https://doi.org/10.33778/kcsa.2023.23.5.065

인간의 습관적 특성을 고려한 악성 도메인 탐지 모델 구축 사례: LSTM 기반 Deep Learning 모델 중심★

정 주 원*

요 약

본 논문에서는 LSTM(Long Short-Term Memory)을 기반으로 하는 Deep Learning 모델을 구축하여 인간의 습관적 특성을 고려한 악성 도메인 탐지 방법을 제시한다. DGA(Domain Generation Algorithm) 악성 도메인은 인간의 습관적인 실수를 악용하여 심각한 보안 위협을 초래한다. 타이포스쿼팅을 통한 악성 도메인의 변화와 은폐 기술에 신속히 대응하고, 정확하게 탐지하여 보안 위협을 최소화하는 것이 목표이다. LSTM 기반 Deep Learning 모델은 악성코드별 특징을 분석하고 학습하여, 생성된 도메인을 악성 또는 양성으로 자동 분류한다. ROC 곡선과 AUC 정확도를 기준으로 모델의 성능 평가 결과, 99.21% 이상 뛰어난 탐지 정확도를 나타냈다. 이 모델을 활용하여 악성 도메인을 실시간 탐지할 수 있을 뿐만 아니라 다양한 사이버 보안 분야에 응용할 수 있다. 본 논문은 사용자 보호와 사이버 공격으로부터 안전한 사이버 환경 조성을 위한 새로운 접근 방식을 제안하고 탐구한다.

Case Study of Building a Malicious Domain Detection Model Considering Human Habitual Characteristics: Focusing on LSTM-based Deep Learning Model

Jung Ju Won*

ABSTRACT

This paper proposes a method for detecting malicious domains considering human habitual characteristics by building a Deep Learning model based on LSTM (Long Short-Term Memory). DGA (Domain Generation Algorithm) malicious domains exploit human habitual errors, resulting in severe security threats. The objective is to swiftly and accurately respond to changes in malicious domains and their evasion techniques through typosquatting to minimize security threats. The LSTM-based Deep Learning model automatically analyzes and categorizes generated domains as malicious or benign based on malware-specific features. As a result of evaluating the model's performance based on ROC curve and AUC accuracy, it demonstrated 99.21% superior detection accuracy. Not only can this model detect malicious domains in real-time, but it also holds potential applications across various cyber security domains. This paper proposes and explores a novel approach aimed at safeguarding users and fostering a secure cyber environment against cyber attacks.

Key words : LSTM, DGA, Human Habit, Typosquatting, Malicious Domain, AUC, ROC curve, Cybersecurity

접수일(2023년 11월 17일), 수정일(2023년 12월 10일), 게재확정일(2023년 12월 29일)

★ 본 연구는 대한민국 정부(산업통상자원부, 방위사업청) 재원으로 국방과학연구소 민군협력진흥원에서 수행하는 국방기술상용화지원사업의 연구비 지원으로 ㈜에이아이스페라와 협업하여 수행되었다. (AI 기반 IP/서버 자동 보안 관제 시스템 개발(22-DC-IN-12), 2022.08~2024.07).

* ㈜모아소프트, 광운대학교 대학원 방산AI로봇융합학과

1. 서 론

악성코드의 생성과 피해는 1980년대 초부터 현재까지 점차 증가하며 발생하고 있다[1][2][3][4]. 악성코드는 컴퓨터 시스템과 네트워크를 침해하여 사용자의 개인 정보나 기업의 기밀 데이터를 노출시킨다. 이에 따라 악성코드 탐지와 분류 기술의 연구는 매우 중요한 과제로 떠오르고 있다.

악성 도메인을 대량 생성하는 한 방식인 DGA (Domain Generation Algorithm)는 인간의 습관적 실수를 악용하는 타이포스쿼팅(typosquatting) 기술의 한 형태로 사용되고 있으며, DGA 악성 도메인을 활용하여 C&C(Command & Control) 서버를 은폐함으로써 악성코드를 전파하고 있다.

