최적화 특징 선택을 활용한 머신러닝 기반 랜섬웨어 분류 방법 연구

전혜민 [1], 최두섭 [2], 임을규 [2]
[1] 한양대학교 정보보안학과
[2] 한양대학교 컴퓨터·소프트웨어학과

crow0506@hanyang.ac.kr, dslab0915@hanyang.ac.kr, imeg@hanyang.ac.kr

A Study on Machine Learning-Based Ransomware Classification methods using Optimized Feature Selection

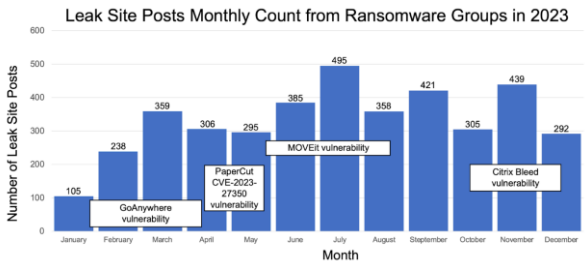
Hye-Min Jeon[1], Doo-Seop Choi[2], Eul Gyu Im[2]
[1]Dept. of Information Security, Hanyang University
[2]Dept. of Computer Science, Hanyang University

요 약

최근 랜섬웨어의 유포 증가로 인한 금전적 피해가 전세계적으로 급증하고 있다. 랜섬웨어는 사용자의 데이터를 암호화하여 금전을 요구하거나, 사용자의 중요하고 민감한 데이터를 파괴하여 사용하지 못하도록 피해를 입힌다. 이러한 피해를 막기 위해 파일의 API calls 이나, opcode 를 이용하는 탐지 및 분류 연구가 활발하게 진행되고 있다. 본 논문에서는 랜섬웨어를 효과적으로 탐지하기 위해 파일 PE 기능 값을 PCA 와 Wrapper 방법으로 데이터 전처리 후 머신러닝으로 학습하고, 학습한 모델을 활용하여 랜섬웨어를 정상과 악성으로 분류하는 방법을 제안한다. 제안한 방법으로 실험 결과 RF 는 98.25%, DT 96.25%, SVM 95%, NB 83%의 분류 정확도를 보였으며, RF 모델에서 가장 높은 분류 정확도를 달성하였다.

1. 서론

랜섬웨어(Ransomware)는 사용자의 데이터를 암호화하여 접근을 불가능하게 만들고, 복구를 위해 돈을 요구하는 악성코드이다. 공격자는 다양한 방법을 통해 공격을 시도하고, 피해자들은 암호화된 데이터를 복구하기 위해 돈을 지불해야 한다.



(그림 1) 2023 랜섬웨어 동향

사이버 보안 기업 팔로알토 네트웍스社의 Ransomware Retrospective 2024 보고서[1]에 따르면 2023 년 랜섬웨어 피해를 입은 기업이 전년 대비 49% 증가했다고 보고하였다. 최근 기존 랜섬웨어 변종 및 새로운 랜섬웨어가 등장함으로써 기업 및 사용자가 지속적인 공격 위협에 노출되어 있는 실정이다. 이렇게 다양한 공격을 탐지하기 위해 파일의 행위 분석을 통해 탐지하거나, 파일이 가지고 있는 특성 정보를 이용하여 탐지하는 방법 등 다양한 연구가 진행되고 있다. 하지만 기존 랜섬웨어 탐지 연구들은 다수의 특성 정보(feature)를 사용하기 때문에 데이터 차원의 복잡도와 증가하는 차원의 저주(Curse of dimensionality)[2] 문제로 인한 분류 정확도 성능이 감소한다. 본 논문에서는 이러한 문제를 해결하기 위해 PCA(Principal Component Analysis) [3]와 Wrapper 방법[4]을 이용해 불필요한 특징들을 제거하여 랜섬웨어를 분류하는 연구를 제안한다.

본 논문에서는 파일 PE 기능 값을 PCA 방법과 Wrapper 방법으로 데이터 전처리 후 머신러닝으로 학습하고, 학습한 모델을 활용하여 랜섬웨어를 정상과 악성으로 분류하는 방법을 제시한다.

본 논문의 구성은 다음과 같다. 2 장에서는 랜섬웨어 탐지 및 분류 관련 연구에 대해 알아보고, 3 장에서는 실험에 활용한 차원 축소 및 특징 선택 알고리즘에 대해 설명한다. 이어서 4 장에서 PCA 방법과 Wrapper 방법을 활용하여 머신러닝 기반의 분류 정확도 실험 결과를 설명하고, 마지막 5 장에서 논문 결론 및 향후 연구에 대해 논의한다.

