R&D 데이터를 활용한 사이버보안 분야 난제 해결 I-KOREA 4.0

정보보호 R&D 데이터 챌린지 2018 대회 설명회

일시 2018. 7. 27.(금) 14:00~17:00

장소 유리앤호텔 Jade 홀

주최 KISA 한국인터넷진흥원

주관 KISA 한국인터넷진흥원 한국정보보호학회 Korea Institute Information Security & Cryptology

기관후원 KISIA 한국정보보호산업협회 CONsortium of CERT CONCERT

기업후원 AhnLab ESTsecurity HAURI SAINT SECURITY

CONTENTS

정보보호 R&D 데이터 챌린지 2018 대회 설명회

대회 소개 및 주요 일정 안내

「정보보호 R&D 데이터 챌린지 2018」 설명회

- 대회 소개 및 주요 일정 안내

2018. 7.27

KISA 한국인터넷진흥원

KISA

Contents

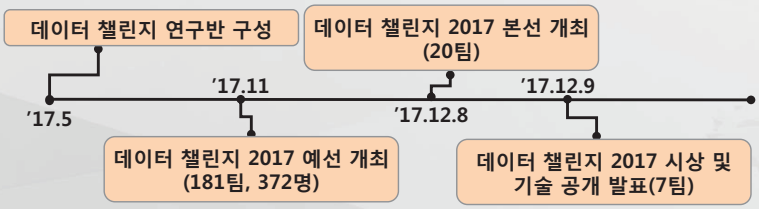
1. 정보보호 R&D 데이터 챌린지 개요

개요

- 대용량의 정보보호 R&D 데이터셋을 활용해 악성코드 탐지, 취약점 자동탐지 AI보안 기술개발 및 개발 기술의 성능을 검증하는 「정보보호 R&D 데이터 챌린지 2018」 개최

추진 경과

- 2017년 정보보호 R&D 데이터셋 구축•공유를 통한 기술개발 역량 개발 지원 및 「정보보호 R&D 데이터 챌린지 2017」 최초 개최('17.11.6~12.8)
- 총 181개팀(372명)이 예선을 거쳐 20팀 본선 진출, 최종 7개 수상팀 선정('17.12.8)



1

2. 정보보호 R&D 데이터 챌린지 2018 소개

정보보호 R&D 데이터 챌린지 2018

- KISA-정보보호학회와 공동으로 ▲AI기반 악성코드 탐지(대학원생,일반), ▲AI기반 취약점 자동 탐지, ▲AI기반 안드로이드 악성앱탐지,▲차량주행데이터 기반 도난탐지 트랙 운영

운영 트랙	AI기반 악성코드 탐지 (대학원생)	AI기반 악성코드 탐지 (일반/제한없음)	AI기반 취약점 자동 탐지	AI기반 안드로이드 악성앱 탐지	차량주향데이터 기반 도난탐지
활용 데이터셋	정상/악성코드 총 50,000개		취약점 포함 바이너리 약 100여개	정상/악성앱 약 11,000개	720km 주행데이터 26시간
시상 (예선)	1위 협회장상, 2위 백신사 사장상				
시상 (본선)	1위 KISA 원장상, 2위 우수상, 3위 장려상			1위 학회장상, 2위 우수상, 3위 장려상	
포상 (본선)	총 2,500만원 상당의 상금 지급 예정				

* 세부 시상 및 포상 내역은 추후 홈페이지를 통해 공개 예정이며, 예산 1~2위 시상은 악성코드 탐지 트랙 대학(원)생 부문에 한해 예정

2

3. 2017년 대비 달라진점

	2017년	2018년
주최/주관	○ 주최: 한국인터넷진흥원, 한국정보보호학회 ○ 주관: 한국인터넷진흥원, 고려대학교	○ 주최: 한국인터넷진흥원 ○ 주관: 한국인터넷진흥원, 한국정보보호학회
트랙운영	○ (KISA) 악성코드 탐지, 악성코드 선제대응 ○ (학회) 악성앱 탐지, 차량 해킹 탐지	○ (KISA) 악성코드 탐지, 취약점 자동 탐지 ○ (학회) 악성앱 탐지, 차량 도난탐지
데이터셋	○ 악성정상코드 30,000개, 지능형악성코드 300개 ○ 악성,정상앱 8,000개, 48시간 주행 데이터	○ 악성정상코드 50,000개, 취약한 바이너리 100여개 ○ 악성,정상앱 11,000개, 720km 주행 데이터
운영방식	○ (예선) 온라인 ○ (본선) 오프라인	○ (예선) 온라인, (본선) 오프라인 ○ 악성코드 탐지 대학(원)생 부문 지역 오프라인 예선 개최
포상(확대)	○ 1위 KISA 원장상 ○ 2위 장려상	○ 1위 KISA 원장상, 2위 우수상, 3위 장려상 ○ 악성코드 탐지 대학(원)생 부문 협회장상 등 신규 포상

3

4. 주요 일정(안) 안내

7월

- 7.16(월)~8.24(금) AI기반 악성코드 탐지 대학(원)생 부문 접수
- 7.27(금)「정보보호 R&D 데이터 챌린지 2018」 설명회

9월~11월

- 9.10(월)~11.9(금) AI기반 악성코드 탐지 일반부문, AI기반 취약점 자동탐지 AI기반 안드로이드 악성앱 탐지, 차량주행데이터 기반 도난탐지 트랙 접수
- 9.10(월)~11.9(금) AI기반 악성코드 탐지 일반부문, AI기반 안드로이드 악성앱탐지, 차량주행데이터 기반 도난탐지 트랙 온라인 예선
- 11.10(토) AI기반 악성코드 탐지 대학(원) 부문 권역별 오프라인 예선
- 11.16(금) 트랙별 본선 진출자 발표
- 11.30(금)~12.1(토) 트랙별 본선(오프라인) 진행

12월

- 12월 중 「정보보호 R&D 데이터 챌린지 2018」 시상식 개최

* 주요 일정 변경시 홈페이지를 통해 공지 예정

4

감사합니다.

