Linux OS 대상 랜섬웨어 탐지 기법 동향

최재민*, 최상훈[1], 박기웅†

세종대학교 SysCore Lab. (대학원생, 연구교수[1])
**세종대학교 정보보호학과 (교수†)

Trends in Ransomware Detection Techniques Targeting Linux OS

Jae-Min Choi*, Sang-Hoon Choi[1], Ki-Woong Park†

* SysCore Lab., Sejong University
**Dept. of Computer and Information Security, Sejong University
(Graduate Student*, Research Professor[1], Professor†)

요 약

클라우드 시스템의 발전으로 클라우드 컴퓨팅 제공 서비스도 크게 발전하였다. 이에따라 많은 기업이 시스템을 클라우드 환경으로 전환하고 있다. 하지만, 클라우드 사용자의 증가에 따라 대부분의 클라우드 시스템에서 사용하고 있는 Linux OS 대상 보안 위협 또한 증가하고 있다. 특히 Windows OS가 주 타겟이었던 랜섬웨어 공격과 그룹이 최근에는 Linux OS를 대상으로 크게 증가하고 있다.이러한 공격을 탐지하기 위한 여러 가지 연구들이 진행되고 있다. 본 논문에서는 이러한 Linux OS 대상 랜섬웨어를 여러 가지 정적, 동적 특징을 이용한 탐지 연구에 대하여 조사하여 분류하고, 이를 기반으로 향후 연구에서 여러 특징을 이용한 분류 연구를 수행해보고자 한다.

I. 서론

최근 클라우드 시스템의 발전으로 AWS와 같은 클라우드 컴퓨팅 제공 서비스도 크게 발전하였다. 이에 따라 많은 기업이 기업 내 시스템을 클라우드 환경으로 전환하고 있다. 클라우드 시스템이 증가함에 따라 대부분의 클라우드 시스템에서 사용하고 있는 Linux OS의 점유율 또한 증가하고있다 [1].

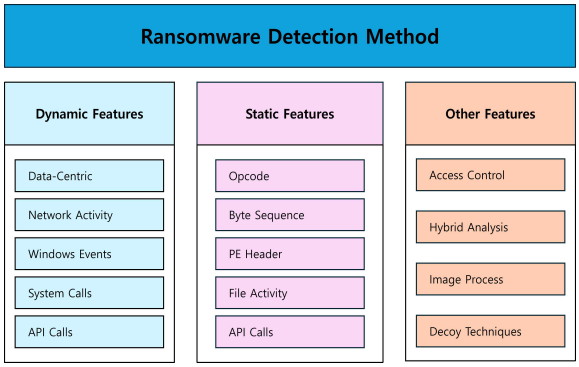
점유율이 증가함에 따라 Linux OS를 대상으로한 위협이 발생하고 있다. 특히 Window OS가 주 타겟이었던 랜섬웨어 공격이 Linux OS를 대상으로 제작되어 배포되고있다 [2]. Kaspersky 보고서에 따르면 2023년과 비교하여 2024년 1분기에는 Linux OS 사용자를 대상으로 공격이 약 130% 이상 증가하였다고 발표하였다 [3].

이러한 Linux OS에서의 랜섬웨어 공격을 대응하기위해 랜섬웨어 탐지 기술에 대한 연구가 진행되고 있다. (그림 1)과 같이 정적 특징을 이용하는 연구와 동적 특징을 이용하는 연구가 진행되고 있다. 이외에도 정적, 동적 특징을 합성하여 탐지하는 하이브리드 탐지 기술, 랜섬웨어의 암호화 순서를 이용한 미끼 탐지 기술, 접근 권한 상승 요청을 기반으로한 탐지 기술등에 대한 연구가 있다.

이러한 Linux OS 대상 랜섬웨어 탐지를 위한 연구들을 조사하여 정적, 동적 특징을 대상으로 분류하여 Linux OS 대상 랜섬웨어 탐지 기술의 동향을 분석하고자 한다.본 논문의 구성은 다음과 같다. 2장에서는 정적 특징을 활용한 연구에 대해 서술하고, 3장에서는 동적 특징을 활용한 연구에 대해 서술한다. 4장에서는 조사한 연구들의 한계점에 대해 서술하고, 5장에서는 결론 및 향후 연구를 서술한다.

† 교신저자: 박기웅 (세종대학교 정보보호학과 교수)

본 논문은 과학기술정보통신부의 재원으로 정보통신기획평가원(IITP)의 정보통신방송기술 국제공동연구(Project No. RS-2022-00165794, 50%), 정보통신방송혁신인재양성사업(Project No.2021-0-01816, 20%), 한국연구재단(NRF) 중견후속연구사업(Project No. RS-2023-00208460, 30%)의 지원을 받아 수행된 연구임.



