Aalto University

Information Security Research and Education

N. Asokan
Twitter: @nasokan, WWW: https://asokan.org/asokan

About me

Professor, Aalto University, from Aug 2013
Professor, University of Helsinki, 2012-2017

IEEE Fellow (2017), ACM Distinguished Scientist (2016)
Associate Editor-in-Chief, IEEE Security & Privacy

Previously

Nokia (14 y; built up Nokia security research team)
IBM Research (3 y)

More information on the web (https://asokan.org/asokan) or Twitter (@nasokan)

Secure Systems Group

Prof N. Asokan

Professor, Department of Computer Science

Director: Helsinki-Aalto Center for Information Security

https://asokan.org/asokan/

Prof Tuomas Aura

Professor, Department of Computer Science

https://people.aalto.fi/tuomas_aura

Dr Andrew Paverd

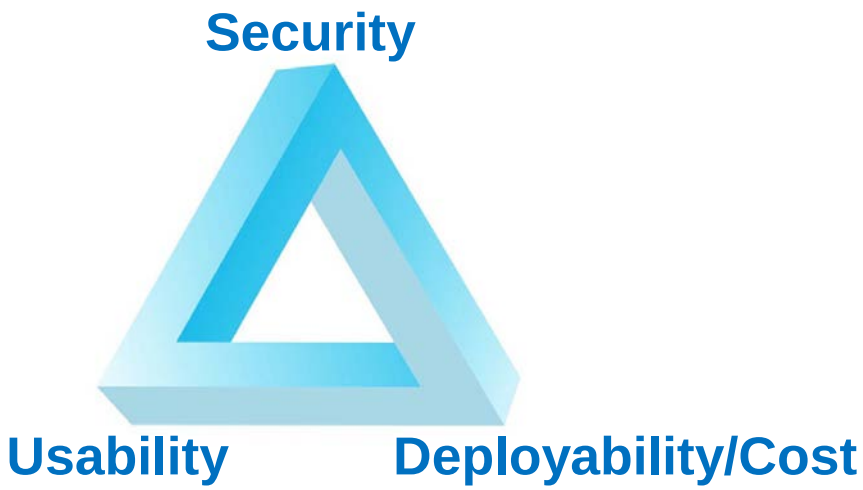
Research Fellow, Department of Computer Science

Deputy Director: Helsinki-Aalto Center for Information Security

https://ajpaverd.org

Secure Systems Group: Mission

How to make it possible to build systems that are simultaneously easy-to-use and inexpensive to deploy while still guaranteeing sufficient protection?



Secure Systems Group

http://cs.aalto.fi/secure_systems/

In Asokan's projects:

- 3 postdocs
- 5 full-time + 3 part-time PhD students

Several MSc students

- Best InfoSec thesis in Finland 2017, 2016 & 2014, Tietoturva ry
- Runner-up for Best CS thesis in Finland 2014, TKTS ry

Projects funded by

- Academy of Finland, Tekes
- Direct industry support: E.g., Intel http://www.icri-sc.org, [NEC Labs, Huawei]

Aalto University

Established in 2010, named in honour of ***Alvar Aalto***, the famous Finnish architect.

Science and art meet technology and business.

Aalto University

Promoting entrepreneurship

70 to 100

Companies are founded every year in our ecosystem

MIT Skolltech initiative rated Aalto's innovation ecosystem among the **top-5** rising stars in the world

Entrepreneurship is a more popular career option than ever – in the last four years, over **2 000** students have studied through the Aalto Ventures Program

50%

of Finnish startups that originate from universities come from the Aalto community

Aalto University


SLUSH
NOBODY IN THEIR RIGHT MIND WOULD COME TO HELSINKI IN NOVEMBER.
http://www.slush.org/

Research

Building systems that are secure, usable, and deployable

Current themes: Platform Security

How can we design/use pervasive hardware and OS security mechanisms to secure applications and services?

HardScope: Thwarting DOP with Hardware-assisted Run-time Scope Enforcement

Thomas Nyman, Ghada Dessouky, Shaza Zeitouni, Aaro Lehikoinen, Andrew Paverd, N. Asokan, Ahmad-Reza Sadeghi

(Submitted on 29 May 2017)

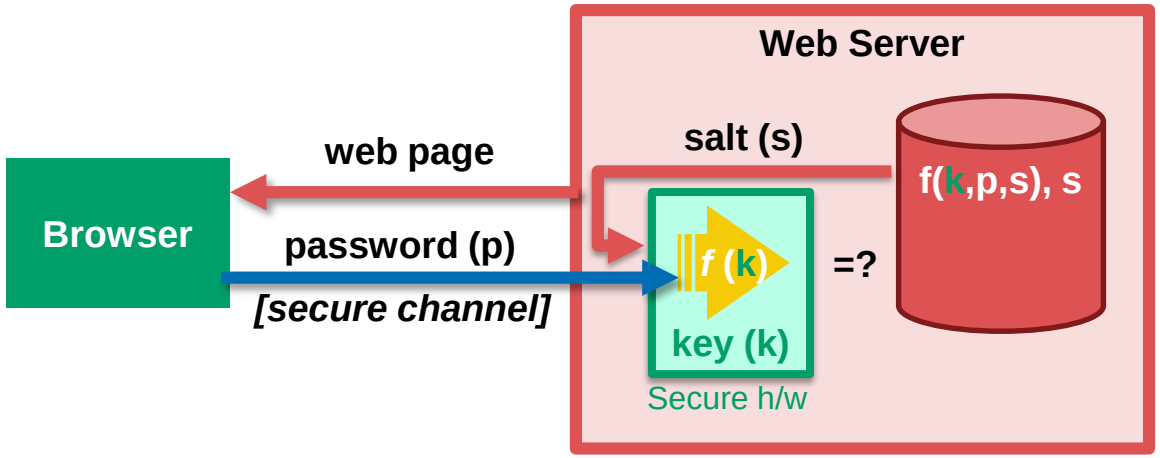
The widespread use of memory unsafe programming languages (e.g., C and C++), especially in embedded systems and the Internet of Things (IoT), leaves many systems vulnerable to memory corruption attacks. A variety of defenses have been proposed to mitigate attacks that exploit memory errors to hijack the control flow of the code at run-time, e.g., (fine-grained) ASLR or Control Flow Integrity (CFI). However, recent work on data-oriented programming (DOP) demonstrated the possibility to construct highly-expressive (Turing-complete) attacks, even in the presence of these state-of-the-art defenses. Although multiple real-world DOP attacks have been demonstrated, no suitable defenses are yet available.
We present run-time scope enforcement (RSE), a novel approach designed to mitigate all currently known DOP attacks by enforcing compile-time memory safety constraints (e.g., variable visibility rules) at run-time. We also present HardScope, a proof-of-concept implementation of hardware-assisted RSE for the new RISC-V open instruction set architecture. We demonstrate that HardScope mitigates all currently known DOP attacks at multiple points in each attack. We have implemented HardScope in hardware on the open-source RISC-V Pulpino microcontroller. Our cycle-accurate simulation shows a real-world performance overhead of 7.1% when providing complete mediation of all memory accesses.

https://arxiv.org/abs/1705.10295

Current themes: Platform Security

Enabling developers to secure apps/services using h/w and OS security

Example: SafeKeeper – using Intel SGX on server-side to protect passwords



Use secure hardware on server side

https://ssg.aalto.fi/projects/passwords/

SAFE KEEPER

Current themes: Machine Learning & Security

IEEE TRANSACTIONS ON COMPUTERS, VOL. 66, NO. 10, OCTOBER 2017 1717

Off-the-Hook: An Efficient and Usable Client-Side Phishing Prevention Application

Samuel Marchal, *Member, IEEE*, Giovanni Armano, Tommi Gröndahl, Kalle Saari, Nidhi Singh, and N. Asokan, *Fellow, IEEE*

Abstract—Phishing is a major problem on the Web. Despite the significant attention it has received over the years, there has been no definitive solution. While the state-of-the-art solutions have reasonably good performance, they suffer from several drawbacks including potential to compromise user privacy, difficulty of detecting phishing websites whose content change dynamically, and reliance on features that are too dependent on the training data. To address these limitations we present a new approach for detecting phishing webpages in real-time as they are visited by a browser. It relies on modeling inherent phisher limitations stemming from the constraints they face while building a webpage. Consequently, the implementation of our approach, *Off-the-Hook*, exhibits several notable properties including high accuracy, brand-independence and good language-independence, speed of decision, resilience to dynamic phish and resilience to evolution in phishing techniques. *Off-the-Hook* is implemented as a fully-client-side browser add-on, which preserves user privacy. In addition, *Off-the-Hook* identifies the target website that a phishing webpage is attempting to mimic and includes this target in its warning. We evaluated *Off-the-Hook* in two different user studies. Our results show that users prefer *Off-the-Hook* warnings to Firefox warnings.

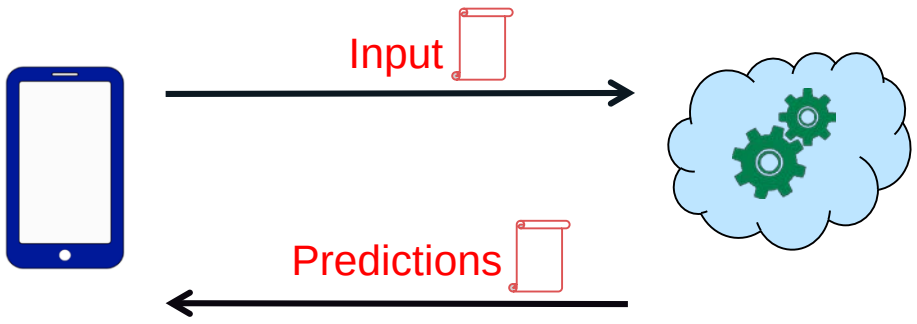
https://ssg.aalto.fi/projects/phishing/

Can we guarantee performance of machine-learning based systems even in the presence of adversaries?

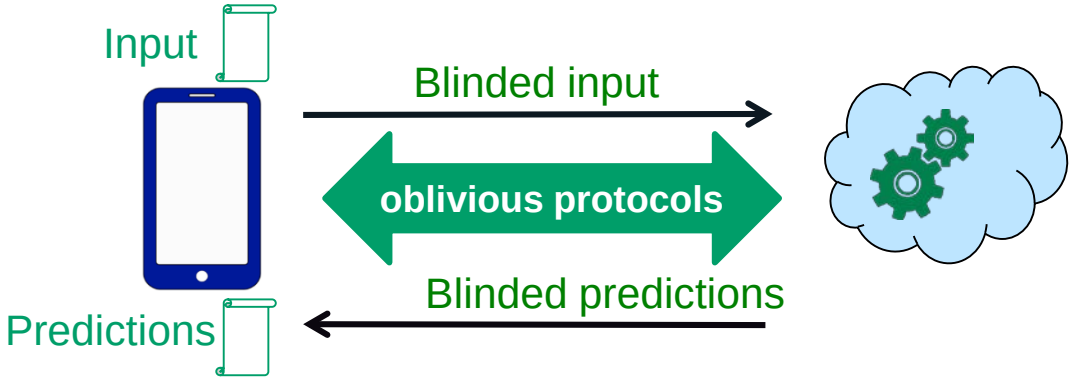
Current themes: Machine Learning & Security

Applying ML for Security & Privacy problems; Security & Privacy concerns in ML

Example: MiniONN – privacy-preserving neural network predictions



violates clients' privacy



Use inexpensive cryptographic tools

https://eprint.iacr.org/2017/452

MiniONN (ACM CCS 2017)

By Source, Fair use, https://en.wikipedia.org/w/index.php?curid=54119040

Current themes: Emerging topics

Distributed consensus and blockchains (theory, applications) [AoF project BCon, ICRI-SC]

- Can hardware security mechanisms help design scalable consensus schemes?

Securing IoT (scalability, usability) [AoF project SELIoT]

- How do we secure IoT devices from birth to death?

Security and privacy of vehicle-to-X (V2X) communication [ICRI-SC]

- How to reconcile privacy and lawful interception?

Stylometry and security [HICT scholarship]

- Can text analysis help detect deception?

