Multinomial malware classification via low-level features

Sergii Banin, Geir Olav Dyrkolbotn

18.07.2018

Long story short

- We use hardware activity produced by malware for malware classification.
- In this paper as hardware activity we take sequence of memory access operations.

Introduction

- Malware is involved in many cases of cyber crimes.
- Different malware analysis and detection techniques exist.
- Dynamic analysis helps to reveal malware functionality.
- It is impossible to avoid execution on hardware.
- Malware categorization is needed to perform appropriate defense and post-attack actions.

State of the art

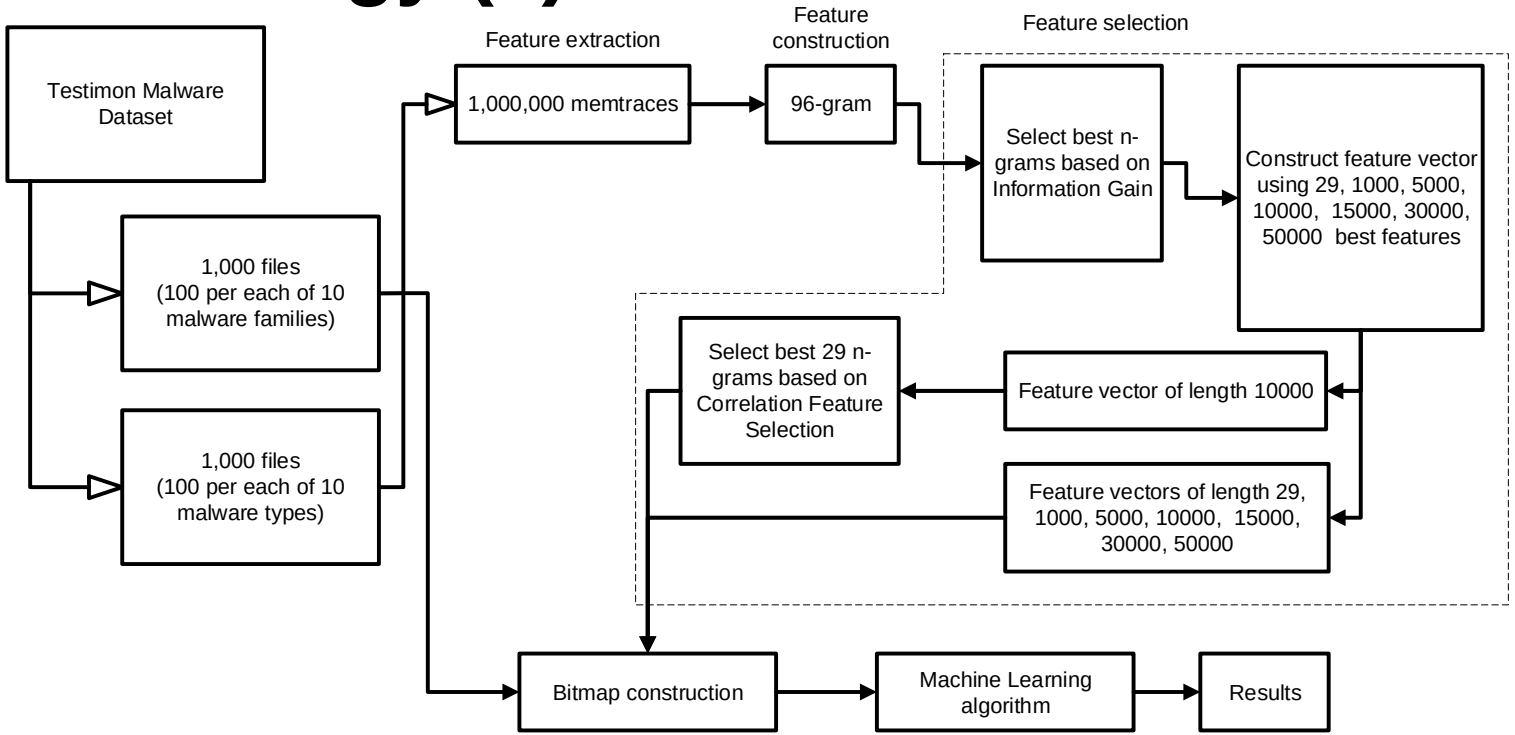
- Low-level features such as: Memory activity, opcodes, file system activity and other hardware-based features were previously used for malware detection. *
- In our previous work (Banin et al., 2016) we showed that it is possible to use memory access patterns for malware detection.

* References are those used in the original paper: (Banin et al., 2016), (Kawakoya et al., 2013), (Khasawneh et al., 2015), (Kirat et al., 2014), (Ozsoy et al., 2016).

Methodology

- Two datasets:
  - 10 malware types (~1000 files)
  - 10 malware families (~1000 files)
- Record 1M of memory access operations (Read and Write):
  - [*RWRRWRWWWR*].
- Split sequence of memory access operations into 96-grams.
- Feature selection.
- Training of ML algorithms.

