개인정보 보호를 위한 신뢰계산기술

서울대학교 전기정보공학부

이병영

byoungyoung@snu.ac.kr

Speaker: 이병영

- Research areas: Hacking, Systems Security, Software Security
  - Microsoft Research, Research Intern (2012 Summer)
  - Google Chrome, Software Engineering Intern (2014 Summer)
  - Purdue University, Assistant Professor (2016-2018)
- Found 100++ vulnerabilities from Windows kernel, Linux kernel, Chrome, Firefox, etc.
- Internet Defense Prize by Facebook and USENIX (2015)
- Three times DEFCON CTF Finalist (2007,2009, and 2011)
- DARPA Cyber Grand Challenge (CGC) Finalist (2016)
- Google ASPIRE Awards (2019)

My Research Areas: Protecting Commodity Systems

Apps

DangNull [NDSS 15]
Expector [WWW 15]
TrackMeOrNot [WWW 16]
MEDS [NDSS 18]
CRFuzz [FSE 20]

CaVer [USENIX Sec 15]
HexType [CCS 17]

NGINX

ASLR-Guard [CCS 15]

OS

Juxta [SOSP 15]
KUP [ATC 16]
Razzer [S&P 19]
uXOM [Security 19]
HFL [NDSS 20]

Morula [S&P 14]
Kenali [NDSS 16]
ExpRace [BlackHat20]

CAB-Fuzz [ATC 17]

Intel® SGX

SGX-ASLR [NDSS 17]
Obliviate [NDSS 18]
Obfuscuro [NDSS 19]
Trustore [CCS 20]

HW

HDFI [S&P 16]

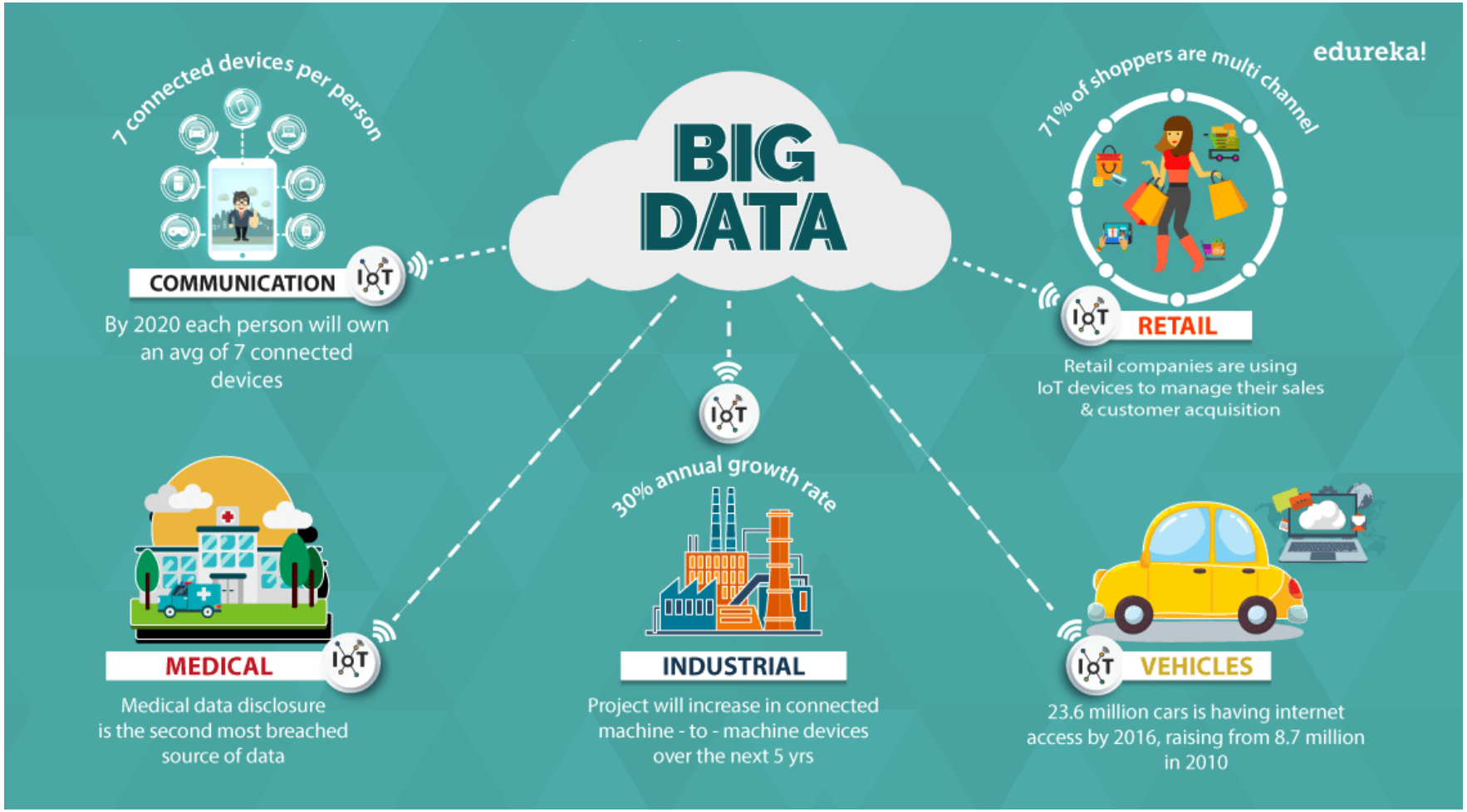
Minion [NDSS 18]