따라서 본 연구에서 타이포스쿼팅을 통해 생성된 DGA 악성 도메인과 인간의 습관적 특성에 의해 접근된 DGA 악성 도메인 간의 상관관계를 분석한다. 이를 통해, DGA 악성 도메인을 효과적으로 탐지하고 분류하기 위한 새로운 LSTM(Long Short-Term Memory) 기반 Deep Learning 모델을 제시한다. 본 연구 방법으로, LSTM 기반 Deep Learning 모델을 통해 도메인을 임의 생성하고, 악성 도메인과 정상 도메인을 식별한다. 이때, 교차 검증을 통해 모델의 신뢰성을 높인다. 최종적으로 ROC 곡선과 AUC 값으로 모델의 탐지 성능을 분석하며, 연구 결과를 다양한 시각화 자료로 제시한다. 이를 통해 기존의 타 연구 모델들 대비 우수한 판별 정확도를 입증하고, 이를 기반으로 향후 연구 방향을 제안한다.

2. 관련 연구

최근 국내·외에서 악성 도메인 탐지에 관한 다양한 연구가 이루어지고 있으나, 이는 여전히 많은 한계점을 지니고 있다. 특히, 악성코드 및 악성 도메인 탐지에 주로 사용되는 시그니처 및 행위 기반 같은 전통적 탐지 방법은 DGA에 효과적으로 대응하지 못한다. 시그니처 방식은 이미 알려진 악성 도메인에만 효과를 보이며, 새로운 악성 도메인이나 변형된 악성 도메인을 감지하는 데 제약이 있다. 행위 기반 방법 또한 다양한 변형과 은폐 기술 발전으로 인해 낮은 탐지 정확도를 보여주고 있다.[5][6][7]

LSTM은 입력 데이터 정보를 장기간 저장하는 셀(Cell)이라는 메모리 유닛이 있어, 악성 도메인 탐지 시 상당히 긴 함수 호출 패턴을 학습하는 능력이 뛰어난 모델이다[7]. DGA 악성 도메인의 특징과 패턴을 학습하여 정확한 판별을 가능하게 하는 이 LSTM 모델을 이용하여 악성 도메인 탐지 연구가 상당수 이루어지고 있으나, 이 또한 DGA의 근본적인 생성 특성을 고려하지 않고 악성 도메인 판별 예측 성능만을 확인하는 연구가 대다수다[8][9].

3. 본 론

3.1 타이포스쿼팅과 인간의 습관적 특성 상관관계

3.1.1 인간의 습관적 특성 정의

표준국어대사전에 의하면 '습관'이란 '어떤 행위를 오랫동안 되풀이하는 과정에서 저절로 익혀진 행동 방식' 혹은 '비교적 고정된 반응 양식'이라고 기술한다.

본 연구에서 인간의 습관적 특성이란 '도메인 주소를 사전에 정확히 확인하지 않고 무의식적으로 웹사이트에 접속하는 행동 양상'이라고 정의한다. 이는 예를 들어, 이메일 링크를 클릭, 또는 검색 엔진에서 관련된 검색 결과 링크를 클릭하거나, 사용자가 직접 도메인을 입력하여 접속하는 행위 등을 포함한 다양한 상황에서 나타날 수 있다.

3.1.2 DGA 및 타이포스쿼팅 정의

DGA(Domain Generation Algorithm)는 '악성코드가 특정 명령 및 제어 서버(C&C 서버)와 통신[10][11][12]할 때 사용할 도메인 주소를 동적으로 대량 생성하는 알고리즘'이다[6].

타이포스쿼팅(typosquatting)(또는 URL 하이재킹(URL hijacking))은 일종의 사회 공학적인 사

이버 공격 기법으로, '정상 도메인과 유사한 이름의 도메인을 등록하여 사용자들이 실수로 정상 도메인으로 오인하게 만드는 수법'을 일컫는다.

타이포스쿼팅은 앞서 정의한 '인간의 습관적 특성'을 악용한 기법이다. 또한 타이포스쿼팅의 한 형태로 DGA가 사용되며, 따라서 DGA는 악성 도메인을 동적으로 생성하여 보안 탐지를 회피하고 사용자들을 혼란에 빠뜨리는 데 활용된다.

