Data-Centric Security in Cloud Era

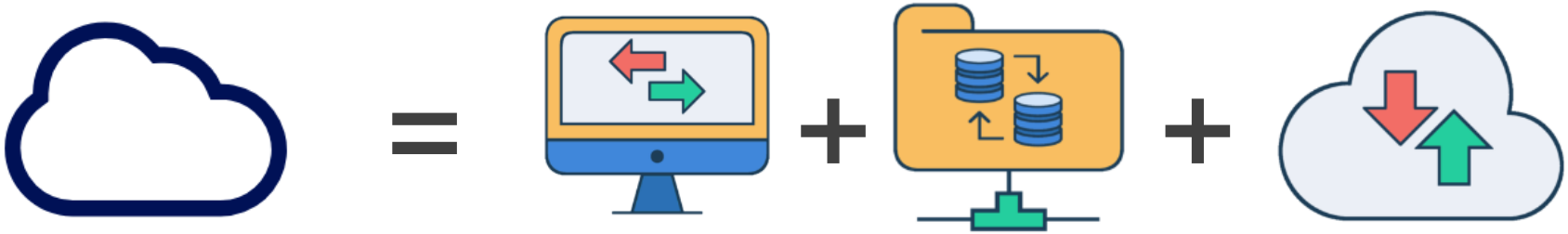
부제: Cryptography @ Inflection Point

삼성SDS 연구소
보안연구팀

SAMSUNG SDS

Agenda

I. Intro: Trends, Risk, Opportunities
II. Compute: Data-in-use Security
III. Storage: Data-at-rest Security
IV. Network: Data-in-transit Security



Agenda

I. **Intro: Trends, Risks, Opportunities**

II. Data-in-use Security

III. Data-at-rest Security

IV. Data-in-transit Security

GDPR(General Data Protection Regulation): 개요

'18년 5월 25일부터 기존 개인정보보호지침(Directive)을 대체하는 법 형식(Regulation)으로 규율되어 있어, EU 모든 회원국에 직접적인 법적 구속력을 가짐

제정 목적/시기	• EU 국가간 개인정보의 자유로운 이동을 보장하고 (제1조 제3항), 정보주체의 개인정보 권리를 강화하고자 (제1조 2항) 제정 • '12.1월 입안 → '16.5월 제정 → '18.5.25일 시행
주요 변동사항	• 적용범위확대: 일반적인 개인정보(이름, 전화번호)외에 온라인식별자(IP/MAC주소, 온라인쿠키), 위치정보, 유전정보 포함 • 정보주체권리 강화 및 동의기준 엄격화
위반 시 과징금	• 중대한 위법 : 전세계 연간 매출 4% 또는 2천만유로 중 높은 금액 예) 동의없이 개인정보 수집, 동의없이 EU 외로 개인정보 전송 등 • 일반적 위법 : 전세계 연간 매출 2% 또는 1천만유로 중 높은 금액 예) 정보 유출시 미신고, 아동의 경우 부모동의 미징구 등

국내 규제동향

국내 개인정보 관련 규제는 글로벌 기존으로도 엄격한 편이나, 최근 규제완화 움직임을 보임

국내 개인정보 보호법

- 「**개요**」 국내 개인정보보호법은 개인정보의 수집부터 폐기까지 전 범위에 걸쳐 사전동의 (Opt-in)방식을 채택했고, 제재규정도 엄격해져 전세계에서도 가장 엄격한 편에 속함
- 「**활용**」 개인정보보호법 상 수집한 개인정보는 수집 목적 범위에서 이용가능하고 (제15조), 목적 외의 용도로 개인정보를 활용하거나 제3자에게 제공해서는 안됨 (제19조)
- 「**보호**」 법 시행령에서는 개인정보를 안전하게 저장·전송할 수 있는 암호화 기술의 적용 또는 이에 상응하는 조치를 적용하도록 명시

최근 금융분야 보안동향 ('18.7)

- 「**기존案**」 비중요 데이터만 클라우드 활용이 가능하고, 개인정보 및 신용정보등은 클라우드 이용 제한
  ※ 개인정보 등을 활용하여 상품/서비스 개발시 클라우드상 AI/분석 인프라 사용불가
- 「**개정案**」 개인신용정보, 고유식별정보를 처리하는 중요시스템도 클라우드 이용가능 ('19年1月 시행예상)
  ※ 국내소재 클라우드로 한정, 금융회사/핀테크기업들이 클라우드를 활용하여 협업이 활성화될 것으로 예상

Cloud Security Alliance (https://cloudsecurityalliance.org)

