안드로이드 악성앱 탐지 TUTORIAL

1 소개

안드로이드 앱에서 추출한 여러가지 특성을 기반으로 악성 앱을 탐지하는 샘플 모델을 소개합니다.

* 튜토리얼에서 소개되는 모델은 참고문헌 중 하나인 Andro-profiler 의 작동 원리를 기반으로 하고 있습니다. 탐지 알고리즘에 대한 상세한 내용은 사이트에 첨부한 논문에서 확인하실 수 있습니다.

** 이번 Data Analysis Challenge 에 제공될 안드로이드 앱은 튜토리얼에서 사용한 데이터셋과 수량 및 파일이 다를 수 있습니다.

2 악성 행위 프로파일링

먼저 안드로이드 앱의 행위를 프로파일링 합니다. 안드로이드 앱을 동적 분석할 수 있도록 도와주는 sandbox 툴인 Droidbox 를 이용하여, 앱의 악성 행위와 관련된 시스템 로그를 추출합니다. 추출한 시스템 로그는 다음과 같이 object 와 operation 에서 정의한 형태로 프로파일링합니다.

2.1 OBJECT

악성 행위의 큰 분류를 의미합니다. 안드로이드 악성 앱 샘플들을 직접 분석하여, 크게 `Telephony|Phone|Network` 로 분류하였습니다.

2.2 OPERATION

악성 행위의 상세한 동작 정보를 의미합니다. Operation 은 다음과 같이 operation-name, operation-target, operation-attribute 의 조합으로 정의됩니다.

```
Operation ::= {Operation-name : {Operation-target : Operation-attribute}}
```

- **Operation-name:** 악성 행위의 종류를 말하며, `Sending SMS|Calling|Sending sensitive information|Converting data` 로 분류합니다.
- **Operation-target:** 악성 행위의 목표물을 말하며, `Premium-rate SMS/number|deviceID|IMEI|IMSI|MCC|MNC` 등이 있습니다.
- **Operation-attribute:** 목표물에서 얻어낸 의미있는 정보를 말합니다. 예를 들어, country code 의 attribute 으로는 Korea 가 있습니다.

2.3 예시

아래 Table 2 는 network object 와 관련된 operation 값의 예시를 보여줍니다. 이 정보를 프로파일링하면 `{Network : {Sending sensitive information : {{IMEI : 357242043237517}, {MCC : 310}, {MNC : 260}, {Location : GPS Coordinates } …, } } }`와 같이 표현할 수 있습니다.

Table 2 Example of mapping of network object

Type	Name	Target	Attribute
Network	Sending sensitive information	Android Id	3531505c0b421c4d
		Device type	Android
		IMEI	357242043237517
		IMSI	310005123456789
		MCC	310
		MNC	260
		OS version	10
		SDK version	2.3.4
		Carrier	Android
		Country code	en
		Location	GPS coordinates
	Converting data	Cipher algorithm	No, DES, AES, Blowfish
		Destination URL	http://my365image.com
		Port	80
		Encoding algorithm	Gzip

3 행위 프로파일을 이용한 안드로이드 앱 분류

3.1 악성 앱 탐지

본 논문에서 사용한 안드로이드 앱 데이터셋에서는 악성 행위의 종류(operation-name)가 총 4 가지—sending of premium-rate SMS, calling of premium-rate number, sending of sensitive information, converting data for transmission—였습니다. 행위 프로파일링을 통해 4 가지 중 어떠한 악성 행위와도 관련된 동작이 없다면 정상앱(benign)으로 분류합니다. 4 가지 중 하나라도 해당하는 행위가 발생하였다면 악성 앱(malware)로 분류합니다.

3.2 악성 앱 분류[1]

본 논문에서는 악성 여부를 탐지한 이후, 추가적으로 악성으로 분류된 앱에 대해 유사도 분석을 통해 패밀리를 분류하였습니다.

먼저, behavior factor 를 악성 행위 종류에 따라 4 가지로 나눕니다.

- Sending premium-rate SMS (SS)
- Calling premium-rate number (CS)
- Sending sensitive information (SIS)
- Converting data (CDS)

각 behavior factor 의 유사도(BFS, similarity of behavior factor) 점수를 Table 3 과 같은 기준으로 부여합니다. 유사도 점수를 부여할 때에는 각 패밀리의 악성 행위를 대표적으로 나타낼 수 있는 행위 프로파일을 기준으로 설정하여 비교합니다. 각 BFS 점수에는 가중치를 적용하여 그 합계로 특정 패밀리와의 유사도 점수를 산출합니다.

$$S = \sum_i w_i \cdot BFS_i \quad where \quad \sum_i w_i = 1$$

[1] 2018 년도 데이터챌린지 AI 기반 안드로이드 악성앱 탐지 트랙에서는 악성앱 패밀리 분류는 필수사항이 아니므로 참고하시기 바랍니다.

Table 3 Similarity metric to apply to each behavior factor

Behavior factor	Behavior target	Similarity metric
Sending SMS	Premium-rate	Binary (0 or 1)
Calling	Premium-rate	Binary (0 or 1)
Sending sensitive information	System information, private information	Jaccard index [0, 1]
Converting data	Destination URL	Modified levenshtein distance [0, 1]
	Cipher algorithm (DES, AES, Blowfish)	Binary (0 or 1)
	Encoding algorithm (Gzip or not)	Binary (0 or 1)

본 논문에서는 여러 번의 실험을 거쳐 가장 성능이 잘 나오는 가중치를 확인하여 적용했습니다. SS 에는 0.33, CS 에는 0.33, SIS 에는 0.21, CDS 에는 0.13 의 가중치를 주었습니다.