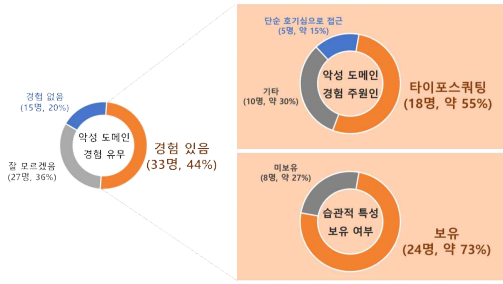
3.1.3 설문 조사 방법

본 연구에서는 타이포스쿼팅으로 생성된 DGA 악성 도메인과, 인간의 습관적 특성에 의해 접근된 악성 도메인과의 상관관계를 분석 목적으로 설문 조사를 진행하였다. 20세부터 60세까지의 75명 성인을 대상으로 악성 도메인 경험 관련 설문 조사를 진행하였고, SPSS를 활용하여 조사 결과에 대해 상관 분석하였다.

3.1.4 설문 조사 분석 결과

총 75명 중 33명(전체의 44%)이 DGA 악성 도메인을 경험했으며, 이 33명 중 18명(약 55%)이 타이포스쿼팅 기법에 따른 DGA 악성 도메인을 경험했다. 또한 습관적 특성을 가지고 악성 도메인에 접근한 사람의 수는 이 33명 중 24명(약 73%)이었다. 자세한 설문 조사 분석 결과는 (그림 1)과 같다.

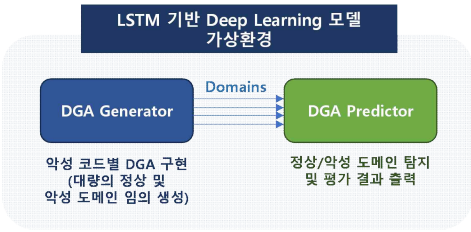
SPSS를 사용하여 타이포스쿼팅을 통해 생성된 DGA 악성 도메인과 인간의 습관적 특성에 따라 접근된 악성 도메인 간의 상관관계를 탐구하였다. 상관 분석 결과, 두 변수 간의 상관계수는 0.625로 산정되었다. 귀무가설은 두 변수 간의 상관관계가 없다는 것으로 설정되었고, 이에 대한 유의확률은 유의수준 0.01보다 작게 나타났다. 따라서, 귀무가설이 기각되고, 두 변수 간의 상관관계가 있다는 대립가설이 채택될 수 있다. 이로써, 타이포스쿼팅으로 생성된 DGA 악성 도메인과 인간의 습관적 특성에 의해 접근된 악성 도메인은 강한 양의 상관관계가 있다고 해석될 수 있다.



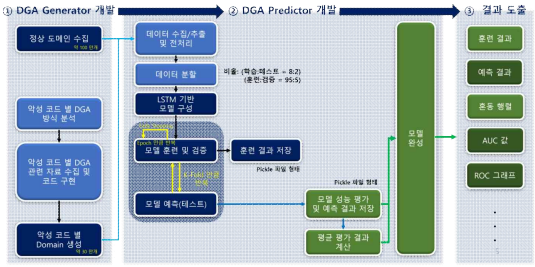
(그림 1) 설문 조사 분석 결과

3.2 연구 방안

인간의 습관적 특성을 고려한 DGA 생성기 및 예측기(DGA Generator 및 Predictor)로 구성된 LSTM 기반 Deep Learning 모델을 구축하였다. 아래 (그림 2)는 가상환경에서 구축된 LSTM 기반 Deep Learning 모델 구조도이고, (그림 3)은 LSTM 기반 Deep Learning 모델 세부 설계 및 개발과정이다. DGA Generator에서 악성코드별 악성 도메인 생성 성격을 고려하여 도메인을 임의 생성하고, DGA Generator에서 생성된 정상 및 악성 도메인을 DGA Predictor의 입력값으로 하여 학습 및 테스트하였다. 악성코드별로 악성 도메인을 라벨링 하여 분류하고, 분류 결과를 최종 AUC 탐지 정확도와 함께 출력하였다. 자세한 연구방안은 3.2.1부터 3.3.2를 통해 설명한다.



(그림 2) LSTM 기반 모델 구조



(그림 3) LSTM 기반 모델 설계 및 개발과정

3.2.1 데이터 수집과 전처리

Malware family 군에서 파생된 14종(banjori, corebot, cryptolocker, dircrypt, kraken, lockyv2, pykspa, qakbot, ramdo, ramnit, simda, matsnu, suppobox, gozi) DGA로부터 30만 개의 악성 도메인 데이터를 생성하고 수집했으며, Alexa Top 1M에서 100만 개 정상 도메인 데이터를 수집했다.