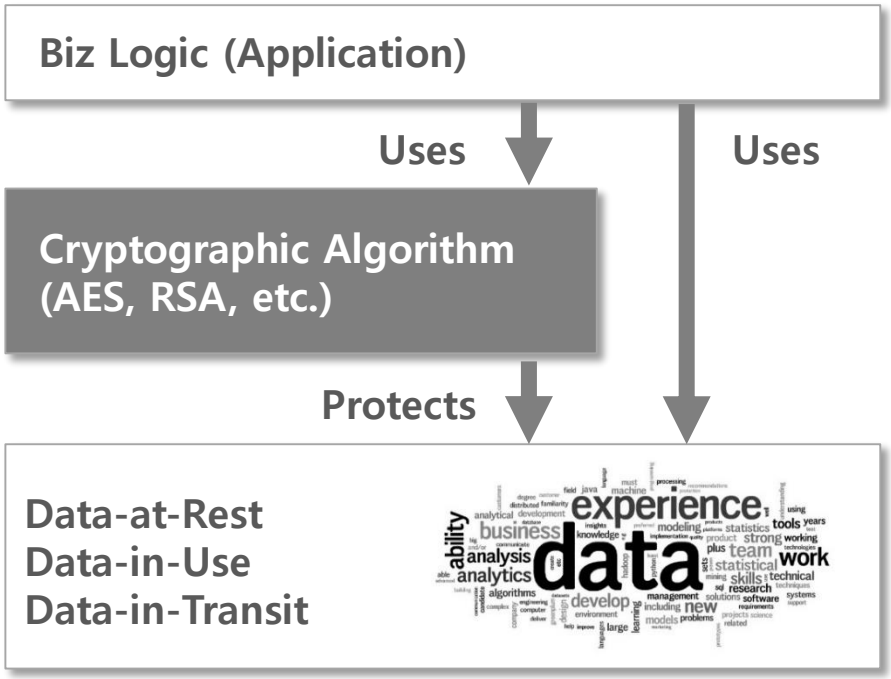
클라우드 보안관련 가장 큰 규모의 Alliance로 다양한 보안기술 및 동향에 대해서 파악가능

400 Corporate Members (80 Chapters Globally & 90,000 Individual Members)

(출처: CSA)

Cryptography in Data Security & Economy

암호기술은 Cloud·IoT·Mobility 등 다양한 환경에서 데이터를 효과적으로 보호하는 것이 가능하고,
최근에는 보호수준에서 벗어나 새로운 비즈니스 기회 (Data Economy) 를 제공



Agenda

I. Intro: Trends, Risk, Opportunities

II. Privacy-Preserving Data Mining
(Data-in-use Security)

III. Data-at-rest Security

IV. Data-in-transit Security

Privacy-preserving Data Mining

최근 데이터의 노출·유출을 원천 차단하면서 데이터를 완전히 활용 가능한 암호·분석기술에 대한 연구개발 진행중

기술 동향

기존 비식별화 기술의 한계

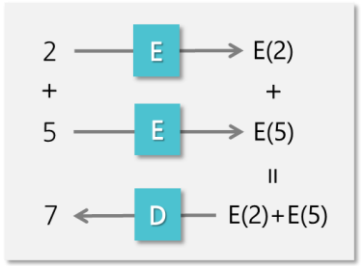
암호기반 분석기술 보유 현황

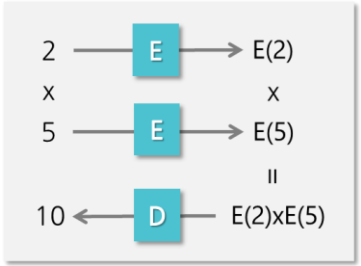
Google, IBM, facebook, Microsoft

- Homomorphic Encryption
- Multi-party Computation
- Zero-knowledge Proof

동형암호 (Homomorphic Encryption)

동형암호는 덧셈 and/or 곱셈을 보존하여 암호화된 상태에서 분석 (Machine Learning, Deep Learning)이 가능





자사는 산학을 통해 세계최고수준*의 동형암호 기술 기반으로 Biz Use Case 확보 진행 中

* '17년 iDASH Genome Privacy & Security Competition 우승
AISIACRYPT2017, EUROCRYPT2018 논문채택

동형암호 Toy Example: Partial Homomorphic Property of RSA

RSA 암호 알고리즘.

e, n : 공개키 d : 비밀키.

(e, n, d 는 일련의 수학적 성질을 만족).

RSA 암호화.

$m^e \bmod n = c$

m: 평문, c: 암호문

RSA 복호화

$c^d \bmod n = m^{ed} \bmod n$

$= m.$

Homomorphic Property of RSA

$(c_1 c_2)^d \bmod n \overset{?}{=} m_1 m_2$

$(c_1 c_2)^d \bmod n$

$= (m_1^e m_2^e)^d \bmod n$

$= (m_1 m_2)^{ed} \bmod n$

$= m_1 m_2$

동형암호 Use Case 1. Cloud기반 데이터 분석

클라우드상에서 암호화된 상태로 데이터분석 서비스 이용 가능하여
사용자는 데이터 노출·유출에 대한 염려없이 클라우드 분석 서비스 이용가능

Cloud 분석 사용자 (데이터 소유)

공개키로 데이터 암호화

비밀키로 분석결과 (모델) 복호화

분석을 위한 암호화된 데이터 전송
(One-way Encryption)

암호화된 분석결과 반환

Cloud 분석 서비스 제공자

암호화된 상태로 데이터 저장

클라우드內 복호화 키가 없어 데이터 유출 원천차단

암호화된 상태로 분석
(예: Logistic Regression)