Attack Mitigation

Vulnerability and Exploitation

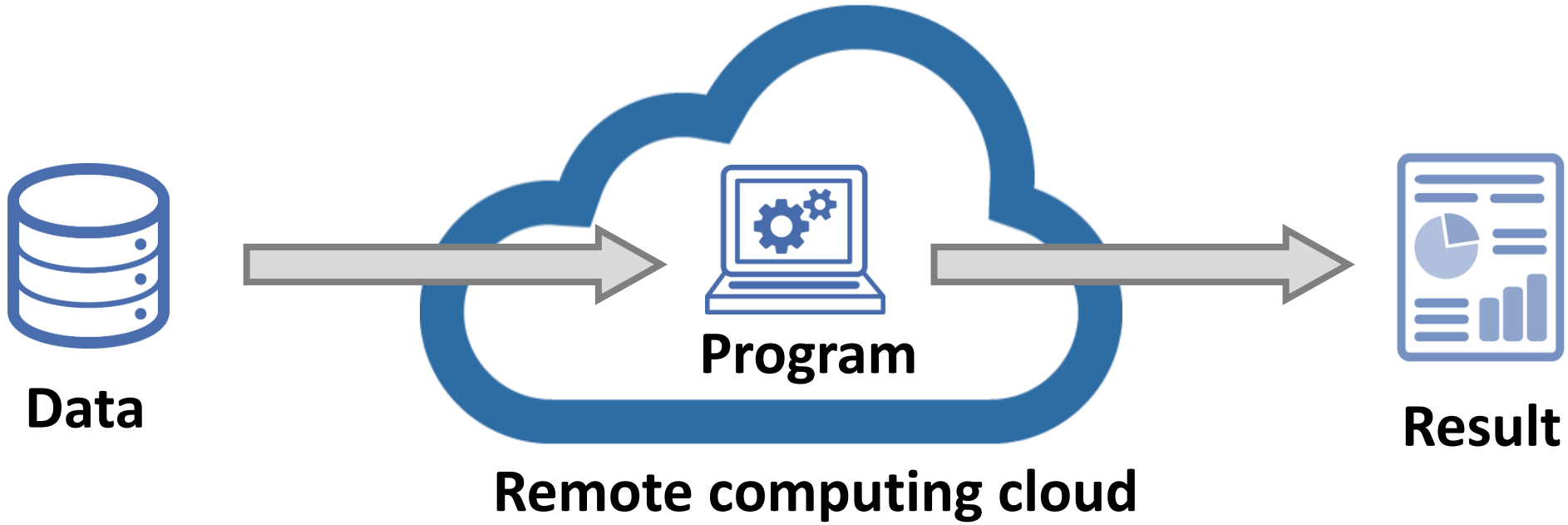
Secure Trusted Computing

The Age of Big Data

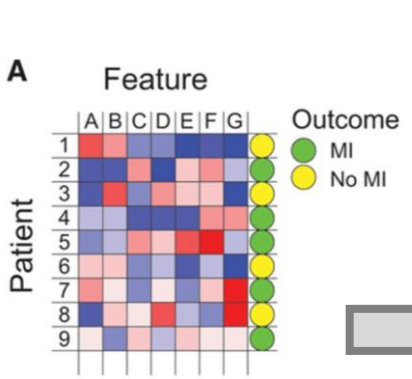


https://www.edureka.co/blog/big-data-applications-revolutionizing-various-domains/