정보보호 R&D 데이터 챌린지 2018 대회 설명회

AI 기반 악성코드 탐지 트랙(대학(원)/일반)

「정보보호 R&D 데이터 챌린지 2018」 설명회

- AI 기반 악성코드 탐지 트랙(대학(원)/일반)

정 소 영 선임

한국인터넷진흥원 보안기술확산팀

2018. 7.27

KISA 한국인터넷진흥원

KISA

Contents

Ⅰ. 개요- 목표 및 참가 부문

트랙 목표

대용량 악성/정상코드 분석과
AI 기반의 악성코드 탐지 알고리즘 개발을 통한
탐지 정확도 향상

참가 부문

대학(원)생 | **일반**

Ⅰ. 개요- 참가 부문 소개

	대학(원)생 부문	일반 부문
참가자격	✓ 전국 대학(원)생 (휴학생 포함) ※ 1~5인 이내 개인 또는 팀 구성	✓ 제한 없음 ※ 1~5인 이내 개인 또는 팀 구성 ※ 팀 대표는 학계, 산업계 등 소속이 있는 자에 한함
예 선	✓ 권역별 오프라인 예선 실시 (예정) ※ 11.10(토) 예정	✓ 온라인 예선 실시 ※ 11.09(금) 24:00 제출 마감
본 선	✓ 오프라인 본선 실시	✓ 오프라인 본선 실시
시상	(예선) 권역별 1위 협회장상 2위 백신사 사장상	(예선) 없음
	(본선) 1위 KISA 원장상 2위 우수상 3위 장려상	(본선) 1위 KISA 원장상 2위 우수상 3위 장려상

Ⅱ. 활용 데이터셋

- ✓ (규모) 정상/악성코드 약 5만개
- ✓ (수집방법) KISA, 안랩, 이스트시큐리티, 하우리, 세인트시큐리티 등 국내 백신사 공동구축
- ✓ (데이터셋 가공) Overfitting 대응을 위해 정상/악성코드 기능·행위 분석 및 분류 결과 기반 데이터셋 구성, 오픈소스 알고리즘, 국내외 R&D 연구결과물, 17년 대회 입상팀 프로그램 등을 활용해 데이터셋 탐지율 사전 테스트 및 결과에 따라 데이터셋 추가 가공

- ✓ (학습 데이터셋) 정상/악성 코드 데이터셋, 정답지(Class label 제공)
- ✓ (예선 데이터셋) 정상/악성 데이터셋(정답지 미포함), 예선 참가 시 해당 데이터셋의 탐지 결과 제출
- ✓ (본선 데이터셋) 정상/악성 데이터셋(정답지 미포함), 대회 1일차, 2일차에 각각 배포되는 데이터셋의 탐지 결과 제출

Ⅲ. 제출 결과물 안내

결과파일

o 탐지결과를 파일명 ID와 정상/악성으로 분류하여 csv파일 형태로 작성
※ ID : MD5 값, Class : 정상코드 1, 악성코드 0

알고리즘 설명문서

o 데이터 분석 및 분류 결과, Feature 추출 및 알고리즘 구성 방법,
수도코드(pseudocode), 예선 데이터 실험과정, 예상 결과, 보완점 등을 작성
※ 8월 중, 홈페이지(www.datachallenge.kr)를 통해 튜토리얼 공개

발표자료

o 알고리즘 설명문서 요약, 본선 데이터 분류 방법, 탐지 결과 등을 포함하여
15분 발표 분량으로 작성
※ 본선 진출자에 한해 제출

IV. 예선 진행방식

대학(원)생 부문

- ✓ 사전접수
  ※ 기간 : '18. 7.16~8.24
- ✓ 권역별* 오프라인 예선 실시(예정)
  * 권역 : 서울 · 강원/ 경기 · 충청/호남/영남/제주
  ※ 오프라인 예선 참가 지역을 선택하는 개념으로, 참가신청서 작성시 선택
  ※ 사전접수 결과에 따라 권역별 오프라인 예선 개최
- ✓ 심사기준 : 탐지율 100%
  ※ 알고리즘 설명문서는 치팅여부 확인용
- ✓ 일정 : 2018.11.10(토)
- ✓ 본선 진출팀 : 홈페이지 공지 및 개별연락

일반 부문

- ✓ 참가접수
  ※ 기간 : '18. 9.10~11.09
- ✓ 결과파일 및 알고리즘 설명문서 평가 실시(온라인)
- ✓ 심사기준: 탐지율 80%, 문서 20%
- ✓ 결과물 접수마감 : 2018.11.9일(금) 24:00
- ✓ 본선 진출팀 : 홈페이지 공지 및 개별연락

V. 본선 진행방식 – 대회 일정

(기간) 11.30(금~12.1(토)

o 1일차 : 대회 개최 선언, 1차 데이터셋 분석, 결과파일 제출
※ 1일차 종료 후 귀가

o 2일차 : 2차 데이터셋 분석, 결과파일 및 발표자료 제출, 발표 심사, 점수 산출 및 순위 발표

Ⅴ. 본선 진행방식 – 운영과정

데이터셋 배포

o 대회 1일차, 2일차 총 2회에 걸쳐 본선 데이터셋 배포

결과파일 접수

o 1일차, 2일차 각각 마지막에 제출한 결과를 최종 탐지 점수로 반영

※ 대회 당일, 점수 보드를 통해 팀명-탐지율 송출

※ 제출 횟수는 홈페이지를 통해 추후 공지

o 1일차 2일차 탐지 점수를 1:1로 반영

발표심사 실시

o 각 부문별 탐지율 상위 3개 팀 대상 발표 진행

※ 평가위원회 의견에 따라 추가 발표가 진행될 수 있음

o 백신사 관계자, 학계 전문가 등으로 구성된 평가위원회에서 심사 실시

점수산출 및 순위 발표

o **탐지율 80%, 발표 점수 20%** 를 반영하여 점수 산출 후, 최종 순위 발표

Ⅵ. 시·포상내역

	대학(원)생 부문	일반 부문
예선	권역별 1위 협회장상 2위 백신사 사장	해당 없음
본선	1위 KISA 원장상 2위 우수상 3위 장려상 ※ 1위팀 지도교수(신청서에 기재시)에게 우수 지도상 수여	1위 KISA 원장상 2위 우수상 3위 장려상
	정보보호 R&D 데이터 챌린지 2018 총 상금 2,500만원 상당	

Ⅶ. 접수기간 및 신청절차

접수기간

대학(원)생 부문	일반 부문
2018.7.16(월) ~8.24(금)	2018. 9.10(월)~11.9일(금)

신청절차

01 datachalleng.kr
- ✓ 정보보호 R&D 데이터챌린지 홈페이지 접속

02 「AI 기반 악성코드 탐지」
- ✓ 트랙 선택

03 신청서류 다운 및 작성
- ✓ 참가신청서
- ✓ 약정서

04 rnd_dataset@kisa.or.kr로 제출
- ✓ 신청서류
- ✓ 소속 확인증명서

※ 대학(원)생 부문은 참가신청서 작성시, 권역별 오프라인 예선 개최를 위해 권역 선택 필수

9

AI 기반 악성코드 탐지 트랙

Q&A

감사합니다.