수집한 도메인 리스트를 딕셔너리(Dictionary) 형태로 저장하고, 각 도메인의 문자열을 정수로 변환하여 LSTM 기반에 모델에 적합하도록 데이터 전처리를 진행했다. 이 과정에선 제로 패딩(Zero-padding) 등의 방법을 활용하여 데이터의 일관된 형태를 유지하고 모델에 적합한 형태로 가공했다.

3.2.2 악성코드별 악성 DGA 도메인 생성 방식 분석과 DGA 생성 모델 설계 및 개발

타이포스쿼팅 기법에 의한 악성 DGA Malware family군 14종의 악성 도메인 생성 방식을 분석하여, DGA 생성 모델(DGA Generator)을 설계 및 개발하였다. 분석을 통한 설계 및 개발 방안은 <표 1>과 같다.

<표 1> 악성코드별 악성 DGA 생성 방식 분석 및 DGA 생성 모델 개발 방안

악성코드별 DGA	DGA 분석 및 개발 방안
banjori	시드 도메인을 기반으로 하여 일련의 규칙에 따라 문자열을 숫자로 매핑하고, 새로운 도메인을 생성
corebot	주어진 시드와 현재 날짜를 기반으로 문자 셋을 초기화하고, 선형 합동 생성기(linear congruential generator)를 사용하여 가변적인 길이의 새로운 도메인을 생성.
cryptolocker	주어진 시드와 현재 날짜를 기반으로 문자 셋을 초기화하고, XOR 및 비트 시프트를 사용하여 가변적인 길이의 새로운 도메인을 생성.
dircrypt	주어진 시드를 기반으로 선형 합동 생성기(linear congruential generator)를 사용하여 가변적인 길이의 새로운 도메인을 생성. 난수 생성기를 클래스로 캡슐화.
kraken	주어진 인덱스, 날짜, 시드, Temp 파일 여부 및 TLD(Top-Level Domain) 세트 번호를 사용하여 새로운 도메인을 생성. 시드 및 Temp 파일 여부에 따라 다른 초기 난수 상태 선택. TLD 목록에서 무작위로 TLD를 선택하여 도메인에 추가.
lockyv2	날짜, 구성 번호 및 도메인 번호를 사용하여 초깃값을 설정하고, 비트 연산과 수식을 적용하여 고정 범위 길이 내의 새로운 도메인을 생성. TLD 목록에서 무작위로 TLD를 선택하여 도메인에 추가.
pykspa	주어진 길이 및 시드를 기반으로, 알파벳 소문자로 이루어진 새로운 도메인을 생성.
qakbot	주어진 날짜, 설정 번호 및 JSON으로 저장된 시드 값을 사용하여 새로운 도메인을 생성. 6에서 12 사이의 무작위 값으로 도메인 길이가 생성. 해당 날짜의 시드 값은 설정된 날짜에서 지난 일수를 계산하여 설정.
ramdo	시드와 도메인 반복자를 사용하고, XOR 및 비트 연산을 이용하여 주어진 수만큼의 새로운 도메인을 생성. 도메인 생성 과정에서 시드와 반복자를 업데이트하여 도메인의 변동성 도입
ramnit	Lehmer random number generator 알고리즘을 사용하여 무작위 정수를 생성하고 일련의 계산과정을 거쳐 시드를 생성. 시드를 사용하여 주어진 개수만큼의 새로운 도메인을 생성. 도메인 생성 과정에서 시드를 계속해서 업데이트하여 도메인의 변동성을 도입.
simda	지정된 자음 및 모음의 문자열 설정. 도메인 수, 도메인 길이, TLD, 도메인 key, base 값을 파라미터로 하여 새로운 도메인을 생성. key와 base 값을 업데이트하여 도메인의 변동성을 도입.

matsnu	단어사전 기반. 특정 날짜를 기반으로 하여 현재 날짜와의 차이를 이용해 지정 개수의 새로운 도메인을 동적 생성. 중복 도메인이 생성되지 않도록 하는 방지 기능 포함.
suppobox	단어사전 기반. 시간에 따라 달라지는 시드값을 기반으로 하여 단어사전에서 단어를 무작위 선택하는 방식 사용. 지정 개수의 새로운 도메인을 생성.
gozi	단어사전 기반. 선형 합동 생성기를 사용하여 난수를 발생시키고, 특정 날짜와 단어사전 목록에 기반하여 새로운 도메인을 생성. 특정 날짜와 지정 날짜와의 차이를 시드 생성에 활용하고, 시드를 사용하여 선형 합동 생성기를 초기화. 지점 범위 내에서 도메인 길이 생성.

