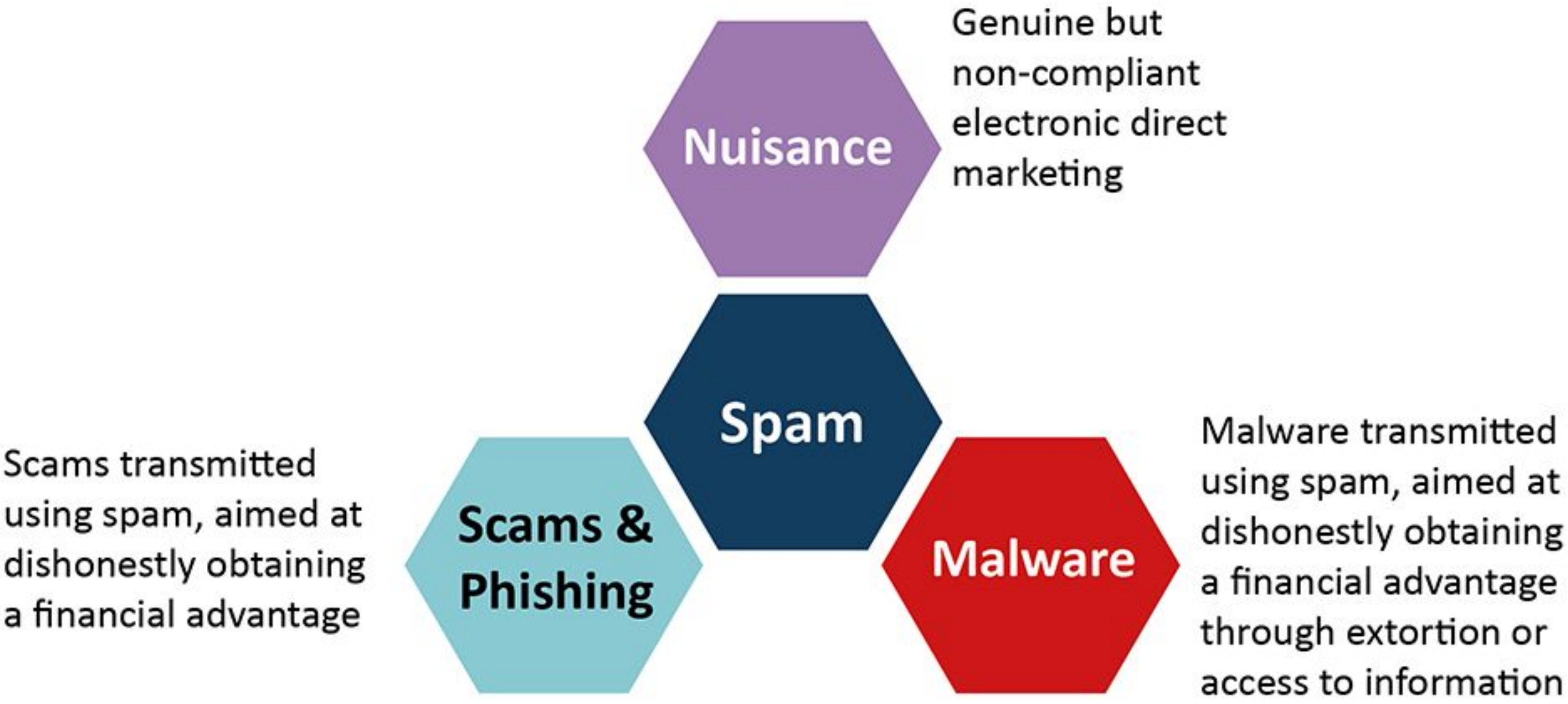
Preventing web-based attacks through Hidden Markov models

Proposed by Martiniuc Alexandru
Coordinator Conf. Dr. Dragoș-Teodor Gavriluț

Table of Contents

Context


Genuine but non-compliant electronic direct marketing
Nuisance
Spam
Scams &
Phishing
Malware
Scams transmitted using spam, aimed at dishonestly obtaining a financial advantage
Malware transmitted using spam, aimed at dishonestly obtaining a financial advantage through extortion or access to information

Problem description - Status quo

“Phishing is the simplest kind of cyberattack and, at the same time, the most dangerous and effective. That is because it attacks the most vulnerable and powerful computer on the planet: the human mind,"