<문의처>
한국인터넷진흥원 정소영 선임
(T. 061-820-1254)
(E. md_dataset@kisa.or.kr)

정보보호 R&D 데이터 챌린지 2018 대회 설명회

AI 기반 취약점 자동 탐지 트랙

「정보보호 R&D 데이터 챌린지 2018」 설명회

- AI 기반 취약점 자동 탐지 트랙

손 경 아 주임

한국인터넷진흥원 보안기술확산팀

2018. 7.27

KISA 한국인터넷진흥원

KISA

Contents

1. AI기반 취약점 자동 탐지 트랙 개요

개요

- 소프트웨어 바이너리의 취약점을 자동으로 탐지 및 공격하는 기술의 성능을 검증하는 정보보호 R&D 데이터 챌린지 2018「AI기반 취약점 자동 탐지」트랙 개최

세부 절차

01	02	03	04	05
신청 접수	학습(연습) 데이터셋 배포	시범대회 진행	본선 진행	시상식 및 세미나
✓ 신청서 접수 및 참가자격 확인	✓ 배포 방식은 추후 공지 예정	✓ 학습(연습) 데이터 추가 배포 및 시범 대회 약 3회 개최	✓ 본선 및 발표 진행	✓ 3개팀 시·포상 ✓ 공개 세미나 개최

2. 활용 데이터셋

취약점이 포함된 소프트웨어 바이너리 약 95개 활용

- 난이도는 상·중·하 로 구분하여 균등 분포, 학습(연습) 데이터셋 및 본선 데이터셋에 랜덤하게 배포
  ※ 난이도 구분 기준은 코드 복잡도 설정, 메모리 프로텍션 우회 기법 설정 등

학습(연습) 데이터셋

- ✓ (규모) 취약점 바이너리 약 45개 활용
- ✓ 학습(연습) 데이터셋을 활용하여 예선 기간 동안 3회의 시범 대회 개최 예정

※ 시범대회 결과를 바탕으로 본선 진출팀 선정

※ 참가팀이 8팀 이하일 경우 별도의 평가 없이 본선 직행

본선 데이터셋

- ✓ (규모) 취약점 바이너리 약 50개 활용
- ✓ 3 round 로 나누어 배포 예정

※ 라운드 당 배포 데이터셋 개수는 추후 공지

2-1. 활용 데이터셋 – 취약점 리스트

취약한 SW 바이너리 데이터셋 개발 과정에서 CWE 주요 취약점 리스트 적용

※ CWE (Common Weakness Enumeration) : 수집된 소프트웨어 취약점의 정의, 설명 등 정식 목록 분류체계

No	CWE-#	Vulnerability Name
1	CWE-77	Improper Neutralization of Special Elements used in a Command ('Command Injection')
2	CWE-78	Improper Neutralization of Special Elements used in an OS Command ('OS Command Injection')
3	CWE-119	Improper Restriction of Operations within the Bounds of a Memory Buffer
4	CWE-120	Buffer Copy without Checking Size of Input ('Classic Buffer Overflow')
5	CWE-121	Stack Based Buffer Overflow
6	CWE-122	Heap-based Buffer Overflow
7	CWE-123	Write-What-Where condition
8	CWE-124	Buffer Underwrite ('Buffer Underflow')
9	CWE-125	Out-of-bounds Read
10	CWE-126	Buffer Over-read
11	CWE-127	Buffer Under-read
12	CWE-128	Wrap-around Error
13	CWE-129	Improper Validation of Array Index
14	CWE-130	Improper Handling of Length Parameter Inconsistency
15	CWE-131	Incorrect Calculation of Buffer Size
16	CWE-134	Use of Externally-Controlled Format String
17	CWE-135	Incorrect Calculation of Multi-Byte String Length
18	CWE-158	Improper Neutralization of Null Byte or NUL Character

[CWE 리스트 예시]

3. 결과물 제출 및 심사기준

1. KEY 파일	✓ 취약점을 공격하여 얻어낸 KEY 파일 제출 ✓ 취약점 바이너리의 난이도를 고려하여 제출한 KEY 파일 점수 부여
2. 알고리즘 설명문서	✓ 취약점 자동 탐지 프로그램의 알고리즘 설명문서 제출 ✓ 자동화, 알고리즘 창의성 평가 ※ 알고리즘 설명문서를 검토하여 편법 사용 등 문제 발견 시, 수상에서 제외될 수 있음
3. 발표자료	✓ 알고리즘 설명문서를 바탕으로 발표자료 제출 ✓ 자동화, 알고리즘 창의성 평가
4. 자동 탐지 결과 보고서	✓ 본선 실행 결과를 바탕으로 12/1 까지 제출 ✓ 정답 근접성 등을 고려하여 심사위원회에서 추가 점수 부여 검토 ※ 예시) Exploit은 실패했지만, Crash 발생 방법이 참신하거나 유일한 경우 등

최종 스코어 (70%) + **발표 점수 (30%)**
(KEY 파일 점수 + 추가 점수)

4. 학습(연습) 데이터셋 활용 시범 대회 소개

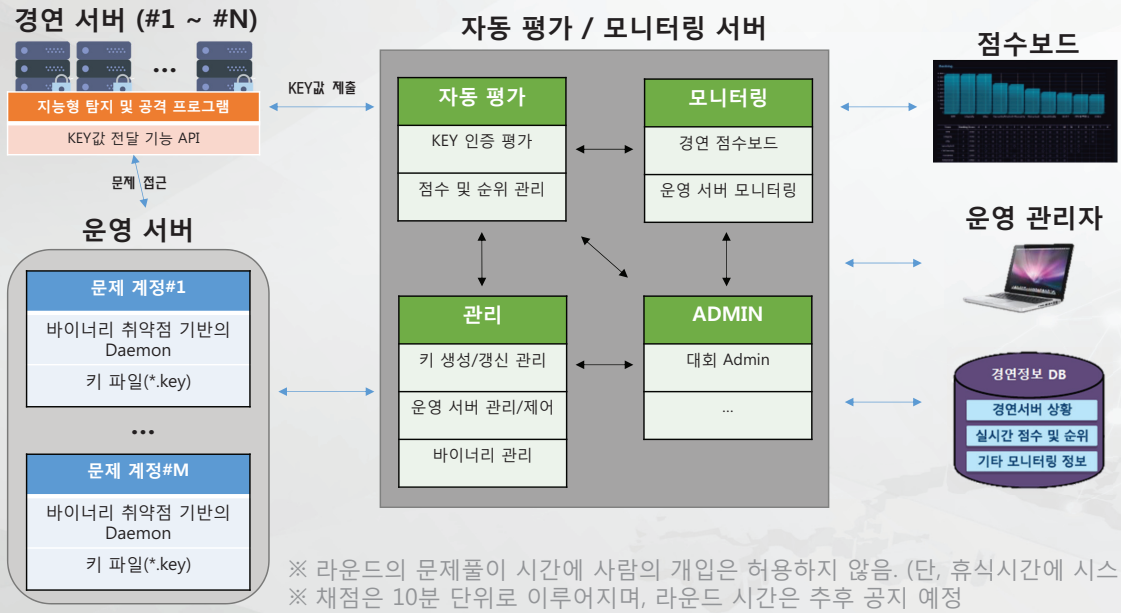
자동 탐지 및 공격 시스템의 성능 평가 및 알고리즘 개선을 위해 3회의 시범 대회 개최 예정