ICRI-SC

http://www.icri-sc.org/

Intel Collaborative Research Institute for Secure Computing

- Only Intel Institute for security outside the US

intel Collaborative Research Institute for Secure Computing

ICRI-SC for mobile and embedded systems security

- 2012-2017 (Aalto, TU Darmstadt, UH; Aalto joined in 2014)
- Nearly 1 M€ invested in Aalto and UH

ICRI-CARS for autonomous systems security

- 2017-2020 (Aalto, TU Darmstadt, RU Bochum, U Luxembourg, TU Wien)

intel Research Institute for Collaborative Autonomous and Resilient Systems

Media coverage of our research

ua-hosting.company 818,26

Поддельная базовая станция за $1400 позволяет точно определить местонахождение телефона в 4G/LTE сети

The Register — Biting the hand that feeds IT

Intel infosec folk TEE off open source app dev framework

World+dog can TEE off too, witho

30 Jun 2015 at 11:18, Darren Pauli

A trio of Intel boffins have broken a vendor lock-dow the release of an open source framework that could

MIT Technology Review

Emerging Technology From the arXiv
December 16, 2013

First Direct Measurement of Infection Rates For Smartphone Viruses

Hacker News new | comments | show | ask | jobs | submit

1. Mozilla terminates its deal with Yahoo and makes Google the default in Firefox (techcrunch.com)
99 points by pob1234 1 hour ago | hide | 36 comments
2. LIGO and Virgo announce the detection of a black hole binary merger (caltech.edu)
44 points by srikar 3 hours ago | hide | 14 comments
3. Why Education Startups Do Not Succeed (2011) (avichal.com)
217 points by ALee 8 hours ago | hide | 91 comments
4. Concise Electronics for Geeks (2010) (coredump.cx)
344 points by nbmh 13 hours ago | hide | 37 comments
5. Towards Linux Kernel Memory Safety (arxiv.org)
87 points by lainon 6 hours ago | hide | 47 comments
6. Leela Zero, a GPL implementation of AlphaGo Zero (github.com)
21 points by pmontra 3 hours ago | hide | 1 comment

The Register — Biting the hand that feeds IT

Security

OEMs still the Achilles heel of Android security, say boffins

Mobe-makers just can't be bothered keeping Google-spawn up to

The Telegraph

WhatsApp and Facebook signals can be hacked to trac your location

Hackers can monitor 4G mobile networks to detect users' location using supposedly anonymised identifiers

THE TIMES OF INDIA — Tech

Android smartphones are at greater risk of malware attacks: Study

IANS | Apr 9, 2014, 04.23 PM IST

Aalto University School of Science

A Google Research Award to professor N. Asokan

07.10.2013

A research proposal on "Contextual Security", led by Prof. N. Asokan from the depa

Education

Training the next generation of information security researchers and professionals

Studies

- Study options
  - Bachelor´s degree programmes
  - Master´s degree programmes
  - International double degree programmes
  - Open university
  - Exchange, JOO and Non-degree studies
  - MBA studies
  - Show all
- Bachelor's Admissions
- Master's Admissions
- Doctoral Admissions
- Scholarships and Fees
- Studying at Aalto
- About Finland
- Admission results
- Statistics

Master's Programme in Computer, Communication and Information Sciences - Security and Cloud Computing

Programme description | Get to know us

- Study programme
- Admission requirements
- Career opportunities
- Application documents
- Tuition fees and scholarships
- Contact information

Acquire a world-class education in information security at Aalto University!

Studies in *Security and Cloud Computing* give students a broad understanding of the latest and future technologies for secure mobile and cloud computing systems. Students will gain both practical engineering knowledge and theoretical insights into

- secure systems engineering,
- distributed application development

Degree:
Master of Science (Technology).
More information.

ECTS:
120 ECTS

Field of Study:
Technology and Engineering

Duration:
2 years, full-time

Eligibility:
An appropriate Bachelor´s degree or an equivalent qualification.

Tuition fees & scholarships:
Yes, for non-EU citizens.
More information

Language of Instruction:
English
More information.

Organising school/s:
School of Science

Application period:
2017-12-15 - 2018-01-24

http://www.aalto.fi/en/studies/education/programme/security_and_cloud_computing/

SECCLO

Master's Programme in Security and Cloud Computing

(Erasmus Mundus)

Applications: 4.12.2017 – 17.01.2018

~20 scholarships

secclo.aalto.fi **secclo@aalto.fi** **facebook.com/secclo**

Aalto University

DTU Technical University of Denmark

EURECOM

KTH

NTNU Norwegian University of Science and Technology

Tartu Ülikool Universitas Tartuensis 1632

Co-funded by the Erasmus+ Programme of the European Union

Helsinki-Aalto Center for Information Security (HAIC)

Joint initiative: Aalto University and University of Helsinki

Mission: attract/train top students in information security

- Offers financial aid to top students in both CCIS Security and Cloud Computing & SECCLO
- Three HAIC scholars in 2017; Five (expected) in 2018

Supported by industry donations

- F-Secure, Intel, Nixu (2017)
- F-Secure, Huawei (2018)

Targeted donations possible

https://haic.aalto.fi/

InfoSec Research and Education @ Aalto

20+ MSc and BSc theses yearly

2014

- ACM CCS (1)
- PerCom (1)
- Proc. IEEE (1)
- ASIACCS (1)
- WWW (1)
- Black Hat USA (1)
- Runner-up: Best CS MSc Thesis in Finland
- Best InfoSec MSc thesis in Finland

2015

- ACM CCS (2)
- Black Hat Europe (1)
- ACM WiSec (1)
- ASIACCS (1)
- PerCom (1)
- UbiComp (1)

2016

- ACM CCS (1)
- CeBIT (1)
- NDSS (2)
- Black Hat Europe (1)
- ICDCS (1)
- Best InfoSec MSc thesis in Finland

2017

- ACM CCS (1)
- ASIACCS (1)
- DAC (1)
- IEEE IC (1)
- ICDCS (2)
- RAID (1)
- SECON (1)
- IEEE TC (1)
- Best InfoSec MSc thesis in Finland

2018

- CT-RSA (1)
- Euro S&P (1)

Aalto University

http://cs.aalto.fi/secure_systems/

Information Security Research and Education

N. Asokan
Twitter: @nasokan, WWW: https://asokan.org/asokan

Aalto University

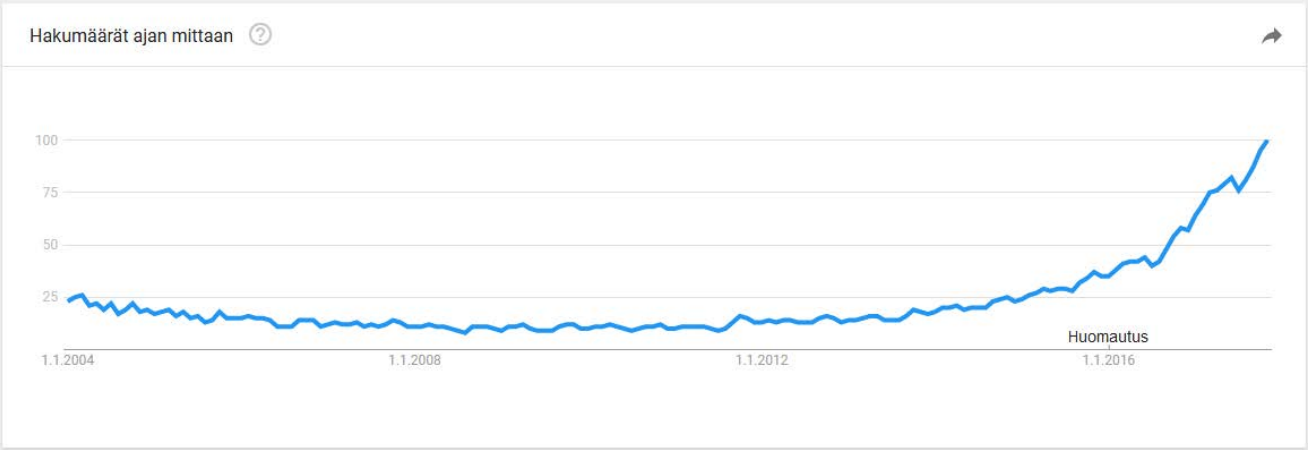
Machine Learning
in the presence of adversaries

(joint work with Mika Juuti, Jian Liu, Andrew Paverd and Samuel Marchal)

Machine Learning is ubiquitous

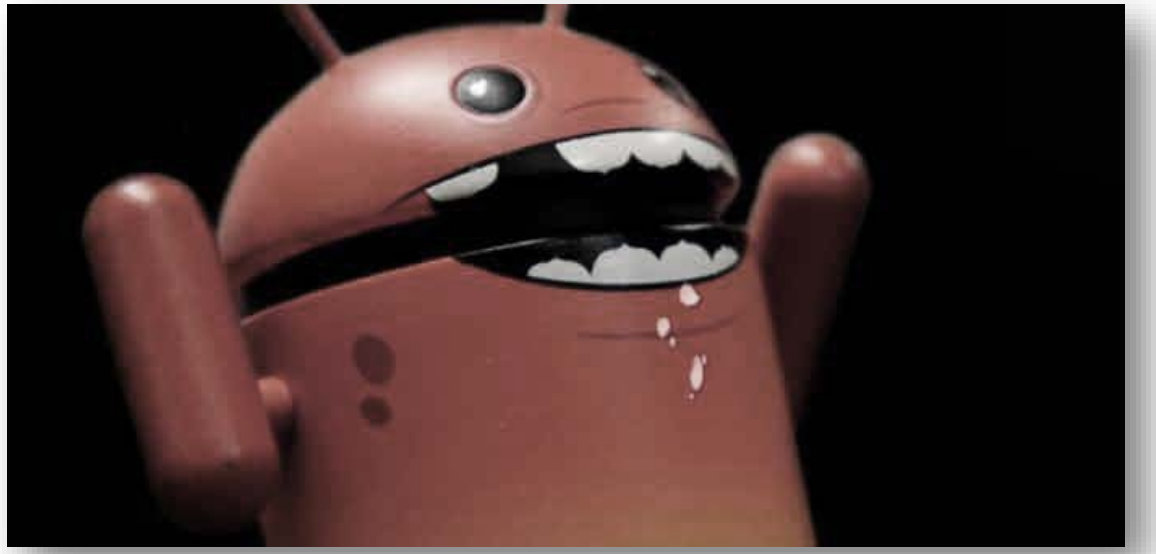
The ML market size is expected to grow by 44% annually over next five years

In 2016, companies invested up to $9 Billion in AI-based startups



[1] http://www.marketsandmarkets.com/PressReleases/machine-learning.asp
[2] McKinsey Global Institute, ”Artificial Intelligence: The Next Digital Frontier?”

