KiSTi 한국과학기술정보연구원
Korea Institute of Science and Technology Information
www.kisti.re.kr

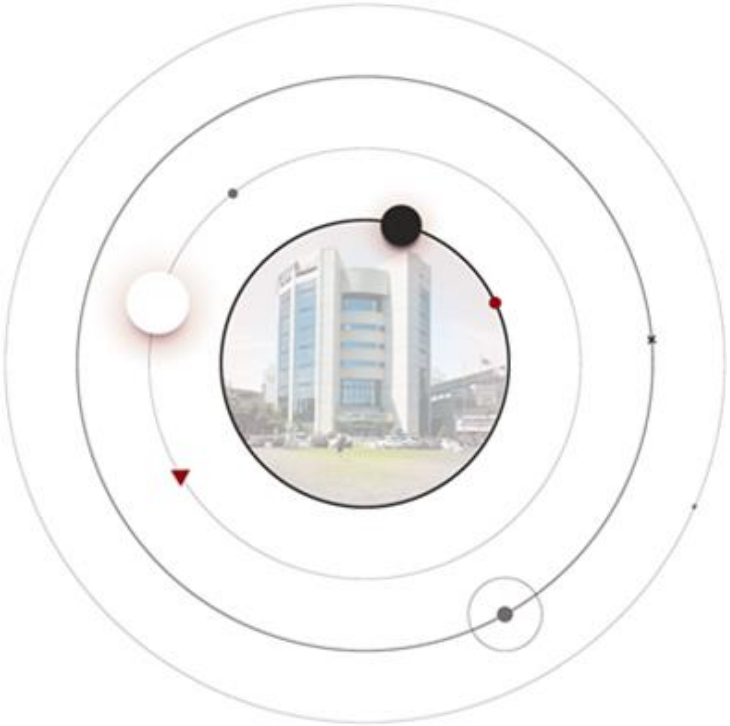
과학기술사이버안전센터
Sience and Technology Cyber Security Center

NetSec-KR 2020

보안관제를 위한 AI 모델 기술 소개

2020. 07. 17

과학기술사이버안전센터 송중석 팀장



01 KISTI 과학기술사이버안전센터 소개 (1/2)

국가·공공 보안관제체계 국가사이버안전센터(NCSC)를 중심으로 분야별 부문보안 관제센터(41개)를 운영

과학기술정보통신부 S&T-CSC 과학기술분야	교육부 교육 사이버안전센터 교육분야	해양수산부 해양수산 사이버안전센터 해양·수산 분야	관세청 관세부문 보안관제센터 관세분야	국세청 국세청 사이버안전센터 국세분야	금융위원회 금융 ISAC 금융분야	경찰청 경찰 정보보호 관제센터 본부 및 지방청
고용노동부 고용노동 사이버안전센터 고용·노동 분야	기상청 기상청 사이버안전센터 기상분야	NCSC(National Cyber Security Center) 국가사이버안전센터 국가 공공분야 전체 총괄			문화체육관광부 문화체육관광 사이버안전센터 문화체육·관광 분야	미래창조과학부 방송통신 사이버안전센터 방송·통신 분야
방위사업청 방위사업청 관제센터 방위사업분야	보건복지부 보건복지 사이버안전센터 보건 · 의료 분야	국가보안기술연구소 보안관제기술 지원센터 정부출연연구소	국방부 사이버사령부 국방분야	안전행정부 사이버침해대응 지원센터 지방자치단체	미래창조과학부 인터넷침해 대응센터 민간분야	기획재정부 재정경제 사이버안전센터 재정 · 경제 분야
안전행정부 정부통합 전산센터(대전) 행정기관	안전행정부 정부통합 전산센터(광주) 행정기관	산업통산자원부 산업통상자원 사이버안전센터 에너지·무역·산업 분야	금융위원회 증권 ISAC 증권분야	특허청 특허청 보안관제센터 특허분야	환경부 환경부 사이버안전센터 환경분야	외교부 외교 사이버안전센터 외교분야

부문보안관제센터 관제체계 (공통) 탐지규칙 기반 장비 활용

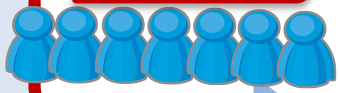
01 KISTI 과학기술사이버안전센터 소개 (2/2)

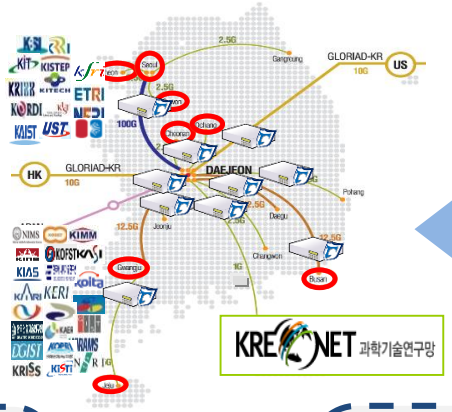
센터 업무 현황

과학기술분야 61개 공공/연구기관에 대한 종합적 · 체계적 정보보호 서비스 제공

일평균 1,700만 건 이상의 위협정보 수집 · 분석

수동분석 중심





보안관제 서비스

보안관제 프로세스

탐 지 → 분 석 → 대 응

SMARTer | VIZCosmos VIZSpacer | 사이버 예/경보 | DNS 싱크홀 | 홈페이지 위·변조 탐지

자체개발 시스템

침해 예방 및 대응 서비스


웹사이트 자가진단

사이버 모의훈련

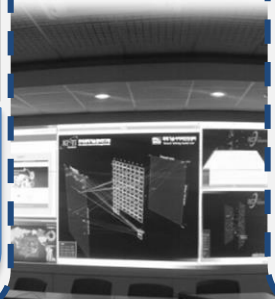
침해사고 현장점검

웹사이트 보안수준 자가진단 솔루션

자체개발 시스템

DDoS 훈련 장비 | 악성메일 훈련 장비

도입 장비



S&T-CSC

서비스 운영 목적의 실용연구·개발 수행

원천기술 자체 연구·개발


가시화 연구


자동분석 연구


악성코드 수집

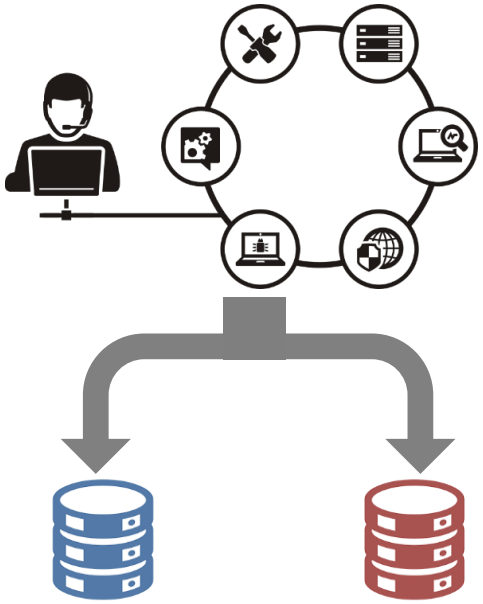

악성행위 수집

보안관제 데이터 개요

- 수집기간 : 2017.01.01.~2018.12.31.

