딥러닝을 이용한 악성파일 탐지기술

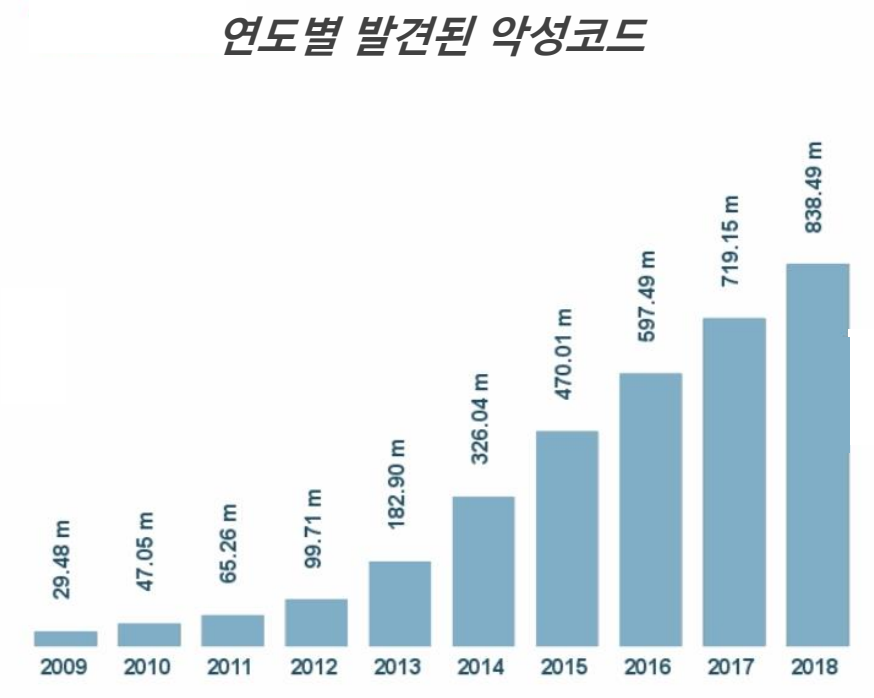
ETRI 한국전자통신연구원
Electronics and Telecommunications Research Institute

딥러닝을 이용한 악성파일 탐지기술

배경 및 필요성 : 악성파일의 증가

- 2018년 기준으로 신규로 발견된 악성파일은 8억3,849만개로 1초에 26.5개가 새롭게 생성
- 랜섬웨어, 금융사기 등 악의적인 목적의 악성파일 증가로 2018년 기준 1,141억 달러 관련 지출 발생
  - 카스퍼스키랩에서 2017년에 발행된 보고서에 의하면, 29.4%의 사용자는 일년에 한 번 이상 악성 코드의 공격을 경험하고 있고, 사이버범죄자의 수는 5년 전에 비해 5배 이상 증가

연도별 발견된 악성코드



* 출처 : AV-test 홈페이지(2018)

2018-2019년 전세계 분야별 보안 지출액(단위: 백만 달러)

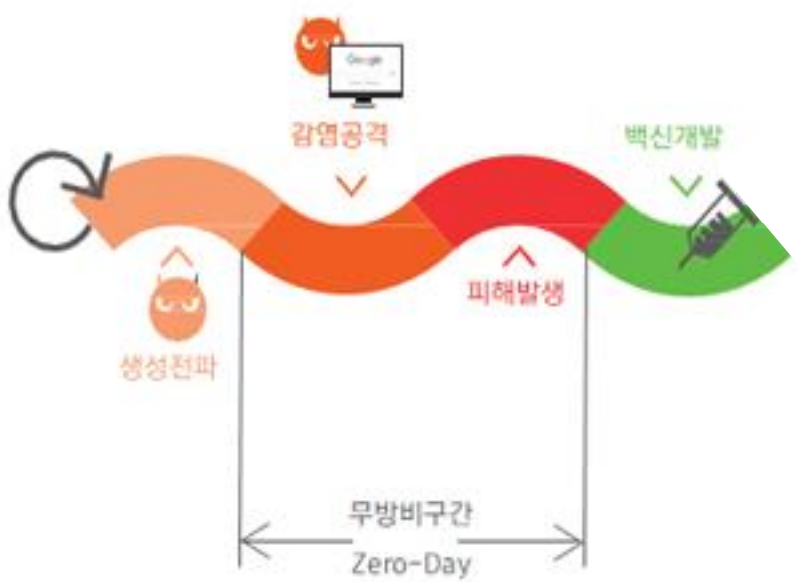
시장분야	2018	2019
애플리케이션 보안	2,742	3,033
클라우드 보안	304	459
데이터 보안	3,063	3,524
IAM	9,768	10,578
인프라 보호	14,106	15,337
통합 리스크 관리	4,347	4,712
네트워크 보안 장비	12,427	13,321
그 외 정보 보안 소프트웨어	2,079	2,285
보안 서비스	58,920	64,237
소비자 보안 소프트 웨어	6,395	6,661
총계	**114,152**	**124,116**

* 출처 : 가트너 (2018)

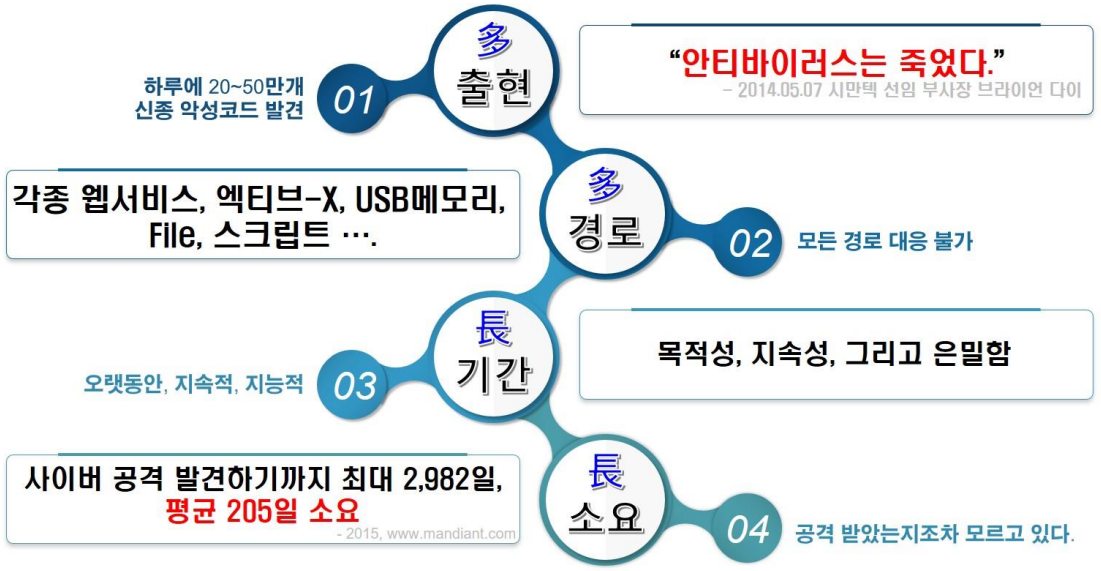
배경 및 필요성 : 악성파일 탐지기술의 한계

- 기존의 악성파일 탐지기술은 보안전문가가 수동으로 패턴을 분석하여 백신을 개발
- 보안전문가가 패턴을 분석할 때 까지 무방비구간이 존재하며 대응까지 평균 205일 소요