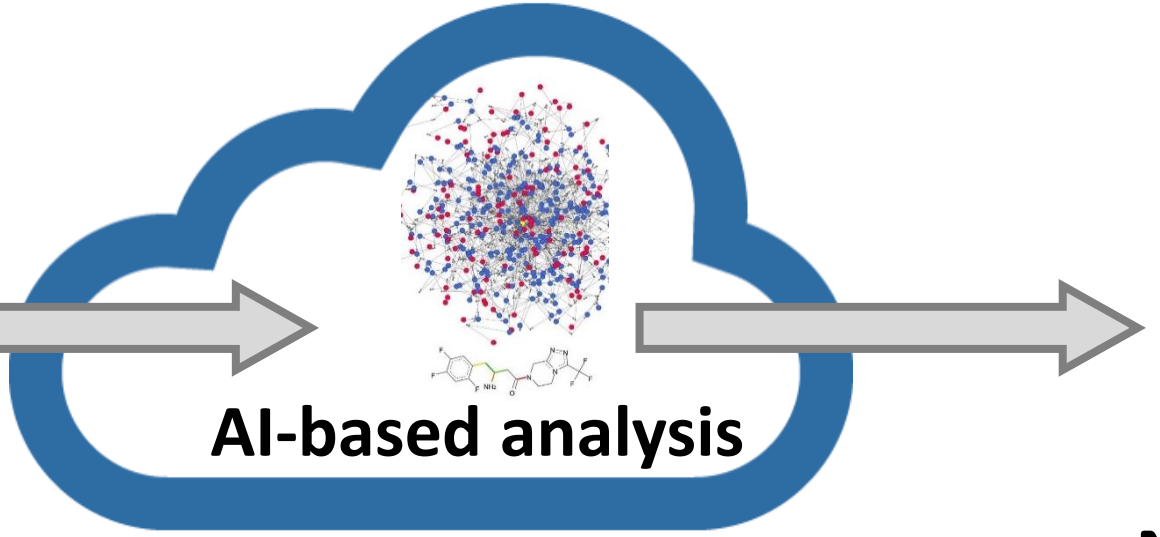
Frameworks for Big Data, AI, ML, and DL



Frameworks for Big Data, AI, ML, and DL



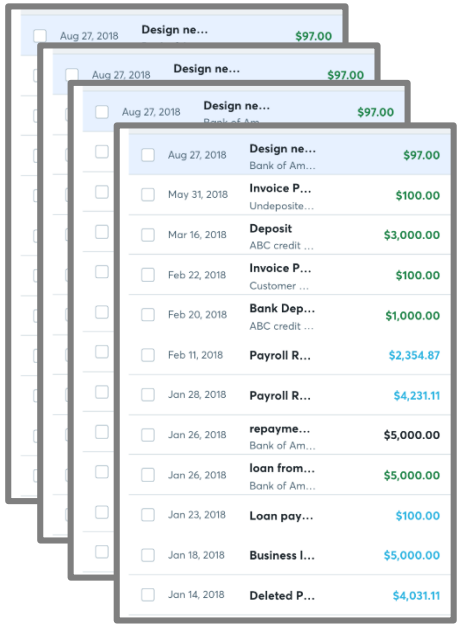
Patient medical records



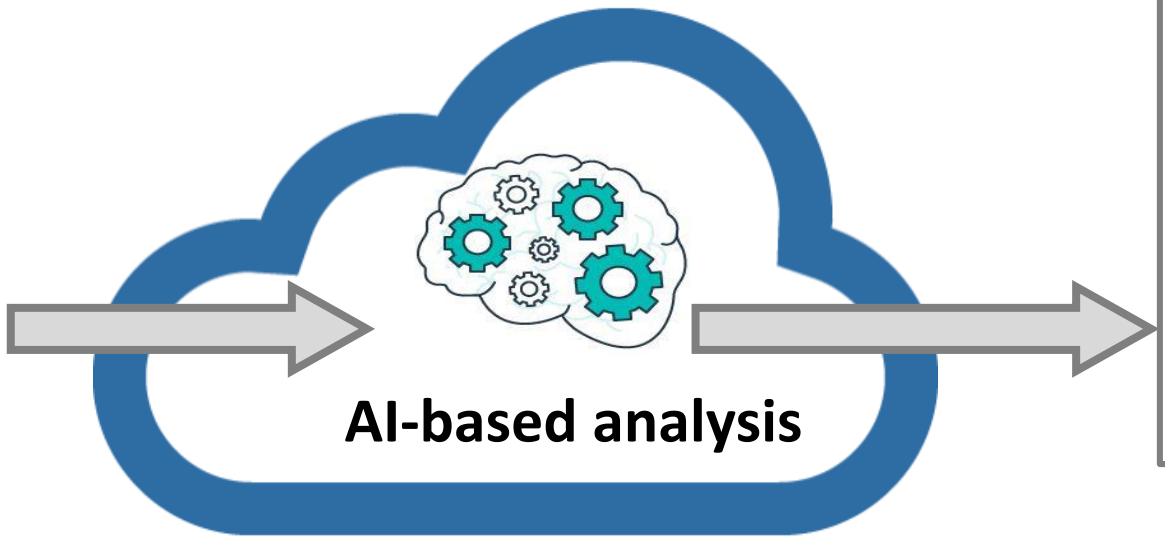
Remote computing cloud

New drug discovery

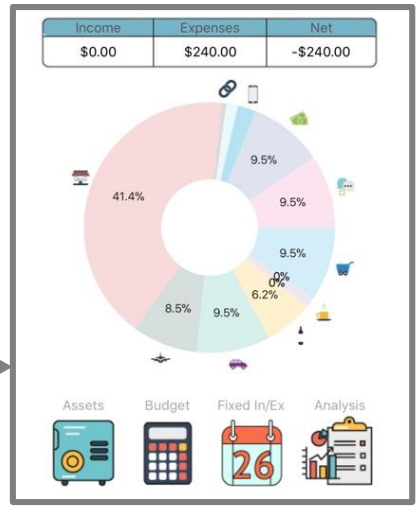
Frameworks for Big Data, AI, ML, and DL



Bank transaction record

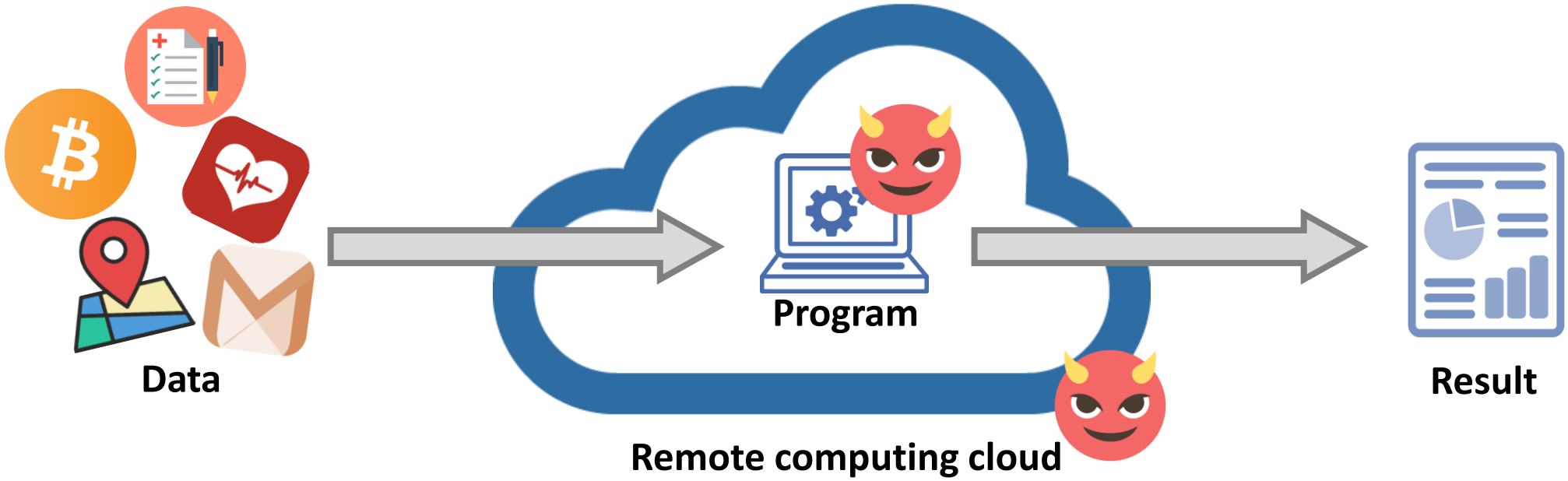


Remote computing cloud



Account book summary with recommendation

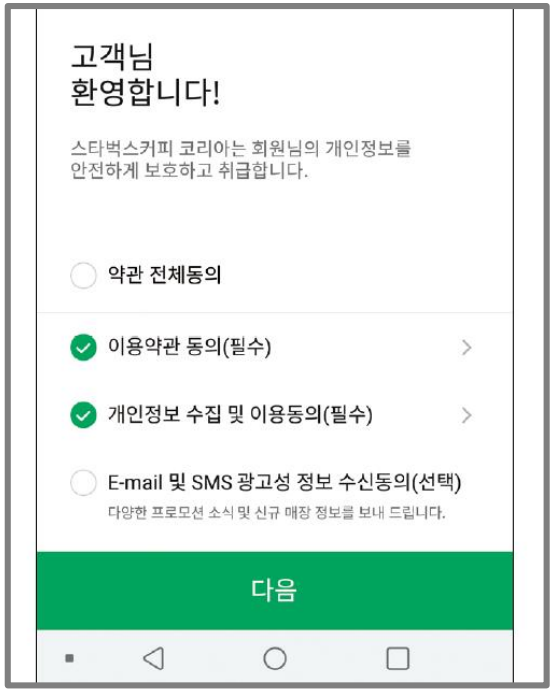
Security and Privacy Threats

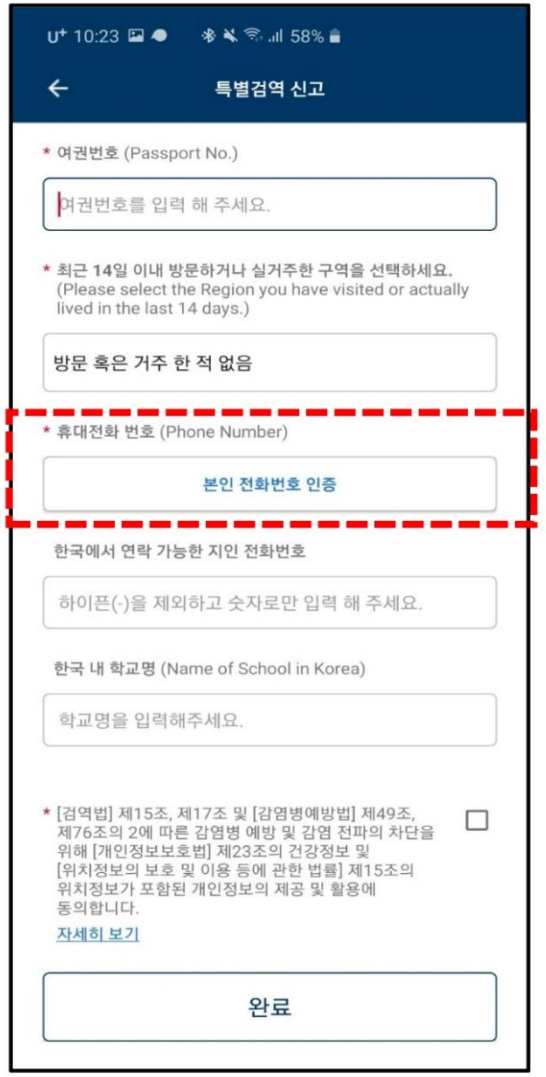


Data anarchy: Users have no control over their data

Challenges: Too strong attack models

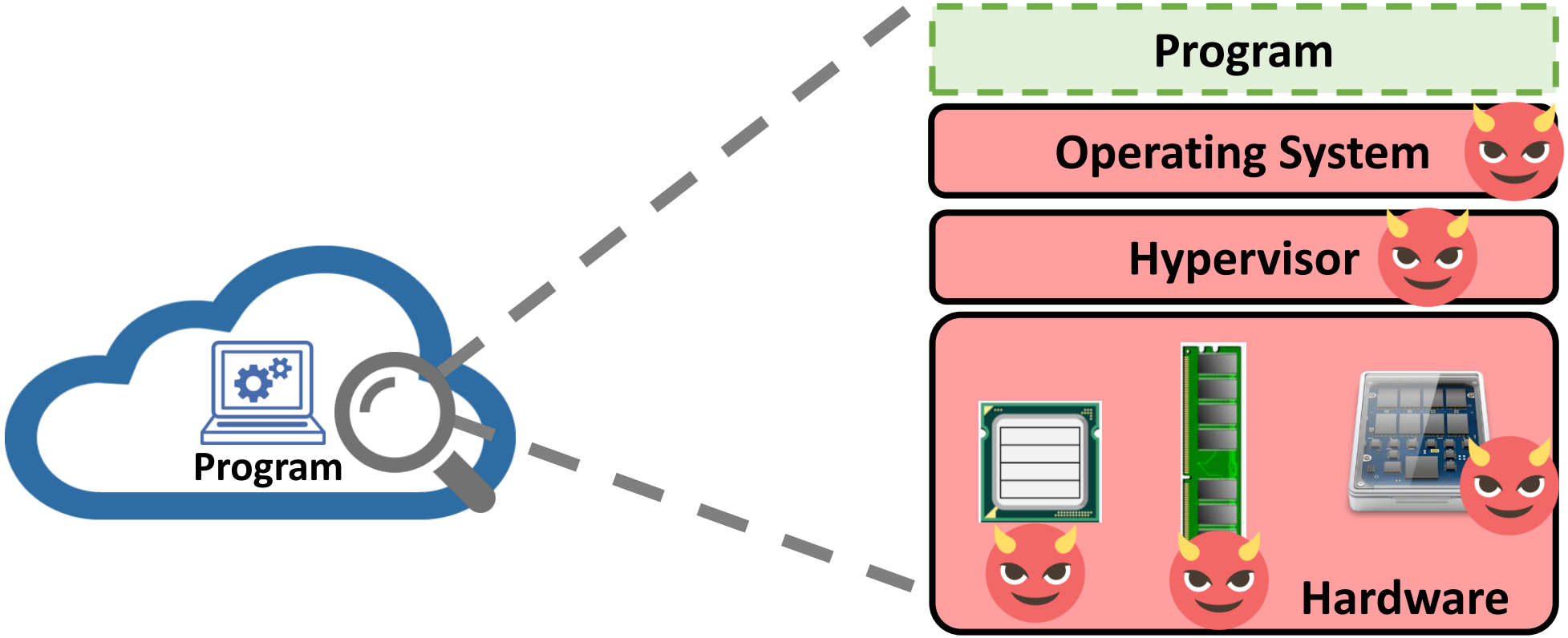
- A program (or program owners) can be malicious
  - A program may promise it would not abuse the data, but there's no technical enforcement