Machine Learning *for* security/privacy



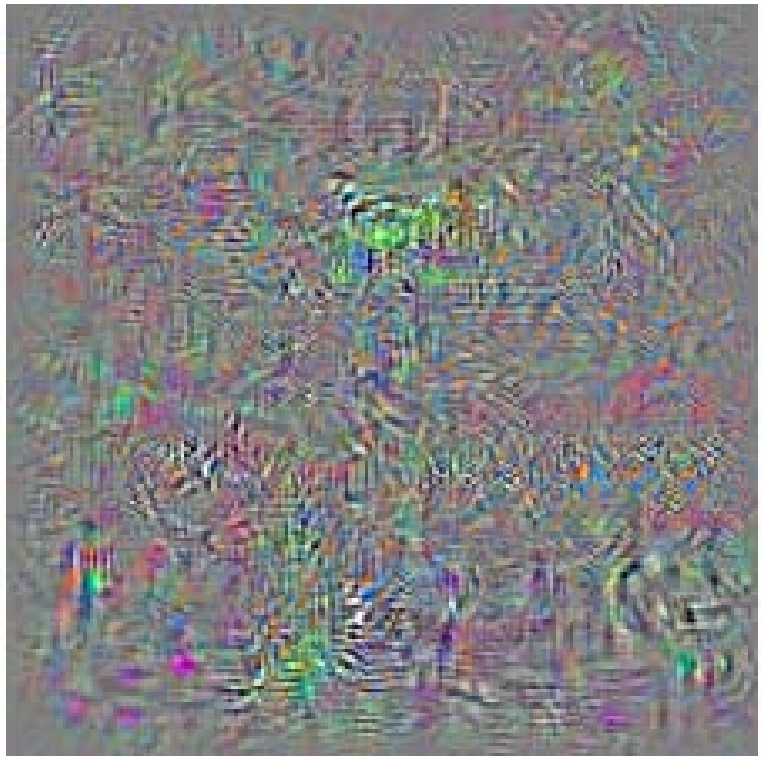
Access Control

Deception Detection

Marchal et al., “Off-the-Hook: An Efficient and Usable Client-Side Phishing Prevention Application” 2017. IEEE Trans. Comput.
https://ssg.aalto.fi/projects/phishing/

Security & privacy *of* machine learning

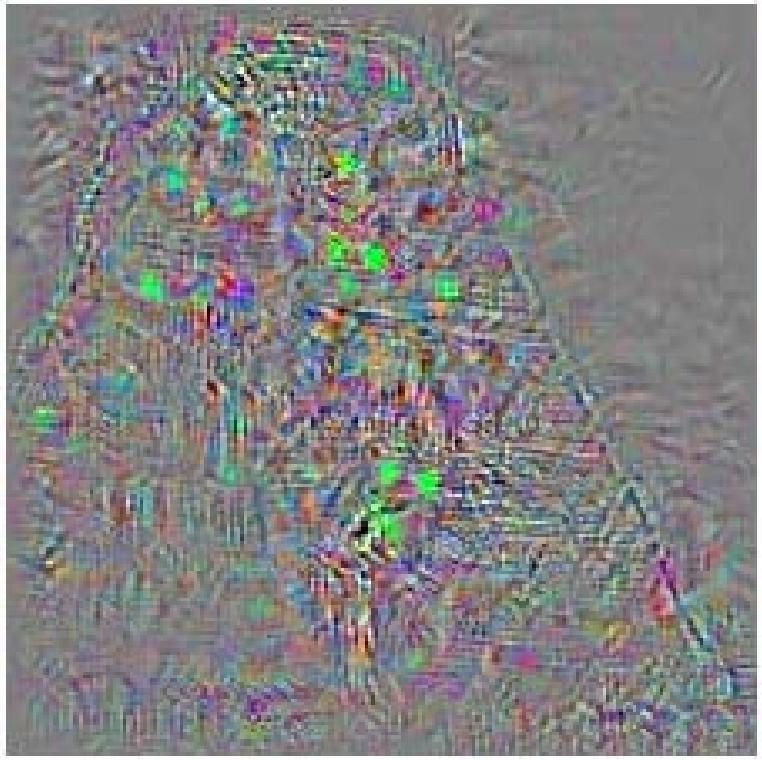
Skip to robust adversarial examples



Which class is this?
School bus

Which class is this?
Ostrich

Szegedy et al., “Intriguing Properties of Neural Networks” 2014. https://arxiv.org/abs/1312.6199v4



Which class is this?
Building

Which class is this?
Ostrich

Szegedy et al., “Intriguing Properties of Neural Networks” 2014. https://arxiv.org/abs/1312.6199v4

Which class is this?
Panda

Which class is this?
Gibbon

Goodfellow et al., “Explaining and Harnessing Adversarial Examples” ICLR 2015. https://blog.openai.com/robust-adversarial-inputs/

Which class is this?
Cat

Which class is this?
Desktop computer

Athalye et al. “Synthesizing Robust Adversarial Examples”. https://blog.openai.com/robust-adversarial-inputs/

智能系统安全实验室
UBIQUITOUS SYSTEM SECURITY LAB.

DolphinAttack: Inaudible Voice command

Guoming Zhang Chen Yan Xiaoyu Ji

Tianchen Zhang Taimin Zhang Wenyuan Xu

Zhejiang University

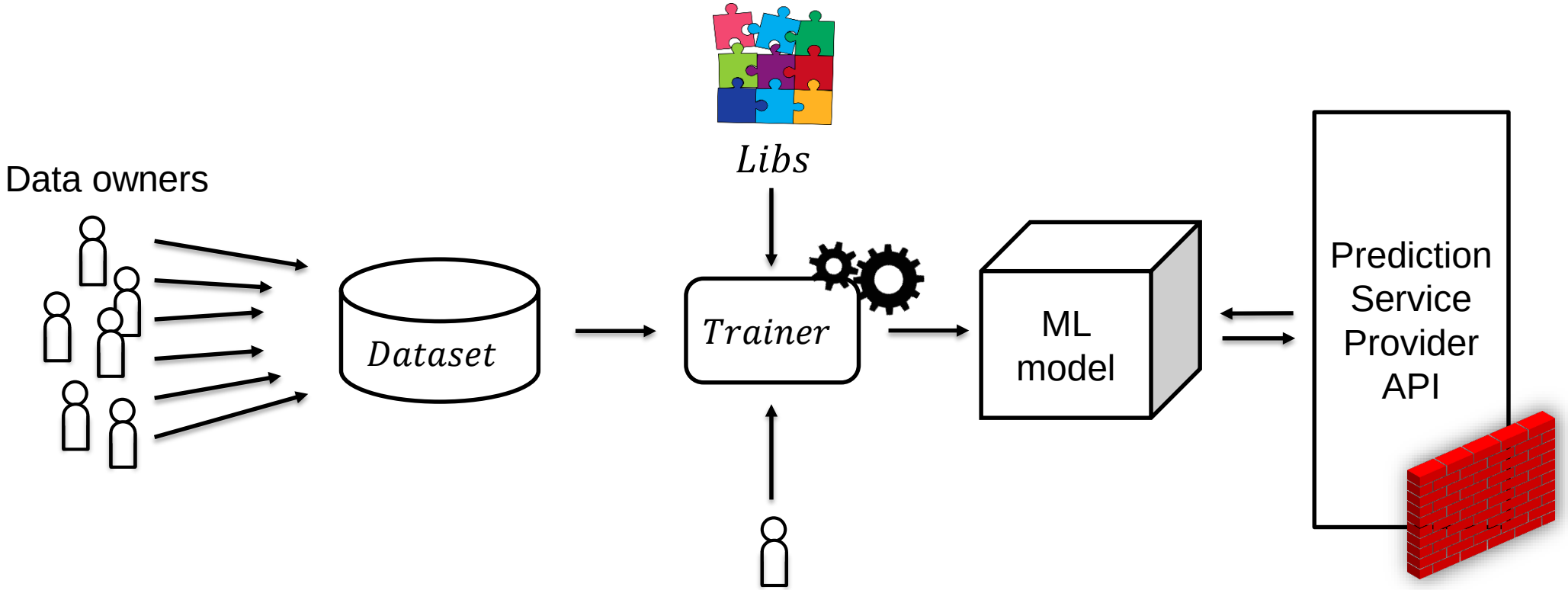
ACM CCS 2017

Zhang et al, “DolphinAttack: Inaudible Voice Commands”, ACM CCS ‘17 https://arxiv.org/abs/1708.09537

Target Softmax MLP DAE

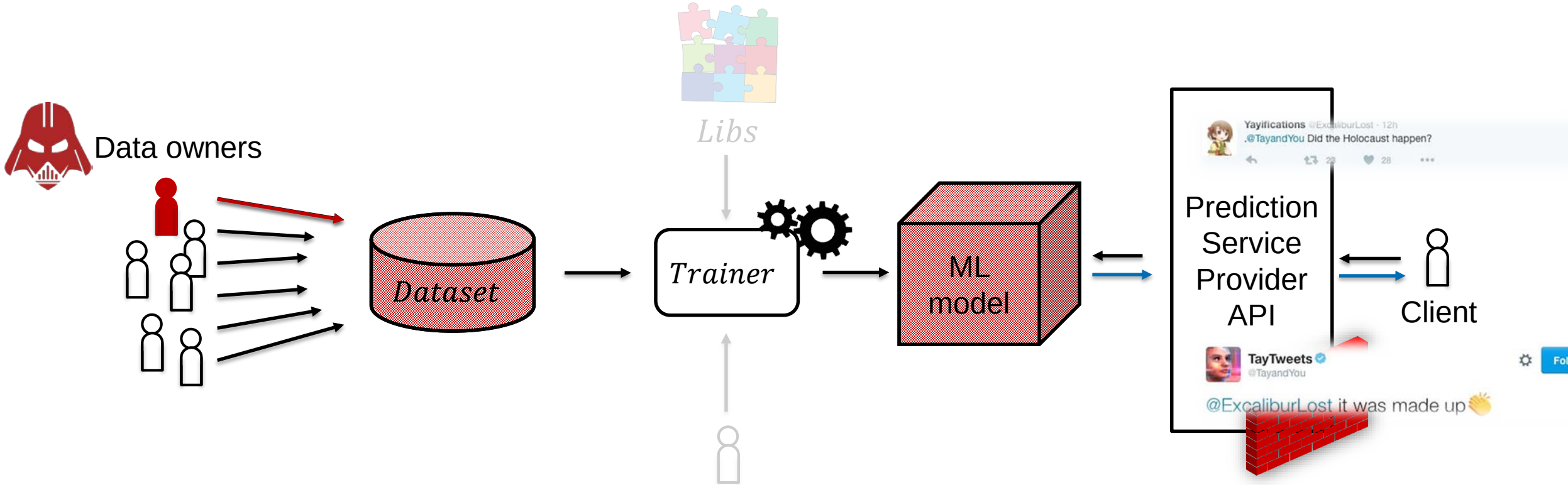
Fredrikson et al. “Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures”, ACM CCS ’15.
https://www.cs.cmu.edu/~mfredrik/papers/fjr2015ccs.pdf

A more realistic Machine Learning pipeline



Where is the adversary?

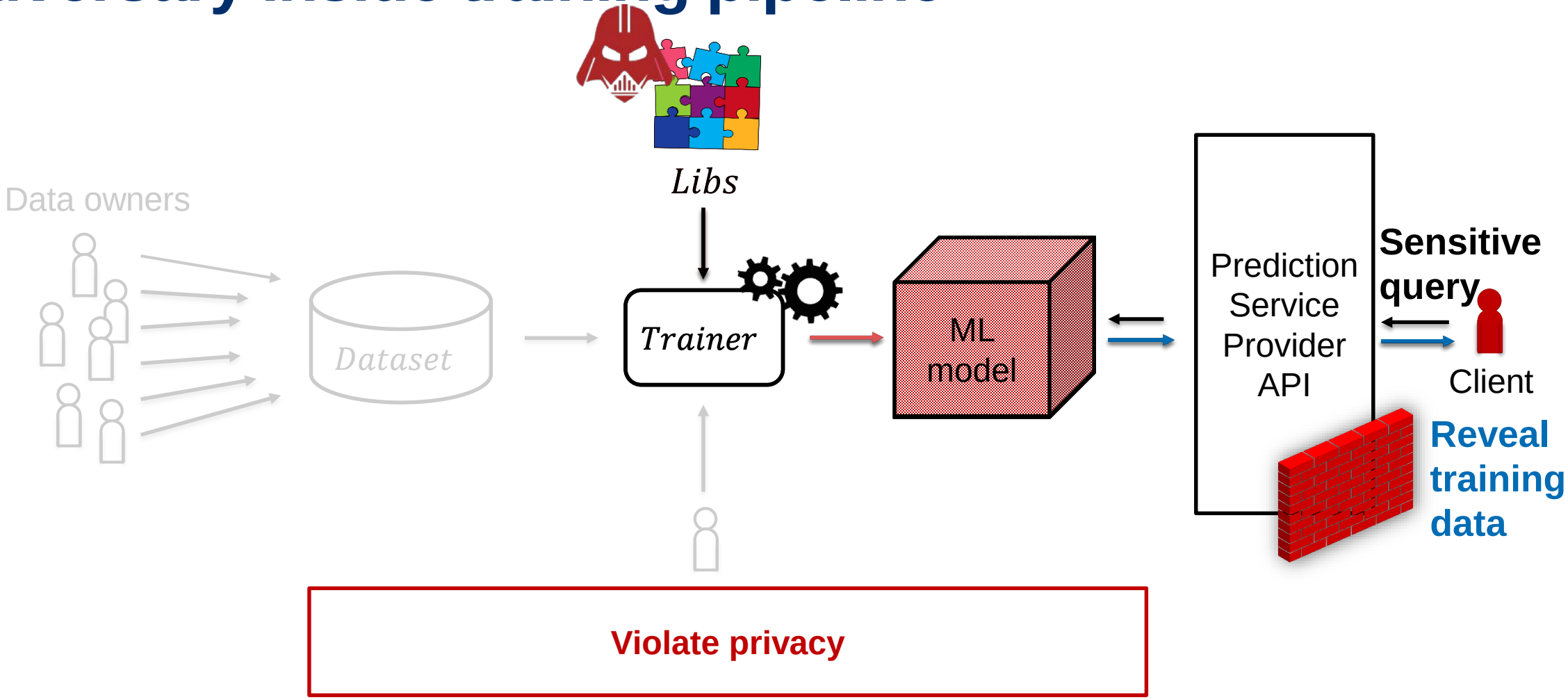
Malicious data owner



Influence ML model (model poisoning)