기존 백신 방식



악성파일 탐지기술의 한계



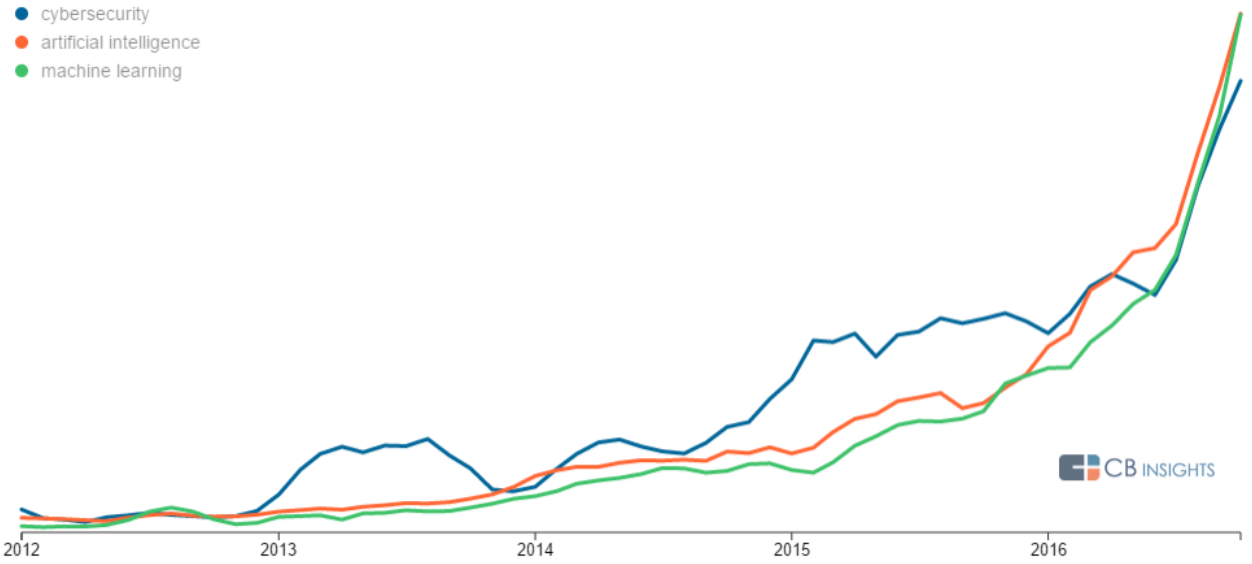
* 출처 : 굿어스 소개자료(2018)

배경 및 필요성 : 딥러닝

- **사이버 보안 관련하여, 인공지능 및 머신러닝을 포함한 미디어 기사가 급증 (상관관계 급증)**

  - CBINSIGHT에 의하면 사이버 공격의 증가가 보안 전문인력의 공급을 앞지르고 있고, 바이러스에 관한 정보가 알려지지 않은 제로 데이터성 공격이 급증함에 따라 사이버 보안 기술에 인공지능과 머신러닝의 도입 필요성이 급증하는 것으로 분석

사이버 보안, 인공지능 및 머신러닝의 상관 관계에 대한 트랜드 분석



* 출처 : Cybersecurity's Next Step: Artificial Intelligence Is Helping Predict, Prevent, And Defeat Attacks, 2016, CBINSIGHT

[참고] 인공지능 기반 보안 솔루션 출시 및 투자 활성화

- **세인트시큐리티, 안랩 등 세계적인 보안회사에서는 인공지능으로 악성파일을 탐지하는 제품을 출시하였으며 2014년에 설립된 보안스타트업 에버스핀은 210억 원 규모의 투자를 유치**