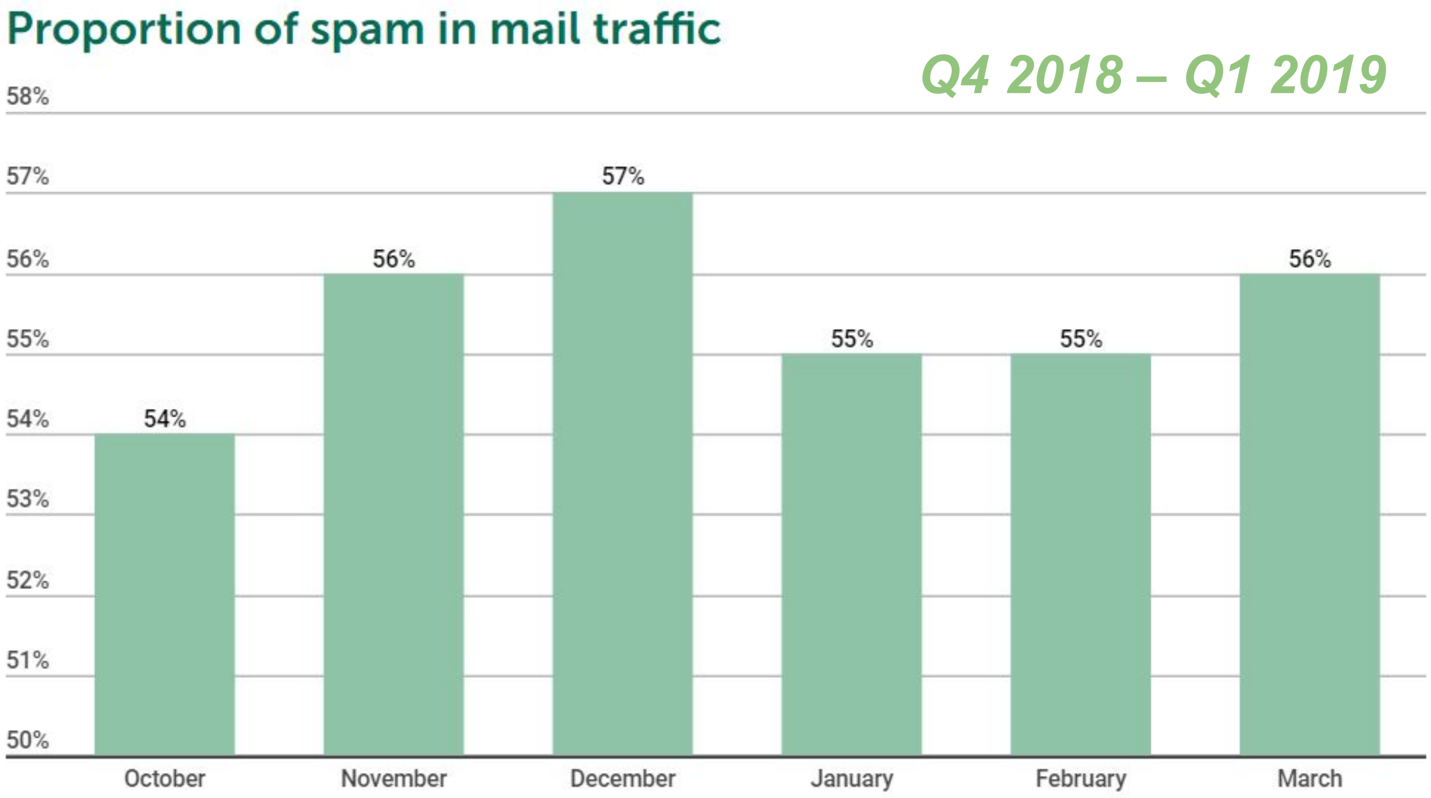
Adam Kujawa
Director of Malwarebytes Labs

- The increasing number of malicious url:

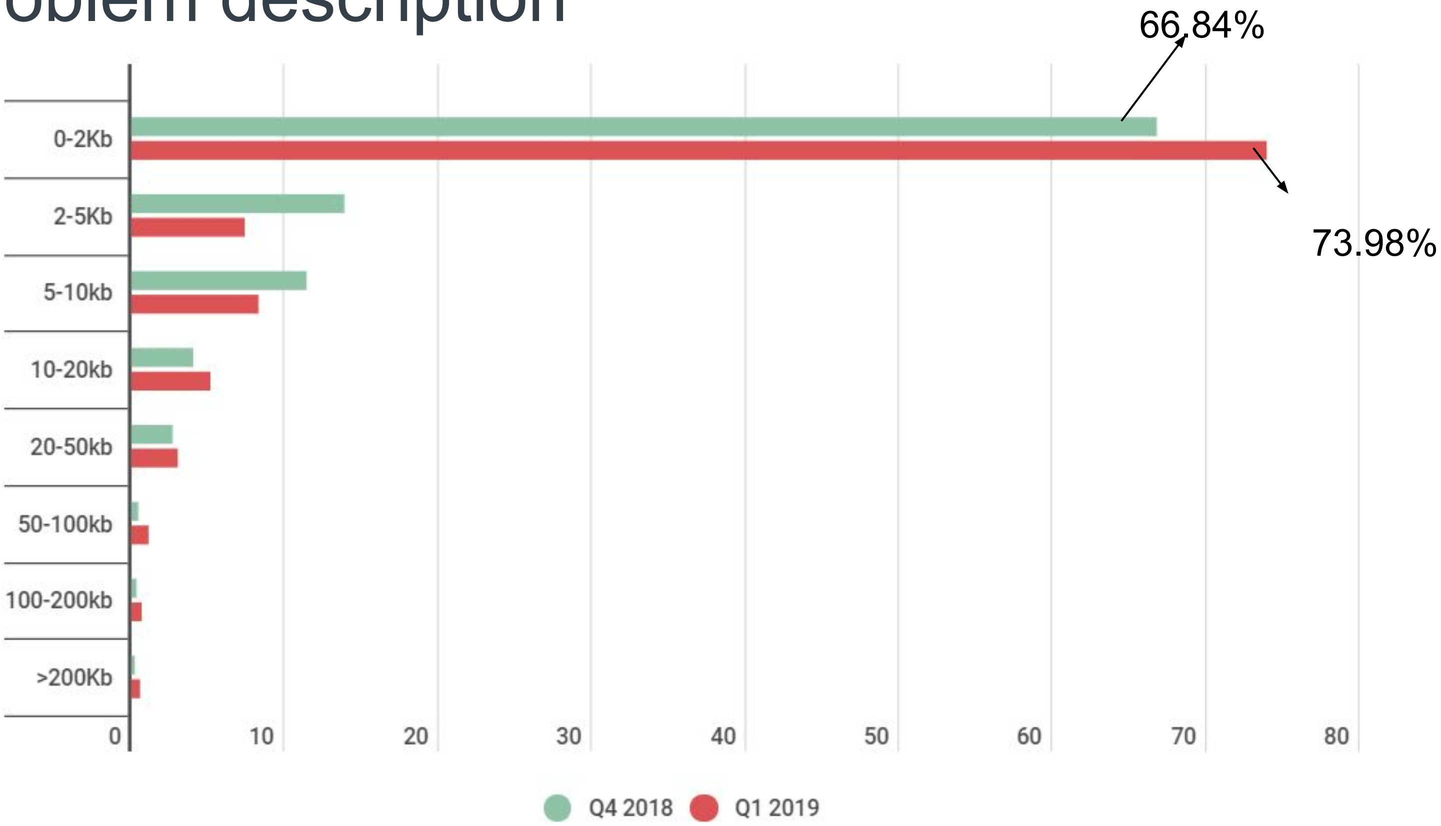
Phishing URLs	Malware URLs
> 35.000 per week	> 6.000 per week