  ※ ‘19, ‘20 데이터 구축 중

- 수집대상 : 61개 보안관제 대상기관
- 보안이벤트 건수 : **약 120억 건**(1,459종)
- 일평균 : **약 1,700만 건**
- 일평균 사고처리 건수 : **3.2건**



중복 제거 및 축약

약 4억 2천만 건

임계치 기반 이벤트 제거
및
데이터 샘플링

**최종 데이터 건 수 :
약 5천 6백만 건**

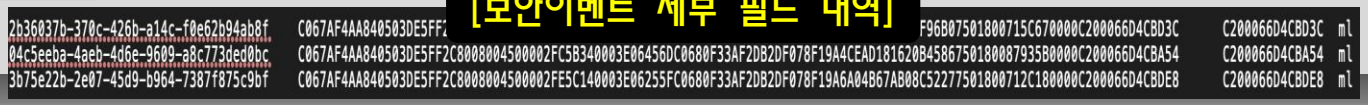
◆ 데이터의 전처리 필요

암호문자, 특수문자, 무의미 문자 등

Payload 분석 및 학습 데이터 품질 향상을 위한 전처리 기술 적용

보안 데이터

[보안이벤트 세부 필드 내역]



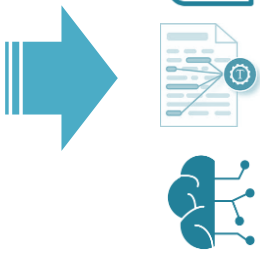
암호화 데이터	불용 문자열	민감 데이터
복호화 불가능으로 인한 불용 처리 필요	페이로드 미보유/무의미 기호 등 불용 문자열 처리 필요	민감 데이터 마스킹 필요 (개인정보(주민번호 등), 기관 IP정보, 관제기관 특정 가능 정보 등)

< Noise 발생 요소 포함 >

전처리 기술 적용

자연언어처리(NLP) 기술 적용 가능성 확보

```
"jsonrpc":"2.0","method":"login","params":
 {
  "login":" ABC@abc.com",
  "pass":"x",
  "agent":"XMRig/2.5.3
  (Windows NT 10.0; Win64; x64)
  libuv/1.14.1 msvc/2017"}}
```



- 보안관제전용 Dictionary 구축 가능
- 보안관제전용 단어 Word embedding 구축 가능
- Payload 정보가공을 통한 AI모델 성능 향상 도모
- ⋮

전처리 과정별 결과

원본데이터

00005E000114009008D891408004500001EB000040003D0670CD81FEAA7CDD9EC126E1BB07E573374DADA789068A80180804D1460000010108 0A69779D850D7CC4E17B226D6574686F64223A22746F7272656E742D6765 74222C226176775 6D656E7473223A7B226669656C643223A5B226942 22C226572726F72222C226572726F7253747269 6E67222C226574612 2C2269734696E6973686564222C2269735374616C6C6564222C226C6566745 6E74696C446F6E65222C226D6574616461615065726356E74436F6D 706C6574652222C22706567273436F6EB6E6563746564222C2270656573 4765747469 6E6746726F6D573222C22706565723565 6E6496E67546F 5573222C22706572636E6E7446F6E65222C22717567565506F73697469 6F6E222C2272617465446F776D6C6F6164222C22726174555706C6F616 4222C22726563686563 6B50726F6772657373222C22736565645261746 96F4D6F6465222C22736565645261746 96F4C696D6974222C2273697A6557 6865 6E446F6E65222C22737461747573222C22747261636 36B657273222C22 646F776E6C6F6164446972222C2275706C6F6164656445766572222C2275 706C6F6164526174696F222C22776562736565647353656E6496E67546F 5573225D2C2269647323A22726563656E746C792D6163746976657227D7D

Step 1

\x00\x00^\x00\x01\x14\x00\x90\x0bM\xb9\x14\x08\x00E\x00\x01\ xeb\x00\x00@\x00-\x06p\xcd\x81\xfe\xaa\xdd\x9e\xc1&\xe1\xbb\ x07\xe5s7M\xad\xa7\x89\x06\x8a\x80\x18\x08\n\xd1F\x00\x00\x0 1\x01\x08\nw\x9d\x85\r\xc4\xe1{"method":"torrent-get","argument s":{"fields":["id","error","errorString","eta","isFinished","isStalled","leftUntilDone","metadataPercentComplete","peersConnected","peersGettingFromUs","peersSendingToUs","percentDone","queuePosition","rateDownload","rateUpload","recheckProgress","seedRatioMode","seedRatioLimit","sizeWhenDone","status","trackers","downloadDir","uploadedEver","uploadRatio","webseedsSendingToUs"],"ids":"recently-active"}}

Step 2

\x00\x00^\x00\x01\x14\x00\x90\x0bM\xb9\x14\x08\x00E\x00\x01\ xeb\x00\x00@\x00-\x06p\xcd\x81\xfe\xaa\xdd\x9e\xc1&\xe1\xbb\ x07\xe5s7M\xad\xa7\x89\x06\x8a\x80\x18\x08\n\xd1F\x00\x00\x0 1\x01\x08\nw\x9d\x85\r\xc4\xe1{"method":"torrent-get","argument s":{"fields":["id","error","errorString","eta","isFinished","isStalled","leftUntilDone","metadataPercentComplete","peersConnected","peersGettingFromUs","peersSendingToUs","percentDone","queuePosition","rateDownload","rateUpload","recheckProgress","seedRatioMode","seedRatioLimit","sizeWhenDone","status","trackers","downloadDir","uploadedEver","uploadRatio","webseedsSendingToUs"],"ids":"recently-active"}}

Step 3

\ \ ^\ \ \ \ \ M\ \ \ \ \ \ \ \ @\ =\ p\ \ \ \ /\ \ \ &\ \ \ \ s7M\ \ \ \ \ \ \ \n\ \ \ \ \ \ \nw\ \ \r\ \ {"method":"torrent-get","arguments":{"fields":["id","error","errorString","eta","isFinished","isStalled","leftUntilDone","metadataPercentComplete","peersConnected","peersGettingFromUs","peersSendingToUs","percentDone","queuePosition","rateDownload","rateUpload","recheckProgress","seedRatioMode","seedRatioLimit","sizeWhenDone","status","trackers","downloadDir","uploadedEver","uploadRatio","webseedsSendingToUs"],"ids":"recently-active"}}

Step 4

\ \ ^\ \ \ \ \ M\ \ \ \ \ \ \ \ @\ =\ p\ \ \ \ /\ \ \ &\ \ \ \ s7M\ \ \ \ \ \ \ \n\ \ \ \ \ \ \nw\ \ \r\ \ {"method":"torrent-get","arguments":{"fields":["id","error","errorString","eta","isFinished","isStalled","leftUntilDone","metadataPercentComplete","peersConnected","peersGettingFromUs","peersSendingToUs","percentDone","queuePosition","rateDownload","rateUpload","recheckProgress","seedRatioMode","seedRatioLimit","sizeWhenDone","status","trackers","downloadDir","uploadedEver","uploadRatio","webseedsSendingToUs"],"ids":"recently-active"}}

Step 5

M @ p s7M
n niw r method torrent-get
arguments fields id error errorString eta isFinished
isStalled leftUntilDone metadataPercentComplete
peersConnected peersGettingFromUs peersSendingToUs
percentDone queuePosition rateDownload rateUpload
recheckProgress seedRatioMode seedRatioLimit sizeWhenDone
status trackers downloadDir uploadedEver uploadRatio
webseedsSendingToUs ids recently-active

Step 6

M @ p s7M
n niw r method torrent-get
arguments fields id error errorString eta isFinished
isStalled leftUntilDone metadataPercentComplete
peersConnected peersGettingFromUs peersSendingToUs
percentDone queuePosition rateDownload rateUpload
recheckProgress seedRatioMode seedRatioLimit sizeWhenDone
status trackers downloadDir uploadedEver uploadRatio
webseedsSendingToUs ids recently-active