각 패밀리를 대표하는 행위 프로파일은 두 가지 방법으로 업데이트할 수 있습니다.

- **Method 1:** 새로운 악성 앱이 특정 패밀리로 분류되었을 때, 해당 패밀리의 대표 프로파일과 추가된 앱의 프로파일의 **교집합(intersection)**을 새로운 대표 프로파일로 업데이트합니다.
- **Method 2:** 새로운 악성 앱이 특정 패밀리로 분류되었을 때, 해당 패밀리의 대표 프로파일과 추가된 앱의 프로파일의 **합집합(union)**을 새로운 대표 프로파일로 업데이트합니다.

4 실험

4.1 사용한 데이터셋

8,840 개의 정상 앱과 643 개의 악성 앱으로 구성된 데이터셋을 사용하여 실험을 진행하였습니다. 정상 앱과 악성 앱의 패밀리 분포는 Table 5 와 같습니다.

Table 5 Malware samples and benign samples for experiments

Category	Family	Quantity	Behavioral characteristics
Malware (643)	AdWo	401	Collect the sensitive information
	AirPush	60	Send SMS and collect the sensitive information
	FakeBattScar	44	Collect the sensitive information
	Boxer	42	Send SMS and collect the sensitive information
	GinMaster	96	Collect the sensitive information
Benign (8840)	Application	7164	Normal application
	Game	1676	Normal game application

4.2 실험 환경

제시한 안드로이드 악성 앱 분류 모델을 구현하여 실험한 환경은 다음과 같습니다.

- Intel(R) Xeon(R) X5660 processor
- 4 GB RAM
- 32-bit Ubuntu 12.04 LTS operating system

4.3 실험 결과

구현한 모델로 악성 앱을 탐지하고, 패밀리를 분류한 결과는 Table 9 와 같습니다.

Table 9 Classification performance for 643 malware and 8840 benign samples

Category	Accuracy			AUC		
	Method 1	Method 2	Crowdroid	Method 1	Method 2	Crowdroid
Malware						
AdWo	*1.00*	*1.00*	0.01	*1.00*	*1.00*	0.49
AirPush	*1.00*	*1.00*	0.00	*1.00*	*1.00*	0.50
Boxer	*1.00*	*1.00*	0.00	*1.00*	*1.00*	0.50
FakeBattScar	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*
GinMaster	*1.00*	*1.00*	0.00	*1.00*	*1.00*	0.49
Benign	*0.97*	*0.97*	0.96	*0.99*	*0.99*	0.52
Average	*0.98*	*0.98*	0.90	*0.99*	*0.99*	0.52

악성 앱의 탐지 정확도는 method 1, method 2 모두 98%로 측정되었습니다. 챕터 3.2 에서 설명한 것과 같이 method 1 은 교집합 방식으로 패밀리 대표 프로파일을 업데이트한 경우이고,

method 2 는 합집합 방식으로 패밀리 대표 프로파일을 업데이트한 경우입니다. 또한, Burguera I, et al.이 2011 년 제안한 Crowdroid 시스템과도 비교하였으며, 성능이 더욱 좋은 것을 확인할 수 있었습니다.

실험한 환경에서의 탐지 및 분류 속도는 55 초/MB 로 측정되었습니다. 안드로이드 에뮬레이터의 부팅 시간은 제외한 속도입니다. 작동 시간의 대부분은 행위 프로파일을 작성하는데 쓰였으며, 악성 앱을 탐지/분류하는 데에는 0.2 초/MB 정도만 소요되었습니다.

5 탐지율 평가 기준

Category		**Actual Class**	
		Malware	Benign
Estimated Class	Malware	*True Positive*	*False Positive*
	Benign	*False Negative*	*True Negative*

True Positive 는 실제 *Malware(True)*를 *Malware(True)*로 정확하게 예측한 상황을 의미하며,
False Positive 는 실제 *Bengin(False)*을 *Malware(True)*로 예측한 상황을 의미합니다.
False Negative 는 실제 *Malware(True)*를 *Bengin(False)*로 예측한 상황을 의미하며,
True Negative 는 실제 *Benign(False)*를 *Benign(False)*로 예측한 상황을 의미합니다.

탐지 결과가 위 표와 같을 때, 전체 결과 중 정상 앱과 악성 앱을 정확히 탐지한 비율을 측정합니다.

$$\text{Accuracy} = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + False\ Negative + True\ Negative}$$

6 REFERENCE

본 튜토리얼에서 소개되는 모델은 참고문헌 중 하나인 Andro-profiler 의 작동 원리를 기반으로 하고 있습니다. 논문 파일은 데이터챌린지 사이트에서 다운로드 받을 수 있습니다.

- Jae-wook Jang, Jaesung Yun, Aziz Mohaisen, Jiyoung Woo, and Huy Kang Kim, "Detecting and classifying method based on similarity matching of Android malware behavior with profile," *SpringerPlus* 5:273, 2016.
- 본 데이터 분석 챌린지에서 제공하는 데이터셋을 기반으로 논문을 작성할 경우, 해당 논문을 반드시 인용해야합니다.