Challenges: Too strong attack models

- Cloud infrastructures can be malicious
  - Clouds include entire computing infrastructure to run a program
  - If any of those is malicious, user's data can be leaked

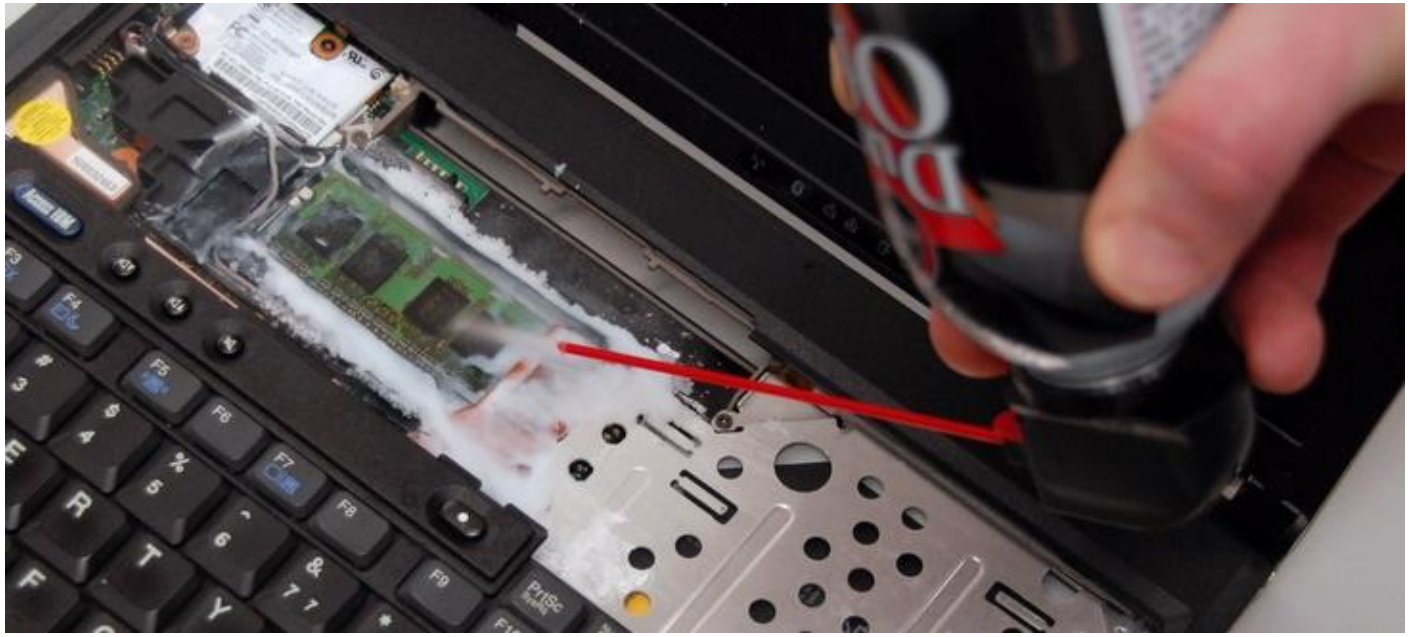


Challenges: Too strong attack models

- Clouds can be malicious
  - Physical attacks make this problem even more challenging
  - System admins can easily pull out the disk to read the data

Challenges: Too strong attack models

- Clouds can be malicious
  - Cold-boot attack: Even DRAM's data can be stolen



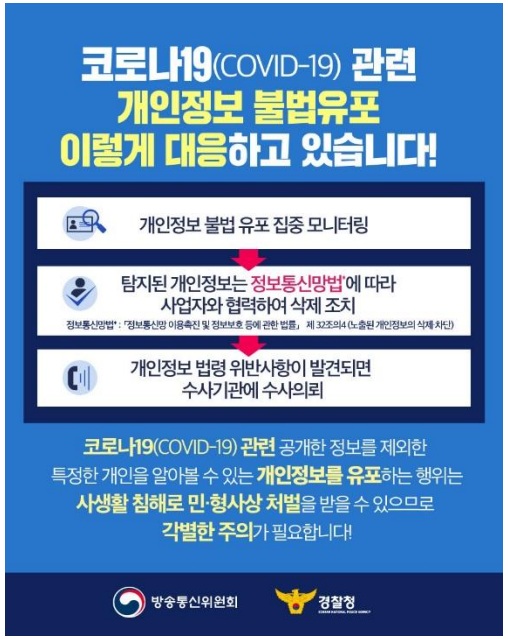
-50˚C: less than 0.2% decay after 1 minute

"Lest We Remember: Cold Boot Attacks on Encryption Keys [USENIX Security 08]"

Fundamental Issue: Data Utility vs. Data Privacy

- **Data utility**
  - Data is the key to truly enable AI/ML/DL services
- **Data privacy**
  - Data contains critical privacy information of users

- How to satisfy both **data utility** and **data privacy**?



Potential Solutions for Data Security

- Data anonymization (데이터 비식별화)
- Differential Privacy (차등보호)
- Homomorphic Encryption (동형암호)
- **Hardware-Assisted Trusted Computing (신뢰계산)**
  - **The most efficient: near to native execution speed**
  - **The most practical: running a generic program**

Data Anonymization (데이터 비식별화)