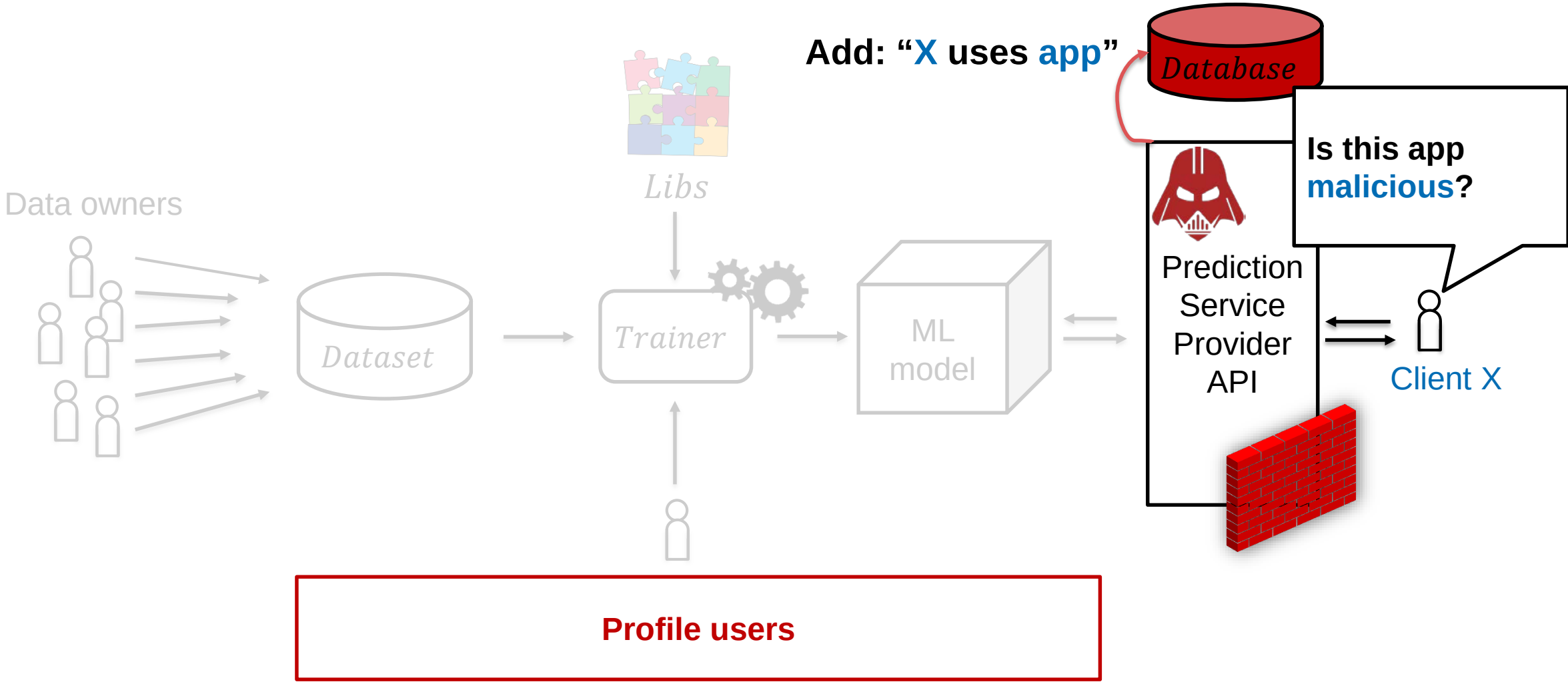
https://www.theguardian.com/technology/2016/mar/26/microsoft-deeply-sorry-for-offensive-tweets-by-ai-chatbot
https://www.theguardian.com/technology/2017/nov/07/youtube-accused-violence-against-young-children-kids-content-google-pre-school-abuse

Compromised toolchain: adversary inside training pipeline



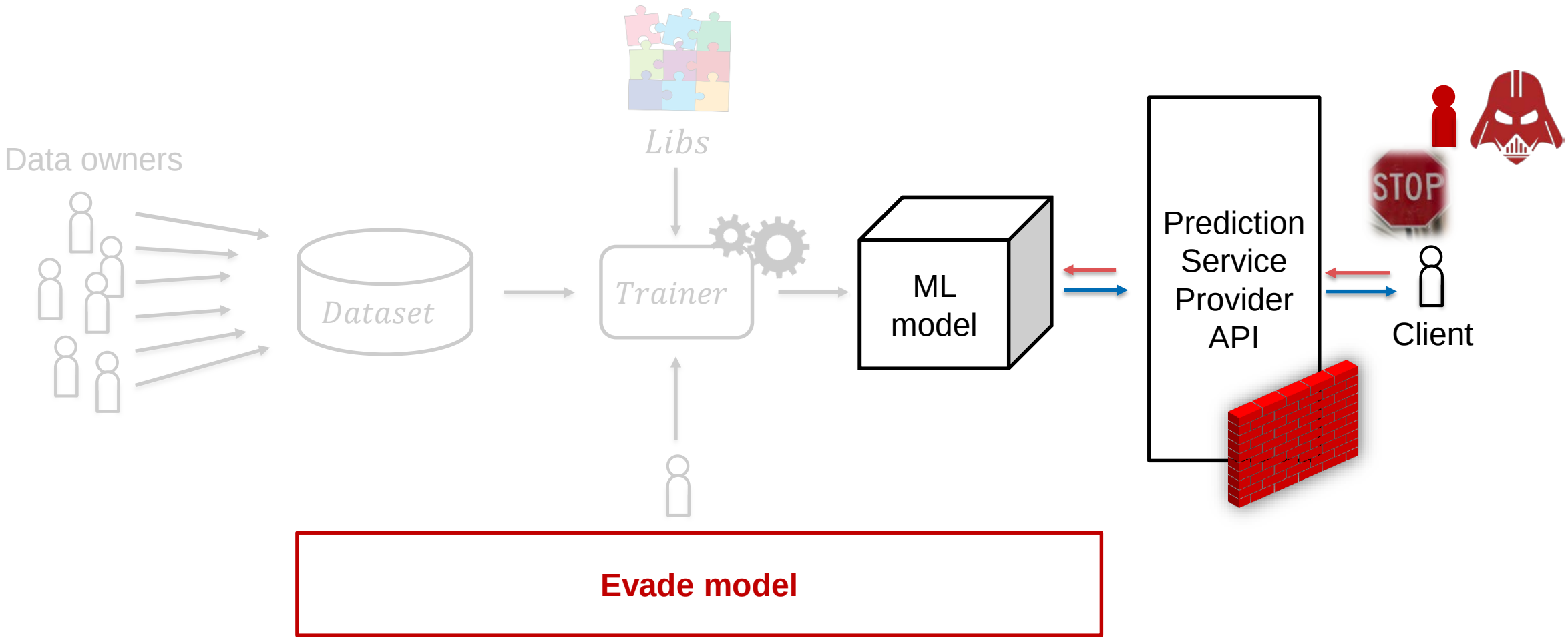
Song et al., “Machine Learning models that remember too much”, ACM CCS ’17. https://arxiv.org/abs/1709.07886
Hitja et al., “Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning”, ACM CCS ’17. http://arxiv.org/abs/1702.07464

Malicious prediction service



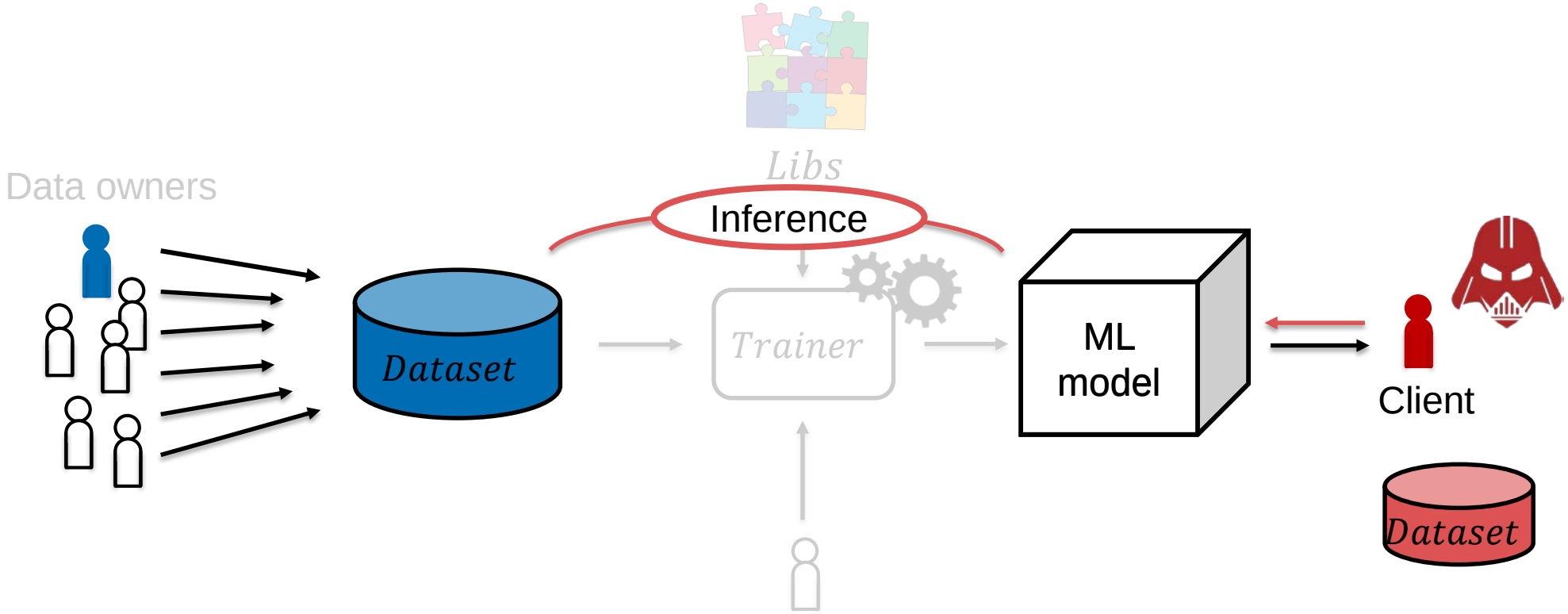
Malmi and Weber. "You are what apps you use Demographic prediction based on user's apps", ICWSM '16. https://arxiv.org/abs/1603.00059

Compromised input



Dang et al., “Evading Classifiers by Morphing in the Dark”, ACM CCS ’17. https://arxiv.org/abs/1705.07535
Evtimov et al., “Robust Physical-World Attacks on Deep Learning Models”. https://arxiv.org/abs/1707.08945
Zhang et al., “DolphinAttack: Inaudible Voice Commands”, ACM CCS ’17. https://arxiv.org/abs/1708.09537