Google Safe Browsing Analysis
01/06/2018 - 01/06/2019

Problem description



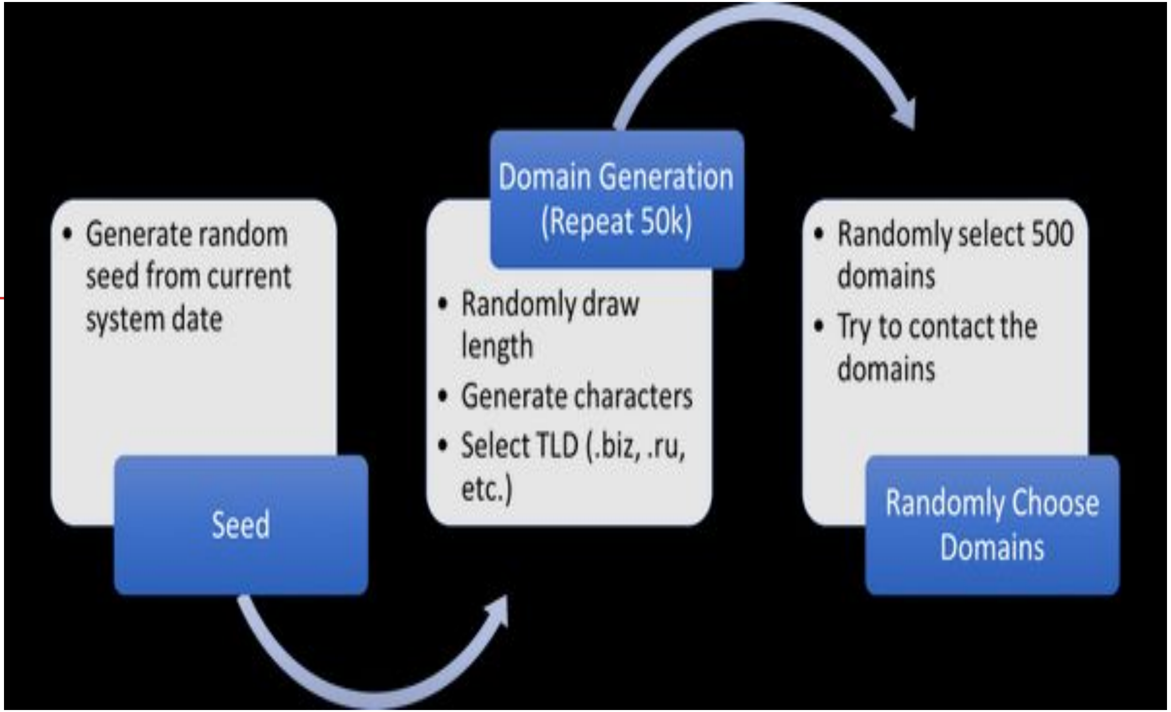
Problem description



Spam email size

Problem description

Worm:Win32 Conficker



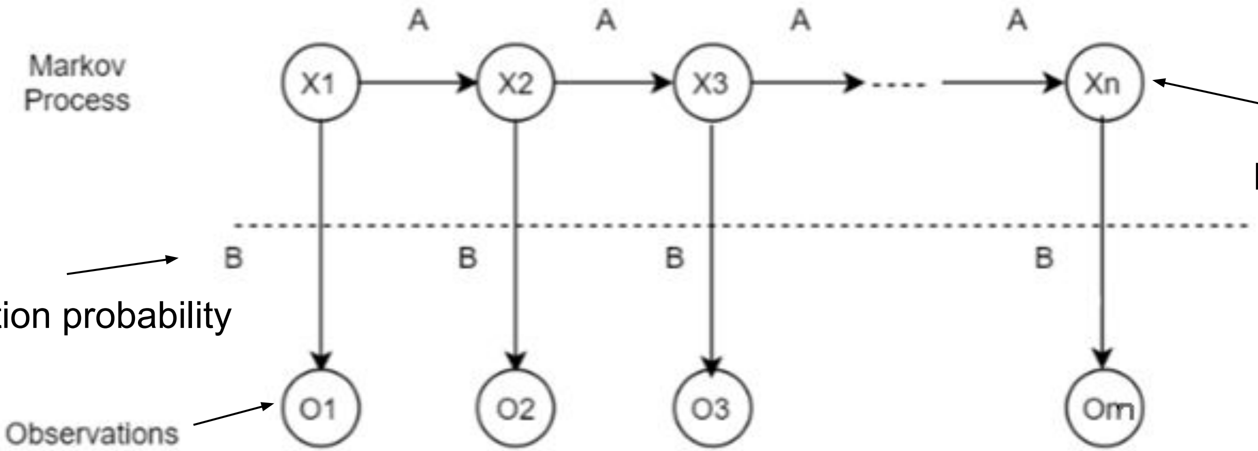
Hidden Markov model

- Markov model vs Hidden Markov model
- HMM representation and parameters

state transition matrix