Step 7

m @ p s7m
n niw r method torrent-get
arguments fields id error errorstring eta isfinished
isstalled leftuntildone metadatapercentcomplete
peersconnected peersgettingfromus peerssendingtous
percentdone queueposition ratedownload rateupload
recheckprogress seedratiomode seedratiolimit sizewhendone
status trackers downloaddir uploadedever uploadratio
webseedssendingtous ids recently-active

Step 8

m p s7m
n niw r method torrent-get arguments
fields id error errorstring eta isfinished isstalled
leftuntildone metadatapercentcomplete peersconnected
peersgettingfromus peerssendingtous percentdone
queueposition ratedownload rateupload recheckprogress
seedratiomode seedratiolimit sizewhendone status trackers
downloaddir uploadedever uploadratio webseedssendingtous
ids recently-active

Step 9

m p s7m
n niw r method torrent-get arguments
fields id error errorstring eta isfinished isstalled
leftuntildone metadatapercentcomplete peersconnected
peersgettingfromus peerssendingtous percentdone
queueposition ratedownload rateupload recheckprogress
seedratiomode seedratiolimit sizewhendone status trackers
downloaddir uploadedever uploadratio webseedssendingtous
ids recently-active

Step 10

m p s7m
n niw r method torrent-get arguments
fields id error errorstring eta isfinished isstalled
leftuntildone metadatapercentcomplete peersconnected
peersgettingfromus peerssendingtous percentdone
queueposition ratedownload rateupload recheckprogress
seedratiomode seedratiolimit sizewhendone status trackers
downloaddir uploadedever uploadratio webseedssendingtous
ids recently-active

Step 11

m p s7m
n niw r method torrent-get arguments
fields id error errorstring eta isfinished isstalled
leftuntildone metadatapercentcomplete peersconnected
peersgettingfromus peerssendingtous percentdone
queueposition ratedownload rateupload recheckprogress
seedratiomode seedratiolimit sizewhendone status trackers
downloaddir uploadedever uploadratio webseedssendingtous
ids recently-active

Step 12

s7m niw method torrent-get arguments fields id error errorstring eta isfinished isstalled leftuntildone metadatapercentcomplete peersconnected peersgettingfromus peerssendingtous percentdone queueposition ratedownload rateupload recheckprogress seedratiomode seedratiolimit sizewhendone status trackers downloaddir uploadedever uploadratio webseedssendingtous ids recently-active

Step 13

s7m niw method torrent-get arguments fields id error errorstring eta isfinished isstalled leftuntildone metadatapercentcomplete peersconnected peersgettingfromus peerssendingtous percentdone queueposition ratedownload rateupload recheckprogress seedratiomode seedratiolimit sizewhendone status trackers downloaddir uploadedever uploadratio webseedssendingtous ids recently-active

보안관제 데이터의 문제점 – 전처리 요약

인지가능 문자로 치환,
암호화 문자 제거 등

URL 암호화 해제 등

불필요 서식/특수문자 제거
(단순, 연속, 패턴, 조합/결합 등)

무의미/개인정보 단어 제거

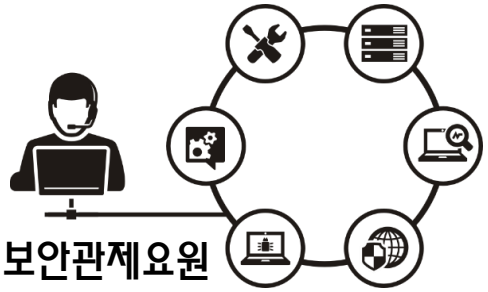
단어추출 대상 범위 확대
(쿠키, 세션 등)

◆ 데이터의 일관성 부족

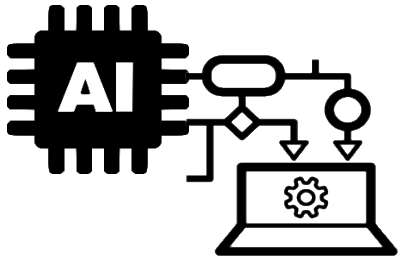
휴먼 에러, 환경 차이, 판단 기준 상이 등

동일한 이벤트(공격 or 정상/오탐)의 경우에도,

- **다른 분석가**에 의한 다른 분석 결과
  분석 기술의 차이, 판단 정책 상이, 분석 대상 누락, 기타 휴먼 에러 등
- **동일 분석가**에 의한 다른 분석 결과
  분석 정보의 추가, 판단 근거/정책 변경, 분석대상 누락, 기타 휴먼 에러 등
- **시간차**에 따른 다른 분석 결과
  특정 시점에서 판단 근거/정책 변경 등
- **내 · 외부 환경차이**에 따른 다른 분석 결과
  공격/피해 시스템의 다른 OS/서비스/포트/네트워크 구성 등
- **Labeling 기준**에 따른 다른 분석 결과
  특정 시점의 공격이벤트와 동일한 前後 이벤트를 추출 시,
  판단 기준(IP, 이벤트명, 페이로드 등) 상이, 前後 시간 정의 상이 등