동형암호 Use Case 1. Cloud기반 데이터 분석

평문데이터 대상 분석과 암호문 데이터 대상 분석의 정밀도 차이가 0.01%이하

PoC 명	PoC 개요 및 목표	분석 방법 및 데이터
중환자 급성 악화 예측	▪ 수술 환자 및 중환자실 환자의 급성 악화를 조기에 발견 할 수 있는 예측하는 모델을 개발하여, 조기 대응 진료에 활용	▪ MIMIC[1]內 수술환자, 중환자실 환자의 생체 징후 데이터 (Vital Sign 등), 환자의 임상 데이터 ▪ 생존률 분석을 위한 Logistic Regression 적용 ▪ 암호화된 데이터로 Training 후 확보한 분석모델에 대해 AUROC[3] 도출: 78.5% (평문데이터 기반 분석時 78.5%)
암환자 예후 예측	▪ 암 환자 예후 예측 모델을 개발하여, 근거 기반 맞춤 진료 지원 체계 수립	▪ SEER[2]內 암 조직 검체에 대한 유전체 데이터, 암 종별 임상 데이터 ▪ 5년 생존률 분석을 위한 Logistic Regression 적용 ▪ 암호화된 데이터로 Training 후 확보한 분석모델에 대해 AUROC를 계산: 70.3% (평문데이터 기반 분석時 70.3%)

1) MIMIC ('Medical Information Mart for Intensive Care') 2001년부터 2012년 사이 Beth Israel Deaconess 메이컬 센터 중환자 6만명의 의료 데이터를 비식별화하여 공개
2) SEER (The Surveillance, Epidemiology, and End Results)는 NCI (National Cancer Institute)의 프로그램으로 1973년부터 2014년까지 미국에서 발생한 암 환자 의료 데이터를 공개
3) AUROC: Area under the Receiver Operating Characteristic

동형암호 Use Case 1. 과학동아 게재

P

질병 연구 패러다임 전환

글 최영준 기자

"의학 연구에서 개인정보 보호는 아주 중요한 문제입니다. 특히 유전자 정보는 매우 민감한 정보인 만큼 필요한 연구에 활용하면서도 안전하게 보호하는 게 필수입니다. 유전자 정보를 암호화시켜서 공유하고 필요한 분석 결과만 얻을 수 있다면 의료와 제약 연구에서 민감한 개인정보를 보호해줄 수 있을 겁니다."

4월 11일 삼성SDS 연구소에서 만난 조지훈 보안연구팀장은 최근 의학 연구와 제약 산업에서 개인정보보호 문제가 화두라며 이 같이 말했다. 정보기술(IT) 기업인 삼성

퓨팅을 이용해 여러 병원이 연구용 유전체 정보를 공유하는 방법이 있지만, 이 역시 정보 유출 우려가 걸림돌이다.

삼성SDS 보안연구팀은 이런 문제의 해법으로 동형암호에 주목했다. 동형암호로 의료 정보를 암호화한 상태에서 분석한 뒤 결과만 확인하면 정보 유출에 대한 걱정 없이 연구를 할 수 있기 때문이다.

환자 생존율 예측에 동형암호 적용

태 변화와 결과(사망 여부)에 대한 정보를 담고 있다. 다른 하나는 미국 국립보건원(NIH) 산하 국립암연구소(NCI)에서 1973~2014년 수집한 연골육종이라는 희귀 암환자 1088명의 상태와 치료 기록, 치료 경과 등을 기록한 데이터다.

연구팀은 두 데이터의 70%를 환자의 상태에 따른 생존율을 예측하는 기계학습(머신러닝) 모델을 개발하는 데 필요한 학습 자료로 썼다. 그리고 나머지 30%는 개발된 모델을 검증하는 데 썼다.

조지훈 삼성SDS 보안연구팀장(왼쪽에서 두 번째)이 3월 15일(현지 시간) 미국 매사추세츠 주 보스턴에서 열린 '동형암호 표준화 국제회의'에 참석해 패널 토의를 하고 있다. 조 팀장은 이 자리에서 "의료와 물류 등 다양한 분야의 데이터 분석에 동형암호를 적용할 수 있을 것"이라고 말했다.

동형암호 Use Case 2. 데이터 결합

데이터의 노출 및 Privacy 이슈에 대한 염려 없이 분산된 데이터를 서로 결합하여 더 정밀한 분석모델 도출 가능

데이터 결합의 필요성

신용평가時 금융권·비금융권 (통신사) 데이터 결합

- 사회초년생 등 금융거래정보가 부족한 개인
- 제2금융권 위험분석 정교화 가능

공동 마케팅

- 광고와 구매간 연관도 분석 (Google社)