- 시범 대회 참가자들의 의견 및 대회 결과 등을 고려해 본선 규칙 확정 예정

- (일정) 10월 17일, 24일, 31일 수요일마다 개최 예정
  - ※ 시범 대회의 일정 및 횟수는 변경될 수 있으며, 신청팀에게 별도 안내 예정
- (1차 학습 데이터셋 배포) 신청접수 후 2~3일 내로 순차 배포
- (1차 시범 대회) 사전에 배포된 1차 학습(연습) 데이터셋을 이용하여 시범 대회 개최
- (2,3차 시범 대회) 추가 학습(연습) 데이터셋을 배포하여 시범 대회 개최

※ 시범 대회 결과를 바탕으로 본선 진출팀 선정
※ 참가팀이 8팀 이하일 경우 별도의 평가 없이 본선 직행

5. 본선 진행방식



※ 라운드의 문제풀이 시간에 사람의 개입은 허용하지 않음. (단, 휴식시간에 시스템 수정 가능)
※ 채점은 10분 단위로 이루어지며, 라운드 시간은 추후 공지 예정

6. 접수기간 및 신청절차

접수기간

2018.9.10(월) ~11.9(금)

신청절차

01 datachalleng.kr
- ✓ 정보보호 R&D 데이터챌린지 홈페이지 접속

02 「AI 기반 취약점 자동탐지 트랙」
- ✓ 트랙 선택

03 신청서류 다운 및 작성
- ✓ 참가신청서
- ✓ 약정서

04 rnd_dataset@kisa.or.kr로 제출
- ✓ 이메일로 신청서류, 소속증명서 제출

AI 기반 취약점 자동 탐지 트랙

감사합니다.

<문의처>
한국인터넷진흥원 손경아 주임
(T. 061-820-1256)
(E. rnd_dataset@kisa.or.kr)

정보보호학회 트랙 소개

2018 정보보호 R&D 데이터 챌린지 설명회
정보보호학회 트랙 소개

고려대학교 정보보호대학원
김 휘 강

2018. 7. 27.

KOREA UNIVERSITY

HCRL
Hacking and Countermeasure Research Lab

CONTENTS

소개

2018년 정보보호학회 담당 트랙 소개

- **주안점 – 연속성, 확장성**

- **2017년의 악성앱 탐지 트랙 ➔ 2018 AI기반 안드로이드 악성앱 탐지 트랙**
  - 안드로이드 악성앱 탐지 분야 트랙의 연속성 유지
  - 기 게재되어 검증된 논문과 dataset 활용, academic follow-up research 활성화
  - 해당 참고 논문 주저자가 직접 문제 출제 담당

- **2017년의 차량 이상징후 탐지 트랙을 고도화 ➔ 2018 차량주행 데이터 기반 도난탐지 트랙**
  - 차량 보안 분야 트랙의 연속성 유지
  - 차량용 IDS 개발 분야 외 “차량 도난탐지”, “개인화 서비스”, “운전자 식별 및 인증” 에 응용 가능
  - 기 게재되어 검증된 논문과 dataset 활용, academic follow-up research 활성화
  - 해당 참고 논문 주저자가 직접 문제 출제 담당

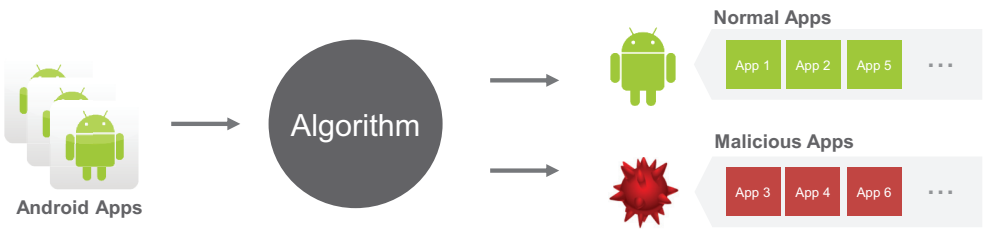
AI기반 안드로이드 악성앱 탐지 트랙

Challenge

AI기반 안드로이드 악성앱 탐지 알고리즘 개발

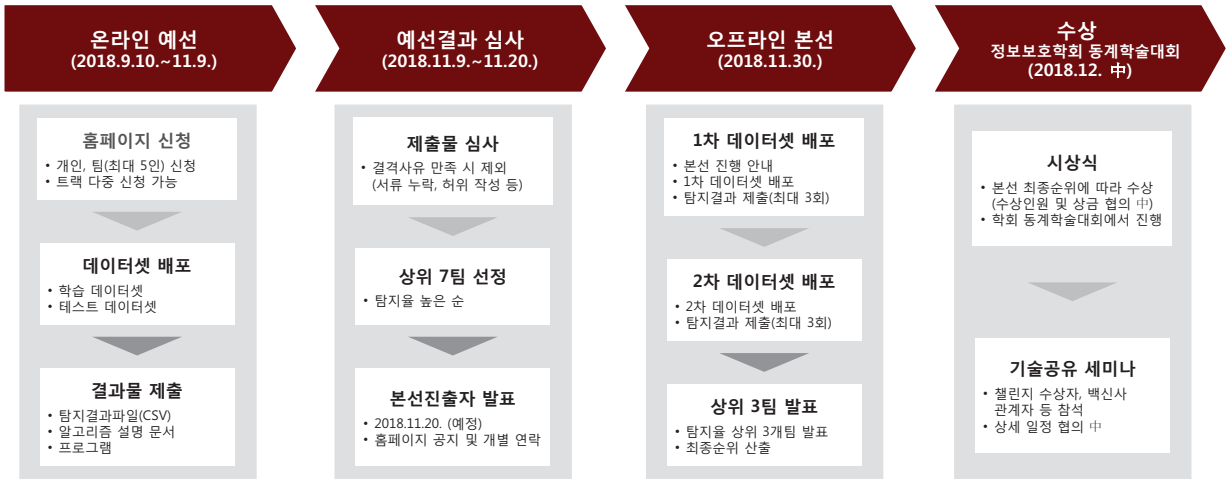
스마트폰 보안 위협에 대처하기 위해서는 수많은 앱에서 악성코드를 탐지해낼 수 있는 자동화된 솔루션이 필요합니다.