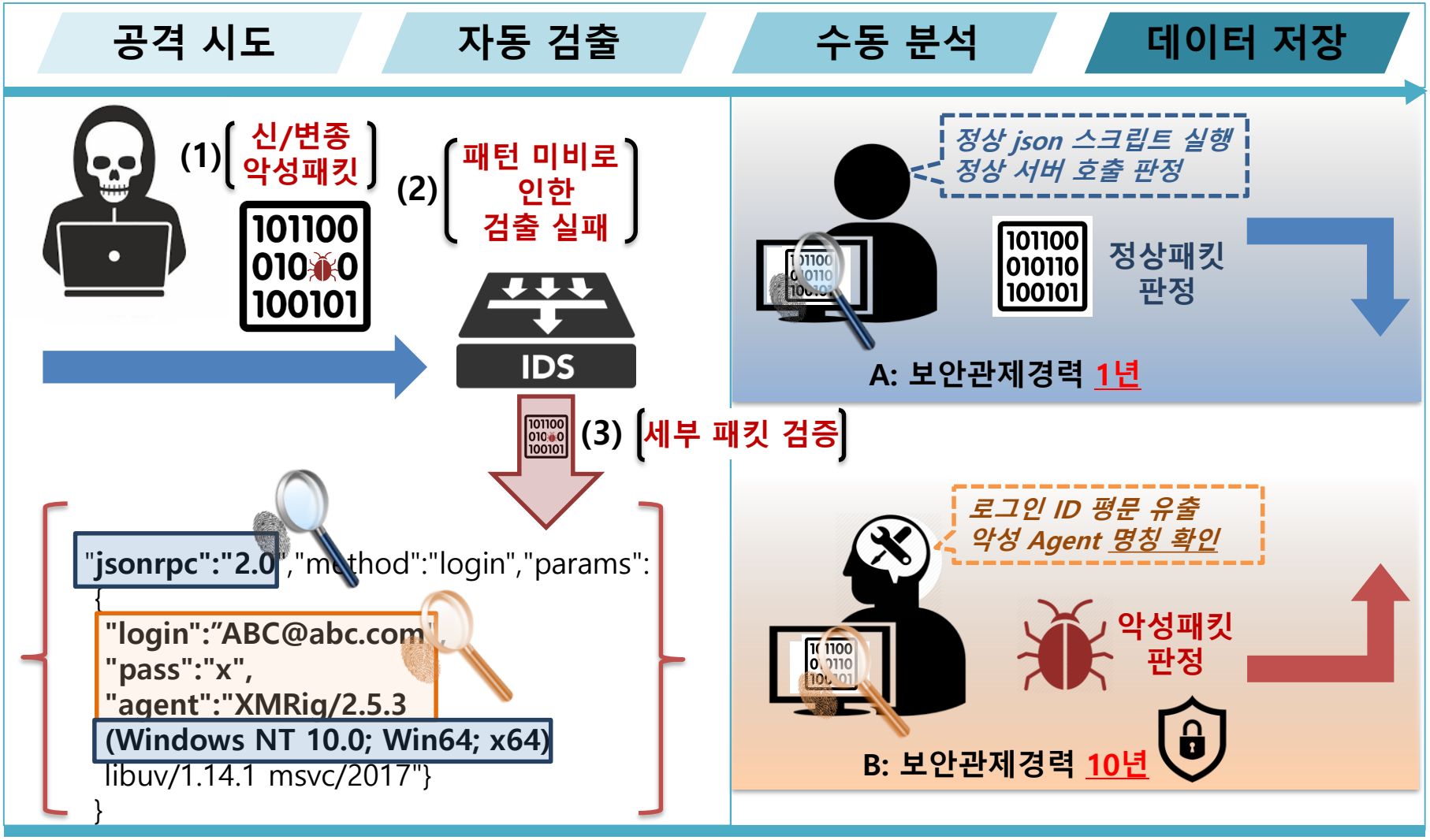
데이터 처리기준 상이
= 데이터 불일치 발생



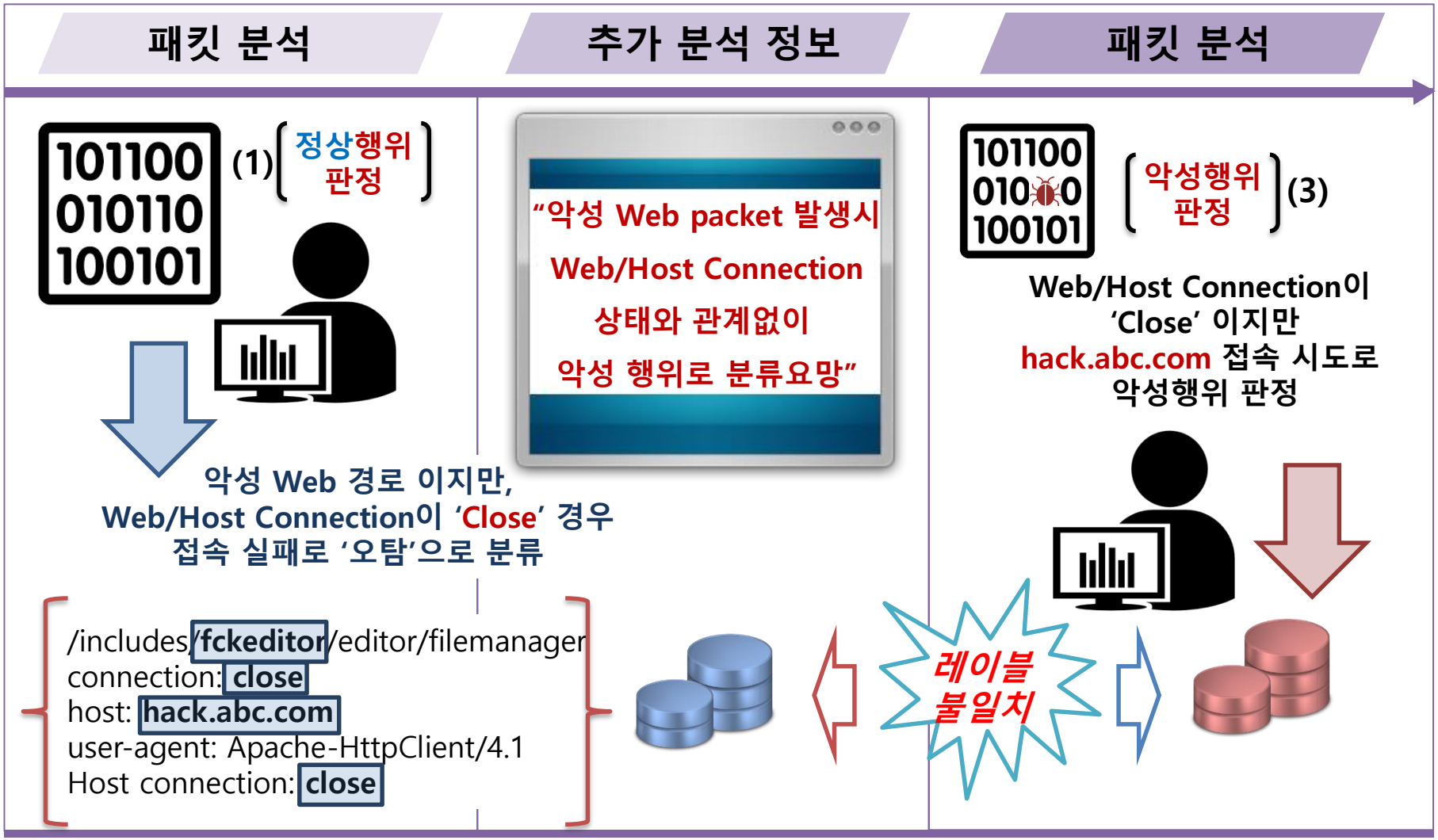
인공지능/머신러닝 모델

○ 다른 분석가에 따른 다른 분석 결과

– 예) 분석 기술의 차이, 판단 정책 상이, 분석 대상 누락, 기타 휴먼 에러 등



○ **동일 분석가에 따른 다른 분석 결과**
- 예) 분석 정보의 추가, 판단 근거/정책 변경, 분석대상 누락 등

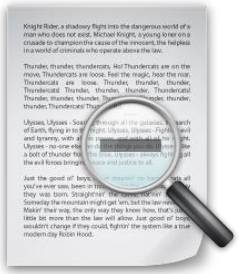


○ **시간차**에 따른 다른 분석 결과
- 예) 특정 시점에서 판단 근거/정책 변경 등

패킷 분석(3월) | 정책 변경 (4월) | 패킷 분석(6월)

(1) [악성행위 판정]

(2) [정책 갱신]



정상 경유지 확인 및 등록
'http://www.abc.com/exe/'