안랩, 머신러닝 기반 차세대 네트워크 통합보안 플랫폼 '안랩 TMS' 출시

김수아 기자 | 승인 2018.05.17 10:11 |

보안 위협 분석을 위해 머신 러닝과 시나리오 기반 상관 분석을 제공

안랩 TMS

* 출처: 2018.05.17일자 인공지능신문 기사『안랩, 머신러닝 기반 차세대 네트워크 통합보안 플랫폼 '안랩 TMS' 출시』발췌

HOME » INVESTMENT

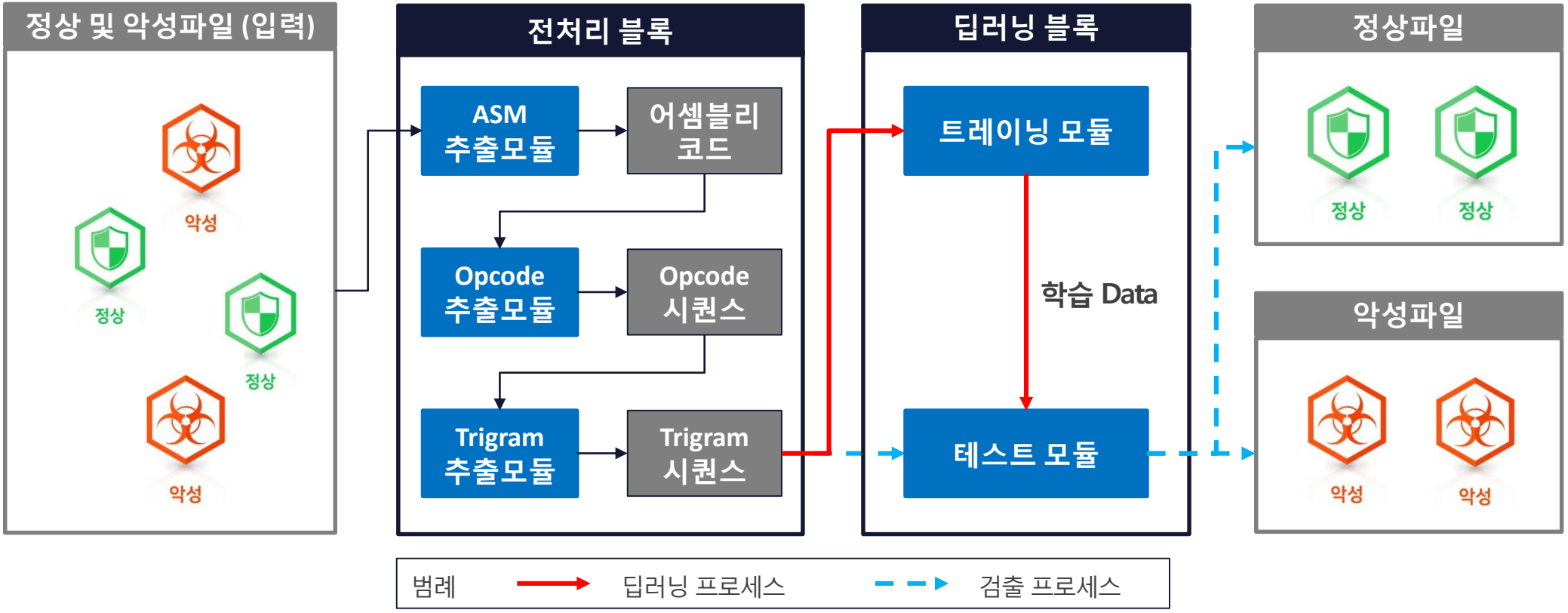
보안 스타트업 '에버스핀', 210억 원 규모 투자유치

김민정 POSTED ON 2018/05/10

* 출처: 2018.05.10일자 플래텀 기사『보안 스타트업 '에버스핀', 210억 원 규모 투자유치』발췌

기술 개요

① 전처리 블록을 통해 정상 및 악성파일에서 Trigram 시퀀스를 추출하고, ② 트레이닝 모듈에서 학습된 데이터를 전송받은 테스트 모듈을 통해 ③ 라벨링되지 않은 전처리 데이터로부터 파일의 정상/악성 여부 판단

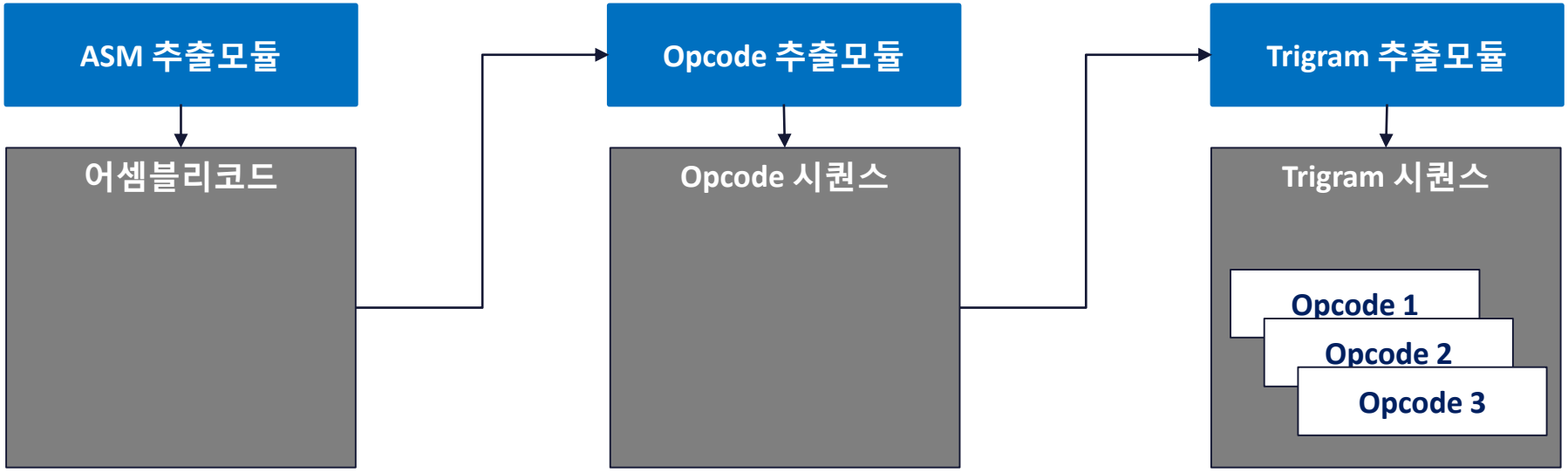


기술의 특장점

- 알려진 악성코드에서 패턴을 추출하는 기존 방식과 달리 딥러닝을 이용하여 알려지지 않은 제로데이 신·변종 악성코드를 탐지 가능

- 실행 파일의 모든 코드를 이용하는 API 시스템콜을 전처리 데이터로 사용하는 방식과 달리 실제 동작에 필요한 Opcode를 전처리 데이터로 사용하여 학습 및 테스트 데이터 생성 속도가 빠름
  - 3개의 연속된 Opcode로부터 트라이그램을 구성하여 트라이그램 시퀀스를 전처리 데이터로 사용

사업화 대상 기술의 전처리 모듈



세부 기술의 특장점

높은 악성파일 탐지율	▪ 기존 패턴기반 안티바이러스 프로그램 대비 19.4% 높은 탐지율 - 딥러닝 기반의 사업화 대상기술의 경우 무작위 악성코드에 대해 99.4% 탐지
신·변종 악성파일 자동탐지	▪ 딥러닝 방식을 이용하여 대용량 데이터에 대한 학습이 용이하여, 패턴이 알려지지 않은 신종, 변종 악성파일을 보안 전문가의 분석 없이 자동탐지 가능
빠른 학습속도	▪ 전처리 방식으로 Opcode만을 선별하여 사용하기 때문에 동적분석형태의 API 시스템콜을 사용하는 기존의 방식에 비해 훨씬 빠르게 학습 및 테스트 데이터 가능
전처리 용이	▪ 기존의 Decision Tree 및 Support Vector Machine (SVM)을 사용하는 머신러닝 방식에 비해 전처리가 용이

기술 경쟁력

- **적은 매개변수로 쉽게 훈련되는 CNN*과 높은 인식률 보이는 RNN*의 장점을 결합하여 99% 탐지가 가능하고, 딥러닝 모델을 통해 다양한 신변종 바이러스를 탐지 가능**

회사명		Product	탐지율	비고
ETRI 한국전자통신연구원 Electronics and Telecommunications Research Institute	ETRI	-	99%	딥러닝
Microsoft	Microsoft	Windows Defender	98.9%	인공지능분류
eset	ESET	NOD32 Antivirus	97.9%	-
KASPERSKY	Kaspersky Lab	Internet Security	98.9%	딥러닝
PC Pitstop	PC pitstop	PC Matic	98.9%	-

* CNN (convolutional neural network), RNN(Recurrent Neural Network)

* 자료: ETRI(5분할 교차검증방식으로 자체시험결과), AV-test 2017년 11월 시험결과

기술완성도(TRL)

▪ TRL 4단계로 성능평가를 완료함

TRL 9	**사업화**	• 본격적인 양산 및 사업화 단계
TRL 8	시작품 인증/표준화	• 일부 시제품의 인증 및 인허가 취득 단계 - 조선 기자재의 경우 선급기관 인증, 의약품의 경우 식약청의 품목 허가 등
TRL 7	Pilot 단계 시작품 신뢰성 평가	• 시작품의 신뢰성 평가 • 실제 환경(수요기업)에서 성능 검증이 이루어지는 단계
TRL 6	Pilot 단계 시작품 성능 평가	• 경제성(생산성)을 고려한, 파일로트 규모의 시작품 제작 및 평가 • 시작품 성능평가
TRL 5	시제품 제작/ 성능평가	• 개발한 부품/시스템의 시작품(Prototype) 제작 및 성능 평가 • 경제성(생산성)을 고려하지 않고, 우수한 시작품을 1개~수개 미만으로 개발
TRL 4	연구실 규모의 부품/시스템 성능평가	• 연구실 규모의 부품/시스템 성능 평가가 완료된 단계 • 실용화를 위한 핵심요소기술 확보
TRL 3	연구실 규모의 성능 검증	• 연구실/실험실 규모의 환경에서 기본 성능이 검증될 수 있는 단계 • 개발하려는 시스템/부품의 기본 설계도면을 확보하는 단계 • 모델링/설계기술 확보
TRL 2	실용 목적의 아이디어/ 특허 등 개념 정립	• 실용 목적의 아이디어, 특허 등 개념 정립
TRL 1	기초 이론/실험	• 연구과제 탐색 및 기회 발굴 단계