데이터 결합 이슈

데이터 결합시 재식별로 인한 컴플라이언스 이슈
혹은 지나친 재식별화로 데이터 결합가치 하락

동형암호 기반 데이터결합 분석

이통사

금융사

Phase 1. 암호화된 데이터 서로 교환

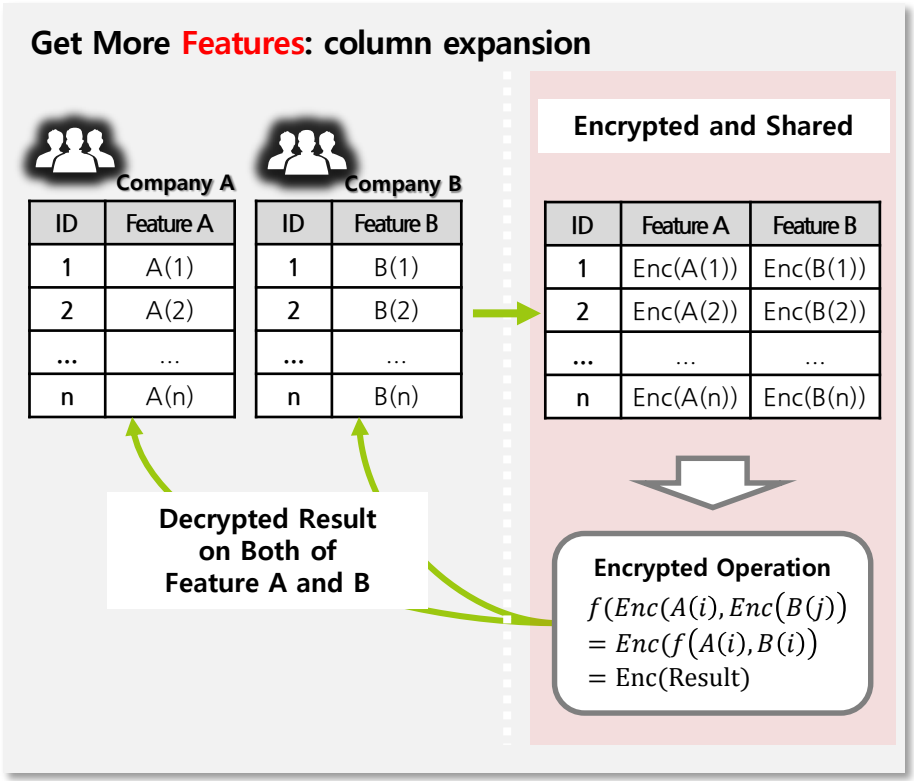
Phase 2. 각자 암호화된 두개의 데이터 셋 기반으로 분석

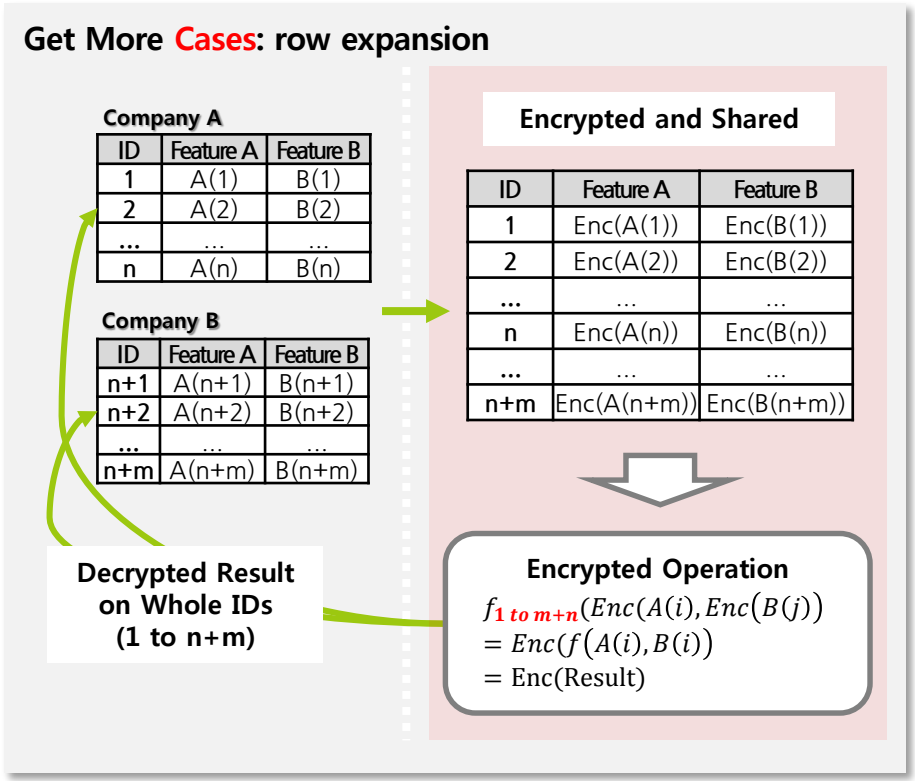
Phase 3. 암호화된 분석결과를 두조직이 협업하여 복호화

※ 암호화된 데이터 및 분석결과는 한조직이 일방적으로 복호화 불가능
※ N개의 조직이 데이터 결합하는 구조로 확장 가능

동형암호 Use Case 2. 데이터 결합

Two different organizations can use each other's data without compromising security and privacy