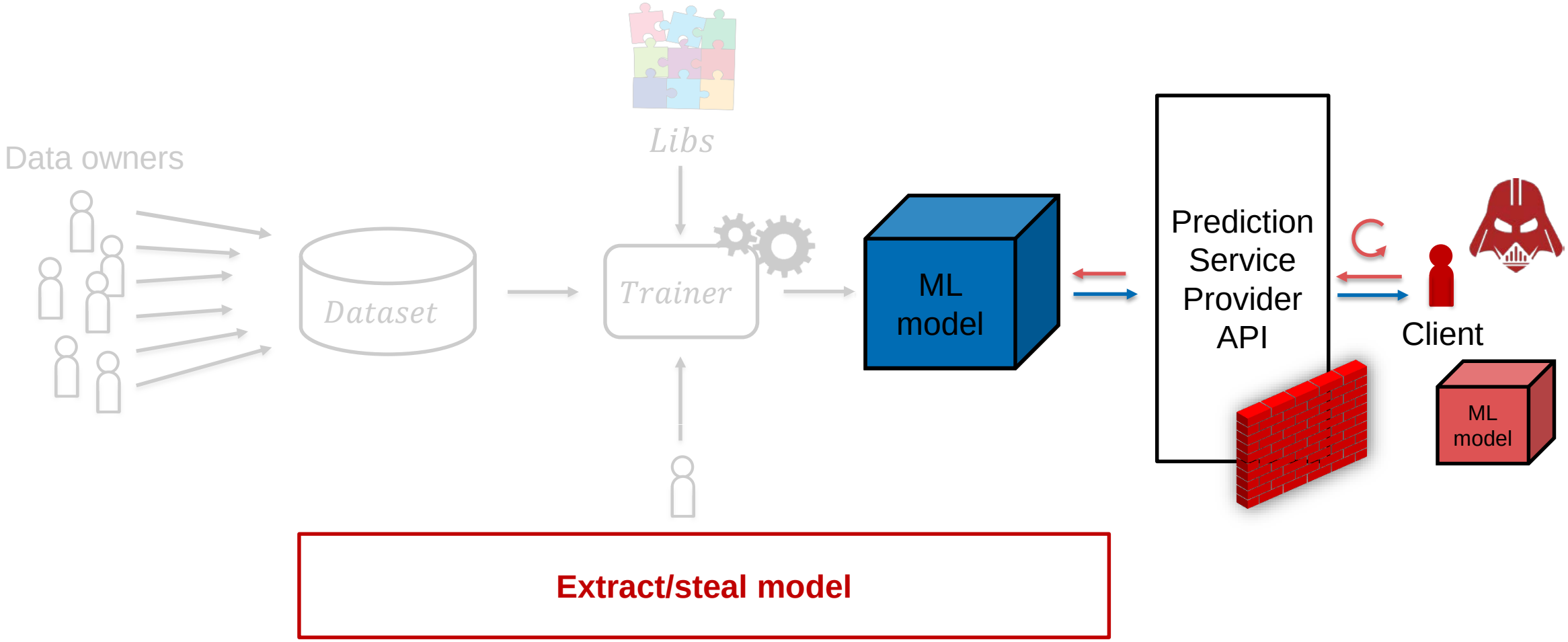
Malicious client



Invert model, infer membership

Shokri et al., “Membership Inference Attacks Against Machine Learning Models”. IEEE S&P ’16. https://arxiv.org/pdf/1610.05820.pdf
Fredrikson et al., “Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures”. ACM CCS’15. https://www.cs.cmu.edu/~mfredrik/papers/fjr2015ccs.pdf

Malicious client



Tramer et al., “Stealing ML models via prediction APIs”, Usenix SEC ’16. https://arxiv.org/abs/1609.02943

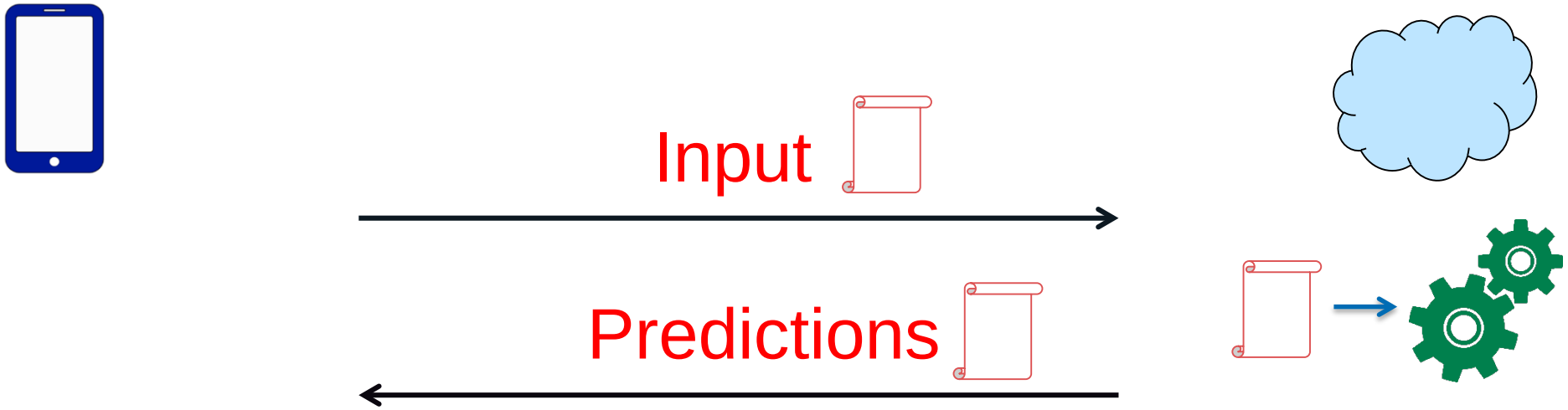
Aalto University

Oblivious Neural Network Predictions via MiniONN Transformations

N. Asokan,
https://asokan.org/asokan/
@nasokan

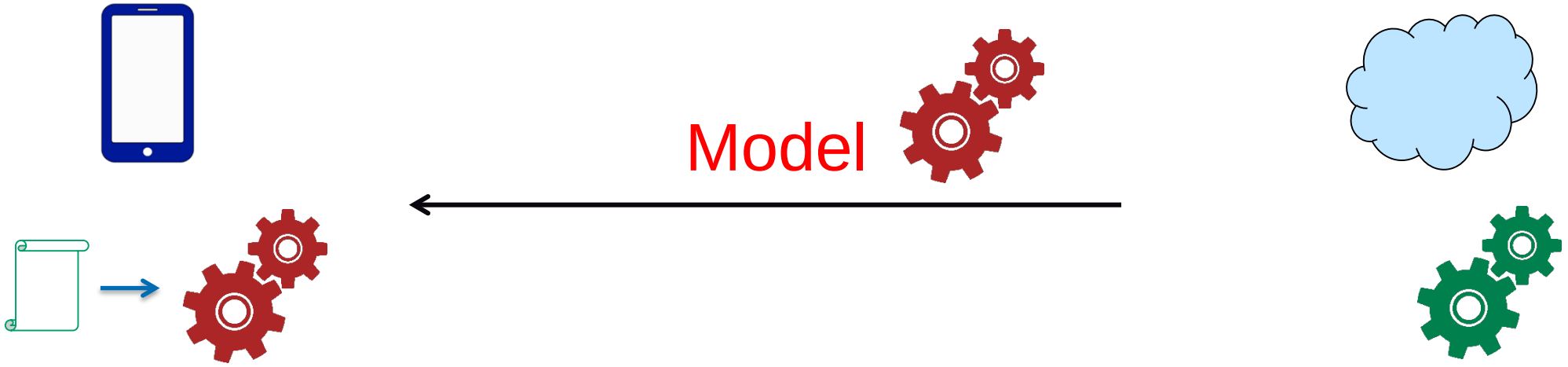
(Joint work with Jian Liu, Mika Juuti, Yao Lu)

Machine learning as a service (MLaaS)



violation of clients' privacy

Running predictions on client-side



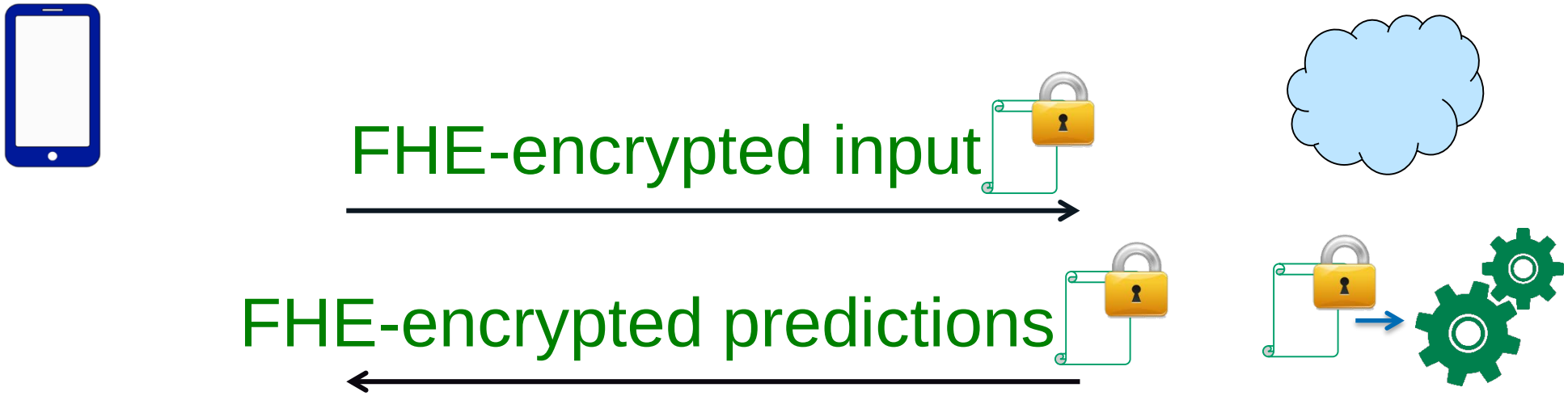
model theft
evasion
model inversion

Oblivious Neural Networks (ONN)

Given a neural network, is it possible to make it oblivious?

- server learns nothing about clients' input;
- clients learn nothing about the model.

Example: CryptoNets

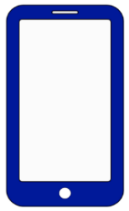


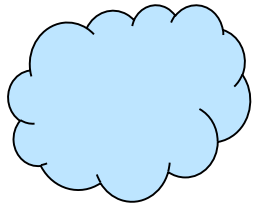
- High throughput for batch queries from same client
- High overhead for single queries: 297.5s and 372MB (MNIST dataset)
- Cannot support: high-degree polynomials, comparisons, …