기술이전 내용 및 지식재산권 현황

기술이전 범위

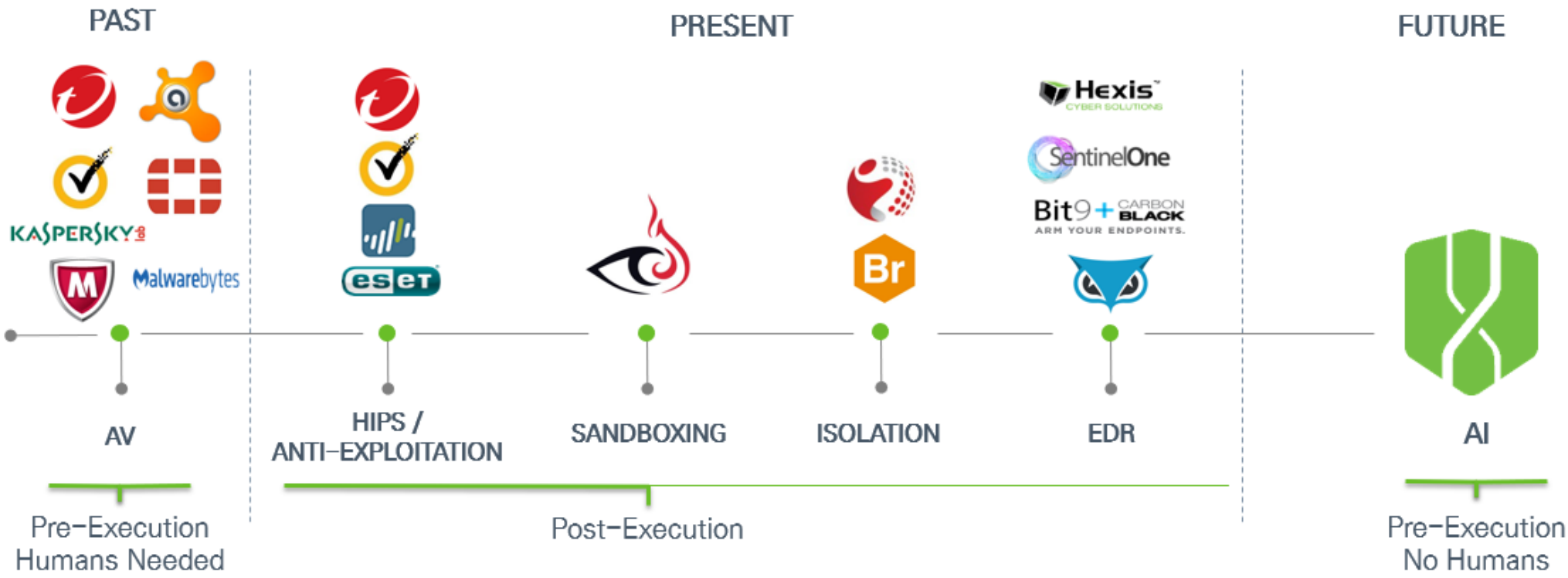
No.	구분	기술이전 범위
1	프로그램	• 딥러닝을 이용한 악성파일 탐지 프로그램
2	기술 문서	• SoA 분석 시스템 구조 설계서 • 딥러닝을 이용한 악성파일 탐지시스템 시험절차결과서

지식재산권 현황

No.	특허번호	특허명	권리현황
1	10-2017-0001184	클라우드 기반으로 악성코드를 탐지하는 장치	출원(공개)
2	10-2017-0010978	네트워크 보안기능 가상화 기반의 클라우드 보안분석장치	출원(공개)
3	10-2017-0130810	파일 이미지를 이용한 악성코드 탐지 방법 및 이를 위한 장치	출원(비공개)

기술 동향

- 딥러닝을 통해 10페타 바이트 이상의 악성코드 데이터를 학습하고 있어 인공지능이 악성코드를 탐지하는 방향으로 진화하고 있음

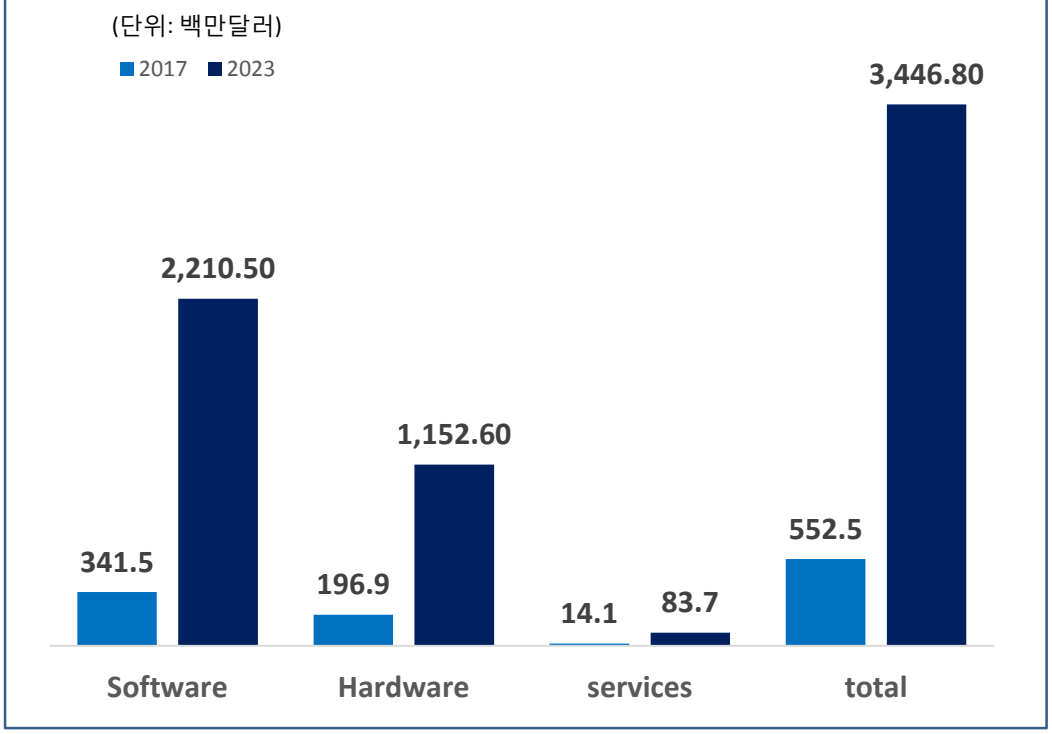


* 출처 : Cylance Protect 소개자료(2017)