✔ **제공된 데이터를 분석하여 악성앱을 탐지할 수 있는 알고리즘을 제시하시기 바랍니다.**



운영방식

온라인 예선(9~11월 초 상시), 오프라인 본선(11/30) 진행



안드로이드 앱 데이터셋

정상앱, 악성앱이 혼합된 안드로이드 앱 데이터셋

□ Andro-profiler dataset

- 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* (2016) 5:273. Available: https://doi.org/10.1186/s40064-016-1861-x

□ **예선: 약 10,000개 안드로이드 앱**

- 학습 데이터셋과 테스트 데이터셋을 약 5:5 비율로 제공
- 학습 데이터셋만 클래스(정상/악성) 정보 제공

□ **본선: 예선에서 제공되지 않은 새로운 안드로이드 앱 2,000개 이상**

- 예선에서 사용되지 않은 정상/악성 앱 제공 (총 2차례)
- 2차 데이터셋 배포 시 1차 데이터셋에 대한 클래스(정상/악성) 정보 제공

구분	사용용도	안드로이드 앱	클래스 정보 여부
예선	학습용	5,000	제공
	테스트용	5,000	-
본선	1차 테스트용	1,000개 이상	2차 배포 시 제공
	2차 테스트용	1,000개 이상	-

* 데이터셋 규모는 조정될 수 있음

상세 진행방식

예선 (온라인)

□ **진행기간**
- 2018. 9. 10. ~ 2018. 11. 9. (상시 접수)

□ **신청방법**
- 데이터 챌린지 홈페이지를 통해 신청 양식 작성 및 제출 (datachallenge.kr)
- 신청자에 한해 예선 데이터셋 다운로드 URL 및 파일 비밀번호 배포

□ **결과물 제출**
- 탐지결과파일(CSV), 알고리즘 설명 문서, 프로그램
- 제출 방법은 사이트를 통해 추후 공지

□ **평가**
- 탐지정확도 100%
- 탐지정확도는 "**평가방법 – 탐지정확도**"를 따름
- 알고리즘 설명 문서와 프로그램은 치팅 여부 검증을 위한 목적으로만 활용

상세 진행방식

본선 (오프라인)

□ **본선 진출자 발표**
- 예선 결과물을 채점하여 탐지정확도 순으로 상위 7팀 선정
- 홈페이지 공지 및 개별 연락(이메일)

□ **진행일자**
- 2018. 11. 30. (금)

□ **본선 진행**
- 1차/2차로 나누어 총 2차례 테스트 데이터셋 배포
- 각 차수 별 최대 3회까지 탐지 결과 제출 가능
- 2차 데이터셋 배포 시 1차 데이터셋에 대한 클래스 정보 제공 → 탐지 모델 추가 학습 및 개선 기회 제공 목적
- 탐지정확도 상위 3개팀에 대해 본선 당일 발표 진행 (10분 발표, 5분 질의)

□ **평가**
- 탐지정확도 80%, 발표 점수 20% 합산
  - 탐지정확도
    - 1차, 2차 탐지정확도의 평균값으로 산정
    - 각 차수의 마지막 제출한 결과를 최종 점수로 반영
  - 발표
    - 문제 해결을 위한 방법론의 논리성, 창의성 위주로 채점

평가방법

탐지정확도(Accuracy)

카테고리		실제결과	
		Malware	Benign
실험결과	Malware	*True Positive(TP)*	*False Positive(FP)*
	Benign	*False Negative(FN)*	*True Negative(TN)*

- True Positive는 실제 Malware(True)를 Malware(True)로 정확하게 예측한 상황을 의미함
- True Negative는 실제 Benign(False)를 Benign(False)로 정확하게 예측한 상황을 의미함
- False Positive는 실제 Benign(False)을 Malware(True)로 예측한 상황을 의미함 (오탐)
- False Negative는 실제 Malware(True)를 Benign(False)로 예측한 상황을 의미함 (미탐)

$$\textit{악성앱 탐지} \quad Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

차량주행 데이터 기반 도난탐지 트랙

Challenge

주행 데이터 기반 차량 도난 탐지를 위한 운전자 분류 알고리즘 개발

✔ 제공된 차량 주행 데이터셋을 기반으로 아래 사항을 분류할 수 있는 알고리즘 및 프로그램을 제시하기 바랍니다.

- 기존 학습된 운전자에 대해 새로운 주행 데이터의 운전자 분류

사례

자동차를 타고 출근하는 K씨는 최근 자신의 차량 근처에 모르는 사람이 자주 보이는 것을 확인했습니다.
그로부터 며칠 후 **K씨는 출근하려고 할 때 자신의 차량이 없는** 것을 알게 되었습니다.

알고 보니 차량 도둑이 K씨의 차량에 접근해 차량을 탈취한 것입니다.
위 상황과 같이 **차량의 원거리 및 직접 접근을 통해 차량 도난이 발생하고 있어, 차량 도난을 효율적으로 탐지하는 방안이 필요합니다.**

제출물

- **프로그램**
  - GitHub 링크 제출 (Public 링크)
- **결과파일**
- **알고리즘 설명 문서**

운영방식

온라인 예선(9~11월 초 상시), 오프라인 본선(11/30) 진행

온라인 예선 (2018.9.10.~11.9.)

- **홈페이지 신청**
  - 개인, 팀(최대 5인) 신청
  - 트랙 다중 신청 가능
- **데이터셋 배포**
  - 학습 데이터셋
  - 테스트 데이터셋
- **결과물 제출**
  - 탐지결과파일(CSV)
  - 기술문서

예선결과 심사 (2018.11.9.~11.20.)

- **제출물 심사**
  - 결격사유 만족 시 제외 (서류 누락, 허위 작성 등)
- **상위 7팀 선정**
  - 탐지율 높은 순
- **본선진출자 발표**
  - 2018.11.20. (예정)
  - 홈페이지 공지 및 개별 연락

오프라인 본선 (2018.11.30.)