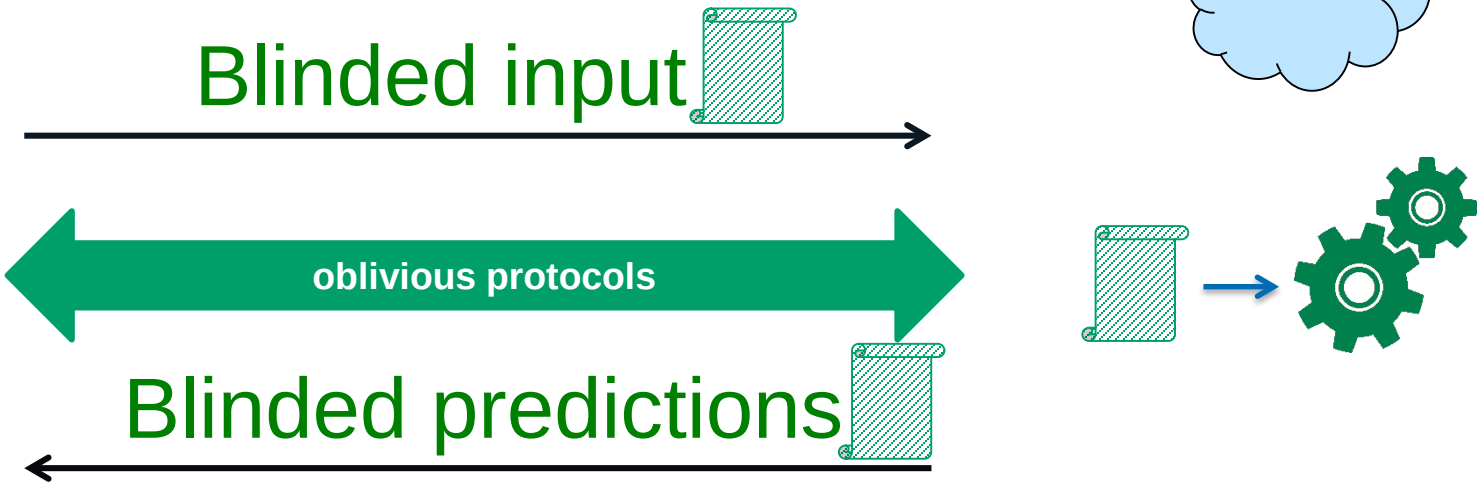
[GDLLNW16] CryptoNets, ICML 2016

MiniONN: Overview

By Source, Fair use, https://en.wikipedia.org/w/index.php?curid=54119040





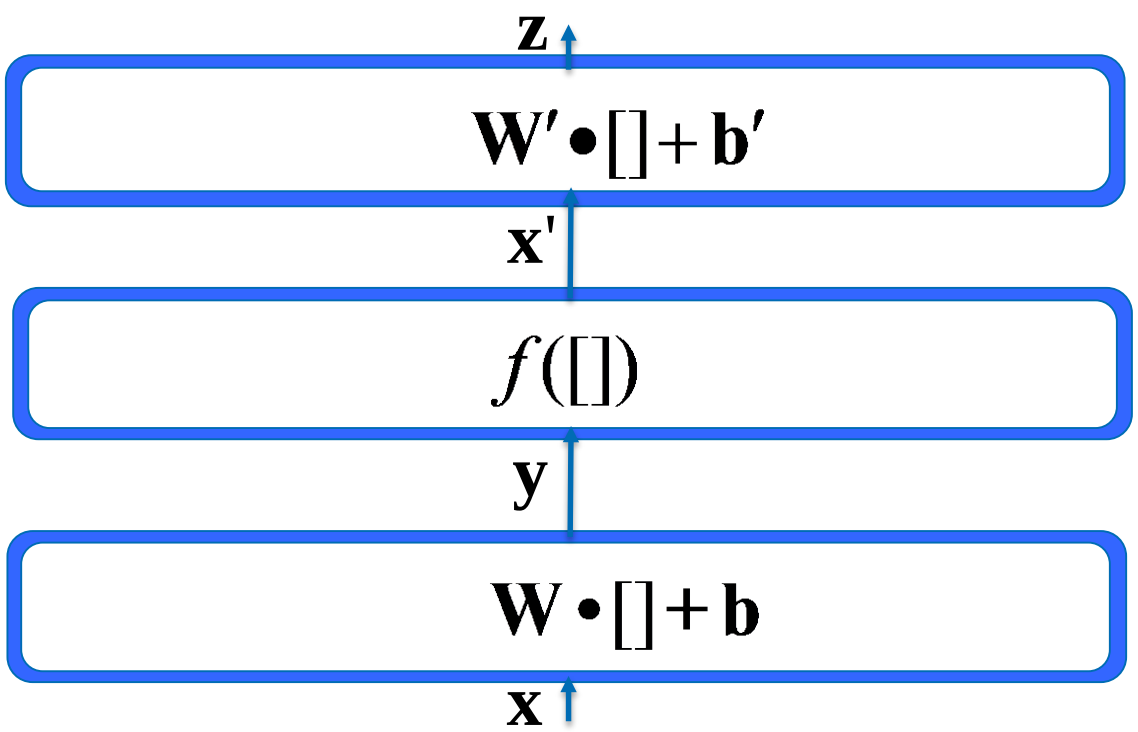


- Low overhead: ~1s
- Support all common neural networks

Skip to performance

Example $\mathbf{z} = \mathbf{W}' \bullet f(\mathbf{W} \bullet \mathbf{x} + \mathbf{b}) + \mathbf{b}'$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \mathbf{W}' = \begin{bmatrix} w'_{1,1} & w'_{1,2} \\ w'_{2,1} & w'_{2,2} \end{bmatrix}, \mathbf{b}' = \begin{bmatrix} b'_1 \\ b'_2 \end{bmatrix}$$

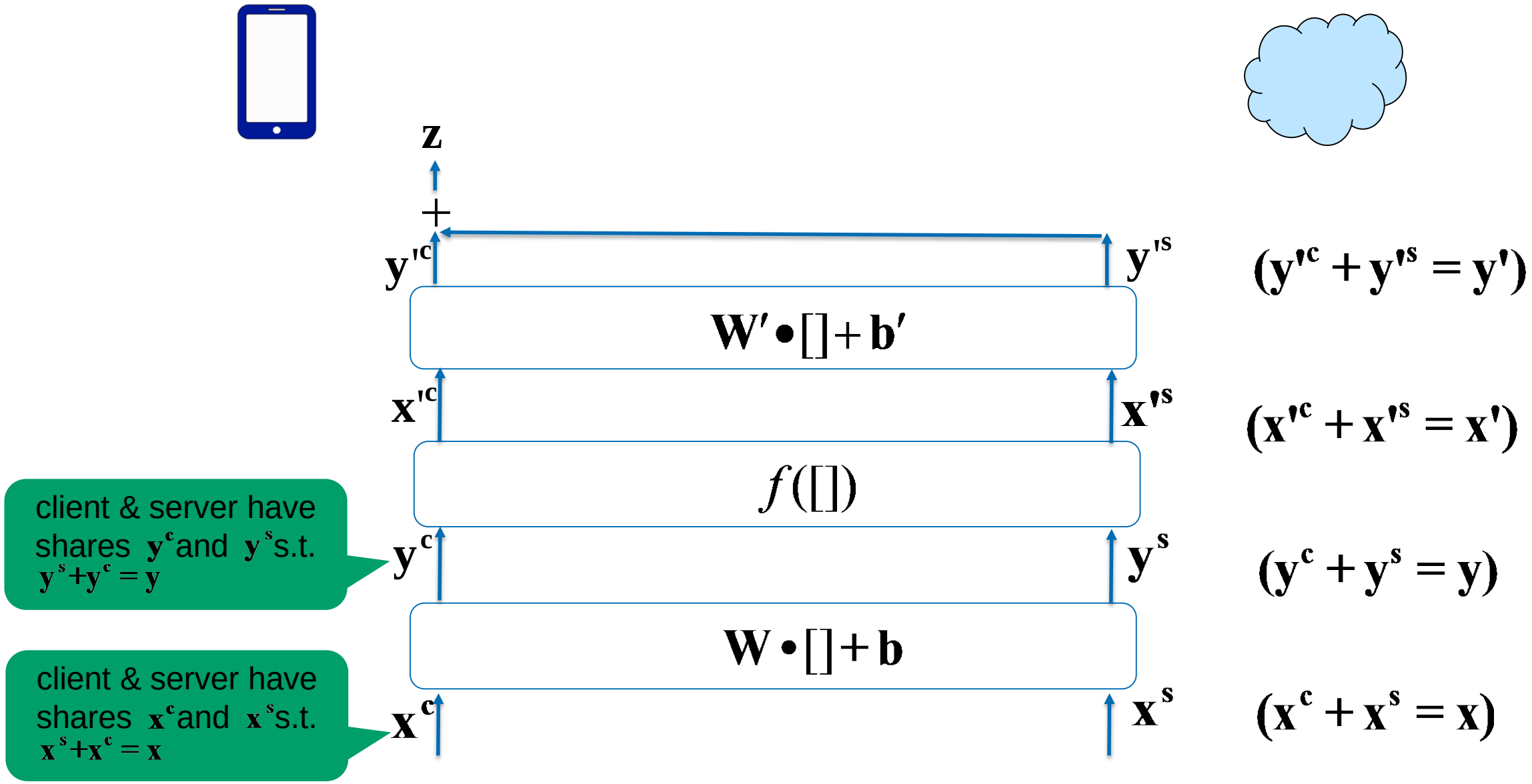


https://eprint.iacr.org/2017/452

All operations are in a finite field Z_N

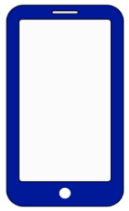
Skip to performance

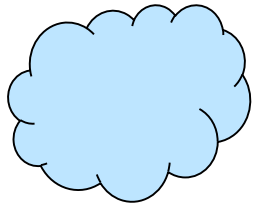
Core idea: use secret sharing for oblivious computation



Use efficient cryptographic primitives (2PC, additively homomorphic encryption)

Secret sharing initial input $\mathbf{x}$





$$x_1^c, x_2^c \xleftarrow{\$} Z_N$$

$$x_1^s := x_1 - x_1^c, \quad x_2^s := x_2 - x_2^c$$

$$\longrightarrow$$

Note that $\mathbf{x}^c$ is independent of $\mathbf{x}$. Can be **pre-chosen**

Oblivious linear transformation $\mathbf{W} \bullet \mathbf{x} + \mathbf{b}$

$$= \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix} \bullet \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix} \bullet \begin{bmatrix} x_1^s + x_1^c \\ x_2^s + x_2^c \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$= \begin{bmatrix} w_{1,1}(x_1^s + x_1^c) + w_{1,2}(x_2^s + x_2^c) + b_1 \\ w_{2,1}(x_1^s + x_1^c) + w_{2,2}(x_2^s + x_2^c) + b_2 \end{bmatrix} = \begin{bmatrix} \boxed{w_{1,1}x_1^s + w_{1,2}x_2^s + b_1} + \boxed{w_{1,1}x_1^c + w_{1,2}x_2^c} \\ \boxed{w_{2,1}x_1^s + w_{2,2}x_2^s + b_2} + \boxed{w_{2,1}x_1^c + w_{2,2}x_2^c} \end{bmatrix}$$

Compute locally by the server

Dot-product

Oblivious linear transformation: dot-product

Homomorphic Encryption with SIMD

$$E(w_{1,1}), E(w_{1,2}), E(w_{2,1}), E(w_{2,2})$$

$$r_{1,1}, r_{1,2}, r_{2,1}, r_{2,2} \xleftarrow{\$} Z_N$$

$$c_{1,1} = E(w_{1,1}x_1^c - r_{1,1})$$

$$c_{1,2} = E(w_{1,2}x_2^c - r_{1,2})$$

$$c_{2,1} = E(w_{2,1}x_1^c - r_{2,1})$$

$$c_{2,2} = E(w_{2,2}x_2^c - r_{2,2})$$

$$c_{1,1}, c_{1,2}, c_{2,1}, c_{2,2}$$

$$D(c_{1,1}), D(c_{1,2}), D(c_{2,1}), D(c_{2,2})$$

$$v_1 = r_{1,1} + r_{1,2} \qquad u_1 = w_{1,1}x_1^c + w_{1,2}x_2^c - (r_{1,2} + r_{1,1})$$

$$v_2 = r_{2,1} + r_{2,2} \qquad u_2 = w_{2,1}x_1^c + w_{2,2}x_2^c - (r_{2,1} + r_{2,2})$$

u + v = W•x^c; Note: **u, v,** and **W•x^c** are independent of **x**.
<**u,v,x^c** > generated/stored in a **precomputation phase**

Oblivious linear transformation $\mathbf{W} \bullet \mathbf{x} + \mathbf{b}$

$$= \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix} \bullet \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix} \bullet \begin{bmatrix} x_1^s + x_1^c \\ x_2^s + x_2^c \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$= \begin{bmatrix} w_{1,1}(x_1^s + x_1^c) + w_{1,2}(x_2^s + x_2^c) + b_1 \\ w_{2,1}(x_1^s + x_1^c) + w_{2,2}(x_2^s + x_2^c) + b_2 \end{bmatrix} = \begin{bmatrix} \boxed{w_{1,1}x_1^s + w_{1,2}x_2^s + b_1} + \boxed{w_{1,1}x_1^c + w_{1,2}x_2^c} \\ \boxed{w_{2,1}x_1^s + w_{2,2}x_2^s + b_2} + \boxed{w_{2,1}x_1^c + w_{2,2}x_2^c} \end{bmatrix}$$

$$= \begin{bmatrix} \boxed{w_{1,1}x_1^s + w_{1,2}x_2^s + b_1} + \boxed{u_1} \\ \boxed{w_{2,1}x_1^s + w_{2,2}x_2^s + b_2} + \boxed{u_2} \end{bmatrix} + \begin{bmatrix} \boxed{v_1} \\ \boxed{v_2} \end{bmatrix}$$