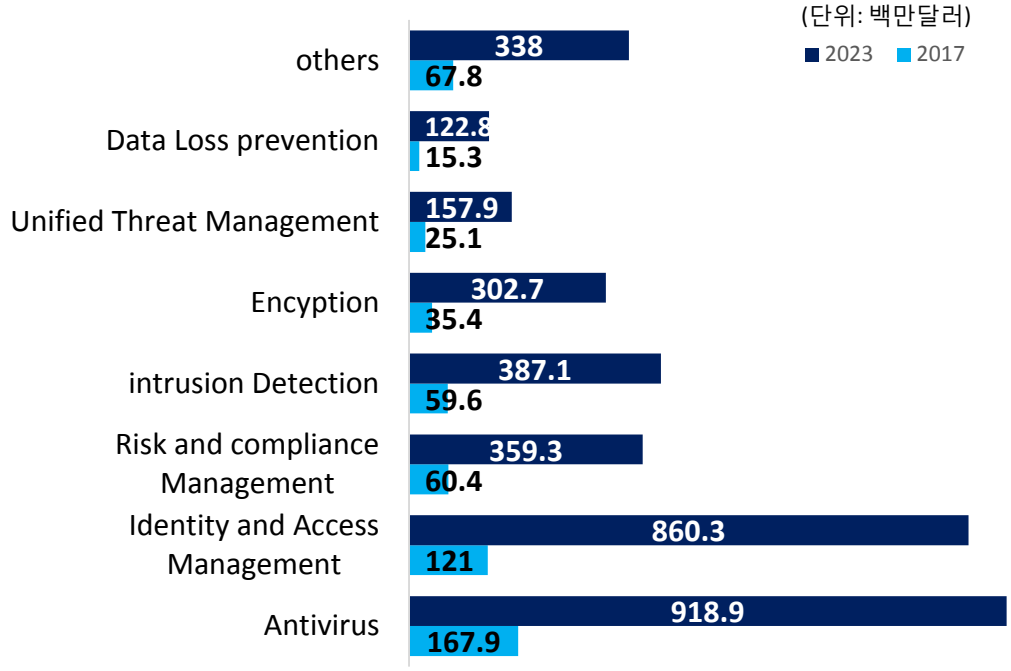
시장 동향

- 딥러닝 보안 시장 규모는 2017년 기준으로 552.5억 달러에서 연평균 35.7%씩 성장하여 2023년에는 3,446.8억 달러 규모로 성장 예상
- 딥러닝 보안기술을 활용한 백신 매출은 2023년 기준 918.9백만 달러로 예측 2017년 대비 5.4배 증가 예상

보안 분야의 딥러닝 마켓(Global)



솔루션별 보안 분야의 딥러닝 마켓(Global)



*출처 : Deep Learning MARKET, 2018, Marketsandmarkets

[참고] 보안위협 동향

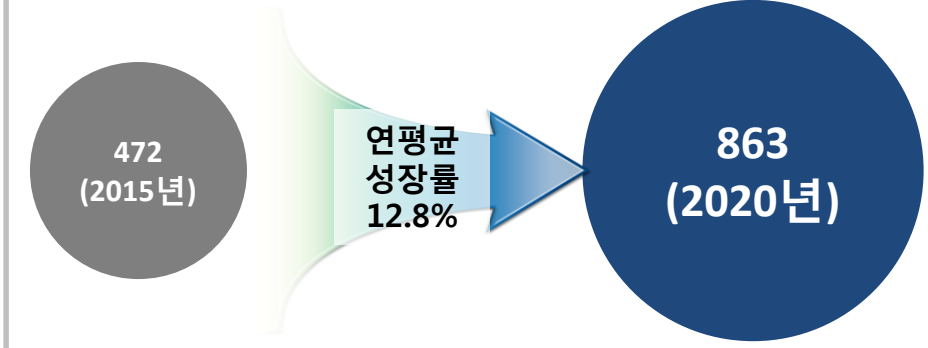
- **향후 3년 안에 50억 개의 IoT 디바이스와 7.8억 개의 데이터가 위협받을 것으로 예상하였으며 이용자의 전문화된 보안능력이 요구**