- Remove personally identifiable information from data
  - While maintaining the data utilization
- k-anonymity
  - Blend each data item with k-1 items having identical column information

microdata

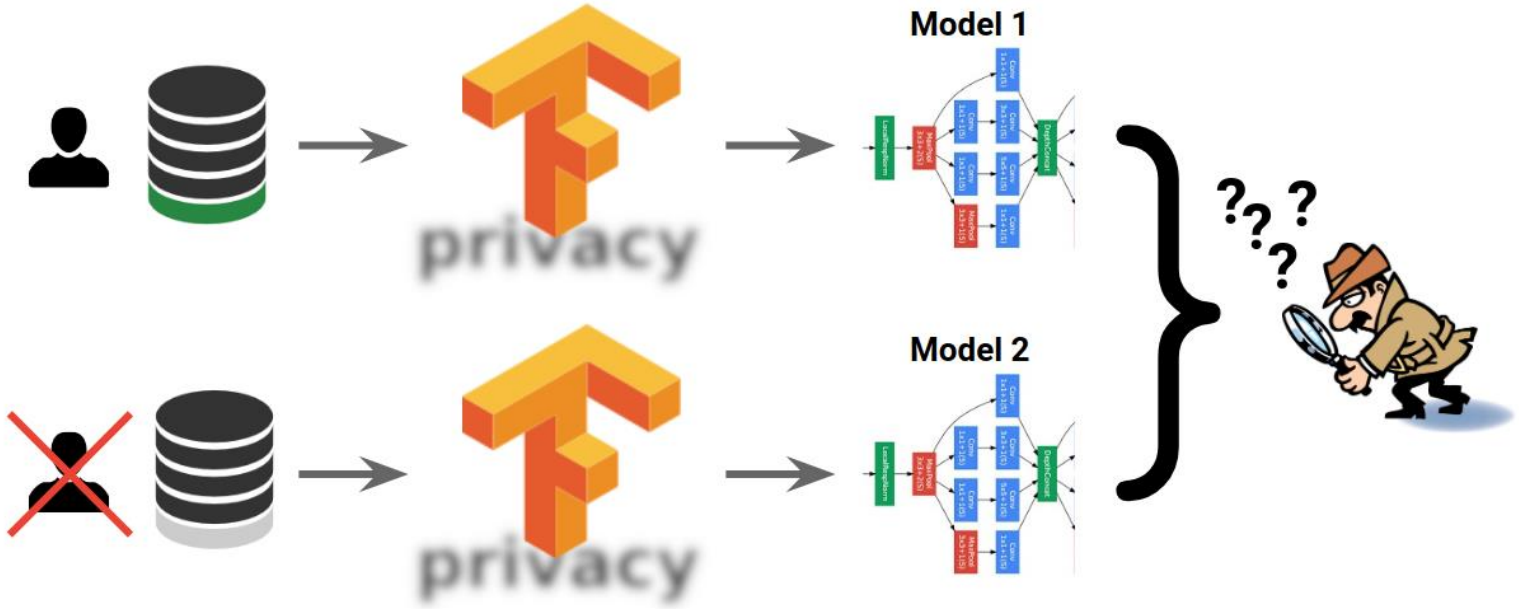
id	Zipcode	Sex	National.	Disease
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

4-anonymous data

id	Zipcode	Sex	National.	Disease
1	130**	<30	*	Heart Disease
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	1485*	≥40	*	Cancer
6	1485*	≥40	*	Heart Disease
7	1485*	≥40	*	Viral Infection
8	1485*	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

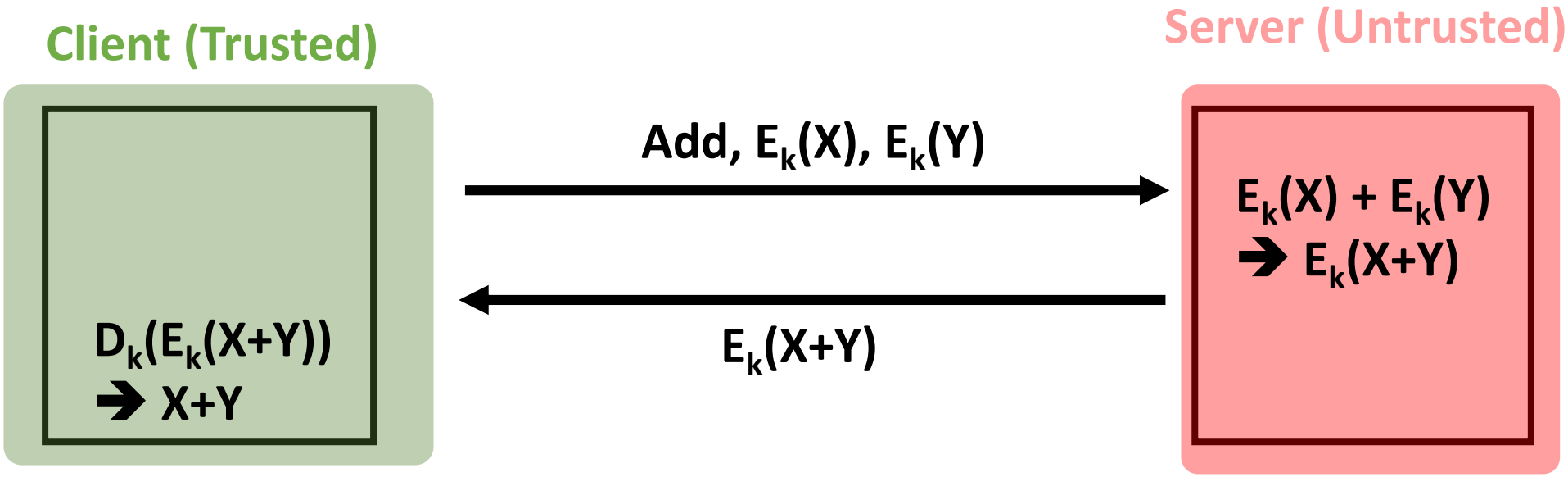
Differential Privacy (차등보호)

- Privacy protection algorithm for a statistical database
- Differential private
  - An observer seeing the output cannot tell if a particular individual's information was used in generating the output



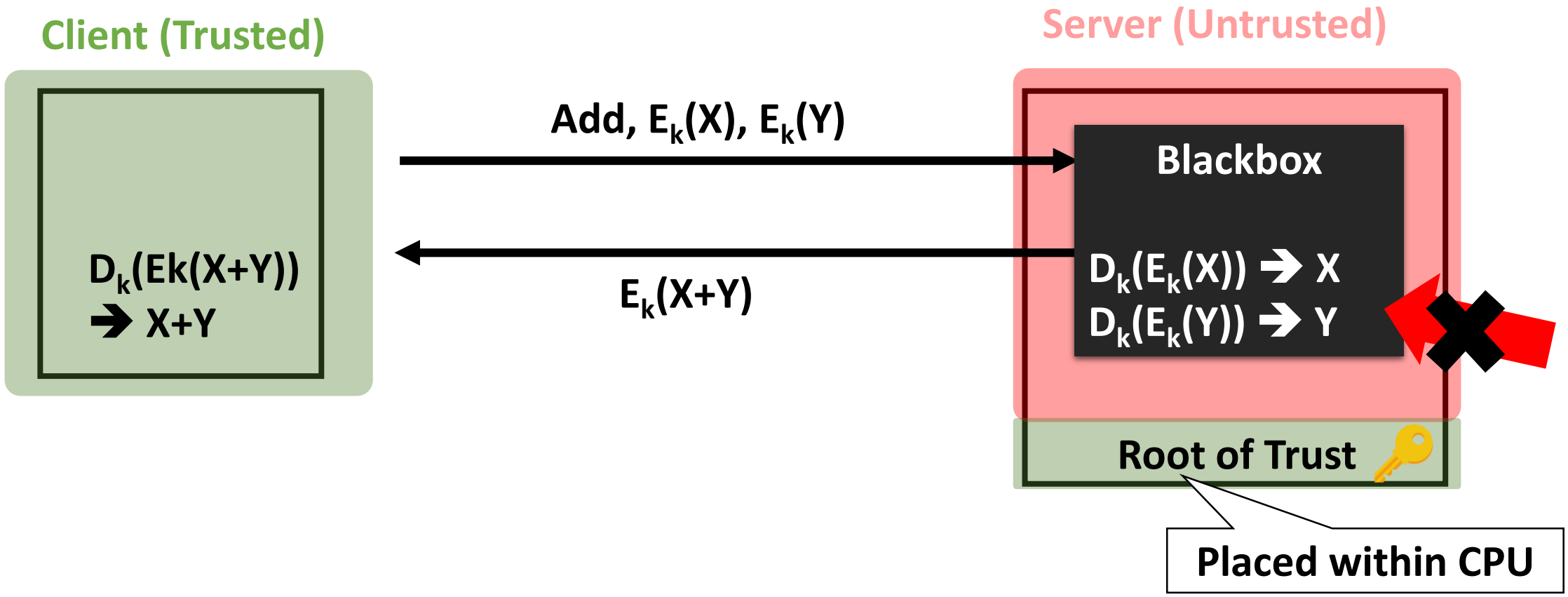
Homomorphic Encryption (동형암호)