Oblivious linear transformation $\mathbf{W} \bullet \mathbf{x} + \mathbf{b}$

$$= \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix} \bullet \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix} \bullet \begin{bmatrix} x_1^s + x_1^c \\ x_2^s + x_2^c \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$= \begin{bmatrix} w_{1,1}(x_1^s + x_1^c) + w_{1,2}(x_2^s + x_2^c) + b_1 \\ w_{2,1}(x_1^s + x_1^c) + w_{2,2}(x_2^s + x_2^c) + b_2 \end{bmatrix} = \begin{bmatrix} w_{1,1}x_1^s + w_{1,2}x_2^s + b_1 + w_{1,1}x_1^c + w_{1,2}x_2^c \\ w_{2,1}x_1^s + w_{2,2}x_2^s + b_2 + w_{2,1}x_1^c + w_{2,2}x_2^c \end{bmatrix}$$

$$= \begin{bmatrix} \boxed{w_{1,1}x_1^s + w_{1,2}x_2^s + b_1 + u_1} \\ \boxed{w_{2,1}x_1^s + w_{2,2}x_2^s + b_2 + u_2} \end{bmatrix} + \begin{bmatrix} \boxed{v_1} \\ \boxed{v_2} \end{bmatrix} := \begin{bmatrix} \boxed{y_1^s} \\ \boxed{y_2^s} \end{bmatrix} + \begin{bmatrix} \boxed{y_1^c} \\ \boxed{y_2^c} \end{bmatrix}$$

Oblivious activation/pooling functions $f(y)$

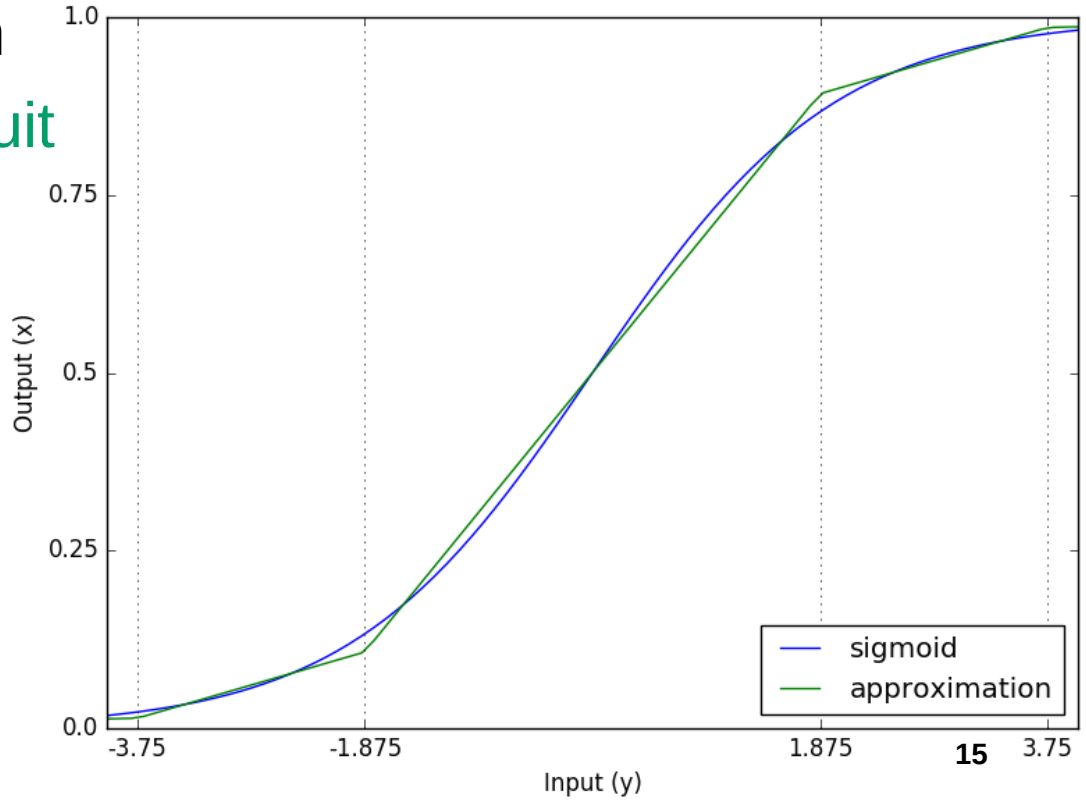
Piecewise linear functions e.g.,

- ReLU: $x := \max(y, 0)$
- Oblivious ReLU: $x^s + x^c := \max(y^s + y^c, 0)$
  - easily computed obliviously by a garbled circuit

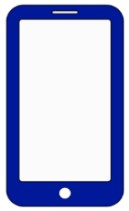
Oblivious activation/pooling functions $f(y)$

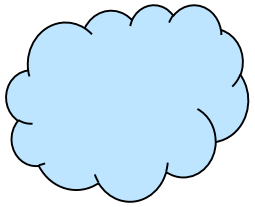
Smooth functions e.g.,

- Sigmoid: $x := 1/(1+e^{-y})$
- Oblivious sigmoid: $x^s + x^c := 1/(1+e^{-(y^s+y^c)})$
  - approximate by a piecewise linear function
  - then compute obliviously by a garbled circuit
  - empirically: ~14 segments sufficient



Combining the final result





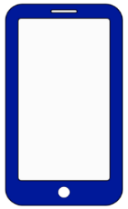
$$y_1^s, y_2^s$$

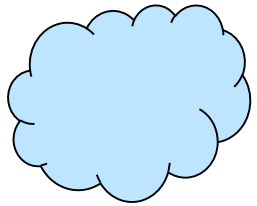
$$y_1 := y_1^s + y_1^c$$

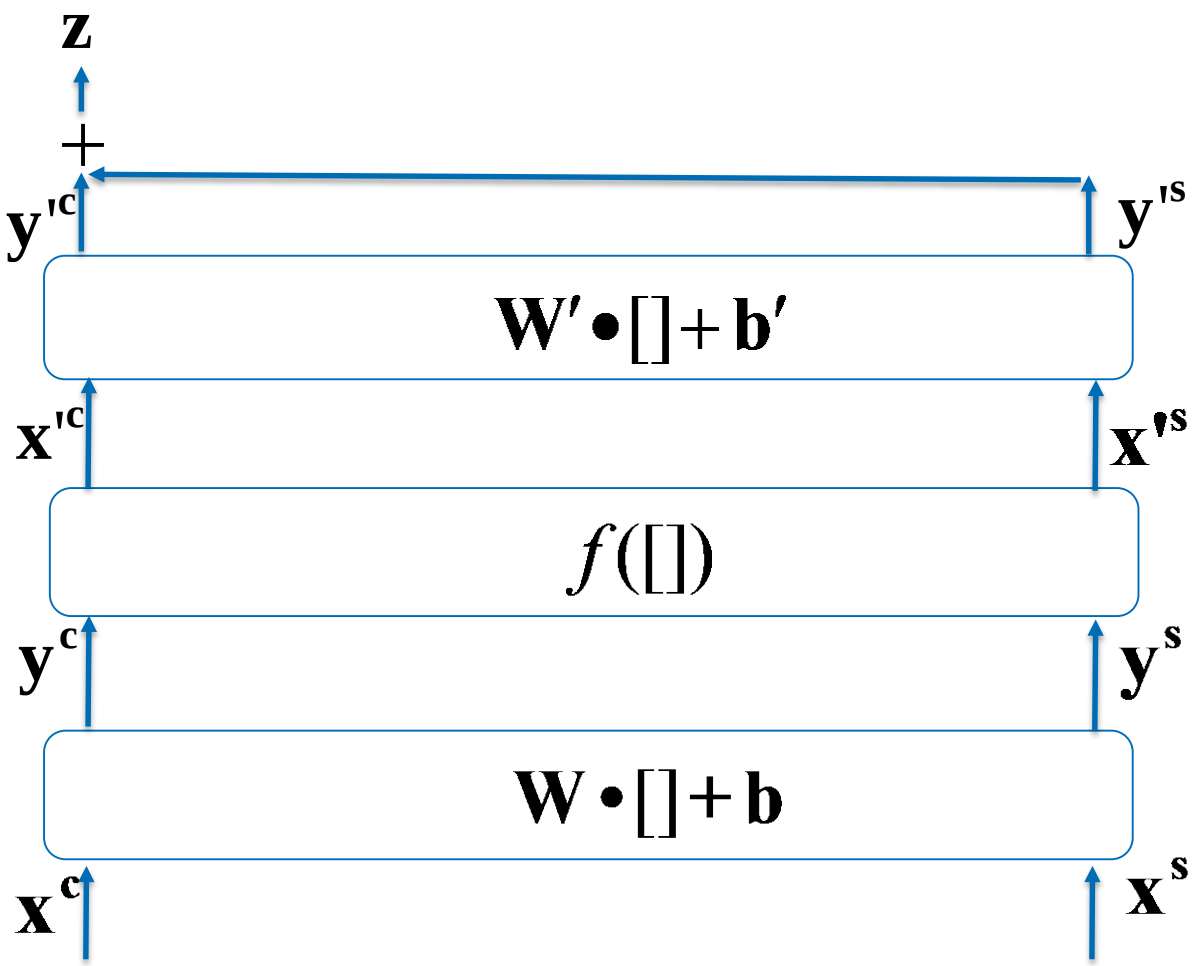
$$y_2 := y_2^s + y_2^c$$

They can jointly calculate max(y_1,y_2)
(for minimizing information leakage)

Core idea: use secret sharing for oblivious computation







$(y'^c + y'^s = y')$

$(x'^c + x'^s = x')$

$(y^c + y^s = y)$

$(x^c + x^s = x)$

Performance (for single queries)

Model	Latency (s)	Msg sizes (MB)	Loss of accuracy
MNIST/Square	0.4 (+ 0.88)	44 (+ 3.6)	none
CIFAR-10/ReLU	472 (+ 72)	6226 (+ 3046)	none
PTB/Sigmoid	4.39 (+ 13.9)	474 (+ 86.7)	Less than 0.5% (cross-entropy loss)

Pre-computation phase timings in parentheses

PTB = Penn Treebank

MiniONN pros and cons

300-700x faster than CryptoNets

Can transform any given neural network to its oblivious variant

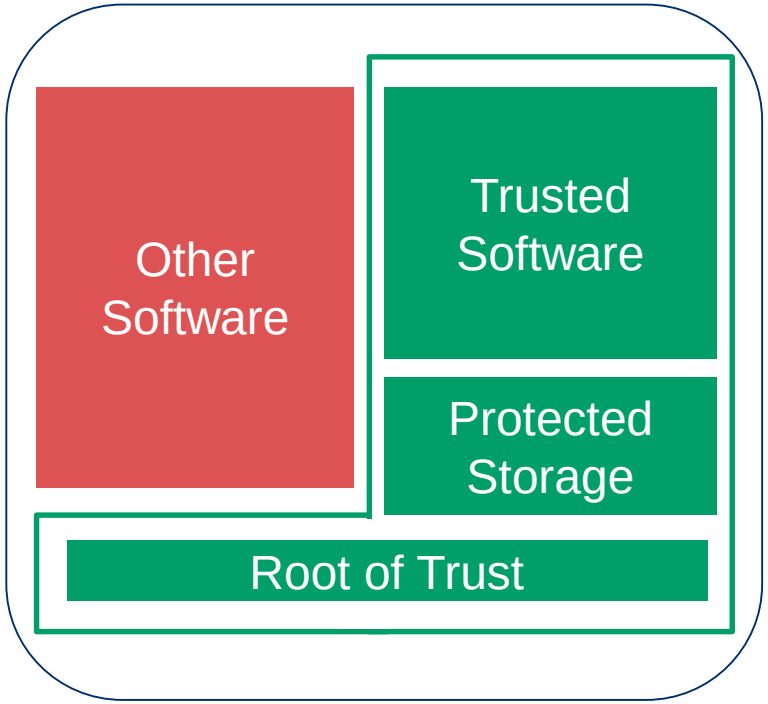
Still ~1000x slower than without privacy

Server can no longer filter requests or do sophisticated metering

Assumes online connectivity to server

Reveals structure (but not params) of NN

Can trusted computing help?