3.2.3 DGA 예측 모델 설계 및 개발

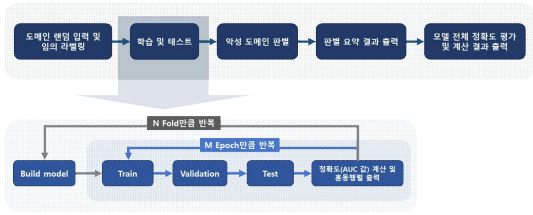
DGA 예측 모델(DGA Predictor)은 Embedding layer, LSTM layer, Dropout layer, Dense layer, Sigmoid function으로 구성되었다.

- Embedding layer: 텍스트 데이터를 고정된 차원의 실수 벡터로 변환하여 입력 데이터를 임베딩 벡터로 변경한다.
- LSTM layer: 입력 데이터의 패턴과 특징을 학습하고 해당 도메인의 악성 여부를 예측하는 데 활용한다. 악성과 정상 도메인은 1(악성)과 0(정상)의 스코어 방식으로 구분된다.
- Dropout layer: 학습 중 과적합 방지를 위해 일부 뉴런을 비활성화(0)하여 정규화한다.
- Dense layer: LSTM layer의 출력을 예측하려는 클래스 수에 맞게 변환하여 1차원의 최종 출력물을 생성한다.

본 모델은 과적합을 방지하기 위해 조기 종료(Early Stopping) 콜백(Callback) 및 Dropout을 활용하고, Sigmoid 활성화 함수를 사용하여 이진 분류를 수행하도록 설계되었다. 또한 Binary Cross-entropy 손실 함수와 경사 하강법 RMSP(Root Mean Square Propagation) 최적화를 사용하여 컴파일되었다.

3.2.4 모델 학습 및 테스트

약 30만 개의 도메인 데이터를 학습8 : 테스트2, 훈련95 : 검증5 비율로 분할하였다. Batch size는 128이며, Epoch는 10으로 설정하였다. 데이터 편향과 오버피팅(over-fitting) 및 언더피팅(under-fitting)을 방지하고 데이터의 활용도를 높이기 위해 5-Fold 교차 검증 방법을 사용하여 모델을 학습 및 테스트하였다. 모델 학습 중에는 모델 성능을 모니터링하여, 모델 성능 향상이 없으면 검증 데이터를 기반으로 조기 종료(Early Stopping)될 수 있도록 조치하였다. 이 접근 방식은 과적합을 방지하면서, 최적의 Epoch에서 모델을 저장함으로써 모델의 일반화 성능을 높인다. 최종적으로 학습 및 테스트 결과는 pickle file 형태로 저장하여 연산 속도를 향상시켰다. (그림 3)에서 전체 학습 및 테스트 구조도를 확인할 수 있으며, 세부 학습 및 테스트 흐름은 (그림 4)와 같다.



(그림 4) 모델 학습 및 테스트 프로세스

3.3 연구 결과

3.3.1 모델 평가 방법

모델의 성능 및 정확도 평가는 테스트 데이터를 기반으로 진행되었다. 혼동 행렬(Confusion Matrix), ROC(Receiver Operating Characteristic) 곡선, 그리고 AUC(Area Under the ROC Curve) 정확도(Accuracy)와 같은 다양한 시각적 평가 지표를 사용하여 모델의 탐지 성능을 분석하였다. (그림 5)는 AUC 결과, (그림 6)은 ROC 곡선 결과, (그림 7)은 혼동 행렬 결과를 보여준다.