hidden states

observation probability matrix



Notation: $\lambda = (A, B, \pi)$ initial state distribution

Hidden Markov model

The 3 basic problems of HMMs

Problem 1 - What is the probabily of the observations sequence?

- *The forward-backward algorithm*

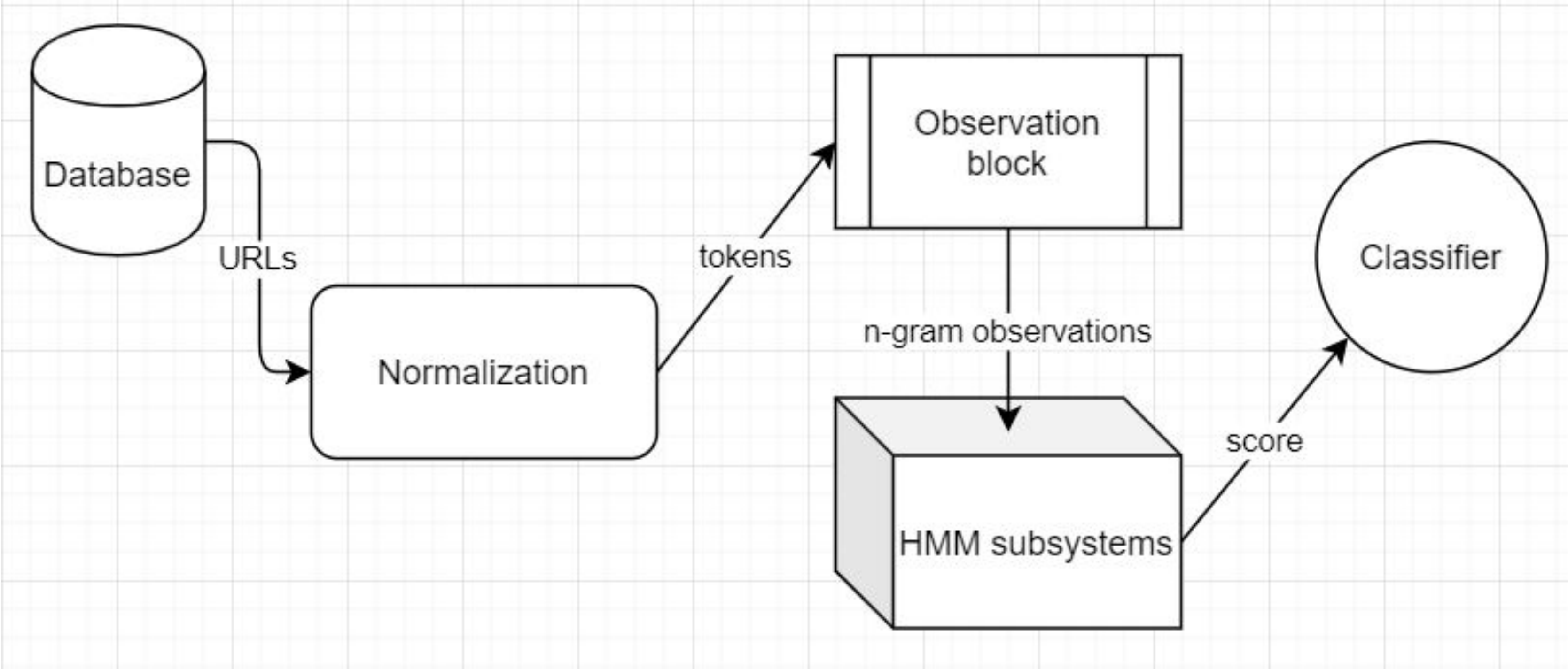
Problem 2 - Optimal hidden states sequence ?