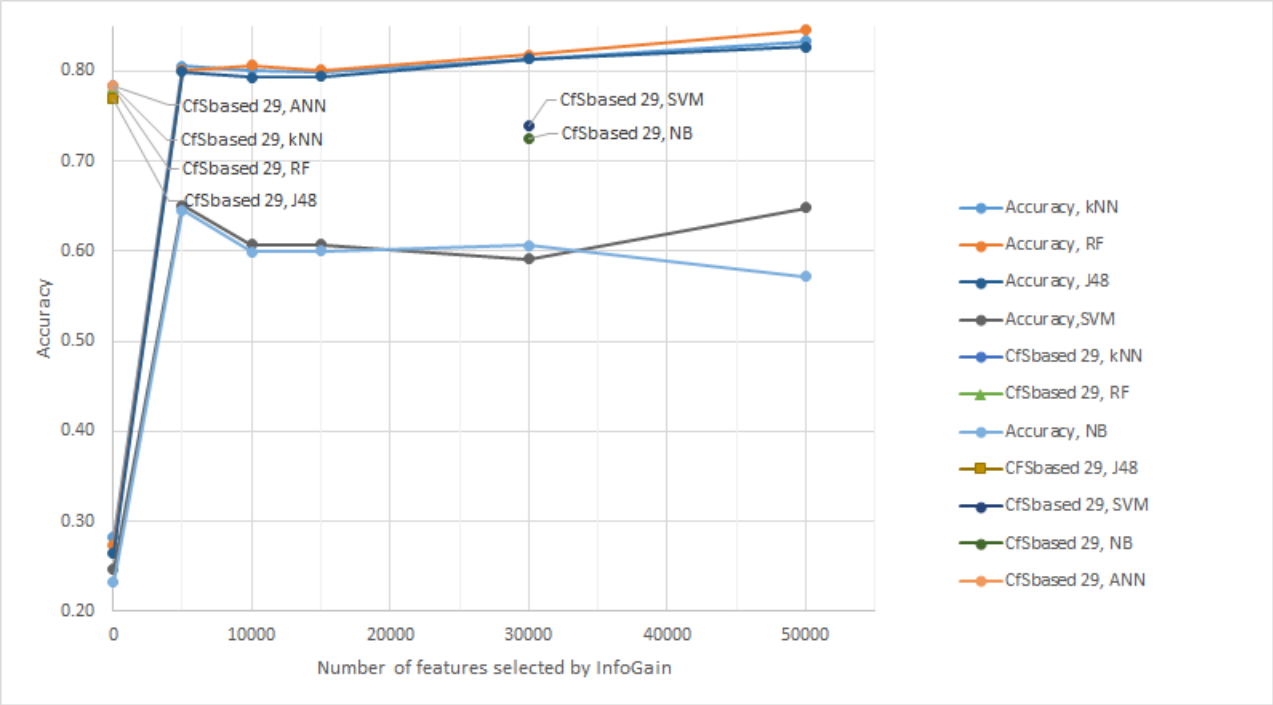
Methodology (2)



Methodology (3)

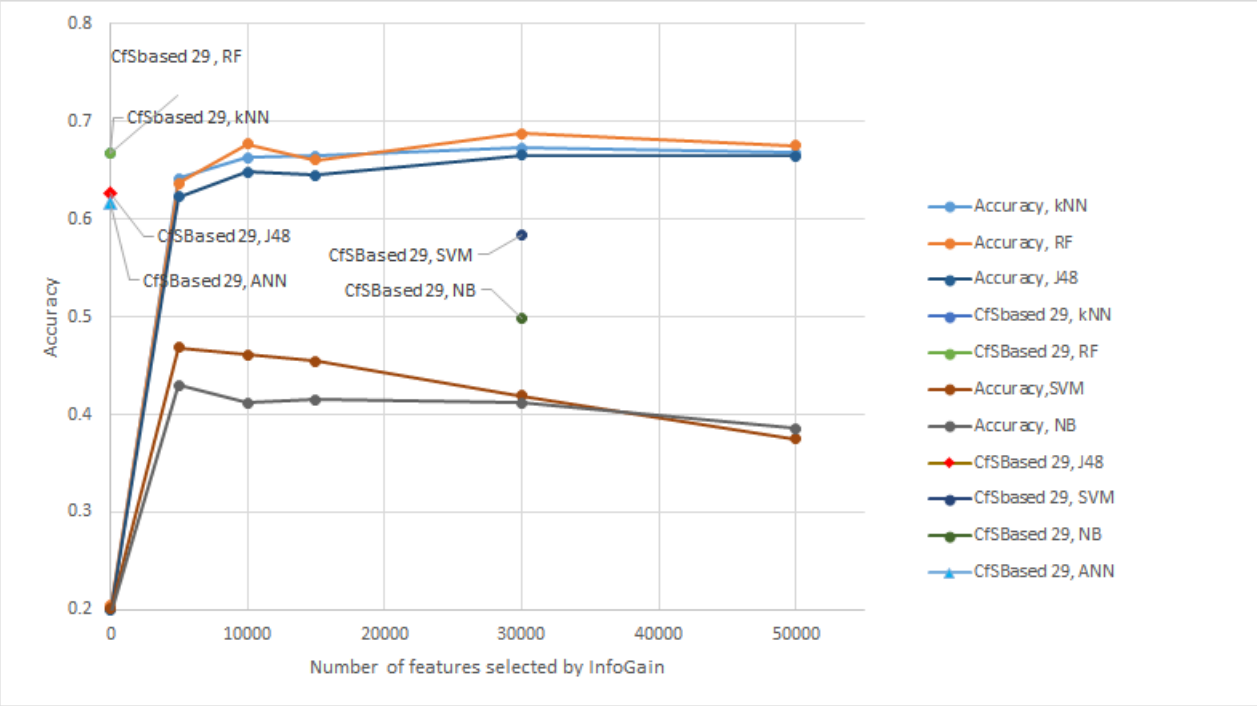
- From raw data we obtained 15M different n-grams for malware families dataset and 6M for malware types dataset.
- We used Information Gain to go down to 50K,30K,15K,10K,5K and 29 features.
- We used Correlation-based feature selection to go down from 10K to 29 features.
- Used k-Nearest Neighbors (kNN), RandomForest (RF), Decision Trees (J48), Support Vector Machines (SVM), Naïve Bayes (NB) and Artificial Neural Network (ANN) Machine Learning algorithms with 5-fold cross-validation.