- **1차 데이터셋 배포**
  - 본선 진행 안내
  - 1차 데이터셋 배포(09:30~)
  - 탐지결과 제출(최대 3회)
- **2차 데이터셋 배포**
  - 2차 데이터셋 배포(13:00~)
  - 탐지결과 제출(최대 3회)
- **상위 3팀 발표**
  - 탐지율 상위 3개팀 발표
  - 최종순위 산출

수상 정보보호학회 동계학술대회 (2018.12. 中)

- **시상식**
  - 본선 최종순위에 따라 수상 (수상인원 및 상금 협의 中)
  - 학회 동계학술대회에서 진행
- **기술공유 세미나**
  - 챌린지 수상자, 관계자 등 참석
  - 상세 일정 협의 中

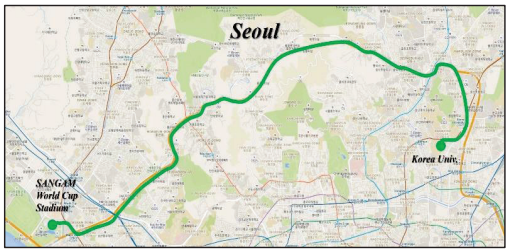
차량 주행 데이터셋

데이터셋 수집

□ **KU-Driving Dataset**

□ **주행 시 차량에 OBD-II 스캐너를 연결하여 주행 데이터를 추출**

- Features : 51개 (차량속도, 스티어링 휠 각도, 연료소모량, 미션오일 온도 등 추출 가능한 전체 feature)
  - 차량 : KIA Soul (S1), Hyundai Sonata (S2)
  - 주행코스 C1 : 한번 주행 시 약 17km (전체 720km)
  - 주행코스 C2 : 한번 주행 시 약 5.5km (전체 220km)



주행구간 (고려대학교 - 상암월드컵경기장) – 주행코스 C1

주행구간 (도심도로) – 주행코스 C2

차량 주행 데이터셋

데이터셋 구성

□ **KU-Driving Dataset**

- Byung Il Kwak, Jiyoung Woo and Huy Kang Kim, "Know Your Master: Driver Profiling-based Anti-theft Method", PST (Privacy, Security and Trust) 2016

□ **예선: 운전자 9명 (A-I), 차종 S1, 주행코스 C1 (전체 720 km 중 운전자별 비율 조정을 위해 610 km의 데이터셋 사용)**

- 분석용 데이터셋의 경우 데이터 분석을 위해 정답지가 포함된 데이터셋을 제공
- 결과 제출시 정답지가 없는 제출용 데이터셋의 사용 결과를 제출해야 함

□ **본선: 운전자 10명 (A-J), 차종 S2 (예선과 다른 차량), 주행코스 C1 (예선과 같은 도로) + 주행코스 C2 (예선과 다른 도심 도로)**

- 분석용 데이터셋의 경우 데이터 분석을 위해 정답지가 포함된 데이터셋을 제공
- 결과 제출시 정답지가 없는 제출용 데이터셋의 사용 결과를 제출해야 함

구분	사용용도	데이터셋 KU-Driving Dataset	운전자	총 주행 거리	차종	주행코스	주행 횟수
예선	예선 분석	1st	A – I (9명)	460 km	S1	C1	3
	예선 제출	1st-test	A – I (9명)	150 km	S1	C1	1
본선	1차 분석	2nd	A – J (10명)	460 km	S2	C1	3
	1차 제출	2nd-test	A – J (10명)	170 km	S2	C1	1
	2차 분석	3rd	A – J (10명)	165 km	S2	C2	3
	2차 제출	3rd-test	A – J (10명)	55 km	S2	C2	1

데이터셋 구성표

상세 진행방식

예선 (온라인)

□ **진행기간**
- 2018. 9. 10. ~ 2018. 11. 9. (상시 접수)

□ **신청방법**
- 데이터 챌린지 홈페이지를 통해 신청 양식 작성 및 제출 (datachallenge.kr)
- 신청자에 한해 예선 데이터셋 다운로드 URL 및 파일 비밀번호 배포

□ **결과물 제출**
- 탐지결과파일(CSV), 알고리즘 설명 문서, 프로그램
- 제출 방법은 사이트를 통해 추후 공지

□ **평가**
- 운전자 분류 정확도 점수 100%로 예선 평가
- 운전자 분류 정확도는 "**평가방법 – 분류 정확도**"를 따름
- 알고리즘 설명 문서와 프로그램은 치팅 여부 검증을 위한 목적으로만 활용

상세 진행방식

본선 (오프라인)

□ **본선 진출자 발표**
- 예선 결과물을 채점하여 탐지정확도 순으로 상위 7팀 선정
- 홈페이지 공지 및 개별 연락(이메일)

□ **진행일자**
- 2018. 11. 30. (금)

□ **본선 진행**
- 1차/2차로 나누어 총 2차례 테스트 데이터셋 배포
  - 본선에서 사용될 데이터셋은 전체 운전자 10명 (A – J)으로 구성
  - 주행코스는 예선과 동일한 코스 C1과 도심내 도로를 주행한 코스 C2로 구성
  - 분석용 데이터셋 – 3회 주행, 제출용 데이터셋 – 1회 주행
- 각 차수 별 최대 3회까지 탐지 결과 제출 가능
- 탐지정확도 상위 3개팀에 대해 본선 당일 발표 진행
  - 10분 발표, 5분 질의

□ **평가**
- 분류 정확도 80%, 발표 점수 20% 합산
  - 분류 정확도
    - 1차, 2차 분류 정확도의 평균값으로 산정
    - 각 차수의 마지막 제출한 결과를 최종 점수로 반영
  - 발표
    - 문제 해결을 위한 방법론의 논리성, 창의성 위주로 채점

평가방법

분류 정확도(Accuracy)

카테고리	운전자	운전자의 전체 주행 데이터수	실제결과	
			일치	비일치
실험결과	A	D_A	T_A	F_A
	B	D_B	T_B	F_B
	...	...	...	...
	J	D_J	T_J	F_J

*운전자 (A-J)는 예선과 본선에서 다르게 조정될 수 있음

- $D_A, D_B, \dots, D_J$ 는 운전자별 전체 주행 데이터 수
- $T_A, T_B, \dots, T_J$ 는 해당 운전자를 맞게 분류한 데이터 수
- $F_A, F_B, \dots, F_J$ 는 해당 운전자를 맞지 않게 분류한 데이터 수

$$\text{운전자 분류 정확도 (}Accuracy\text{)} = \frac{T_A + T_B + \cdots + T_J}{D_A + D_B + \cdots + D_J}$$

Thank You

KOREA UNIVERSITY

HCRL Hacking and Countermeasure Research Lab

정보보호 R&D 데이터 챌린지 2018 대회 설명회

대회 참여 경험 공유
–정성균 개인팀–

정보보호 R&D 데이터 챌린지 2017
대회 참여 경험 공유
-정성균 개인팀-

31th Master Course **Sungkyun Jung**

Digital Forensic Research Center
Center for Information Security Technologies
Korea University