Agenda

I. Intro: Trends, Risk, Opportunities

II. Data-in-use Security

III. DB Encryption (Data-at-rest Security)

IV. Data-in-transit Security

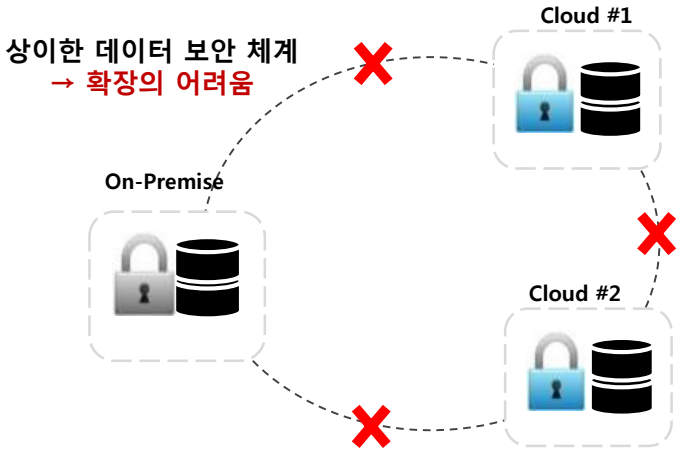
Cloud向 Database Encryption

특정 Cloud에 종속될 위험성을 제거하기 위해 On-prem과 Cloud를 단일한 보안체계로 지원하는 솔루션 사용 필요

<Gartner, Hype Cycle for Data Security, 2017>

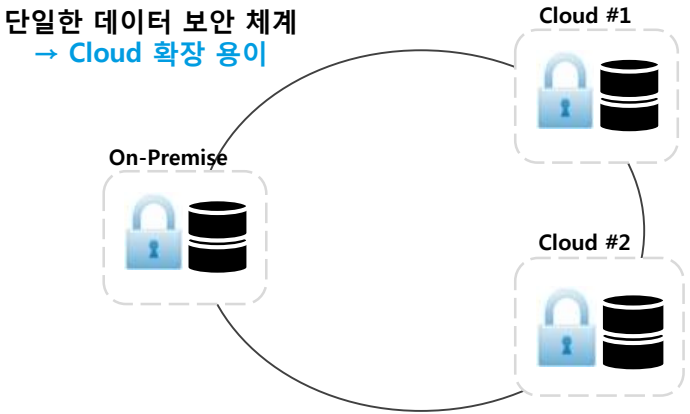
Pain Points

On-prem/Hybrid/Multi Cloud에 대한 단일한 보안관리체계(암호화 키, 정책, 배포 등) 관리의 어려움



SDS DB암호화 솔루션

SW기반 DB암호화 솔루션으로 On-prem/Hybrid/Multi Cloud 환경에 대한 단일한 보안관리체계 제공

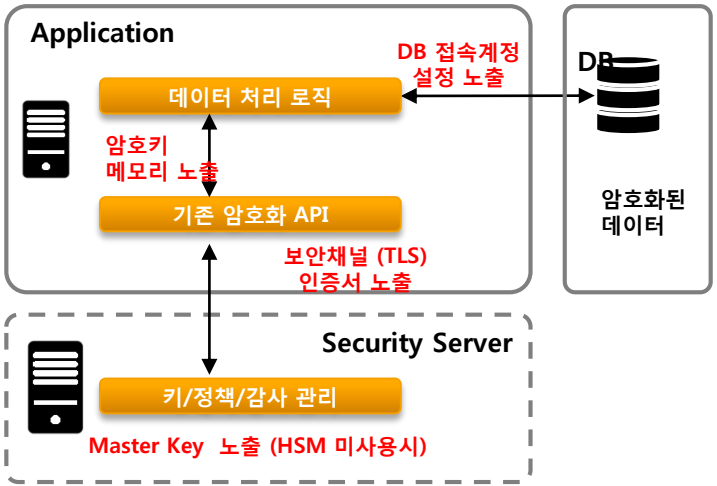


SDS솔루션 특장점

SW기반 해킹방지 암호기술을 적용하여 高보안의 데이터 암호화 솔루션 지원 (H/W의존성 탈피 가능)

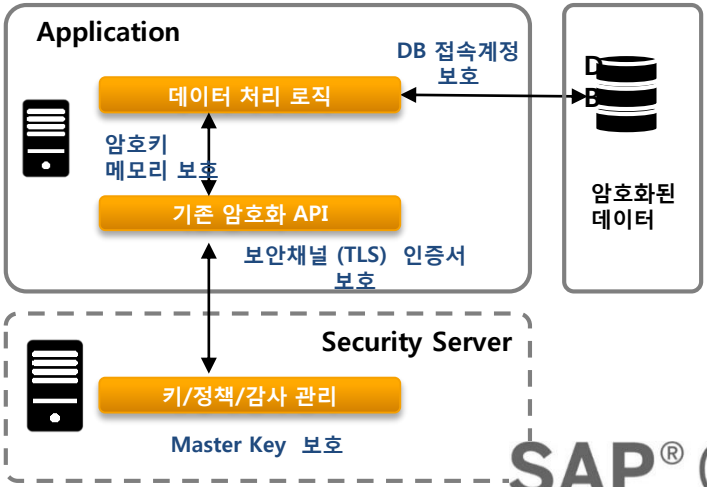
AS-IS: 他 암호화 솔루션

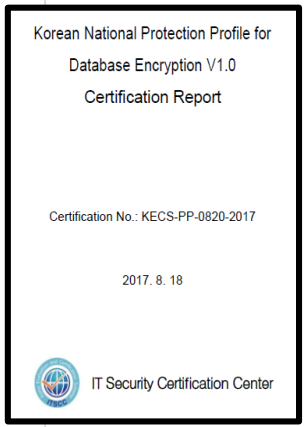
기존 데이터 암호화 솔루션은 클라우드 환경에서
다양한 침해 시나리오에 노출 가능