- *The Viterbi algorithm*

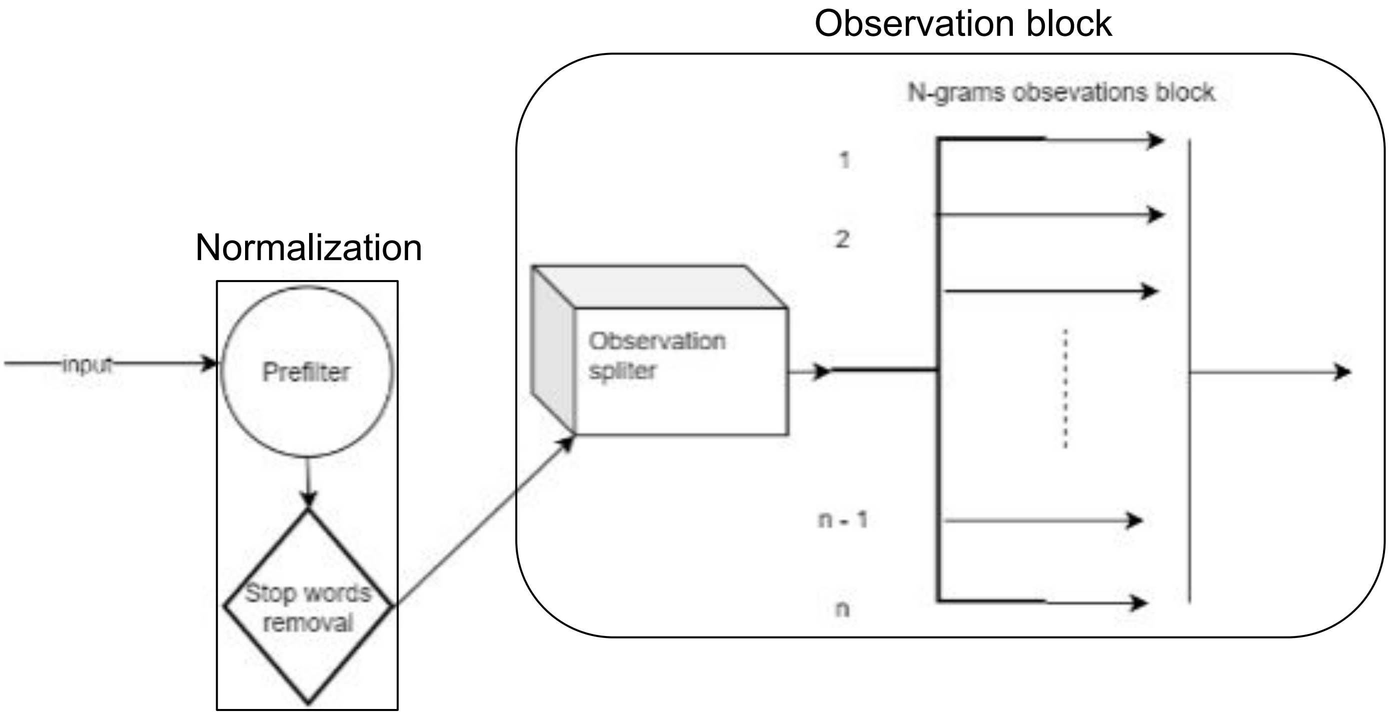
Problem 3 - How to training the model?