Digital Forensic Research Center
Center for Information Security Technologies in Korea University

목차

- **Content**
  - Motivation
  - Experience of the competition
  - Machine Learning In Security
  - Q & A

Motivation

- **Machine Learning에 대한 호기심**
  - AlphaGo vs. Lee Sedol
    - Google DeepMind 개발자들의 바둑에 대한 이해도
    - 인간이 생각지도, 시도해보지도 못한 새로운 접근법
    - AlphaGo의 패배
- **온전히 개인 관심사에 기초한 프로젝트 수행 의지**
  - 익숙한 분야에서의 응용
    - PE Malware Detection
  - 2017 정보보호 R&D 데이터 챌린지 개최
    - 악성코드(PE File Format) 탐지 트랙 참여

Experience of the competition

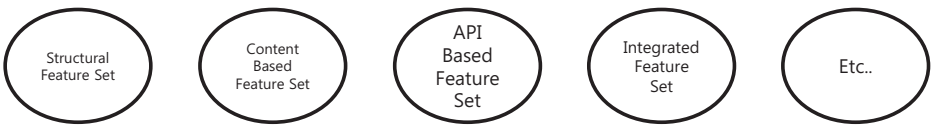
- **ML 기반 PE Malware 탐지 관련 논문 (2010 ~ 2017) 다수 참조**
  - 현행 연구 파악 및 개선 방안 모색
- **주요 한계점**
  - 학습에 사용된 Dataset의 전처리 부족
    - Malware 및 Benign의 개수 차이가 매우 큼
    - 중복된 바이너리 다수
  - Non-practical approach
    - Benign 및 Malware의 분석 방해 기법, 탐지 우회 기법에 대해 고려하지 않음
    - 절대적인 Decompressor를 Model에 포함
  - 제한적인 Feature 사용
  - 모든 Paper에서의 주된 Appeal 요소는 높은 탐지율
    - High Accuracy = Best Model ?

Experience of the competition

- **복합적 Feature 사용**
  - Issue
    - 다양한 Feature 추출
      - 특정 Feature에만 의존할 경우, 해당 정보가 거짓(Noise)인 상황에 대응할 수 없음
      - 최대한 많은 Feature를 추출하여 복합적 분석 필요
  - Challenges
    - Curse of dimensionality
      - Feature의 개수를 늘린다는 것은 차원 증가를 의미
      - 데이터의 차원이 증가할수록 모델 추정에 필요한 샘플 데이터의 개수가 기하급수적으로 증가 -> 효율적인 학습 기대 어려움

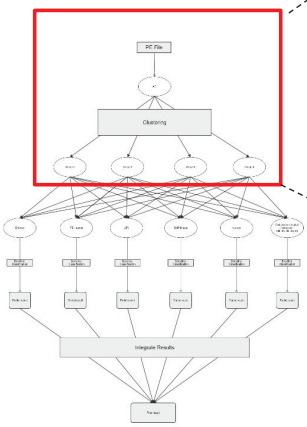
Experience of the competition

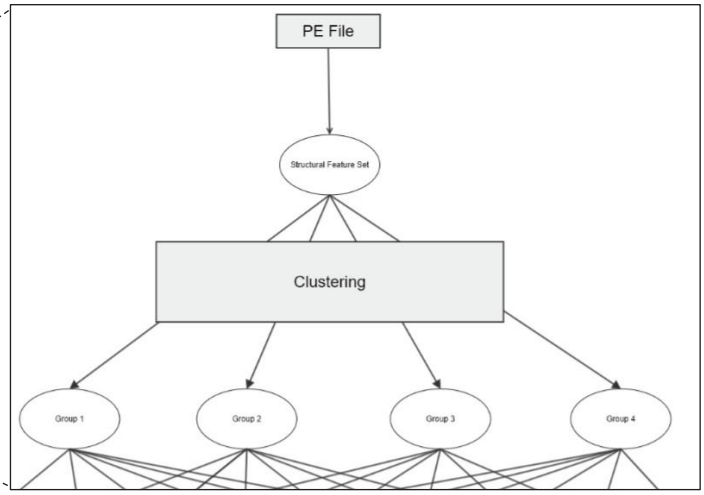
- **복합적 Feature 사용**
  - Solution
    - Multiple feature sets
      - 많은 Feature들을 성질이 비슷한 것들끼리 set으로 구성하여 독립적으로 학습
      - Curse of dimensionality
        - 학습 대상은 단일 feature set, 차원의 증가 없이 모델 추정 가능
      - 복합적 Feature의 활용 문제
        - Feature set 마다 가지는 고유의 성질을 반영 가능
        - 독립적인 Feature set의 학습 결과가 종합되어 최종 추론



Experience of the competition

- **Phase 1 - Clustering**

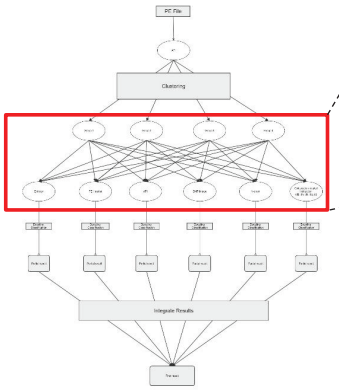


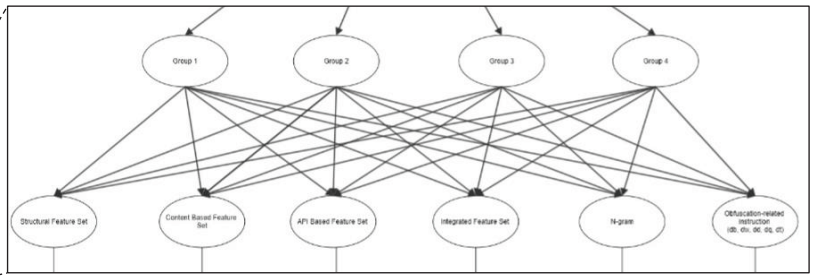


- **Clustering**
  - Train : Dataset에서 유사한 PE 파일들을 구조적인 특징을 기반으로 Grouping 하여, Centroid 값을 도출
  - Test : 여러 Centroid 값 중 가장 가까운 Group과 매핑
  - 더욱 세분화 된 학습 및 추론 가능 (Phase 2에서 부연 설명)