- Computation over encrypted data
  - Example: Client wants to offload the computation, X+Y



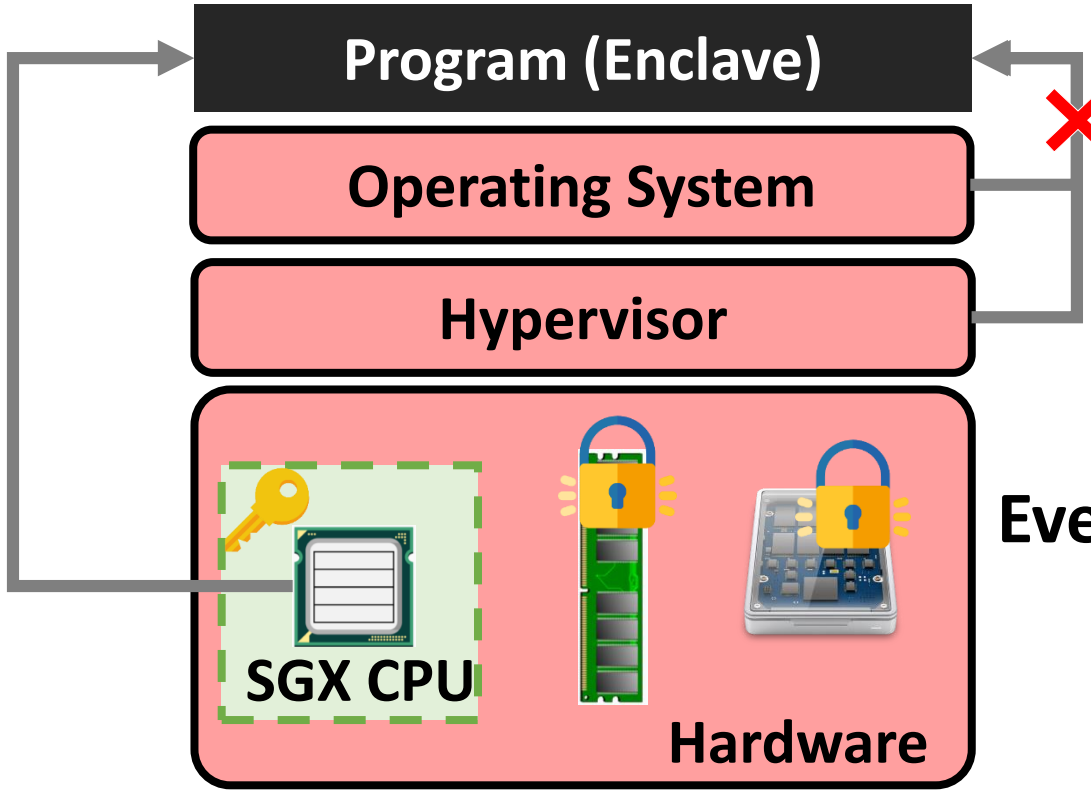
Hardware-Assisted Trusted Computing (신뢰계산)

- Trusted computation by placing a small root of trust in hardware



Intel SGX: Data Security Feature for the Future

Hardware-protected execution region



Cannot see a program

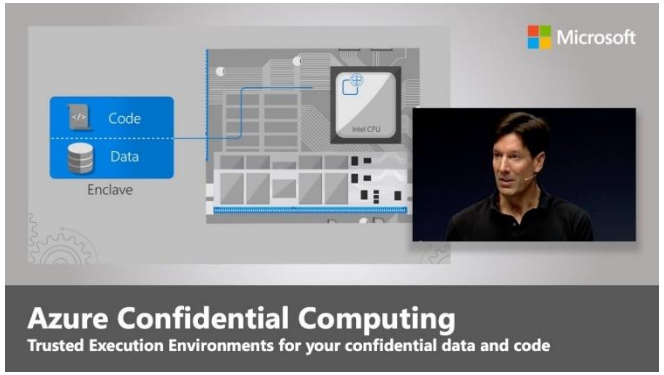
Everything is encrypted

Most Intel CPUs today are shipped with SGX support.

Intel SGX: already market available

- Most of consumer-grade Intel CPUs are shipped with SGX support

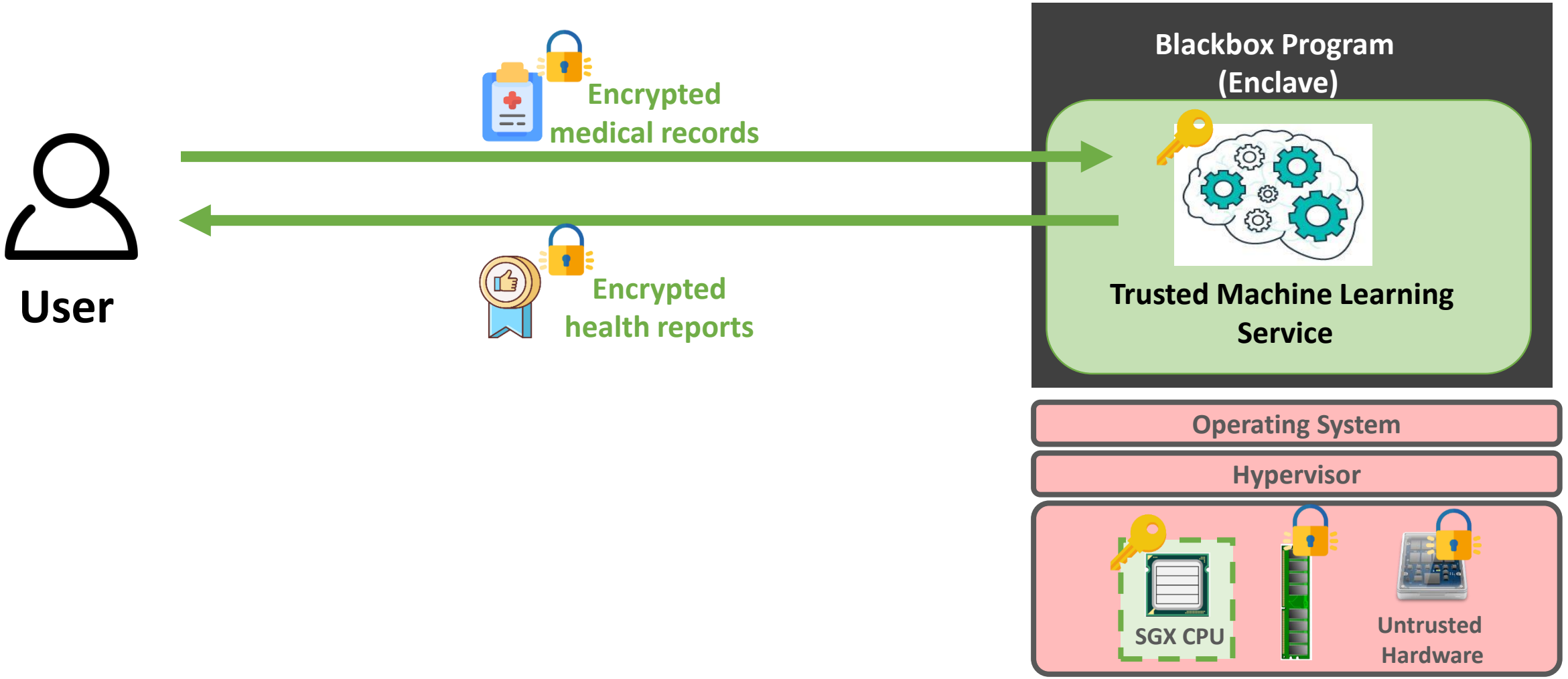
- Strong demands on SGX features from cloud providers
  - Growing security needs for trusted computing
    - Observing EU GDPR and any (expected) national regulation
  - Azure Confidential Computing is already available (since 2020 May)
    - SGX-based secure cloud services