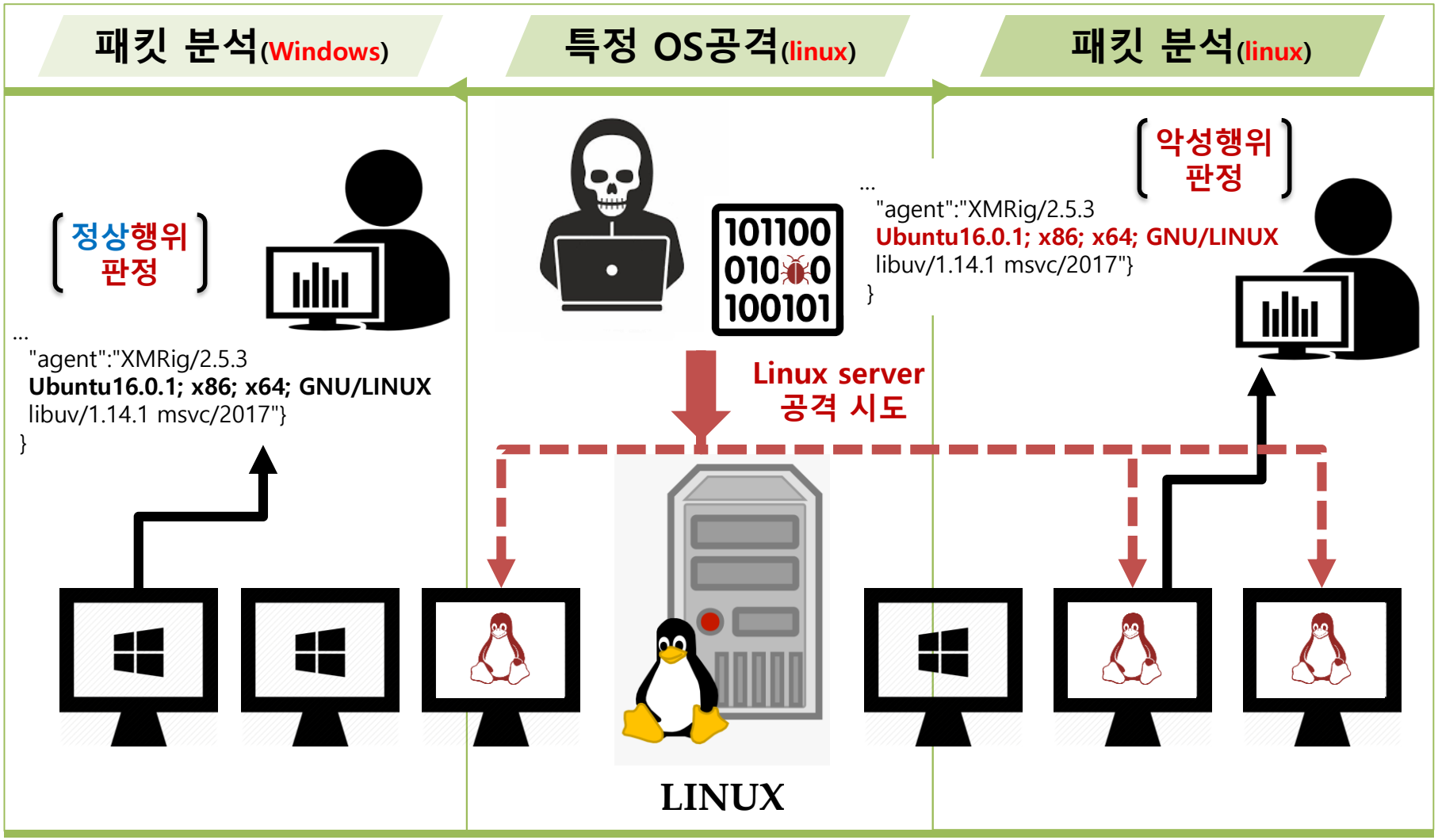
[정상행위 판정] (3)

악성 경유지 등록 주소 ➔ 정탐 DB

Accept: image/png, image/svg+xml
Referer: **http://www.abc.com/exe/**
Accept-Language: ko-KR
User-Agent: Mozilla/5.0
Accept-Encoding: gzip, deflate
Host: hack.abc.com
If-Modified-Since: Tue, 01 Oct

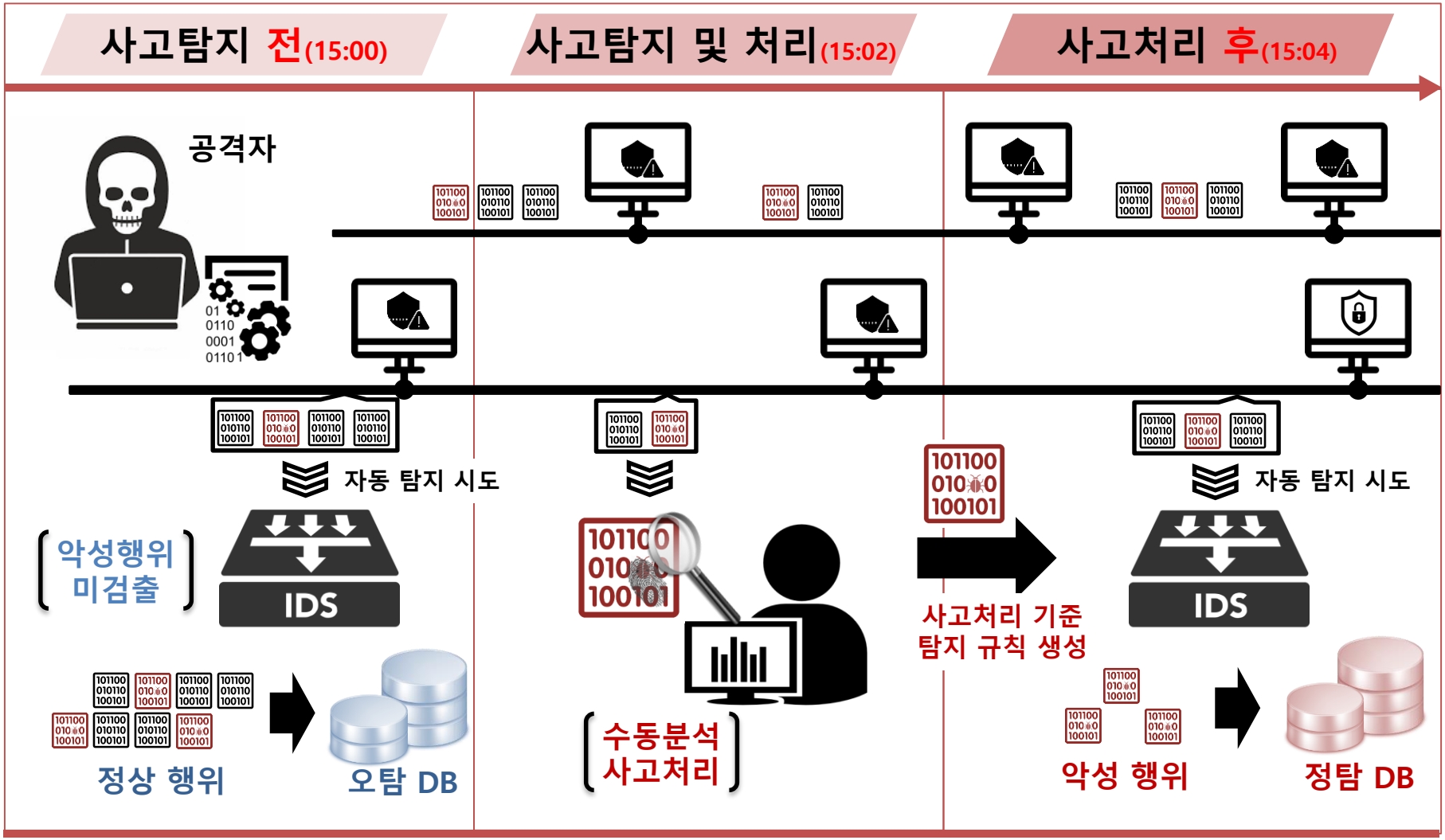
정상 경유 행위 ➔ 오탐 DB

○ **내 · 외부 환경차이**에 따른 다른 분석 결과
- 예) 공격/피해 시스템의 다른 OS/서비스/포트/네트워크 구성 등



○ Labeling 기준에 따른 다른 분석 결과

– 예) 특정 시점의 공격이벤트와 동일한 前後 이벤트를 추출 시,
판단 기준(IP, 이벤트명, 페이로드 등) 상이, 前後 시간 정의 상이 등



◆ 보안관제 데이터 일관성 확보 개념 및 절차

3-Steps 데이터 일관성 확보 과정

사고처리 기반 악성 행위 추출

사고처리/지원처리 이력으로부터
악성행위 데이터(정탐셋) 추출

- 사고처리/지원처리 이력 기록 재검토
- 정탐 데이터 선별 (Baseline 데이터)
- 사고처리 후 회신된 판별 데이터 재정렬

악성행위 데이터 기반 유사도 검출

악성 행위 데이터와
기존 DB간 유사도 계산

- 유사도 검출 비교대상 특징(feature) 선정
  - Src IP, Dst IP, Port, Direction, Payload(header 제외)
- 통계기반 유사도 검출 알고리즘 활용 계산
  - (TF-IDF, Cosine 유사도, Pearson 계수 등)