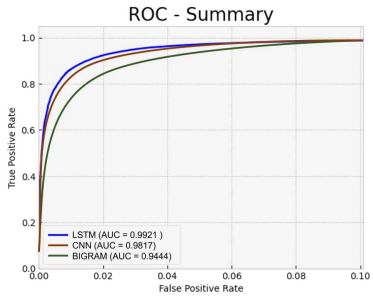
3.3.2 모델 평가 결과

LSTM 기반 Deep Learning 모델이 AUC 99.21% 이상의 정확도를 달성했음을 (그림 5)와 같이

확인했다. 이는 DGA의 근본적인 생성 특성을 고려하지 않고 단순히 악성 도메인 탐지만을 고려한 CNN이나 BIGRAM 모델의 성능보다 최소 1% 이상 차이 값을 보여주는 의미 있는 결과이며, 모델 비교 결과는 (그림 6)에서 확인 가능하다. (그림 7)에서, 혼동 행렬(Confusion Matrix)에서 TP (True Positive)와 TN(True Negative)의 값이 뚜렷하게 높고, FP(False Positive)와 FN(False Negative)의 값이 상대적으로 매우 낮게 나타나 악성 도메인의 정확한 탐지가 이루어졌다는 것을 자세히 확인 가능했다.

```
================================================================================
최종 lstm 모델 AUC 값: 0.9921
================================================================================
```

(그림 5) AUC 결괏값



(그림 6) ROC 결과 그래프

```
- CONFUSION MATRIX(val):
[[26661  1599]
 [  796 26918]]

- CONFUSION MATRIX(%):
[[94.3418259  5.6581741 ]
 [ 2.87219456 97.12780544]]
```

(그림 7) 혼동 행렬 결괏값

4. 결론 및 제언

4.1 결과 해석과 의의

본 연구 결과로 타이포스쿼팅과 인간의 습관적 특성이 악성 도메인 생성과 밀접한 관련성을 갖는다는 것을 확인했다. 이는 악성코드 제작자들이 사용자의 습관적 특성을 파악하여 DGA 도메인을 생성하고, 이를 통해 악성코드의 Command & Control (C&C) 서버를 숨기고 악성코드를 전파하는데 DGA 도메인을 활용한다는 점을 시사한다. 이 연구의 중요성은 보안 분야에서의 새로운 접근 방식과 기술적 발전을 제시했다는 데 있다. 악성코드의 지속적인 증가와 변화 속에서, 기존 방법론의 한계를 넘어서 인간의 습관적 특성과 악성 도메인 생성의 관계를 이해하는 새로운 시각을 제시함으로써, 더욱 효과적인 보안 대응책을 마련하는 데 이바지한다. 또한, 인간의 습관적 특성과 도메인 패턴을 고려하여 설계된 LSTM 기반 Deep Learning 모델은 AUC 99.21% 이상의 탁월한 악성 도메인 탐지 정확도를 나타냈다. 이와 같은 연구 결과는 보안 업계에서의 악성 도메인 탐지 기술을 개선하고, 사용자들을 보다 효과적으로 보호하는 데 기여할 수 있다.

4.2 제시 모델의 한계점 및 향후 연구 방향 제안

데이터의 수량과 다양성은 모델의 성능에 결정적인 영향을 미치는 요소이다. 따라서 다양한 종류의 악성 및 정상 도메인 데이터를 대량으로 확보하고 학습시켜야 한다. 이를 통해 모델의 결함을 보완하고 오류를 최소화함으로써 모델의 성능과 완성도를 향상시킬 수 있다.

향후 연구에서는 다양한 데이터셋을 활용하고, 실제 환경에서의 적용 가능성을 고려하여 실시간 악성 도메인 탐지 시스템을 개발해야 한다. 본 연구에서 제시한 모델을 기반으로 악성 도메인을 이미지로 변환하고 학습된 악성 도메인과의 유사도를 분석한다면, 이메일로 유입되는 다양한 형태의 악성 도메인을 실시간으로 신속하게 감지할 수 있을 것이다. 이러한 접근 방식은 보다 안정적이고 효율적인 사이버 보안 시스템의 구축과 현실적인 위협에 대응하는 방안을 모색하는 데 큰 도움이 될 것이다.

참고문헌

[1] 연세대학교 산학협력단, “데이터마이닝 기반 악성코드 변종그룹 식별방안 연구보고서”, 한국인터넷진흥원, December, 2016.

[2] 김정욱, “악성도메인 IP 주소 추적을 이용한 효과적인 보안관제 방안 연구”, 고려대학교 정보경영공학전문대학원, 2009.

[3] 한국데이터산업진흥원, “Dicon Report/컴퓨터 바이러스”, April, 2006.

[4] 장진호, 임채현, 이태진, “머신러닝 기반 악성 도메인 분석을 위한 핵심 feature 추출 연구”, 한국통신학회, pp. 1528-1529, June, 2022.