- *The Baum-Welch algorithm*

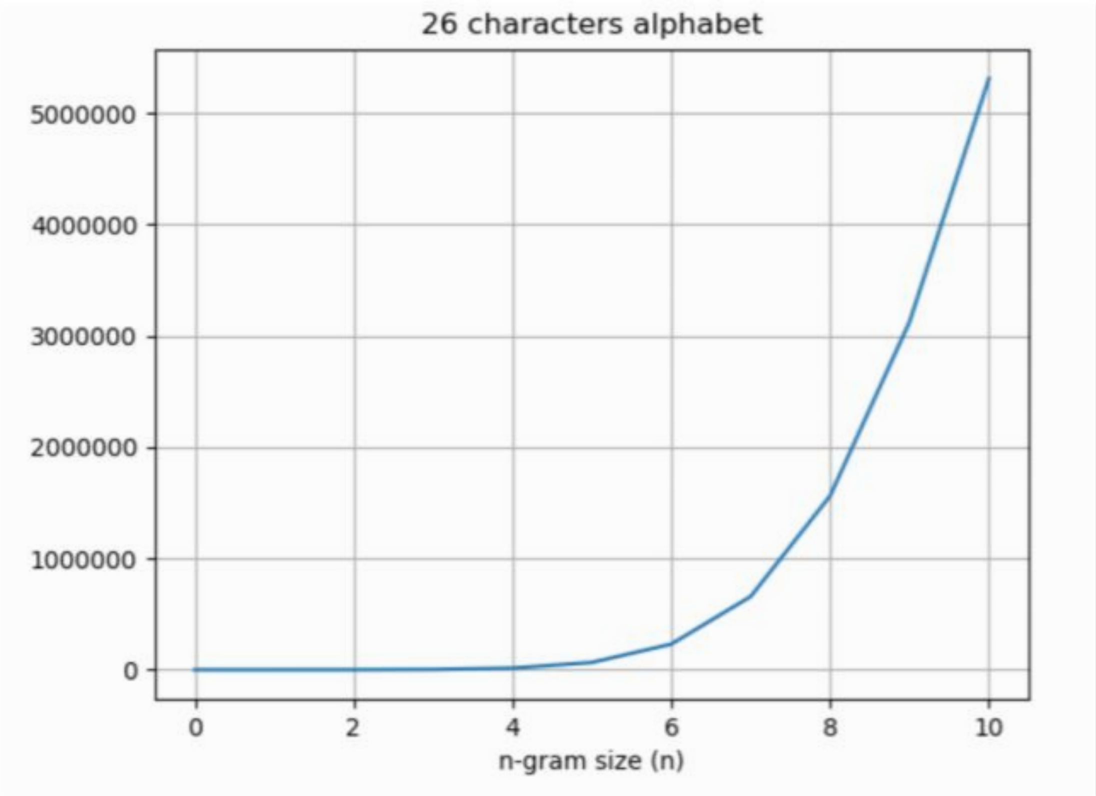
Proposed solution - System diagram



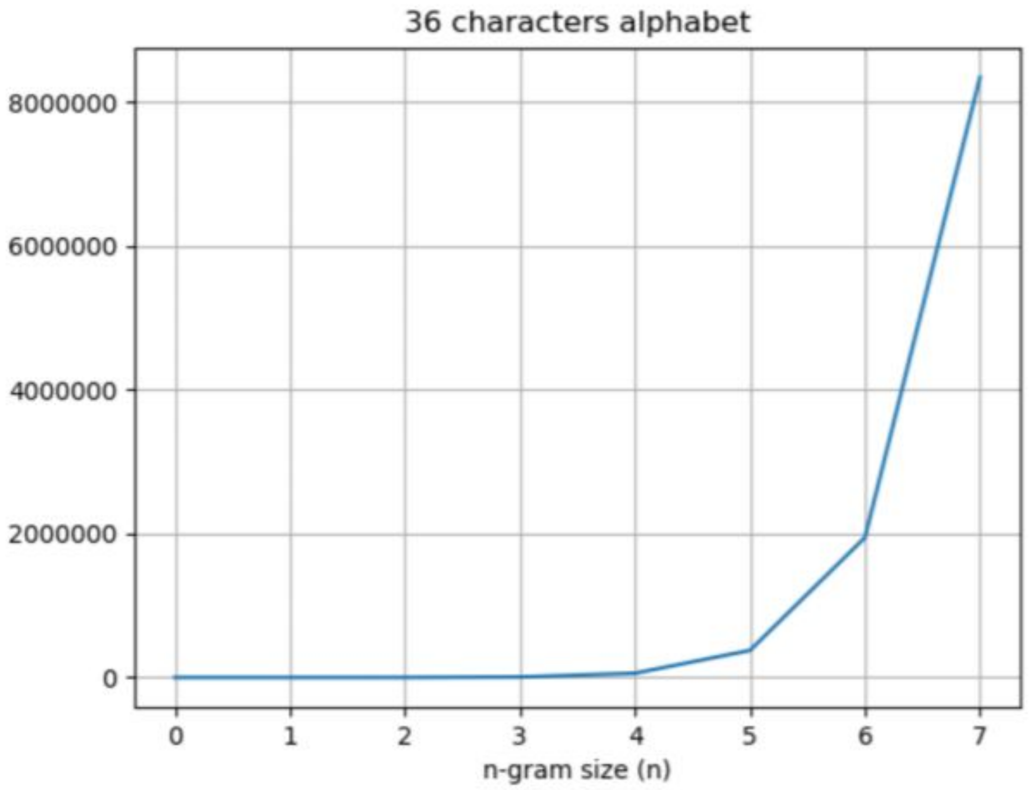
Proposed solution - Flow 1



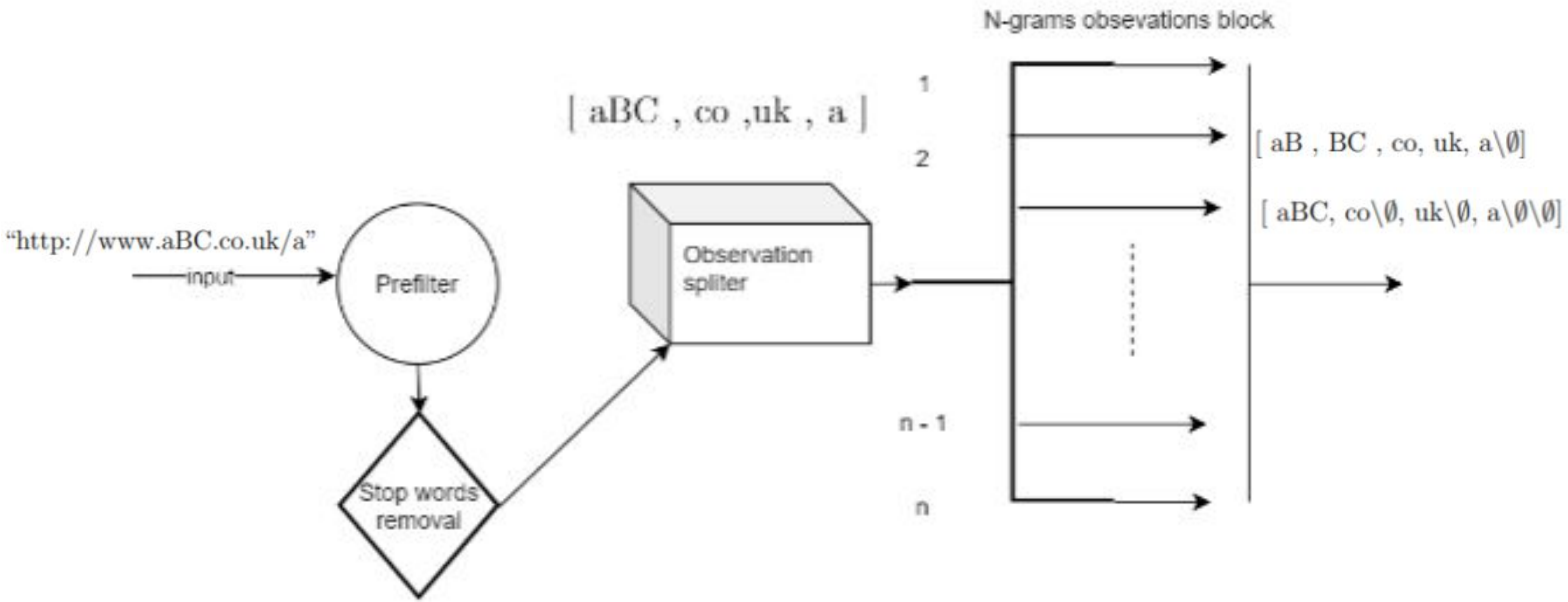
Proposed solution



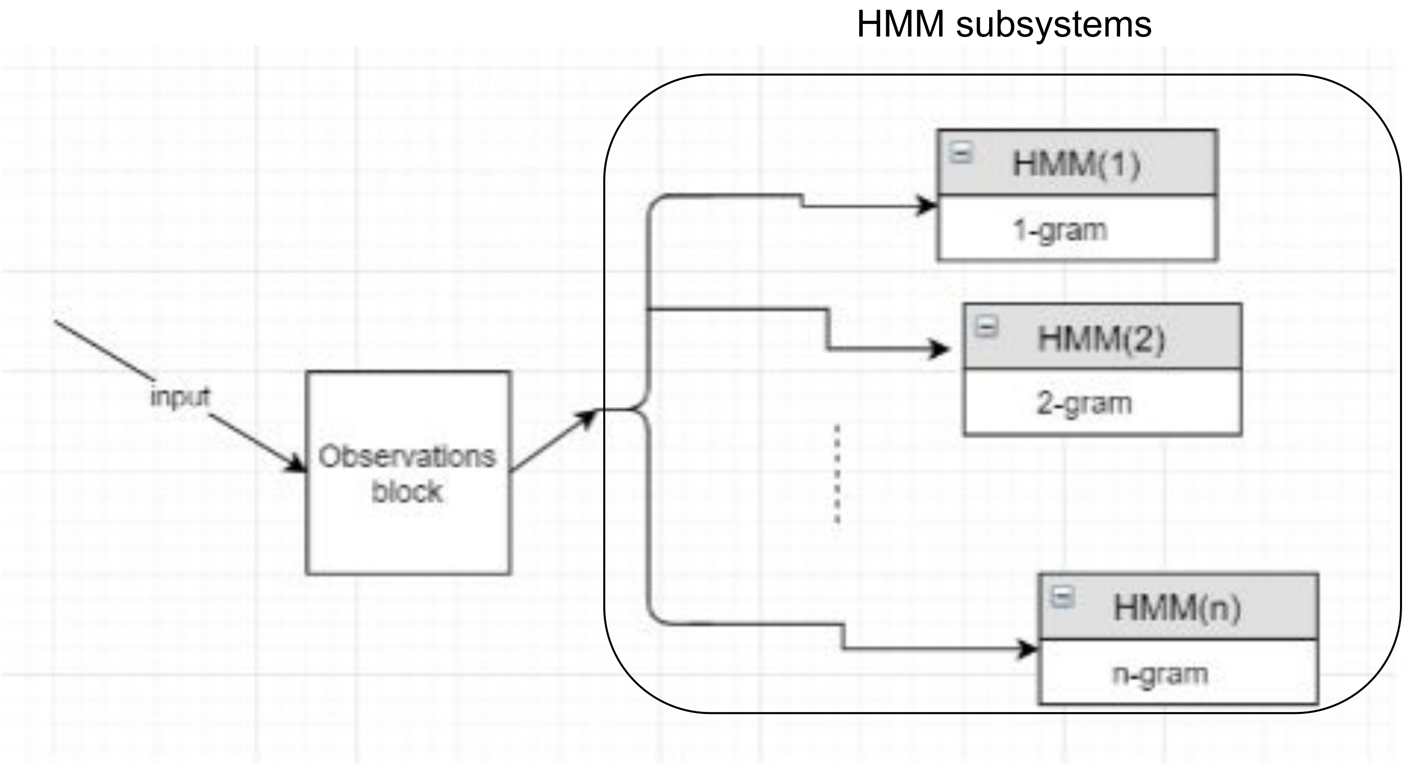
Proposed solution

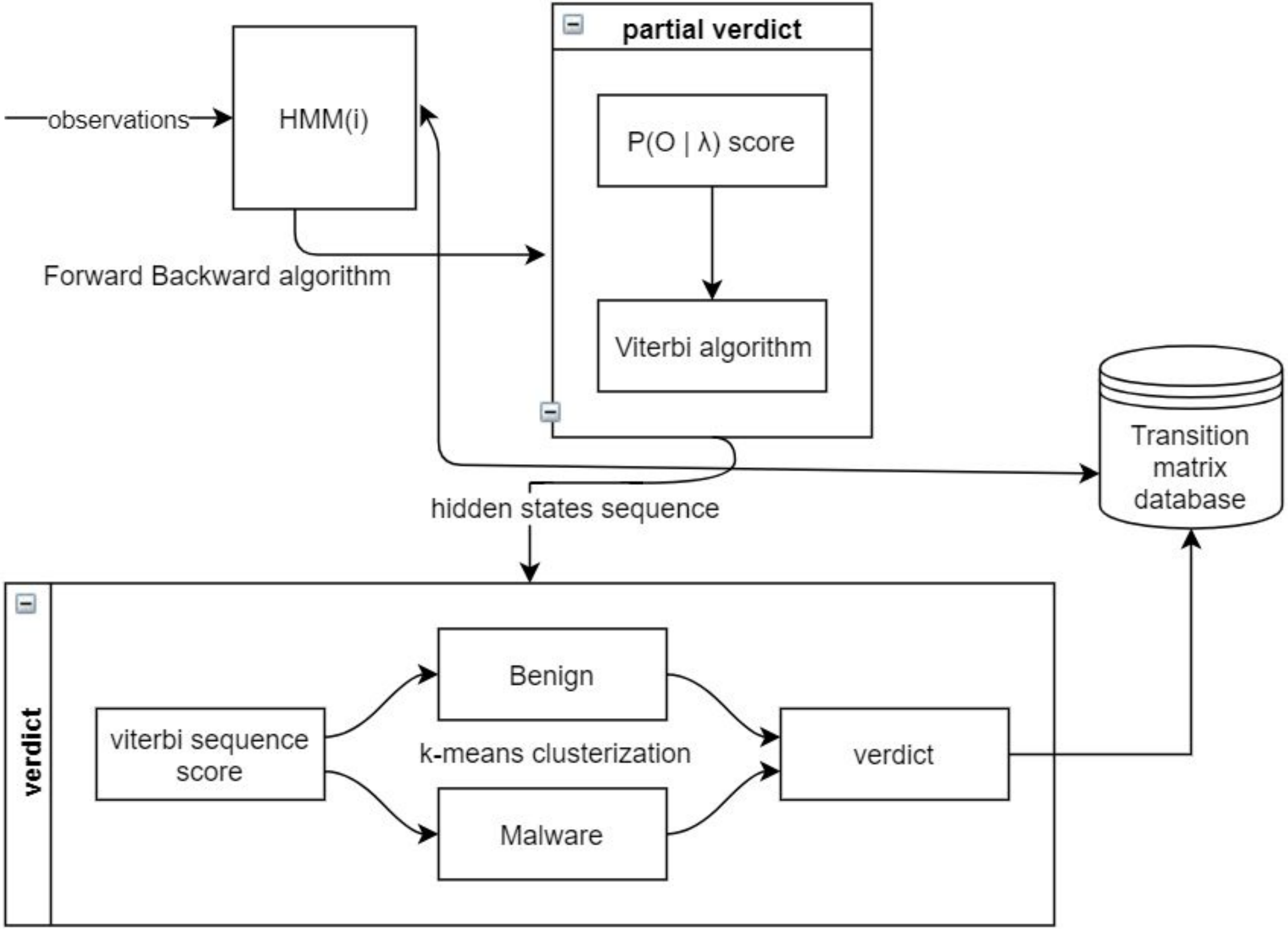


Proposed solution



Proposed solution - Flow 2




observations
HMM(i)
Forward Backward algorithm
partial verdict
P(O | λ) score
Viterbi algorithm
Transition matrix database
hidden states sequence
verdict
viterbi sequence score
Benign
k-means clusterization
Malware
verdict

Results

- ❖ Two subsytems:

  - ❑ Subsystem 1 : Frequency based static Hidden Markov model

    ➔ Subsystem 1.1 : trained with alphabet [a-z0-9\∅]*