Results



Accuracy for malware families dataset

Results (2)



Accuracy for malware types dataset

Results (3)

Classification performance for families and types datasets.

Number of features	Accuracy for families						Accuracy for types					
	kNN	RF	J48	SVM	NB	ANN	kNN	RF	J48	SVM	NB	ANN
29	0.282	0.274	0.265	0.246	0.232	0.271	0.201	0.204	0.2	0.201	0.198	0.206
5000	0.806	0.802	0.800	0.651	0.646	N/A	0.642	0.637	0.623	0.468	0.430	N/A
10,000	0.802	0.807	0.793	0.607	0.599	N/A	0.663	0.678	0.648	0.461	0.412	N/A
15,000	0.800	0.802	0.795	0.607	0.600	N/A	0.665	0.661	0.645	0.455	0.415	N/A
30,000	0.814	0.818	0.814	0.591	0.606	N/A	0.673	**0.688**	0.666	0.419	0.412	N/A
50,000	0.833	**0.845**	0.827	0.648	0.572	N/A	0.668	0.675	0.665	0.375	0.386	N/A
CfSbased 29 features	**0.784**	0.781	0.769	0.740	0.724	0.783	**0.668**	**0.668**	0.626	0.584	0.498	0.617

Results (4)

- More features not always provide better accuracy.
- In general, classification accuracy was higher for malware families dataset.
- Malware **families** assigned according to *particular* functionality, while malware **type** - according to *general* functionality.
- We performed additional statistical and context analysis.

Analysis

- We ran additional cross-validations to record per-category classification accuracy.
- For samples from initial categories (families and types) we used information about their subcategories (types and families respectively).
- We analyzed how per-category classification accuracy is affected by subcategories.

Analysis (2)

class	acc.	unalike.	entropy	subN
agent	0.56	0.23	2.43	8
vbinject	0.59	0.98	0.08	2
obfuscator	0.64	0.98	0.08	2
hupigon	0.69	0.88	0.34	2
vb	0.75	0.36	1.83	8
small	0.84	0.73	0.92	7
vundo	0.88	0.94	0.22	3
renos	0.91	1.00	0.00	1
onlinega.	0.99	1.00	0.00	1
zlob	0.99	0.90	0.29	2

(a) Families

class	acc.	unalike.	entropy	subN
worm	0.43	0.02	5.69	63
pws	0.54	0.06	4.50	40
trojan	0.54	0.12	4.14	37
trojandr.	0.62	0.22	3.35	26
backdoor	0.67	0.11	4.18	40
trojanspy	0.71	0.27	2.92	22
trojando.	0.74	0.27	2.75	20
virtool	0.77	0.24	2.53	15
virus	0.81	0.02	5.42	55
rogue	0.86	0.31	2.08	9

(b) Types

Analysis (3)

- In general: less diverse (in terms of statistics) category brings higher accuracy.
- But there are exceptions:
  - Family Vbinject: Low entropy – low accuracy.
  - Type Virus: High entropy – high accuracy.
- What happens: our approach is not capable of generalizing over a certain types of functionality.
- Why: to be discussed in future work.

Conclusions

- Memory access patterns can be used for malware classification.
- Achieved accuracy of 0.845 and 0.688
- It is possible to reduce a feature space by several orders of magnitude.

Future work

- Perform deeper context analysis.
- Perform misclassification analysis.
- Explain which functionality of malware creates certain memory access patterns.
- Apply our approach to more robust datasets.

Questions?

- sergii.banin@ntnu.no