TO-BE: SDS SW기반 암호화 솔루션

자사 독자 SW기반 해킹방지 기술을 적용하여 암호키의
저장/사용시 보호





SAP® Certified
Integration with SAP S/4HANA®

Agenda

I. Intro: Trends, Risk, Opportunities

II. Data-in-use Security

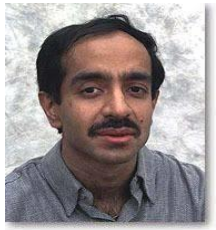
III. Data-at-rest Security

IV. Post-Quantum Crypto

(Data-in-transit Security)

Attacks 양자컴퓨터와 양자알고리즘을 이용한 공격

양자컴퓨터가 현실화되면 적용 가능한 양자기반 효율적인 알고리즘들이 제안되어 있으며, 이들 중 암호해독에 큰 영향을 미치는 알고리즘들이 존재함



양자 알고리즘

1 Shor Algorithm [1994]

소인수분해, DLP 문제를 P-time에 풀어냄

$O\left(e^{1.9(\log N)^{\frac{1}{3}}(\log\log N)^{\frac{2}{3}}}\right)$

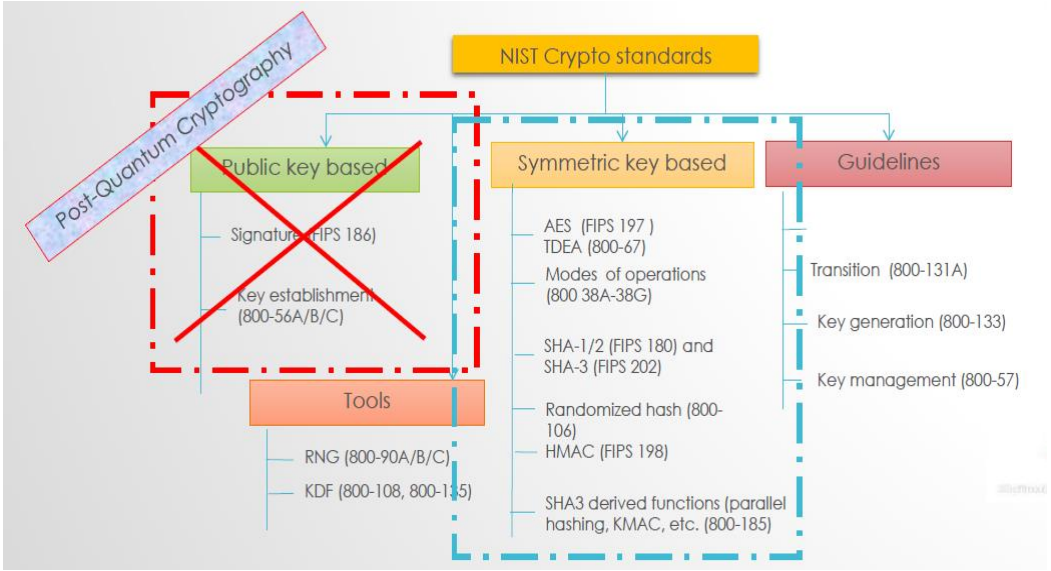
$\rightarrow O((\log N)^3)$

2 Grover Algorithm [1996]

주어진 함수의 특정 출력값에 대응되는 입력값을 효율적으로 찾아냄

$O(N) \rightarrow O(\sqrt{N})$

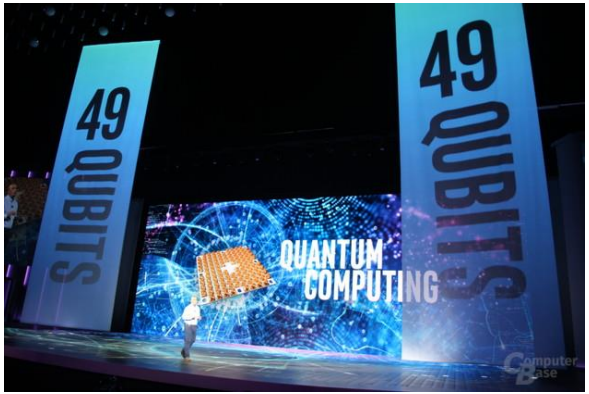
Impact on Crypto

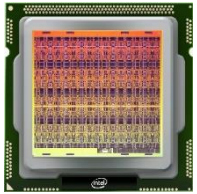


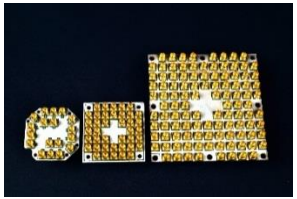
Quantum Computers

양자컴퓨팅은 예상보다 훨씬 빠른 속도로 발전하고 있음 (적어도 알려진 바로는...)