HITL 기반 정합 기준 유사도 확립

Human In The Loop 기반
분류 기준 유사도 임계치 검증

- 유사도 비율 별 정/오탐 데이터 분류
  - (90%→60%, 10% 단위)
  - 90% 80% … 60%
- 상위 비율에서 하위 비율로 비교 수행을 통한 최적 유사도 임계치 검출

◆ 보안관제 데이터 일관성 확보 방법론

Step1. 사고처리 기반 악성 행위 데이터 추출 결과 및 통계

- **2017 ~ 2018 사고처리 건수(정탐 판단 건 수) : 총 1062건**

※ 학습데이터로 사용하기 위해 선별한 정탐 건 수

Step2. 악성 행위 데이터 기반 유사도 검출 방법

악성행위 데이터 기반 유사도 검출

악성 행위 데이터와
기존 DB간 유사도 계산

유사도 검출 비교대상
특징(feature) 선정

Src IP, Dst IP, Port, Direction, Payload(header 제외)

통계기반 유사도 검출
알고리즘 활용 계산

(TF-IDF, Cosine 유사도, Pearson 계수 등)

Feature	Src IP	Dst IP	Port	Direction	Payload
1	192.168.0.1	192.168.9.1	80	IN	C067AF4AA840503DE5FF2C8008004500002FBF3E400...
2	192.168.10.1	192.168.21.3	80	IN	BDSFF4AA840503DE5FF2C8008004500002FBF3E400...
3	192.168.0.9	192.168.9.1	80	IN	C067AF4AA840503DE5FF2C8008004500002FBF3E412...

선정

+ −
× ÷

기준 이벤트
(악성 행위)

각 특징 간
유사도 계산

비교 이벤트

192.168.0.1 ⟷ 192.168.0.9

계산

BDSFF4AA840503DE5FF2C8008004500002FBF3E400... ⟷ C067AF4AA840503DE5FF2C8008004500002FBF3E412...

◆ 보안관제 데이터 일관성 확보 방법론

Step2. 악성 행위 데이터 기반 유사도 검출 방법 – 유사도 계산

TF-IDF & Cosine 유사도 활용 방법의 예

- 이벤트명 : Web-PAT-Apache_Struts(CVE17-5638).17030804@
- 전체 문서의 개수 : 10,000개
- 전체 단어의 수 : 3,560개
- 기존 보안이벤트들을 활용하여 TF-IDF Matrix 생성

3,560개

	get	http	Accept-encoding	identity	…	keep-alive
Event 1	1	4	1	1	…	0
Event 2	1	4	1	1	…	0
Event 3	1	4	1	1	…	0
Event 4	1	4	1	0	…	1
Event 5	0	1	0	0	…	0
⋮					…	
Event 10,000	0	1	1	1	…	0
DF	2,219	9,990	5,560	5,079	…	2,953
IDF	$\log(\frac{10,000}{2,219+1})$ = 0.6536	$\log(\frac{10,000}{9,990+1})$ = 0.0004	$\log(\frac{10,000}{5,560+1})$ = 0.2548	$\log(\frac{10,000}{5,079+1})$ = 0.2941	…	$\log(\frac{10,000}{2,953+1})$ = 0.5295

◆ 보안관제 데이터 일관성 확보 방법론

Step2. 악성 행위 데이터 기반 유사도 검출 방법 – 유사도 계산

TF-IDF & Cosine 유사도 활용 방법의 예

- 전체 이벤트 10,000개의 TF-IDF값 작성
  - Ex) Event 1을 기준으로 **http의 TF = 4**, **http의 IDF = 0.0004**, **TF-IDF = 4*0.0004 = 0.0016**

TF-IDF Matrix

	get	http	Accept-encoding	identity	…	keep-alive
Event 1	0.6536	0.0016	0.2548	0.2941	…	0
Event 2	0.6536	0.0016	0.2548	0.2941	…	0
Event 3	0.6536	0.0016	0.2548	0.2941	…	0
Event 4	0.6536	0.0016	0.2548	0	…	0.5296
Event 5	0	0.0004	0	0	…	0
⋮	⋮	⋮	⋮	⋮	…	⋮
Event 10,000	0	0.0016	0	0	…	0

◆ 보안관제 데이터 일관성 확보 방법론

Step2. 악성 행위 데이터 기반 유사도 검출 방법 – 유사도 계산

TF-IDF & Cosine 유사도 활용 방법의 예

- 전체 이벤트 10,000개의 TF-IDF값 작성
  - Ex) Event 1을 기준으로 **http의 TF = 4**, **http의 IDF = 0.0004**, **TF-IDF = 4*0.0004 = 0.0016**
- 사고처리 이력이 있는 보안이벤트 선별

	get	http	Accept-encoding	identity	…	keep-alive
Accident Event 1	0.6536	0.0016	0.2548	0.2941	…	0
⋮	⋮	⋮	⋮	⋮	…	⋮
Accident Event n	1.3072	0.0004	0	0.5842	…	0

	get	http	Accept-encoding	identity	…	keep-alive
Event 1	0.6536	0.0016	0.2548	0.2941	…	0
Event 2	0.6536	0.0016	0.2548	0.2941	…	0
Event 3	0.6536	0.0016	0.2548	0.2941	…	0
Event 4	0.6536	0.0016	0.2548	0	…	0.5296
⋮	⋮	⋮	⋮	⋮	…	⋮
Event 10,000	0	0.0016	0	0	…	0

사고처리

사고처리

◆ 보안관제 데이터 일관성 확보 방법론

Step2. 악성 행위 데이터 기반 유사도 검출 방법 – 유사도 계산

TF-IDF & Cosine 유사도 활용 방법의 예

- 전체 이벤트 10,000개의 TF-IDF값 작성
  - Ex) Event 1을 기준으로 **http의 TF = 4**, **http의 IDF = 0.0004**, **TF-IDF = 4*0.0004 = 0.0016**
- 사고처리 이력이 있는 보안이벤트 선별
- 선별된 보안이벤트 n개와 나머지 이벤트들($10{,}000-n$개) 간의 유사도를 측정

	get	http	Accept-encoding	identity	…	keep-alive
Accident Event 1	0.6536	0.0016	0.2548	0.2941	…	0
⋮	⋮	⋮	⋮	⋮	…	⋮
Accident Event n	1.3072	0.0004	0	0.5842	…	0

	get	http	Accept-encoding	identity	…	keep-alive
Event 1	0.6536	0.0016	0.2548	0.2941	…	0
Event 2	0.6536	0.0016	0.2548	0.2941	…	0
Event 3	0.6536	0.0016	0.2548	0.2941	…	0
Event 4	0.6536	0.0016	0.2548	0	…	0.5296
⋮	⋮	⋮	⋮	⋮	…	⋮
Event 10,000	0	0.0016	0	0	…	0