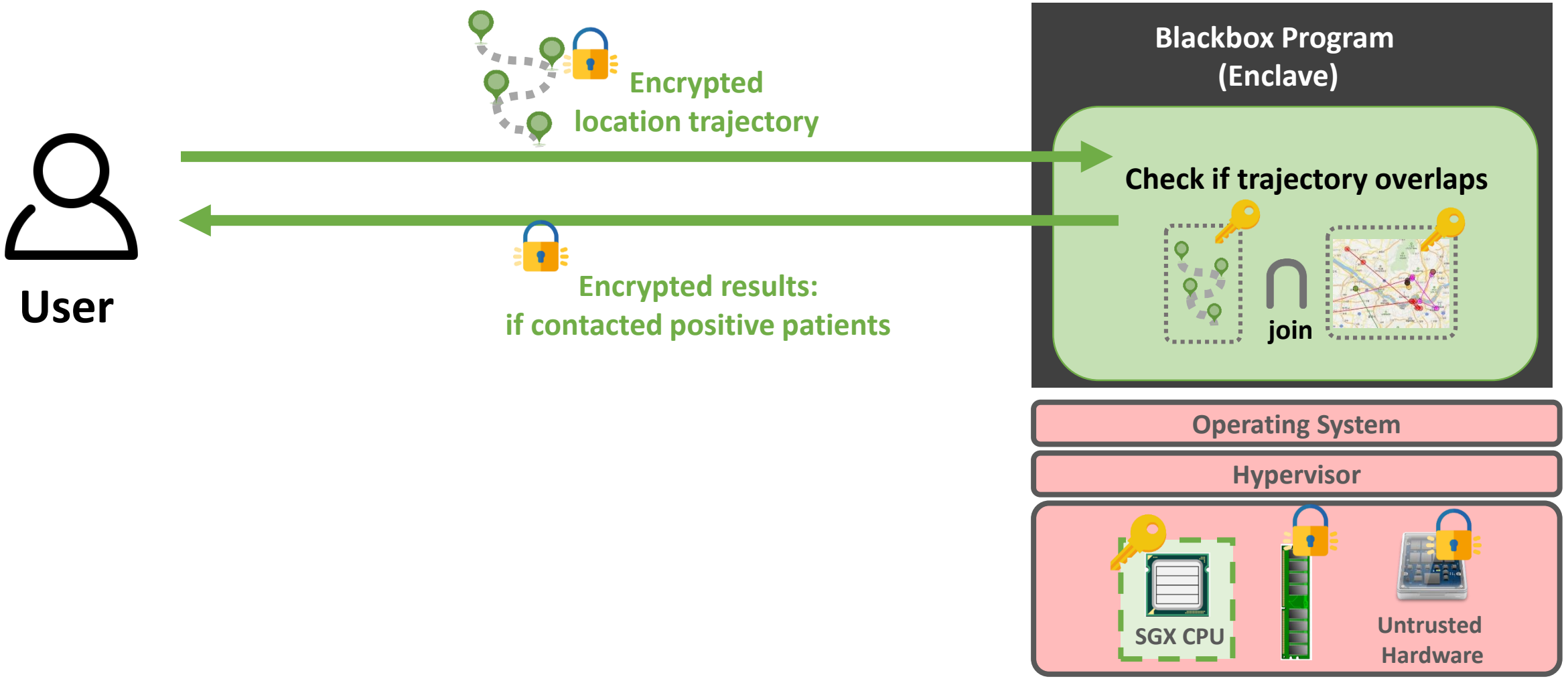
Truly Secure Applications with Intel SGX

- Trusted Machine Learning
  - 예제: 안전한 AI 기반 건강관리 서비스
- Trusted Private Join
  - 예제: 개인정보를 보호하는 코로나바이러스 환자 동선 확인
- Trusted Network Middleware/Server
  - 예제: 안전한 화상회의 아키텍쳐 (Zoom, Google Meet)

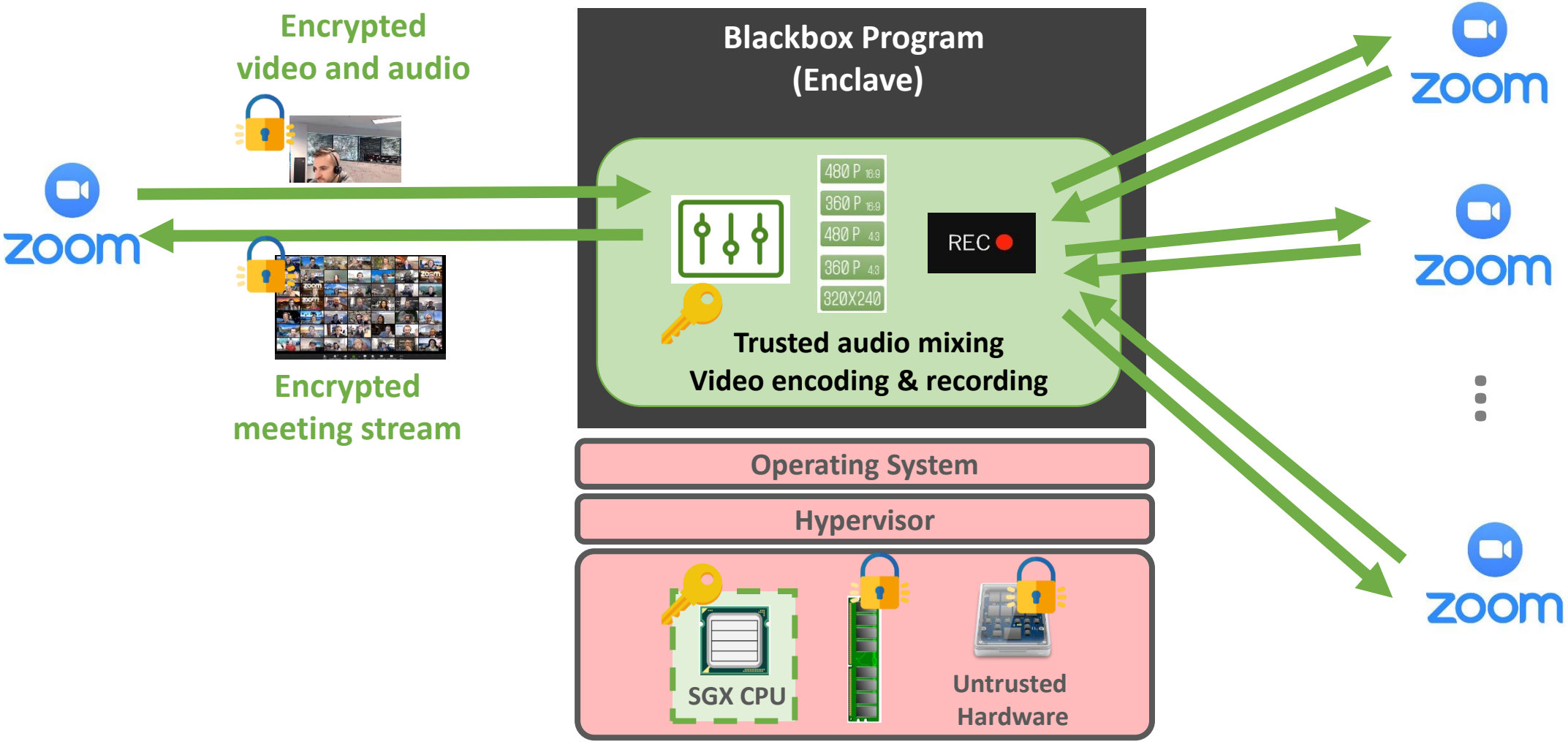
Trusted Machine Learning: Health Prediction