* 자료 : 주요보안전망 재가공

시장 동향 : 정보보안 시장전망

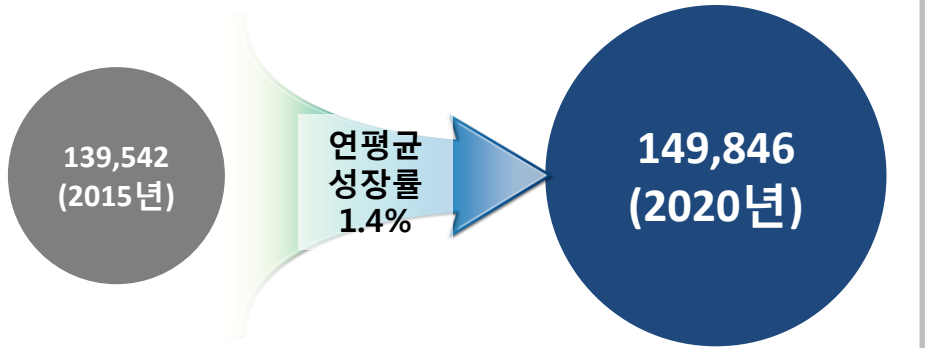
지식정보 보안

* 단위 : 조



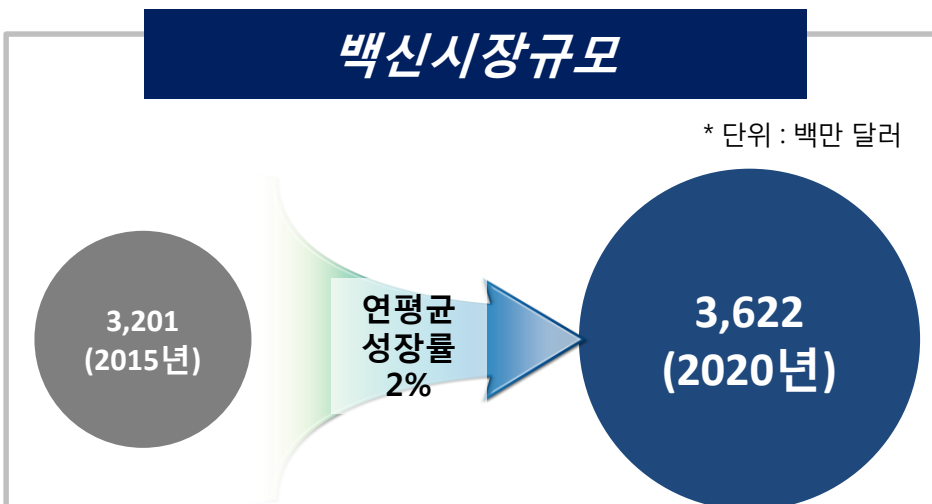
사이버 공격 탐지/대응

* 단위 : 억 원



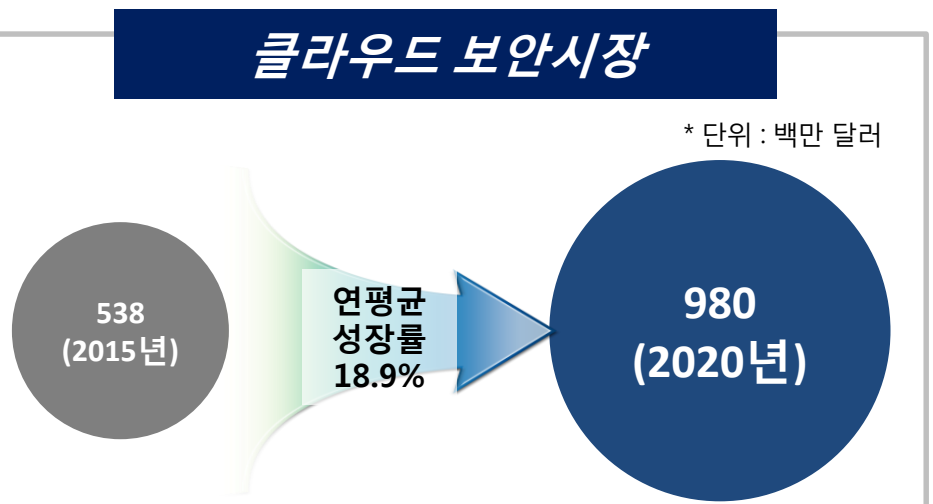
백신시장규모

* 단위 : 백만 달러



클라우드 보안시장

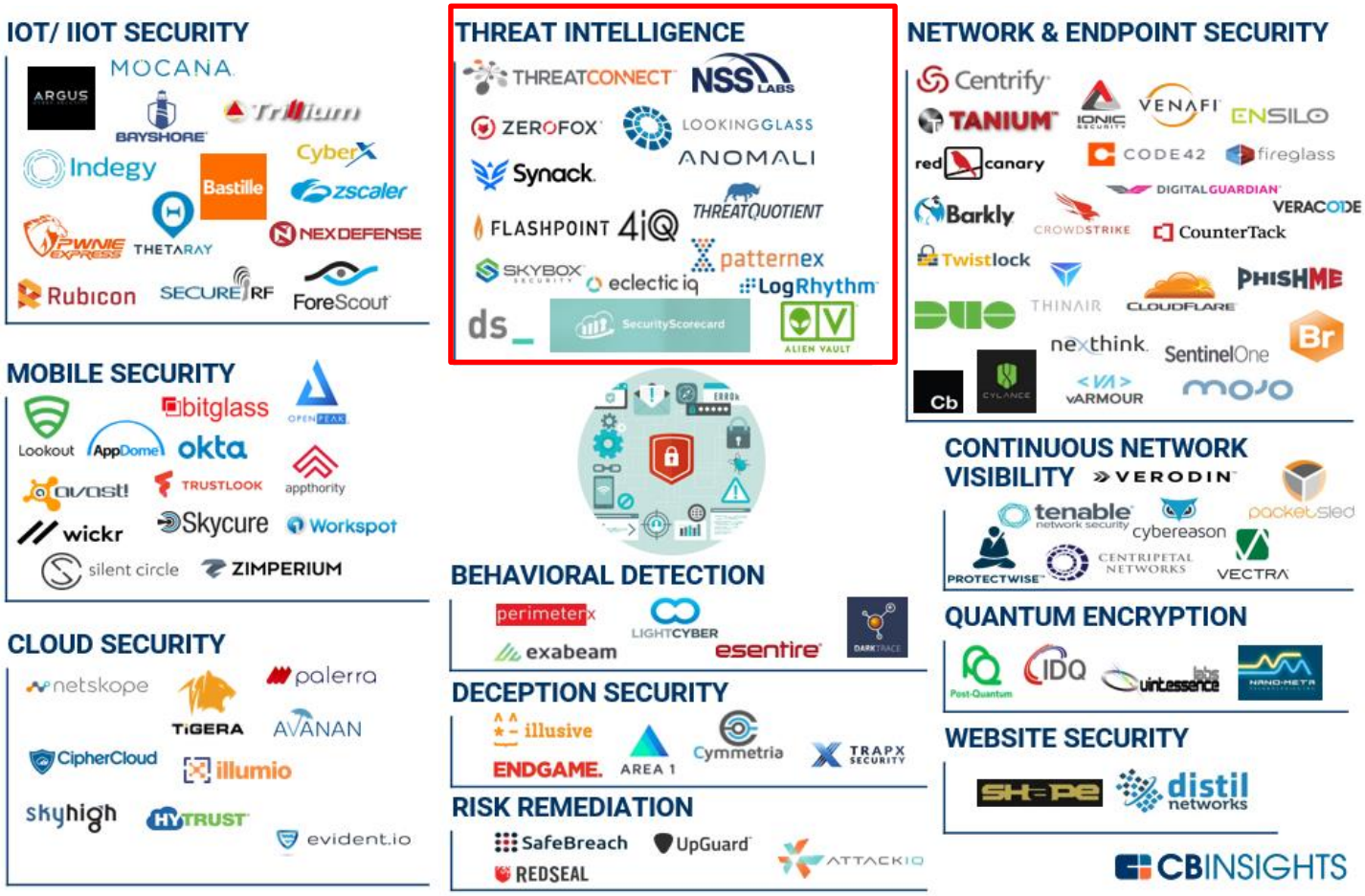
* 단위 : 백만 달러



* 출처 : Gartner(2016)

시장 동향 : 정보보호 기반 비즈니스 현황

- **Cyber Security 관련 스타트업은 총 658개이며 575개의 회사에서 매출이 발생하고 있어 높은 수익가능성을 보임**



* 출처 : Cbinsight, From IIoT Security To Quantum Encryption: 106 Cybersecurity Startups In A Market Map (2016)

[참고] 국내외 스타트업 현황

- **투자 유치를 통해 성장하고 있는 국내외 스타트업의 경우 신규 악성코드를 탐지하기 위해 딥러닝 기술을 보유하고 있으며, 대부분 금융, 클라우드 분야로 사업 진출**