cos similarity

◆ 보안관제 데이터 일관성 확보 방법론

Step2. 악성 행위 데이터 기반 유사도 검출 방법 – 유사도 계산

TF-IDF & Cosine 유사도 활용 방법의 예

- $similarity = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$

- Accident Event1 & Event 3 유사도
  - $similarity = \frac{0.6536*0.6536+0.0016*0.0016+0.2548*0.2548+0.2941*0.2941}{\sqrt{0.6536*0.6536+0.0016*0.0016+0.2548*0.2548+0.2941*0.2941}*\sqrt{0.6536*0.6536+0.0016*0.0016+0.2548*0.2548+0.2941*0.2941}} = \mathbf{1}$
- Accident Event1 & Event 4 유사도
  - $similarity = \frac{0.6536*0.6536+0.0016*0.0016+0.2548*0.2548+0.2941*0+0*0.5296}{\sqrt{0.6536*0.6536+0.0016*0.0016+0.2548*0.2548+0.2941*0.2941}*\sqrt{0.6536*0.6536+0.0016*0.0016+0.2548*0.2548+0*0+0.5296*0.5296}} = \mathbf{0.7360}$

	get	http	Accept-encoding	identity	…	keep-alive
Accident Event 1	0.6536	0.0016	0.2548	0.2941	…	0

	get	http	Accept-encoding	identity	…	keep-alive
Event 3	0.6536	0.0016	0.2548	0.2941	…	0
Event 4	0.6536	0.0016	0.2548	0	…	0.5296

cos similarity

◆ 보안관제 데이터 일관성 확보 방법론

Step2. 악성 행위 데이터 기반 유사도 검출 방법 – 유사도 계산

데이터 유사도 계산 결과 예

[정탐 기준 데이터]

```
78c6adf7-ff6a-40e4-8187-4eafe145e123  C067AF4AA840503DE5FF2C8008004500002F35AD40003E06D573C0680
F33AF2DB2DF078F19A797244F2F7104E3C650180071622C0000C200066D4CBCAA  C200066D4CBCAA  ml
```

유사도 계산 수행

[데이터 유사율: 65%]

```
2b36037b-370c-426b-a14c-f0e62b94ab8f  C067AF4AA840503DE5FF2C8008004500002FBF3E40003E064BE2C06
80F33AF2DB2DF078F19A39F5C658A3EF96B07501800715C670000C200066D4CBD3C  C200066D4CBD3C  ml
```

[데이터 유사율: 99%]

```
78c6adf7-ff6a-40e4-8187-4eafe145e123  C067AF4AA840503DE5FF2C8008004500002F35AD40003E06D573C0680
F33AF2DB2DF078F19A797244F2F7104E3C650180071622C0000C200066D4CBCAB  C200066D4CBCAA  ml
```

[데이터 유사율: 100%]

```
78c6adf7-ff6a-40e4-8187-4eafe145e123  C067AF4AA840503DE5FF2C8008004500002F35AD40003E06D573C0680
F33AF2DB2DF078F19A797244F2F7104E3C650180071622C0000C200066D4CBCAA  C200066D4CBCAA  ml
```

◆ 보안관제 데이터 일관성 확보 방법론

Step3. 전문가 검증 기반 정합 기준 유사도 확립

HITL 기반 정합 기준 유사도 확립

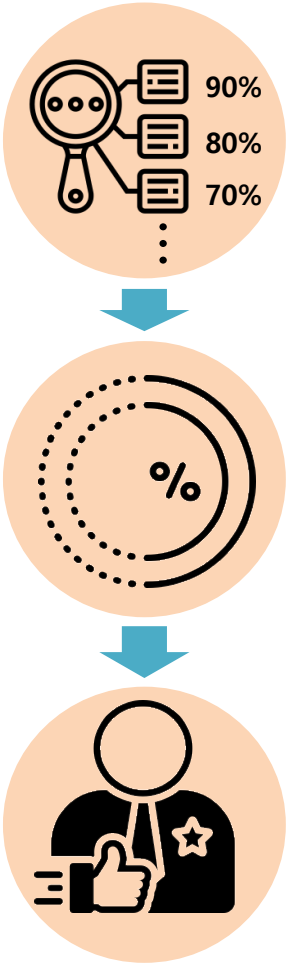
Human In The Loop 기반
분류 기준 유사도 임계치 검증

유사도 비율 별
정/오탐 데이터
분류

(90%→60%, 10% 단위)

90% 80% … 60%

상위 비율에서 하위 비율로
비교 수행을 통한
최적 유사도 임계치 검출



유사도 비율별 정/오탐 데이터 분류 정렬

- 분류 단위: 90% → 60%, 10% 단위
  (추후 5%단위로 정밀화 예정)

전문가 검증 기반 유사도 임계치 검증

- 10%단위 감소율로 이벤트 정/오탐 여부 전수 검사
- Payload 내역을 포함한 전체 필드 검토
- 최적 (최소 노이즈) 성능이 보장되는 유사도 선택
- 전문가 Cross-check를 통한 신뢰성 확보

재 레이블링을 통한 일관성 확보 수행

- 80% 기준 (이상) 데이터 정/오탐 레이블 변경 수행
- 사후 분석(재 레이블링) 수행을 통한 지속적인 일관성 확보
- 주기적인 유사도 비율 검증 수행

◆ 보안관제 데이터 일관성 확보 방법론

Step3. 전문가 검증 기반 정합 기준 유사도 확립

임계치(80%) 설정 근거

- TF-IDF & Cos 유사도 계산 후 각 이벤트 유사도 별 전문가(관제요원) 전수 검사
- 페이로드를 구성하는 단어의 길이가 길어지면 글자 수로는 많은 차이가 있지만 단어의 개수는 차이가 적어 TF-IDF & Cos 유사도에서는 큰 차이가 나지 않음
  - Case 1 & Case 2의 Cos 유사도는 70%이상 (7개의 단어 중 5개 일치)
  - 이러한 부작용 최소화 및 Cos 유사도의 장점을 극대화하기 위하여 80% 선택

CASE 1

dd, eval, base64_decode, post, z0

COS 유사도 70% 이상

CASE 2

dd, eval, base64_decode, post, z0, z0, qgluav9zzxqoimrpc3bsyxlfzxjyb3jziiwimcipo0bzzxrfdgltzv9saw1pdcgwkttac2v0x21hz2ljx3f1b3rlc19ydw50aw1l

◆ 보안관제 데이터 일관성 확보 방법론

Step3. 전문가 검증 기반 정합 기준 유사도 확립

***CASE: 유사도 탐색 결과를 신뢰할 수 없는 경우**

CASE 1

Uhjvz3jhbsbnyw5hz2vyaa

- “Uhjvz3jhbsbnyw5hz2vyaa==“ → “Program Manager”(base 64 decoding)
- Trojan의 감염신호로 추정되나, 정상일 가능성도 존재