IBM 50 qubit (2017.11)	Intel 49 qubit (2018.01)	Google 72 qubit (2018.03)







Post Quantum Cryptography (PQC, 양자내성암호)

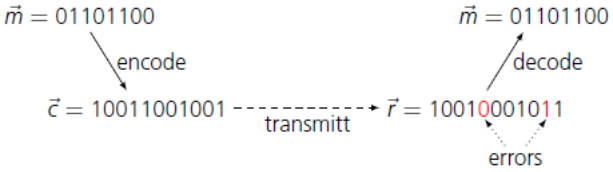
NIST(미국표준국)과 유럽에서는 양자컴퓨터와 양자암호알고리즘에 안전한 PQC 표준화를 진행 中, 그러나 온도차가 존재

* **Initial recommendations** [PQCRYPTO Project, 2015] : https://pqcrypto.eu.org/index.html

Code-based Cryptography [`78]

Public-key cryptosystem based on **error-correcting codes**

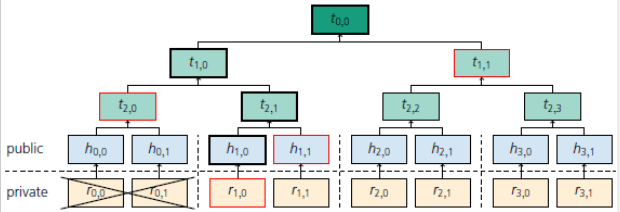
- **McEliece system***
- Niedereiter system



Hash-based Cryptography [`79]

Signatures are built **around hashes**, have security proofs that only rely on **the security of a hash function**

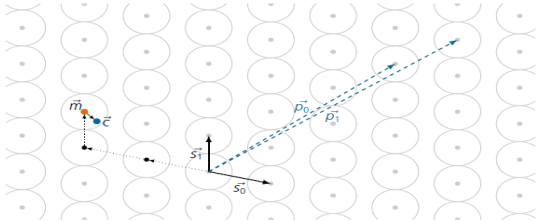
- **XMSS***
- **SPHINCS***



Lattice-based Cryptography [`98]

Public-key/signature cryptosystem based on **SVP and CVP** in lattices

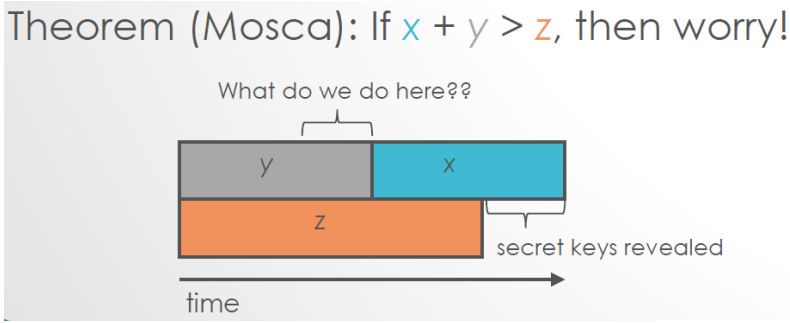
- NTRU
- GGH(Goldrech-Goldwasser-Halevi)



※ Others: Multivariate Cryptography [`88], Isogeny-based Cryptography [`11], ….

우리에게 남아있는 시간은?

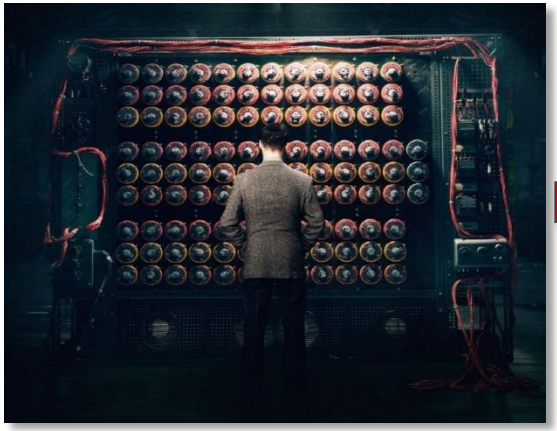
실제 암호알고리즘을 공격할 수 있는 양자컴퓨터의 출현 전부터 이에 대한 대비가 필요



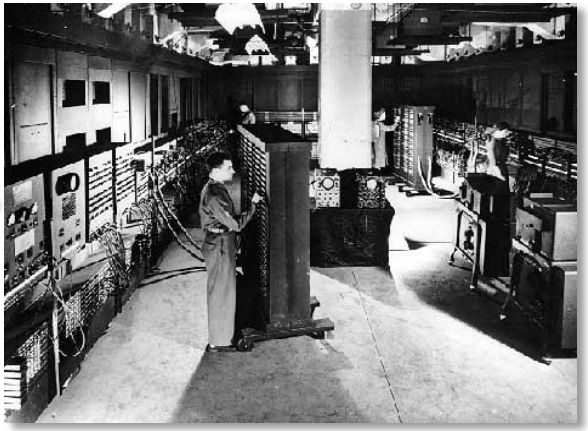
M. Mosca:
e-Proceedins of 1st ETSI Quantum-Safe Cryptography Workshop, 2013

1. **x years** – security shelf-life (How long must our **secrets remain secret**?)
2. **y years** – migration time (How long will **a full transition to PQC** take?)
3. **z years** – collapse time (When do we expect **a large-scale QC** to exist?)