[5] 김영호, 이현종, 황두성, “엔트로피 시계열 데이터 추출과 순환 신경망을 이용한 IoT 악성코드 탐지와 패밀리 분류”, KIPS Transactions on Software and Data Engineering, Vol. 11, No. 5, pp. 197-202, 2021.

[6] 김휘강, “인공지능을 활용한 사이버위협 모니터링 기술”, 고려대학교 정보보호대학원, 2021.

[7] 이윤석, “파일 접근 행태의 LSTM 학습을 활용한 악성 코드 탐지 기법”, The Journal of Korean Institute of Information Technology, Vol. 18, No. 2, pp. 25-32, February, 2020.

[8] Cengiz Acarturk 외 5명, ‘Malicious Code Detection: Run Trace Output Analysis by LSTM’, IEEE Access, Vol. 9, 2021.

[9] Muhammad Shoaib Akhtar, Tao Feng, ‘Detection of Malware by Deep Learning as CNN-LSTM Machine Learning Techniques in Real Time’, MDPI, November, 2022.

[10] 류소준, “악성코드 C2통신 상세분석”, 한국인터넷진흥원, 2019.

[11] 김동현, 김강석, “N-gram을 활용한 DGA-DNS 유사도 분석 및 APT 공격 탐지”, Journal of The Korea Institute of Information Security & Cryptology, Vol. 28, No. 5, pp. 1141-1151, October, 2018.

[12] Bin Yu, Daniel L. Gray, Jie Pan, Martine De Cock, Anderson C. A. Nascimento, “Inline DGA Detection with Deep Networks”, IEEE International Conference on Data Mining Workshops (ICDMW), pp. 683-692, 2017.

[13] 김용식, “LSTM 활용 제어시스템 이상징후 탐지 향상”, 고려대학교 정보보호대학원, August, 2020.

[14] 김보람, 권소연, 김유빈, 김은결, 이광재, “도메인 이름 및 부가정보를 활용한 기계학습 기반 보안 QR코드 스캐너 개발”, 한국통신학회, pp. 992-993, 2023.

[15] 황현정, 김강석, “효율적인 이상 탐지를 위한 적대적 도메인 적응 기법”, Journal of Digital Contents Society, Vol. 24, No. 2. pp. 369-378, February, 2023.

[16] 심유진, “2021년 상반기 악성코드 은닉사이트 탐지 동향 보고서”, 한국인터넷진흥원, January, 2022.

[17] 김영준, 이재우, “URL 주요특징을 고려한 악성URL 머신러닝 탐지모델 개발”, 한국정보통신학회논문지, Vol. 26, No. 12, pp. 1986-1793, December, 2022.

[18] 정일옥, 신덕하, 김수철, 이록석, “N-gram을 활용한 DGA 기반의 봇넷 탐지 방안”, Journal of Information and Security, Vol. 22, No. 5, pp. 145-154, December, 2022.

[19] 이승현, 문종섭, “명령 실행 모니터링과 딥러닝을 이용한 파워셸 기반 악성코드 탐지 방법”, Journal of The Korea Institute of Information Security & Cryptology, Vol. 28, No. 5, pp. 1197-1207, October, 2018.

[20] 배장성, 이창기, 최선오, 김종현, “Skip-Connected LSTM RNN을 이용한 악성코드 탐지 모델”, Journal of KIISE, Vol. 45, No. 12, pp.1233-1239, December, 2018.

[21] 조영복, “딥러닝 기반의 R-CNN을 이용한 악성코드 탐지 기법”, Journal of Digital Contents Society, Vol. 19, No. 6, pp. 1177-1183, 2018.

[22] Jonathan Woodbridge, Hyrum S. Anderson,

Anjum Ahuja, Daniel Grant, “Predicting Domain Generation Algorithms with Long Short-Term Memory Networks”, Cornell University, Vol. 1, November, 2016.

〔저 자 소 개〕

정 주 원 (Ju-won Jung)
2018년 8월 순천향대학교 컴퓨터공학과/경영학과 학사
2023년 9월 ~ 현재 광운대학교 대학원 방산AI로봇융합학과 석사 과정
2019년 (재)한국기원 교육보급팀 연구원
2020년 4월 ~ 현재 ㈜모아소프트 AI/Data Science 연구소 선임 연구원
email : wopalw@naver.com