➔ **원본 페이로드가 너무 짧아 추가적인 분석 필요**

CASE 2

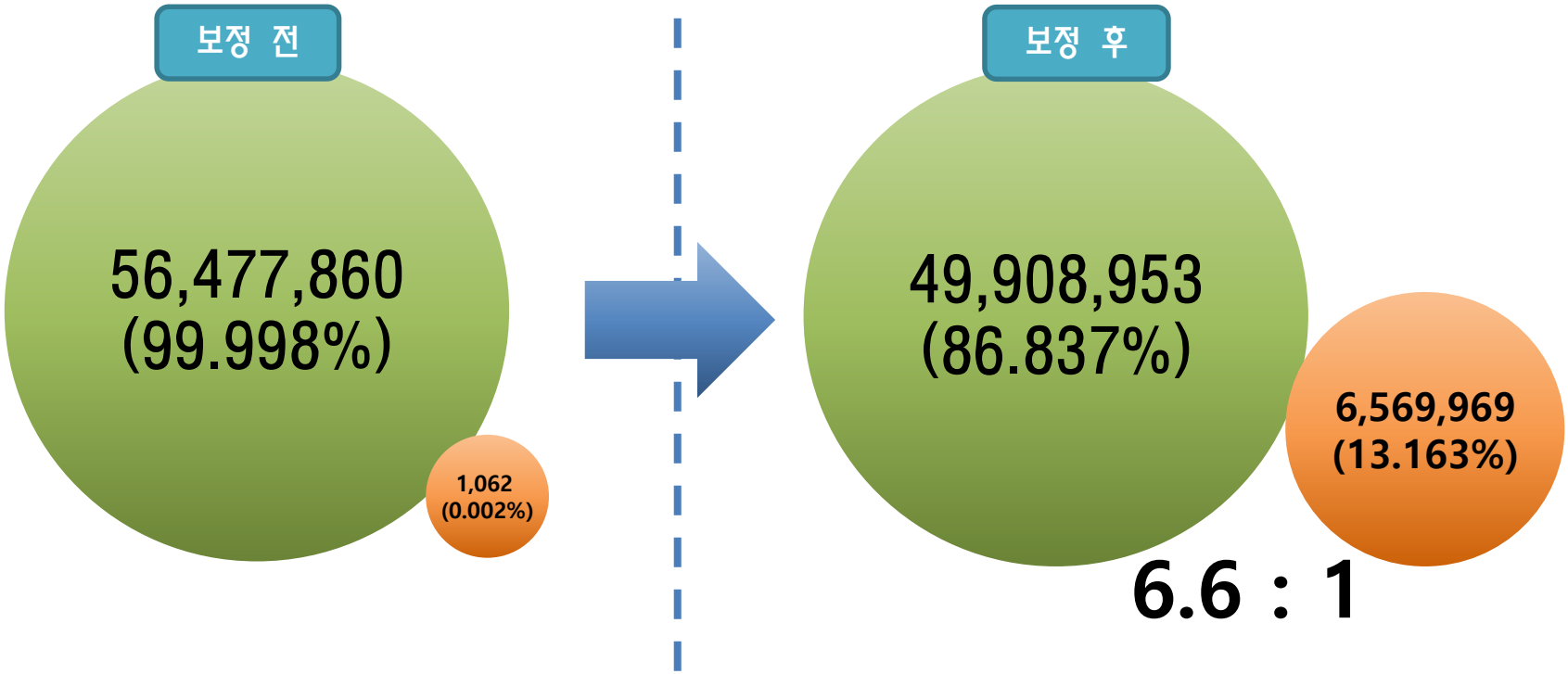
Get, http, host, accept, text, html, accept-encoding, deflate, gzip, identity, user-agent, mozilla, windows, nt, rv, 9.0.1, gecko, firefox, 9.0.1

- 정상 통신에서도 나올 수 있는 단어들

➔ **NLP기술 적용을 통한 추가적인 분석 필요**

◆ 구축 학습 데이터 현황 및 통계

정합 데이터 분포 (예시: 2017~2018년 데이터)



- 정탐 학습 데이터 큰 폭으로 증가(약 6500% 이상)
- 정/오탐 DB 일관성 확보 (교차 사고처리 2000건 이상 보정)

◆ 학습 데이터 유효성 및 품질 검증

ML기반 데이터 학습 및 성능 비교

[정합 전 실험 데이터]

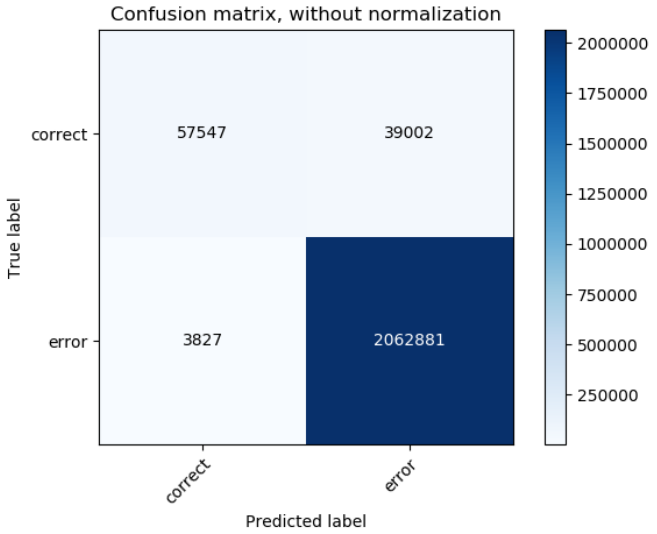
[정합 후 실험 데이터]

[학습 대상 특징(feature): 15종 필드]

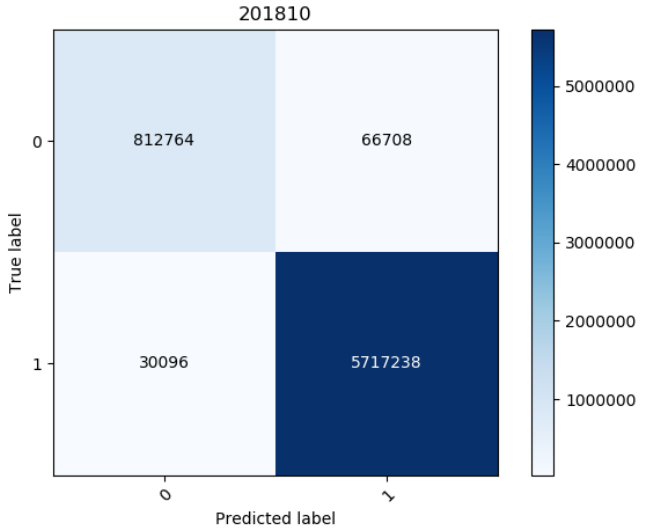
Source IP	Destination IP	Source Port	• • • *보안 상 생략	Payload Word DF

◆ 학습 데이터 유효성 및 품질 검증

실험결과 – 혼동 행렬 (Confusion matrix)



정합 전



정합 후

모델	정확도	정밀도	재현율	F1-score
정합 전	98.02%	93.76%	59.6%	72.8%
정합 후	98.23%	96%	95.42%	94.3%

**동일 실험 조건은 아니나, 정합 후 일관성 확보로 인한 성능향상 확인 가능*

- **대용량 보안관제 ML/AI 학습 데이터 구축**
  - 자동화 된 보안관제 AI모델 학습을 위한 **대용량 데이터 확보**
  - 이벤트 특성 및 Taxonomy 분석을 통한 **학습 데이터 정제**
  - 실제 보안관제 환경을 고려한 데이터 **레이블링 규칙 정의**
- **고성능 자동 보안관제 AI 모델 구축 및 운용**
  - **고품질 학습데이터** 기반 AI모델 학습 및 성능 강화
  - **강화 학습** 기반 자동 학습 모델 플랫폼 설계
  - **실 환경 운용** 테스트 및 적용 시도
- **신변종 공격 탐지용 ML/AI 모델 개발**
  - 대용량 학습데이터를 활용한 **확률/통계 기반 이상치 검출**
  - 이벤트 특성을 고려한 **파라미터 설정 및 학습 방법** 정의