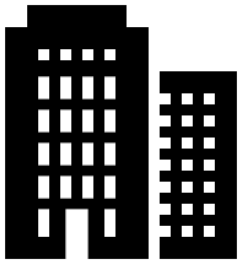
<표> 국내외 보안 관련 스타트업 현황

기업명		HashiCorp	Securosis	Area 1 Security	㈜에버스핀	센스톤	(주)에스이웍스
본사		San Francisco (미국)	Arizona (미국)	San Francisco (미국)	Seoul (한국)	Seoul (한국)	Seoul (한국)
사업분야		▪ Cloud Infrastructure ▪ Private Cloud	▪ Cloud Security	▪ Cyber Security ▪ Network Security	▪ Cloud Security ▪ API/SDK	▪ AI ▪ Natural Language ▪ Speech Recognition	▪ AI /Deep learning ▪ Natural Language ▪ STT/TTS
설립년도		2012	2009	2013	2014	2015	2012
최근 매출		$3.8M	$30.4M	$2.3M	8억원	5억원	3억원
고용인력		325	11	74	30	13	16
투자현황	횟수	4	1	4	3	3	3
	금액	**$174.2M**	**$2.5M**	**$57.5M**	**271억**	**21억**	**73억**
기타				Google Analytics, WordPress, G Suite의 보안 담당	IBK기업은행 (국책은행) APP보안 서비스 계약 체결	다양한 간편인증 솔루션을 제공	클라우드 기술 기반의 보안·암호화 분야에서 API/SDK 형태로 지원

* 자료 : crunchbase (https://www.crunchbase.com) 및 벤처스퀘어 재구성

투자 동향

- 정보보안 분야의 경우, 1,445회의 투자유치가 이루어졌으며, 투자유치 금액은 총 95억 달러 규모
- 합병된 기업은 29개이며, 247개 투자자가 투자 참여
- IPO 기업은 9개로, 공모금액은 132.5백만 달러 수준



658
Number of Organizations

Apr 5, 2013
Average Founded Date

29
Number of Acquired

247
Number of Investments

1,445
Number of Funding Rounds

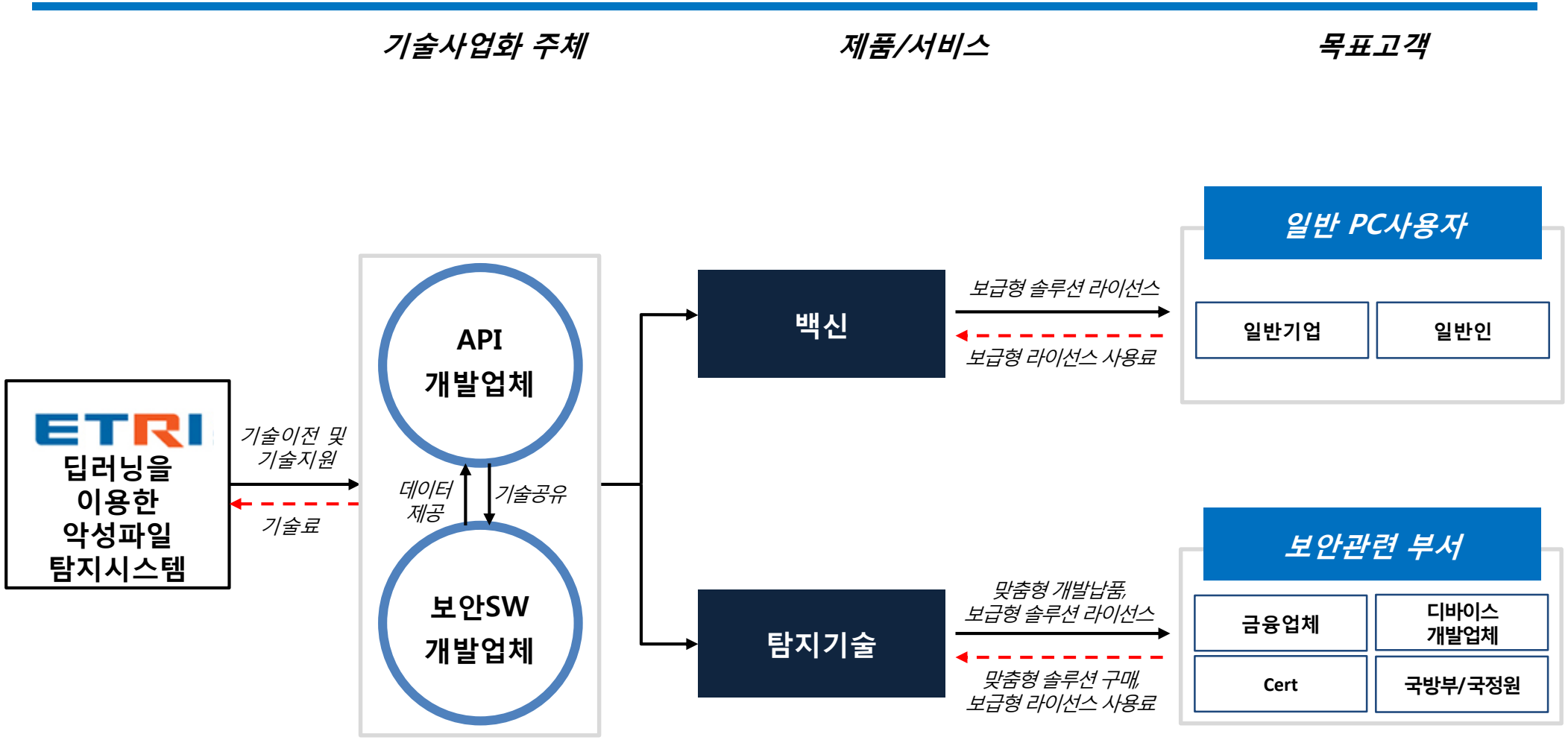
$9.5B
Total Funding Amount

9
Number of IPOs

$132.5M
Total Amount Raised in IPO

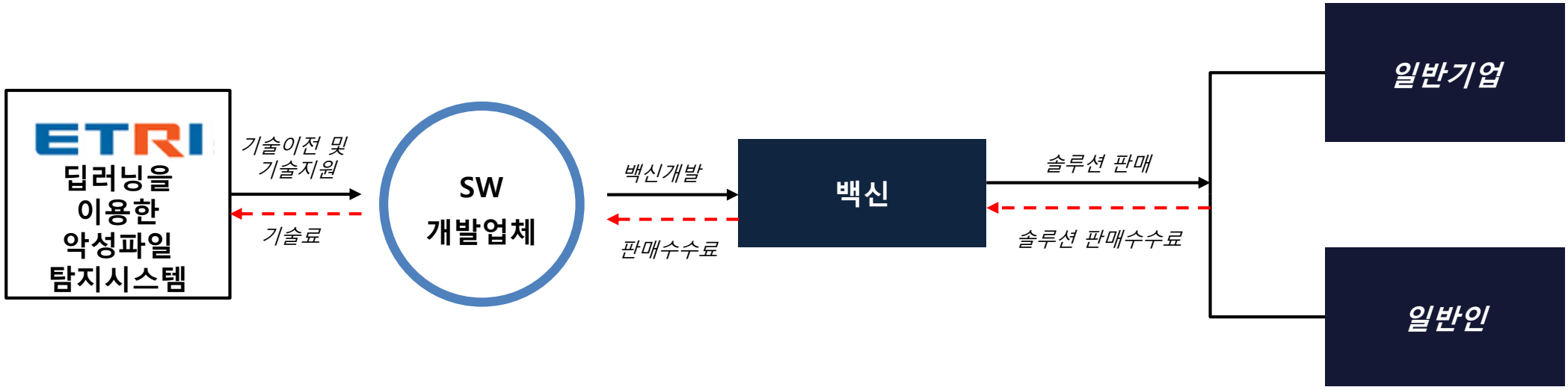
* 자료 : crunchbase (https://www.crunchbase.com) 재구성