2. 관련 연구

Nahid Ebrahimi Majd 외 1 인은 랜섬웨어의 PE 기능 값을 특징으로 사용하여 분류하는 연구를 진행하였다. PE 기능 값들이 많기 때문에 특징 선택 기법인 필터, Wrapper, 임베디드를 사용하여 중요도가 높은 특징을 선택하여 머신러닝, 딥러닝 알고리즘을 이용하여 실험을 진행하였다. 사용한 모델은 Decision Tree, Random Forest, Logistic regression, Naive Bayes, Support Vector, K-Nearest Neighbors, Extream Gradient Boosting, MLP, CNN 이고, 제일 높은 F1-score 를 보인 모델은 RF 모델이다[5].

Bin Zhang 외 5 인은 opcode 의 임베디드 N-gram 을 활용한 패치 기반 CNN 과 자기 주의 네트워크를 통한 랜섬웨어 분류 방법을 제안하였다. opcode 시퀀스를 N-gram 으로 변환한 후, 이를 다수의 패치로 분할하여 각 패치에 대해 자기 주의 기반 CNN 을 사용하여 실험을 진행하였다. 실험 모델은 SA-CNN, NB, DT, KNN 을 사용하였고, 각 모델의 정확도는 87.6% ,57.3%, 85.4%, 85%를 보였다[6].

Mohammad Masum 외 5 인은 랜섬웨어 탐지 및 분류를 위한 머신러닝 알고리즘을 적용한 특성 선택 기반 프레임워크를 제안하였다. Z -점수 표준화 기법을 사용하여 변수를 비슷한 척도로 변환하고, 변동성 임계값 및 분산 팽창 계수(VIF)를 이용한 특성 선택 방법을 적용하여 낮은 변동성 특성과 높은 상관 관계를 가진 특성을 제거하여 실험을 진행하였다. 사용한 모델은 DT, RF, NB, LR, NN 이며 각 모델의 정확도는 98%, 99%, 35%, 96%, 97%를 보였다[7].

Houria Madani 외 3 인은 다양한 신경망 모델을 사용하여 랜섬웨어를 탐지하고 분류하는 방법을 제안하였다. 손상되거나 난독화된 파일의 정확한 라벨링 및 제외를 보장하기 위한 전처리 후 다양한 신경망 모델을 사용해 실험을 진행하였다. 실험 모델은 인공 신경망(ANN), 합성곱 신경망(CNN), 순환 신경망(RNN)을 사용하였고, 각 모델의 정확도는 91%, 94%, 79%를 보였다[8].

3. 차원 축소 및 특징 선택 알고리즘

3.1. PCA (Principal Component Analysis)

주성분 분석(PCA)는 상관관계가 있을 수 있는 변수 세트를 주성분이라고 하는 더 적은 수의 상관관계가 없는 변수로 변환하는 차원 축소 기술이다. PCA 는 필수적으로 필요한 정보를 유지하면서 데이터 세트의 기능 수를 줄여 학습 알고리즘의 성능을 향상시킬 수 있으며, 데이터의 패턴을 식별하고, 변화를 강조하여 새로운 패턴을 탐지할 수 있다. PCA 를 이용하여 데이터셋이 복잡한 다차원 구조를 나타내는 경우가 많은 랜섬웨어 탐지에 유용하고, 계산 효율성이 향상된다.

3.2. Wrapper method

Wrapper 방법은 특정 예측 모델의 성능을 기반으로 기능을 기능의 하위 집합을 평가한다. 그리고 이 접근 방식을 사용 선택하는 기능 선택 기술이다. 기능의 본질적인 속성에 의존하는 필터 방법과 달리 이 방법은 실제로 모델을 훈련하여 하면 Wrapper 방법을 계산하는데 오버헤드가 있지만 선택 프로세스가 모델에 맞게 조정되므로 성능이 좋아질 수 있다.

.

3.3. 최종 선택된 랜섬웨어 특성 정보

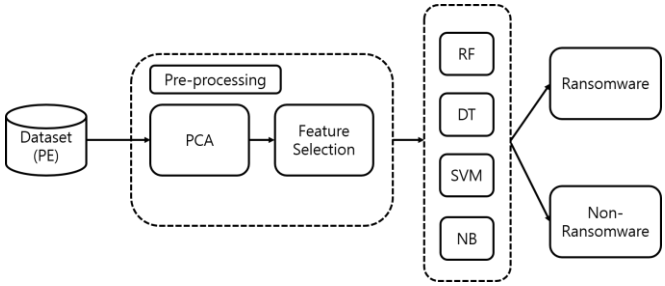
기존 랜섬웨어 분류 실험에는 총 57 개의 특징을 사용하는데 3 장에서 설명한 PCA 와 Wrapper 방법을 통해 [표 1]과 같이 총 21 개의 특징을 최종 선택하여 실험에 사용하였다.