(그림 1) Ransomware Detection Method by Features

II. Static Features를 활용한 연구

2.1 Binary Opcode 분석을 통한 탐지

Blue Eton 외 4인은 Linux OS 대상 랜섬웨어를 RNN 순환 신경망을 활용하여 (그림1)의 Static Features 중 하나인 Binary Opcode 순서 정보를 분석하여 기존의 시그니쳐 기반 탐지를 회피하는 랜섬웨어를 보다 효과적으로 탐지할 수 있는 접근 방식을 제안하였다 [4]. 해당 연구에서는 Binary Opcode 순서 정보를 추출하기위해 각 샘플 실행파일을 디스어셈블 후 명령어를 Opcode에 매핑하였다. 이후 해당 Opcode 순서 정보를 길이에 맞게 축소와 패딩을 하여 학습 데이터로 활용하였다. 이후 LSTM(Long Short-Term Memory) 3중 레이어를 사용하여 RNN 순환 신경망을 구축하였다. 해당 레이어를 통하여 80/10/10 분할을 활용하여 학습을 진행하였다. 이때, 랜섬웨어 샘플은 샘플 개수가 적기 때문에, 손실함수에 클래스 가중치를 적용하였다. 최종적으로 학습한 모델을 이용하여 Linux OS 대상 랜섬웨어를 F1-Score 약 92.7%로 Binary Opcode를 활용하여 정적 데이터를 토대로 효과적으로 탐지하였다.

2.2 블록 엔트로피 분석을 통한 탐지

Seungkwang Lee 외 4인은 FUSE를 사용하여 암호화 작업을 블록 레벨 모니터링을 통해 Linux OS 대상 랜섬웨어를 탐지하는 블록 레벨 모니터링 시스템을 제안하였다 [5]. 해당 연구에서는 (그림1)의 Static Features 중 하나인 Opcode를 활용하여 엔트로피를 추출한 뒤 탐지를 진행한다. 해당 연구에서는 랜섬웨어의 블록 레벨 모니터링 시스템을 위해 커널 코드를 수정하지 않고 휴리스틱 규칙을 찾기위해 FUSE(Filesystem in Userspace)를 활용한 'Rcryptect'를 정의한다. 우선, 샘플 파일의 각 블록에 대한 주파수 테스트 결과를 분석한 뒤 이를 고 엔트로피 블록과 저 엔트로피 블록 2개의 집합으로 나타낸다. 이후 엔트로피를 기반으로 휴리스틱 규칙을 작성한다. 해당 휴리스틱 규칙은 다음과 같다. "연속된 n개의 블록 중 90% 이상이 높은 엔트로피를 갖는다면 해당 블록은 암호화 함수의 결과일 가능성이 높다." 해당 규칙을 기반으로 n이 고 엔트로피 블록, m 이 연속된 블록 개수 일 때 (n,40) 쌍으로 쌍을 생성하였다 이는 실험 결과 최적의 오탐/미탐을 제거 할 수 있는 쌍이다. 이후 FUSE를 활용한 Rcryptect 파일시스템에서 약 13%의 오버헤드로 동작중인 랜섬웨어를 탐지하였다.

III. Dynamic Features를 활용한 연구

3.1 HPCs, System Call, Network Traffic 분석을 통한 탐지

Archit Gajjar 외 7인은 HPCs(Hardware Performance Counters), System Call, Network Traffic을 분석하여 동시에 활용함으로써 기존의 FPGA 기반 ML 모델보다 효과적으로 탐지할 수 있는 XGBoost 가속기를 제안하였다 [6]. 해당 연구에서는 (그림1)의 Dynamic Features에 해당하는 HPCs, System Call, Network Traffic을 분석하여 동시에 학습에 활용함으로써 오탐과 미탐을 줄였다. 해당 데이터는 각각 백그라운드에서 실행 중인 정상, 랜섬웨어 HPC 데이터를 수집하고 이를 각 HPC 그룹에 매칭한다. System Call 데이터 특징은 기존의 연구 [7]에서 추출하였고, Network Traffic은 온라인 사기를 예측하기 위한 오픈 소스 데이터셋과 4Paradigm의 데이터를 수집하였다. 이후 해당 데이터를 이용하여

70/30 분할을 하여 학습을 진행하였다. 학습 시 기존 XGBoost 모델에서 하이퍼파라미터 튜닝을 진행하였다. 최종적으로 학습한 모델을 이용하여 Linux OS 대상으로 약 0.95의 정확도를 달성하였고, 1000개의 샘플에서 CPU 대비 약 65.8배, GPU 대비 약 4.1배의 속도 향상을 보여주었다.