비즈니스 모델 Overview



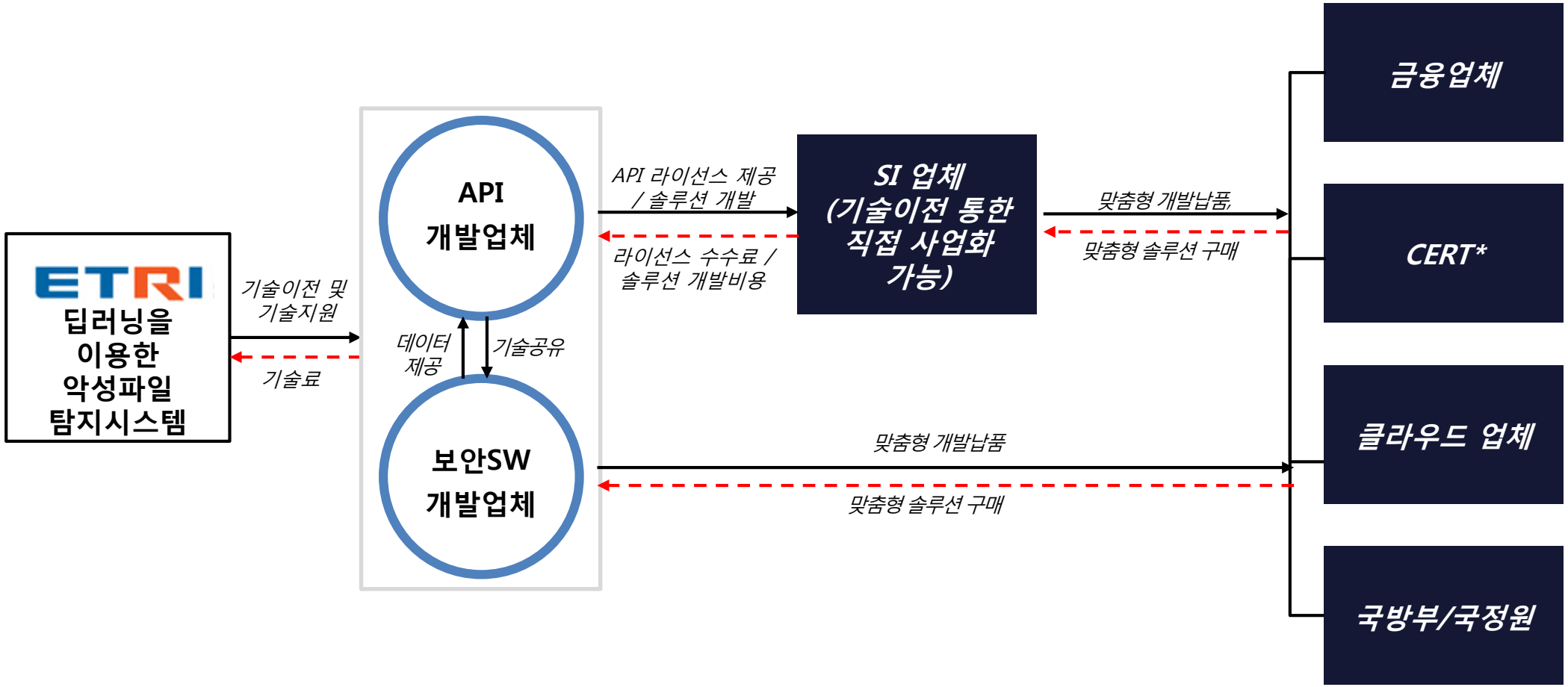
BM ① 일반 PC사용자를 위한 백신개발

- 일반 기업을 대상으로 안티바이러스 제품 개발 및 해당 솔루션 판매



BM ② 전문 보안을 요하는 기관 맞춤 탐지솔루션

- 금융, 기업보안 등 특정 분야 맞춤형 탐지 API 혹은 탐지 솔루션 개발 및 판매
- 다양한 개인 정보를 보유하고 있는 클라우드 및 공공기관 대상 솔루션 판매



* CERT(computer emergency response team), 컴퓨터 비상대응팀

ETRI 개발기술 도입 통한 사업화 프로세스

사업화 유망기술 (TRL 4)

ETRI
한국전자통신연구원

- 정상파일 및 악성파일 전처리 기술
- 악성파일탐지를 위한 딥러닝모델 트레이닝 기술
- 악성파일탐지를 위한 딥러닝모델 테스트 기술
- 관련 모듈

기술이전 1개월

딥러닝을 이용한 악성파일탐지 기술전수

- 시스템 설계서/ 개발문서/ 기술문서 제공 /
  연구자 인적 기술지도/개발자용 지침서

수요기업 필요역량

- 수억개 이상의 정상파일 및 악성파일의 수집능력
- 소프트웨어 구현 및 디버깅을 위한 개발 환경 구축
- 기업 자체 커스터마이징 개발예산 확보
- 기계학습/딥러닝 학습DB, 노하우 확보
- 출연연구소 공동 R&D 수행경험

BM 선정 / 구체화(1개월)

기업 맞춤형 BM 구체화

- 사업화 아이디어 선정
- 목표시장 별 서비스 수요 파악
- 수요 맞춤형 BM 수립

상용화 개발(3개월)

BM 기반 제품/서비스 개발

- 기업 맞춤형 솔루션 개발 추진
- 제품/서비스 상용화 Test 실시 (연구자 자문)

사업화(4개월)

BM 적용 제품 양산/판매

- 솔루션 연계 서비스 개발
- 홍보 및 마케팅 위한 산업진흥기관 지원사업 연계
- 솔루션 기술보호 위한 신규 IP 확보

ETRI 기업지원 프로그램

(참고) 정부 추가개발 사업화 지원 사업 안내(기술이전 조건부)

사업명	기술이전사업화 지원사업	R&D 재발견 프로젝트	중대형복합기술사업화지원사업
공고기관	연구개발특구진흥재단 (www.innopolis.or.kr)	한국산업기술진흥원 (www.kiat.or.kr)	과학기술일자리진흥원 (www.compa.re.kr)
사업비 (2018년 기준)	1년 기준 2억 원	1년 기준 4억 원	1년 기준 7.5억 원
신청자격	특구 내 공공연구기관 기술도입기업 또는 추천기술 도입기업	NTB 사이트 등록된 공공연구기관 기술도입 기업	Tech-BM 검증지원사업 통한 경쟁
공고시기 (2018년 기준)	2월 또는 3월	3월	2월

기술이전 문의

ETRI 사업화협력실

042-860-1804 / hominkim@etri.re.kr