Experience of the competition

- **Phase 2 – Train multiple feature sets in each group**

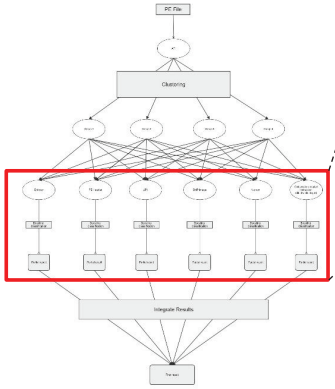


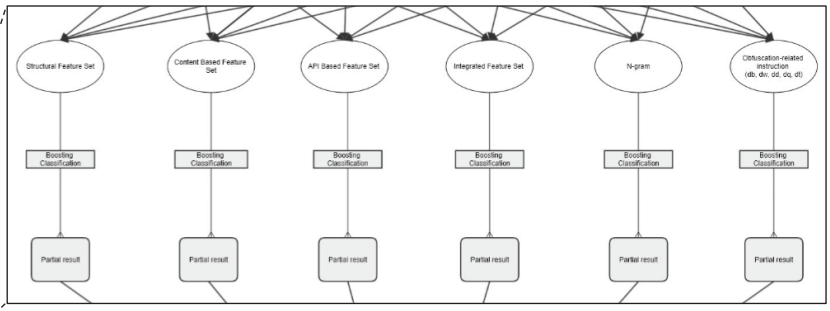


- **Train multiple feature sets in each group**
  - Group 각각이 정의 된 Feature set을 독립적으로 학습
    - Group 마다 각 Feature set에 부여하는 가중치가 달라짐
    - 즉, Input PE File의 특징(Packing, Obfuscation, Action, etc..)에 따라 Feature set의 중요도를 평가
  - 특정 Feature set이 효력이 없을 경우(정확도가 낮을 경우), 가중치가 낮아지며 다른 Feature set에서 대응
    - 기존 연구에서의 제한적인 Feature 사용으로 인한 문제를 보완

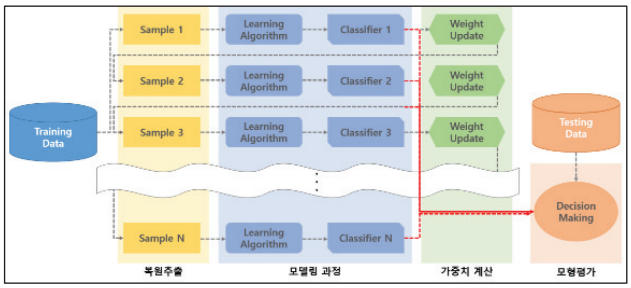
Experience of the competition

- **Phase 3 – Boosting classification in each feature set**



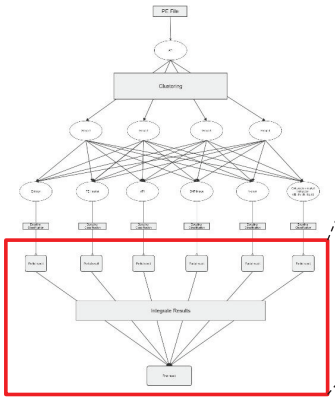


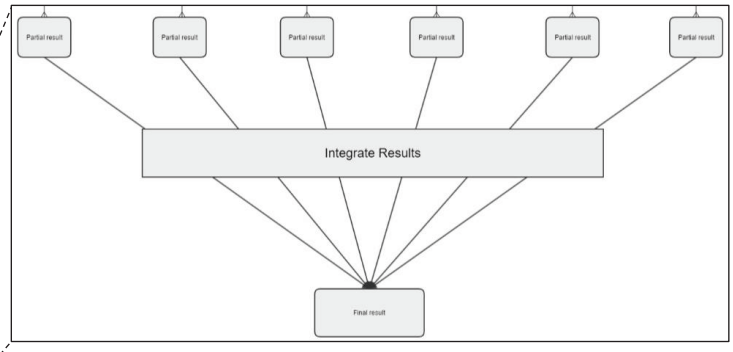
- **Boosting classification in each feature set**



Experience of the competition

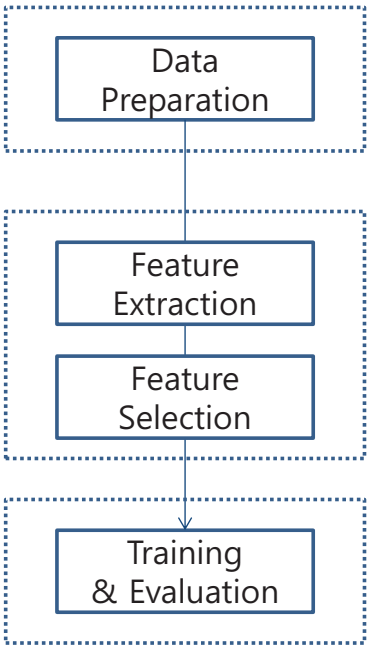
- **Phase 4 – Final result according to importance of feature set**





- **Final result according to importance of feature set**
  - Train 단계에서 Feature set의 Boosting classification의 정확도를 기준으로 비교하여 Group 별 Partial result의 가중치를 결정
  - 최적화 된 가중치를 각 Partial result(True or False)에 반영하여 Final result를 추론

Machine Learning In Security



- 가장 중요하고 가장 어려운 단계
  - 아무리 훌륭한 Feature set과 Classifier를 사용하더라도 학습 데이터가 잘못 구성돼 있다면, 결코 좋은 Model을 생성해낼 수 없음
  - Ex) 잘못된 Labeling, 편향된 Dataset, 단순한 Dataset
- Training Set을 어떻게 구성하느냐에 따라 Model의 방향성이 달라짐
  - Modeling을 하는 사람의 주관에 의해 결정돼야 함(정답 X)
  - Adware, Riskware, PUP/PUA = Malware or Benign?

- 보안 분야의 ML은 타 분야에 비해 Noise 데이터(Obfuscation, Packing, Encryption, Etc...)를 다루게 될 확률이 매우 높음
  - Benign은 기술 보호를 위해 정보를 은닉
  - Malware는 탐지 우회를 위해 정보를 은닉
- 학습 및 판단 대상은 Binary file이 아닌, Binary file을 해석한 결과
  - 보안 관점의 Binary file 분석 기술이 기반 되어야 함
  - 은닉 기법에 대한 동향 파악 및 대응 방안 연구 필요
  - 서로 다른 Binary file에 대해 상이한 해석 결과가 필요

- 쉽게 활용 가능한 ML Library (Scikit-learn, tensorflow, keras, etc...)
  - 적은 수의 Code로도 Training & Evaluation 가능
  - 효율적인 다양한 Classifier 제공
- Evaluation 방법에 대한 고민 필요
  - Training이 잘 되었는지 어떻게 확인해야 하는가?

Thank you for listening