3.2 IRP 분석을 통한 탐지

Soyar Hargreaves 외 4인은 IRP(Input/Output Request Packet)을 분석하여 랜섬웨어 특유의 행동 패턴을 추출한다. 이후 해당 데이터를 이용하여 Random Forest 모델을 통하여 Linux OS 대상 랜섬웨어를 탐지하는 방법을 제인하였다 [8]. 해당 연구에서는 (그림1)의 Dynamic Features에 해당하는 IRP 데이터를 정상적인 파일 IO, 문서 편집, 시스템 업데이트 와 랜섬웨어의 파일 암호화, 대량 파일 접근, 공격적인 파일 쓰기를 캡처하였다. 이때, 데이터를 관련 없는 작업들을 제거하고 시간적 관계 파악을 위한 시계열 정렬 처리를 진행하였다. 이후 해당 데이터의 특징을 추출하여 특징 벡터로 사용하였다. 파이썬 Scikit-learn 라이브러리를 사용하여 Random Forest 모델을 구현하였고, 해당 모델의 하이퍼파라미터 튜닝을 진행하였다. 위의 데이터를 학습하였고 교차 검증 기법을 통해 검증을 진행하였다. 최종적으로 해당 모델은 약 94.7%의 정확도로 Linux OS 대상 랜섬웨어를 탐지하였다.

IV. 기존 연구의 한계점

Linux OS 대상 랜섬웨어 탐지기법에 대한 연구에 대하여 조사를 수행함에 있어 대부분의 랜섬웨어 탐지 기법 연구는 동적 특징을 이용한 연구들이었다. 이는 더 이상 정적 특징만으로는 암호화와 난독화된 랜섬웨어를 탐지해내기 어려움을 의미한다. 즉, 암호화와 난독화된 랜섬웨어를 보다 효과적으로 탐지하기 위한 동적 특징을 이용한 탐지 기법이 주로 연구되고 있다. 하지만, 동적 특징은 자동화 추출에 어려움이 있다. 이는 안티 디버깅과 구동 환경에 문제가 발생하여 랜섬웨어가 실행이 안되는 경우가 존재한다. 동일한 Linux OS임에도 불구하고 버전에 따라 사용하는 패키지, 라이브러리가 모두 달라지는 Linux OS의 특성으로 인하여 이러한 문제점이 발생한다.

V. 결론 및 향후 연구

본 논문에서는 Linux OS 대상 랜섬웨어를 탐지하기 위한 연구들을 조사하여 정적, 동적 특징을 이용하는 2가지의 분류로 나누어 조사하였다. 정적 특징과 동적 특징을 활용하는 두 개의 탐지 기법 모두 각각의 한계점이 있고 두 연구는 서로 상호보완이 가능하다. 따라서, 앞으로는 정적 특징과 동적 특징을 융합한 하이브리드 탐지 기법에 대한 연구가 필요하다. 향후 연구에서 는 Linux OS 대상 랜섬웨어 탐지 기법에서의 특징들을 기반으로 분류를 진행하고자 한다.

[참고문헌]

[1] Statista : Market share held by the leading computer (desktop/tablet/console) operating systems worldwide from January 2012 to August 2024

[2] Kaspersky : Attacks on virtualization systems and Linux servers

[3] Kaspersky : Cybercriminals double exploitation of Linux vulnerabilities

[4] Blue, Eton, et al. “Ransomware Detection on Linux Operating System Using Recurrent Neural Networks with Binary Opcode Analysis.” OSF Preprints, 23 Sept. 2024.

[5] Seungkwang Lee, Nam-su Jho, Doyoung Chung, Yousung Kang, Myungchul Kim, “Rcryptect: Real-time detection of cryptographic function in the user-space filesystem,”, Computers & Security, vol. 112, pp. 102512, January, 2022.

[6] Gajjar, Archit and Kashyap, Priyank and Aysu, Aydin and Franzon, Paul and Choi, Yongjin and Cheng, Chris and Pedretti, Giacomo and Ignowski, Jim, “RD-FAXID: Ransomware Detection with FPGA-Accelerated XGBoost”, 2024 ACM Trans. Reconfigurable Technol. Syst., Aug, 2024.

[7] S.H. Kok, Azween Abdullah, NZ Jhanjhi, “Early detection of crypto-ransomware using pre-encryption detection algorithm.” Journal of King Saud University-Computer and Information Sciences, vol. 34, pp. 1984-1999, 2022.

[8] Soyar Hargreaves, Rosalind Montalvo, Leopold Santana, et al. “Ransomware Detection in Linux File Systems Using Random Forests on IRP Data.”, Authorea. October, 2024.