    ➔ Subsystem 1.2 : trained with alphabet [a-zA-Z0-9\∅]*

Results

- Subsystem 1.1:

Type	Se	Sp	TN	TP	Acc	FN	FP
Unigram	70.15%	69.78%	7015	6978	70.13%	29.15%	30.22%
Bigrams	69.8%	70.70%	6980	7070	70.25%	30.2%	29.3%
Trigrams	71.78%	71.80%	7178	7180	71.79%	28.22%	28.2%

Results

- Subsystem 1.2:

Type	Se	Sp	TN	TP	Acc	FN	FP
Unigram	70%	70.87%	7000	7087	70.435%	30%	29.13%
Bigrams	71.9%	70.9%	7190	7090	71.4%	28.1%	29.1%
Trigrams	72.12%	71.21%	7212	7121	71.665%	27.88%	28.79%

Proposed solution

❖ Two subsytems:

❑ Subsystem 2 : Dynamic Hidden Markov model

➔ Subsystem 2.1 : trained with alphabet [a-z0-9\∅]*

➔ Subsystem 2.2 : trained with alphabet [a-zA-Z0-9\∅]*

Results

- Subsystem 2.1:

Type	Se	Sp	TN	TP	Acc	FN	FP
Unigram	99.82%	97.72%	9772	9982	99.72%	2.28%	0.18%
Bigrams	100%	99.4%	10000	9940	99.7%	0%	0.3%
Trigrams	88.92%	86.91%	8892	8691	89.35%	11.08%	13.09%

Results

- Subsystem 2.2:

Type	Se	Sp	TN	TP	Acc	FN	FP
Unigram	99.05%	90.90%	9905	9090	94.975%	0.95%	9.1%
Bigrams	100%	91.14%	10000	9114	95.57%	0%	8.86%
Trigrams	97.22%	93.12%	9722	9312	95.179%	2.78%	6.88%

Conclusions

Type	Se	Sp	TN	TP	Acc	FN	FP
Unigram	99.82%	97.72%	9772	9982	99.72%	2.28%	0.18%
Bigrams	100%	99.4%	10000	9940	99.7%	0%	0.3%
Trigrams	88.92%	86.91%	8892	8691	89.35%	11.08%	13.09%

- prefilter
- postfilter
- future work

Future work

- Increase n-grams size
  - ➔ n = 4,5 …
  - ➔ more unknown information
- Increase number of hidden states
  - ➔ N = 2 …10
  - ➔ suspicion score

Q & A

Thank you!