<표 1> 최종 선택된 랜섬웨어 특성

Num.	Feature Name	Importance
1	ImageBase	0.253449
2	VersionInformationSize	0.148055
3	ResourcesMinSize	0.086986
4	SectionsMaxEntropy	0.079506
5	MajorLinkerVersion	0.070027
6	SectionsNb	0.036329
7	ResourcesMinEntropy	0.036272
8	MajorOperatingSystemVersion	0.035503
9	SectionsMinEntropy	0.033402
10	SizeOfStackReserve	0.028395
11	ResourcesMaxEntropy	0.024011
12	Characteristics	0.023735
13	SectionsMinRawsize	0.022685
14	CheckSum	0.019320
15	MajorSubsystemVersion	0.018882
16	MinorImageVersion	0.015633
17	DllCharacteristics	0.015223
18	ResourcesMaxSize	0.014940
19	Subsystem	0.013936
20	SizeOfImage	0.012687
21	SectionsMinVirtualsize	0.011023

상위 5개의 특징이 각각 무엇인지 설명하면 ImageBases는 실행 가능한 이미지가 불러와지는 메모리의 기본 주소이다. VersioninformationSize는 실행 파일의 버전 정보 크기를 의미한다. ResourcesMinSize는 PE 파일 내 리소스 섹션의 가장 작은 크기를 의미한다. SectionMaxEntropy는 PE 파일의 엔트로피를 의미한다. MajorLinkerVersion은 실행 파일을 생성하는 데 사용된 링커의 주요 버전 번호를 의미한다.

4. 랜섬웨어 분류 방법론



(그림 2) 랜섬웨어 분류 방법론

4.1. 데이터 전처리

머신러닝 모델을 학습하기에 앞서, 주어진 데이터를 학습하기에 적절한 형식으로 전처리하였다. 먼저 PCA 알고리즘을 통해 데이터셋의 변수를 주성분으로 알려진 더 작은 상관 관계가 없는 변수 세트로 변환하여 중요한 분산을 유지하면서 차원성을 줄인다. PCA 감소에 이어 RFE(Recursive Feature Elimination)[9]를 사용하는 Wrapper 방법을 적용하여 기능 세트를 개선하였다.

Wrapper 방법은 모델 교육 및 성능 평가의 반복 프로세스를 통해 기능의 하위 집합을 평가함으로써 모델 정확도를 최대화하는 하위 집합을 식별한다. 이 접근 방식은 기능 선택을 위한 적응형 메커니즘을 제공하여 데이터셋이 랜섬웨어 분류 실험에 효율적인 성능을 내도록 도와준다.

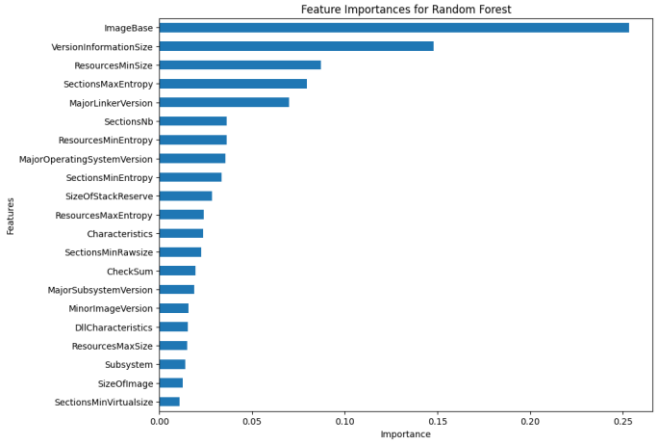
4.2. 실험 환경

실험은 Anaconda 의 Jupyterlab 에서 진행하였으며[10], 데이터셋은 랜섬웨어 PE 특징으로 구성되어 있다[11]. 데이터셋의 분포는 정상 2,000 개, 악성 2,000 개로 구성되어 있다. 표 1 은 실험을 진행한 환경 및 소프트웨어 버전을 보여준다.

<표 2> 실험 환경

OS	Windows 10
CPU	Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz
Memory	16GB
Jupyterlab	3.5.3
Python	3.8.16
Anaconda	4.12.0