Trusted Private Join: Covid-19 Proximity Check

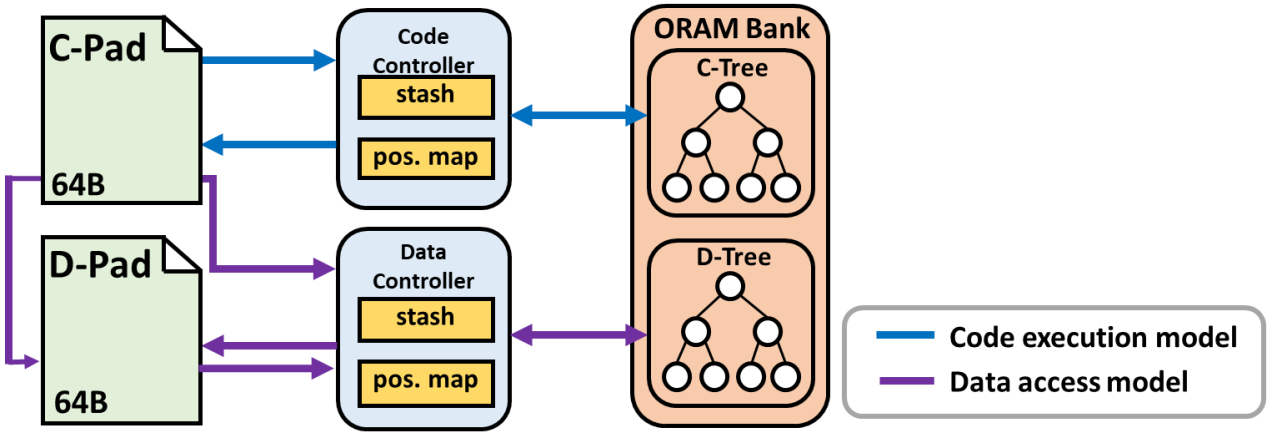


Trusted Network Server: Trusted Online Meeting

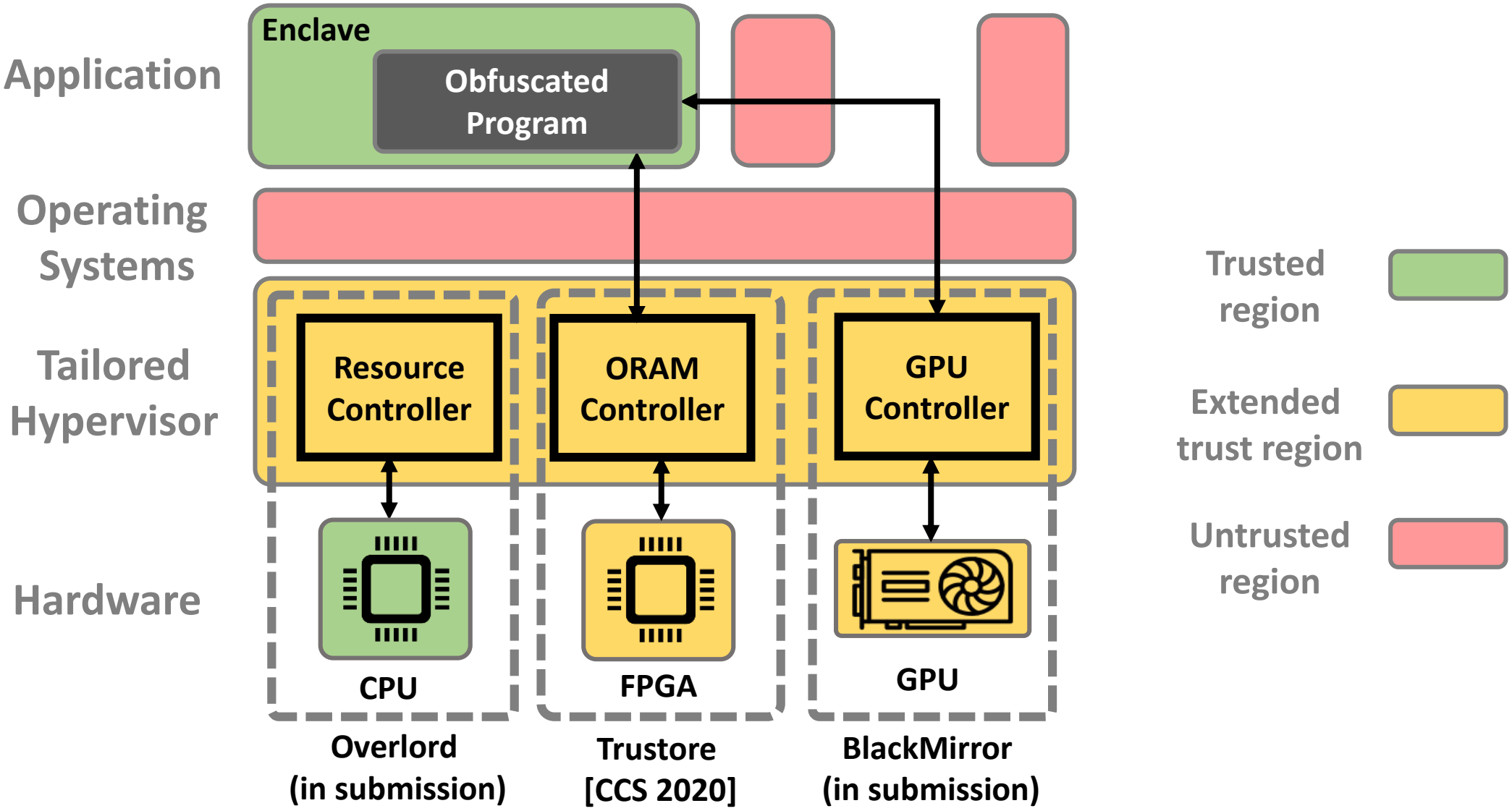


Side-Channel Resistant Intel SGX

- **Obliviate [NDSS 2018]**
  - ORAM-based file systems to prevent side-channel attacks
  - All file accesses are performed with ORAM
- **Obfuscuro [NDSS 2019]**
  - Program obfuscation on Intel SGX
  - All programs always exhibit the same control/data flows (using ORAM)



Enabling Practical Services for Intel SGX



Conclusion

- Protecting the data is crucial in the age of big data
- Trusted computing opens up new opportunities towards truly secure services
  - With systematic and technical security assurance

감사합니다

서울대학교 전기정보공학부
이병영
byoungyoung@snu.ac.kr