History of Security & Computing: 양지 vs. 음지



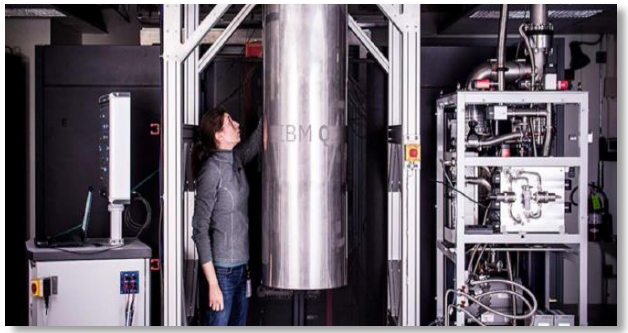
Colossus (1943년)

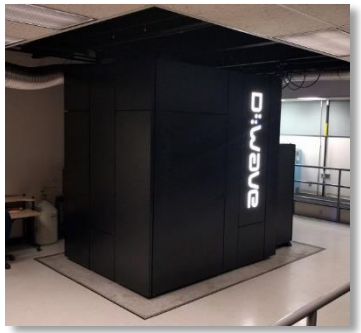


ENIAC (1947년)

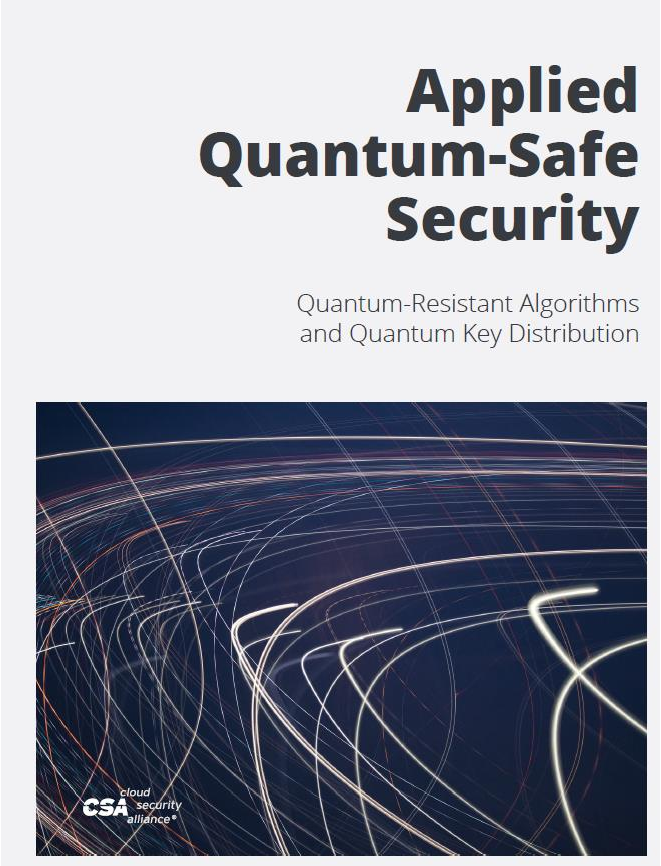
IBM PC (1970년대초)

?

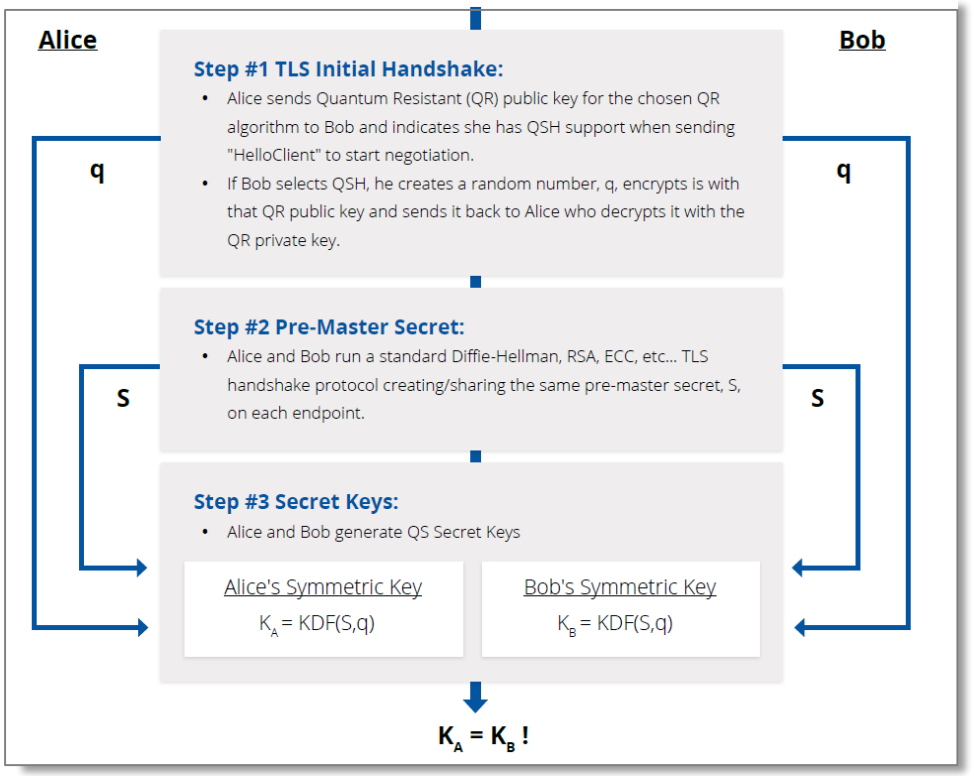




PQC기반 Secure Channel Hybrid Approach



Quantum-safe Hybrid Protocol <출처: Cloud Security Alliance>



SAMSUNG SDS

Realize your vision

www.samsungsds.com