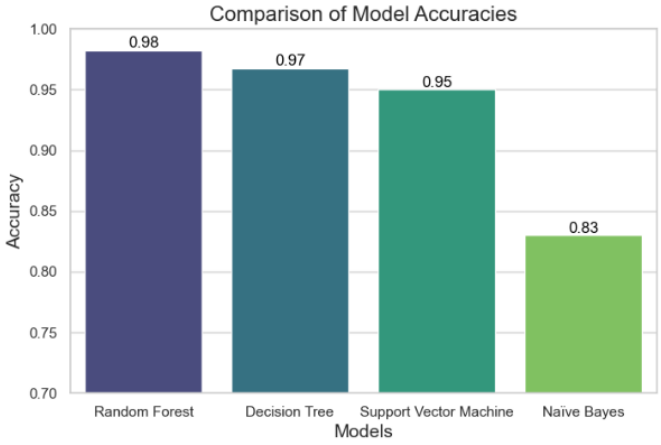
4.3. 실험 결과



(그림 3) PE 기능 값 중요도

[그림 3]은 PCA 와 Wrapper 방법을 사용하여 랜섬웨어를 분류할 때 사용한 특징 그래프이다. 중요도는 각 특징이 머신러닝 모델의 정확도를 향상하고 불확실성을 줄이는데 얼마나 영향을 미치는지에 따라 계산한다. X 축은 정상, 악성으로 분류할 때 사용한 특징들의 중요도를 의미하고, Y 축은 실험에서 사용한 특징이다.

특성 중요도 값이 높을수록 분류 결과에 더 큰 영향을 미치는 것과 직접적인 관계가 있으며, 이는 랜섬웨어를 정확하게 분류하는데 중요하다는 것을 나타낸다.



(그림 4) 실험 모델 정확도 비교

<표 4> 정상, 악성 분류 정확도

Model	RF	DT	SVM	NB
Accuracy	**98.25%**	**96.75%**	**95%**	**83%**
F1-score	**97.68%**	**95.39%**	**93.37%**	**81.29%**
Precision	**98.13%**	**96.85%**	**95.89%**	**83.28%**
Recall	**97.88%**	**95.16%**	**93.44%**	**81.75%**

실험에서 사용한 모델로는 Random Forest, Decision Tree, Support Vector Machine, Naïve Bayes 을 사용하여 정상과 악성을 분류하는 실험을 진행하였다.

논문에서 제안한 데이터 전처리 방법 및 실험 결과 RF 모델은 98.25%, DT 96.25%, SVM 95%, NB 83%의 정확도를 보였으며, RF 모델이 가장 높은 정확도를 보였다.

5. 결론 및 향후 연구

본 논문에서는 파일 PE 기능 값을 PCA 와 Wrapper 방법으로 데이터 전처리 후 머신러닝으로 학습하고, 학습한 모델을 활용하여 랜섬웨어를 정상과 악성으로 분류하는 방법을 제시하였다. 실험 결과 RF 는 98.25%, DT 96.25%, SVM 95%, NB 83%의 탐지 정확도를 보였으며, RF 모델에서 가장 우수한 성능을 보였다.

향후 연구에서는 랜섬웨어의 정확한 탐지 및 분류를 위해 패킹 여부를 확인하여 패킹되지 않은 데이터셋을 중심으로 추가 실험을 진행할 계획이다. 또한, 본 연구에서는 총 4,000 개의 데이터셋을 실험에 사용하였는데, 일반화 성능 및 분류 정확도를 향상시키기 위해 랜섬웨어 데이터를 추가로 수집하여 실험을 진행할 예정이다.

6. Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2022R1A4A1032361)

참고문헌

[1] PaloAltoNetworks, Unit42. “Ransomware Retrospective 2024: Unit 42 Leak Site Analysis”. https://unit42.paloaltonetworks.com/unit-42-ransomware-leak-site-data-analysis/

[2] Köppen, Mario. "The curse of dimensionality." In proceedings of the *5th online world conference on soft computing in industrial applications (WSC5)*. Vol. 1. 2000.

[3] Maćkiewicz, Andrzej, and Waldemar Ratajczak. "Principal components analysis (PCA)." *Computers & Geosciences* 19.3 (1993): 303-342.

[4] Kohavi, Ron, and George H. John. "Wrappers for feature subset selection." Artificial intelligence 97.1-2 (1997): 273-324

[5] Majd, Nahid Ebrahimi, and Torsha Mazumdar. "Ransomware Classification Using Machine Learning." In proceedings of the *2023 32nd International Conference on Computer Communications and Networks (ICCCN)*. IEEE, 2023

[6] Zhang, Bin, et al. "Ransomware classification using patch-based CNN and self-attention network on embedded N-grams of opcodes." *Future Generation Computer Systems* 110 (2020): 708-720.

[7] Masum, Mohammad, et al. "Ransomware classification and detection with machine learning algorithms." In proceedings of the *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022.

[8] Madani, Houria, et al. "Classification of ransomware using different types of neural networks." *Scientific Reports* 12.1 (2022): 4770.

[9] Guyon, Isabelle, et al. "Gene selection for cancer classification using support vector machines." *Machine learning* 46 (2002): 389-422.

[10] Anaconda, Available: https://www.anaconda.com/

[11] RansomwareDetection, Available: https://github.com/mudimathur2020/RansomwareDetection