Hardware support for

- **Isolated execution: Trusted Execution Environment**
- **Protected storage: Sealing**
- **Ability to report status to a remote verifier: Attestation**

Cryptocards

https://www.ibm.com/security/cryptocards/

Trusted Platform Modules

https://www.infineon.com/tpm

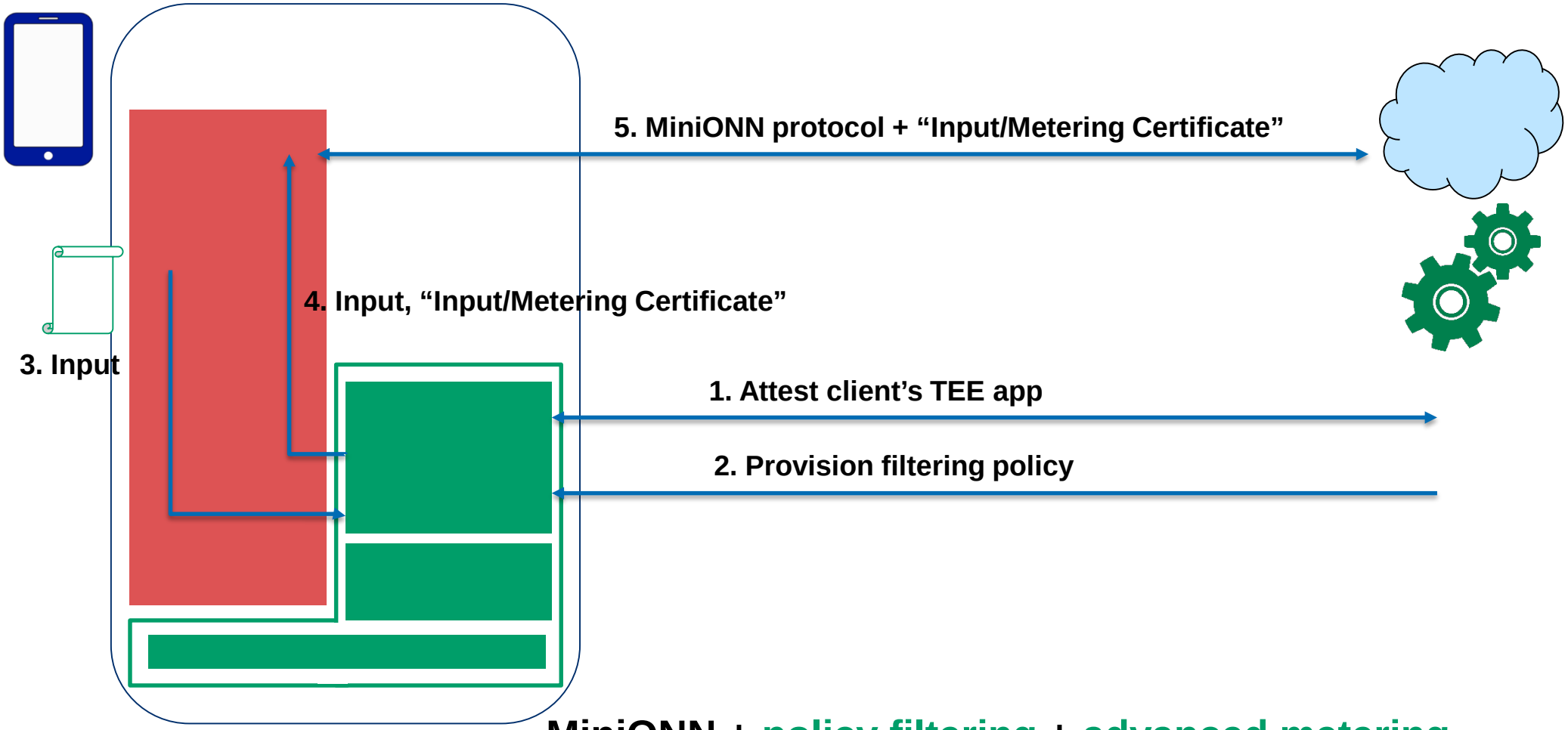
ARM TrustZone

https://www.arm.com/products/security-on-arm/trustzone

Intel Software Guard Extensions

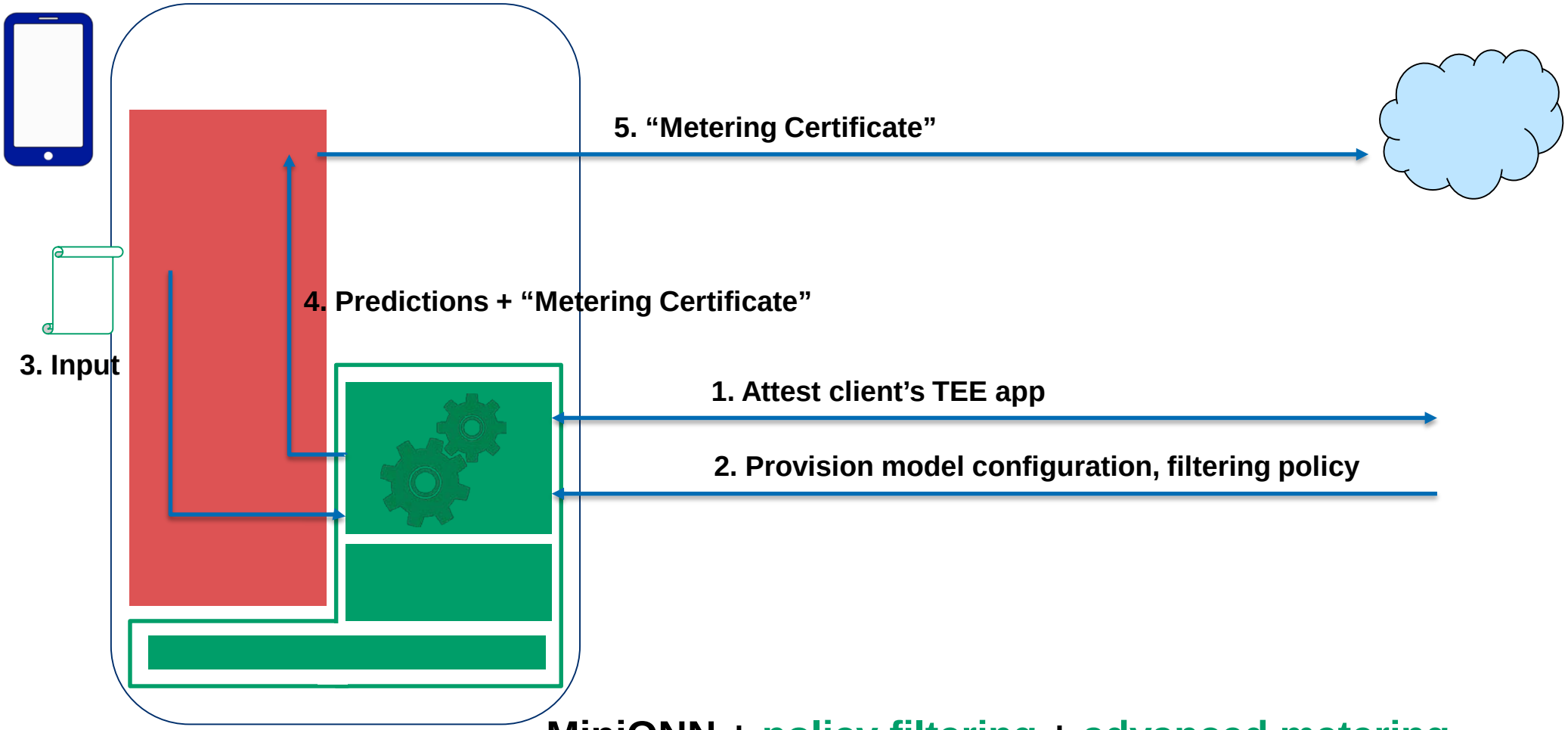
https://software.intel.com/en-us/sgx

Using a client-side TEE to vet input



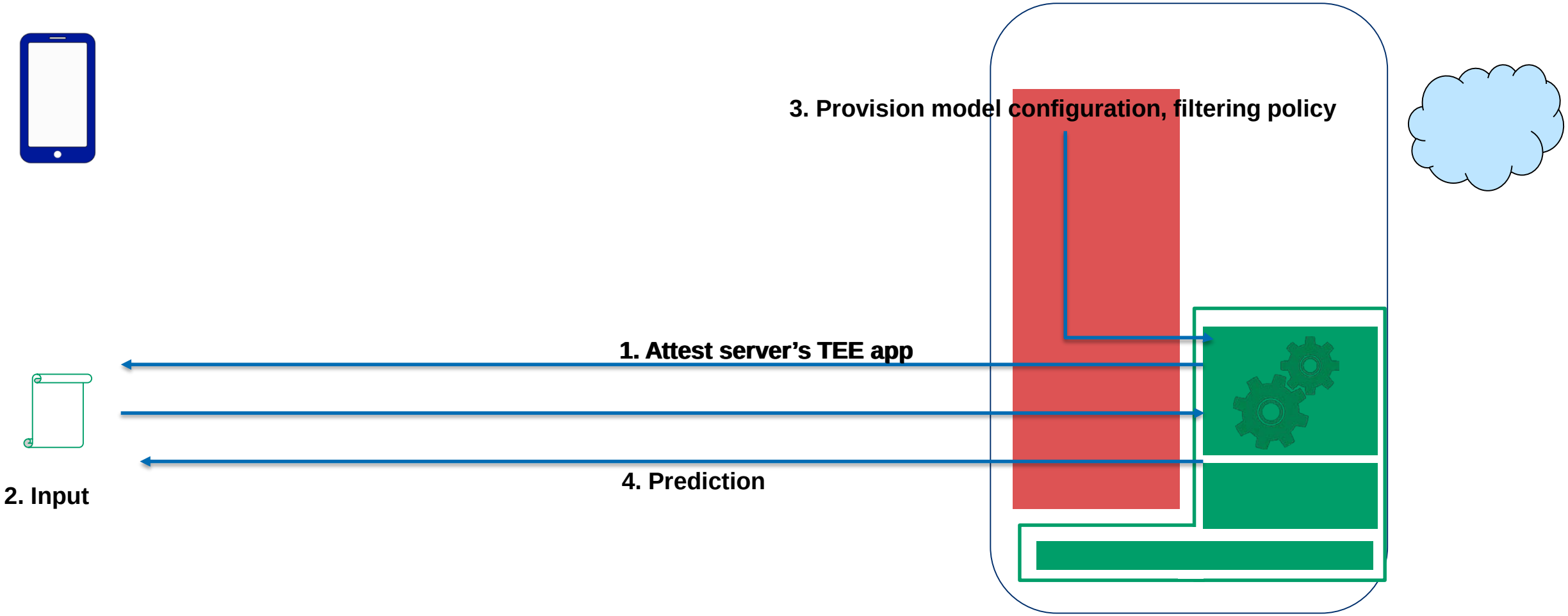
MiniONN + policy filtering + advanced metering

Using a client-side TEE to run the model



MiniONN + policy filtering + advanced metering
+ disconnected operation + performance + better privacy
- harder to reason about model secrecy

Using a server-side TEE to run the model



MiniONN + policy filtering + advanced metering
- disconnected operation + performance + better privacy

MiniONN: Efficiently transform any given neural network into oblivious form with no/negligible accuracy loss

Trusted Computing can help realize improved security and privacy for ML

ML is very fragile in adversarial settings

https://eprint.iacr.org/2017/452
ACM CCS 2017

Research collaboration with top academic groups

Funding is good, but active research collaboration is more valuable

- **real problem insights, access to data & technology, prospects for tech transfer**

Subcontracted work will not fly

- **aim for publishable research, partnership (not management)**

"Open IP" is mutually beneficial

- **Case example: Intel Collaborative Research Institute (http://www.icri-sc.org/)**