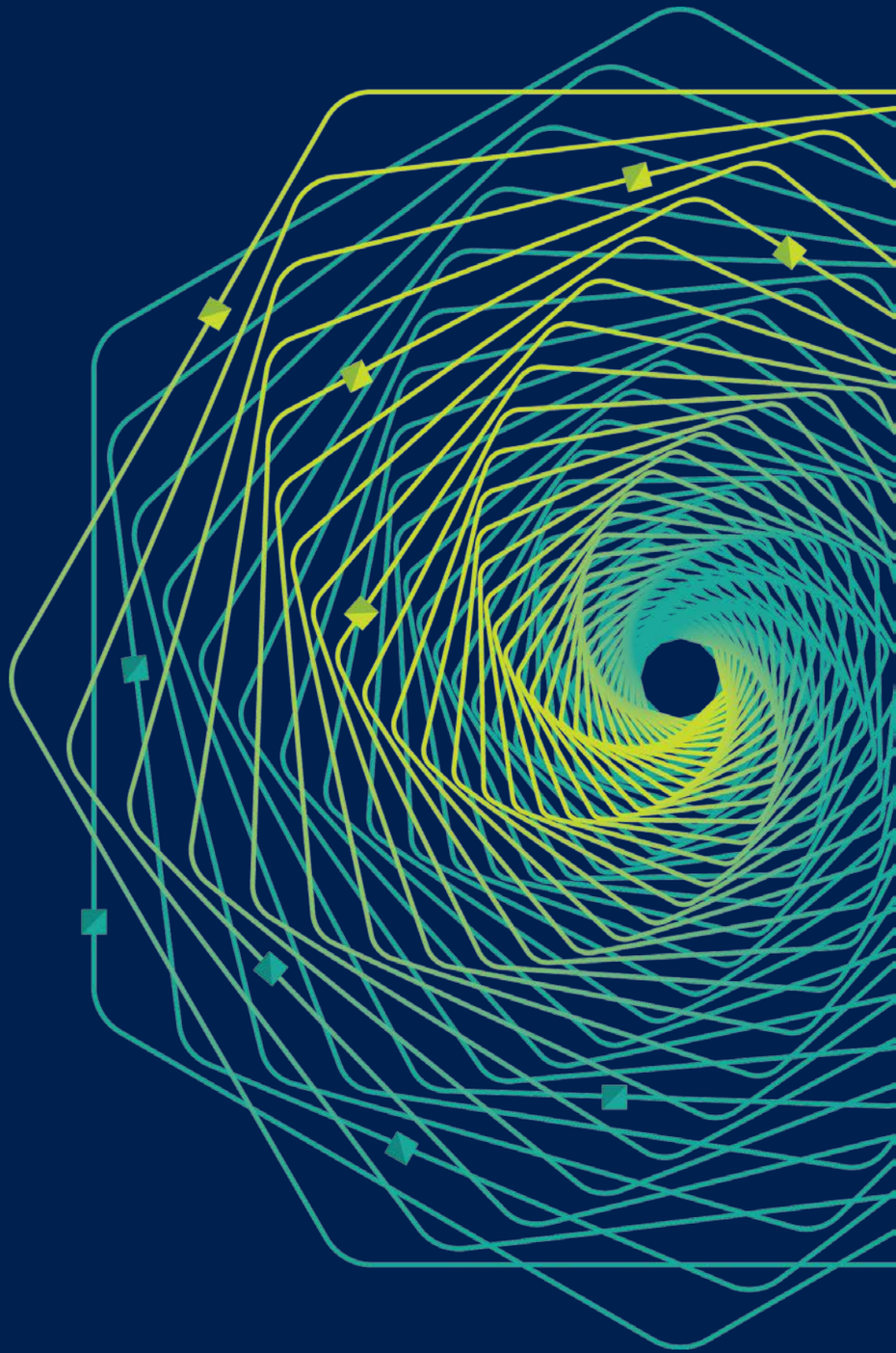
Microsoft

Research Faculty Summit 2018

Systems | Fueling future disruptions



Oasis: Privacy-Preserving Smart Contracts at Scale

Dawn Song
Professor, UC Berkeley
Founder and CEO, Oasis Labs

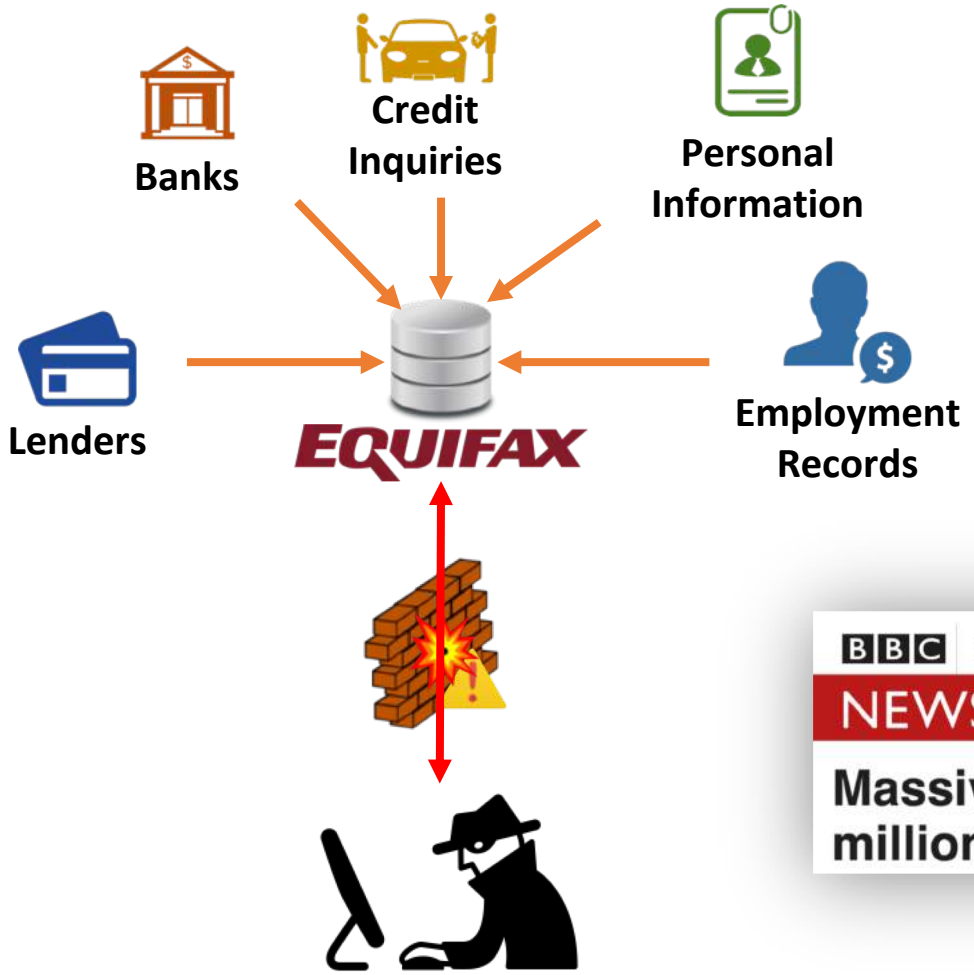
The Value of Data Analytics and Machine Learning

Data analysis and machine learning has many applications, huge potential impact

“Data is the New Oil”

What are biggest problems affecting data today?

Data breaches are becoming more common

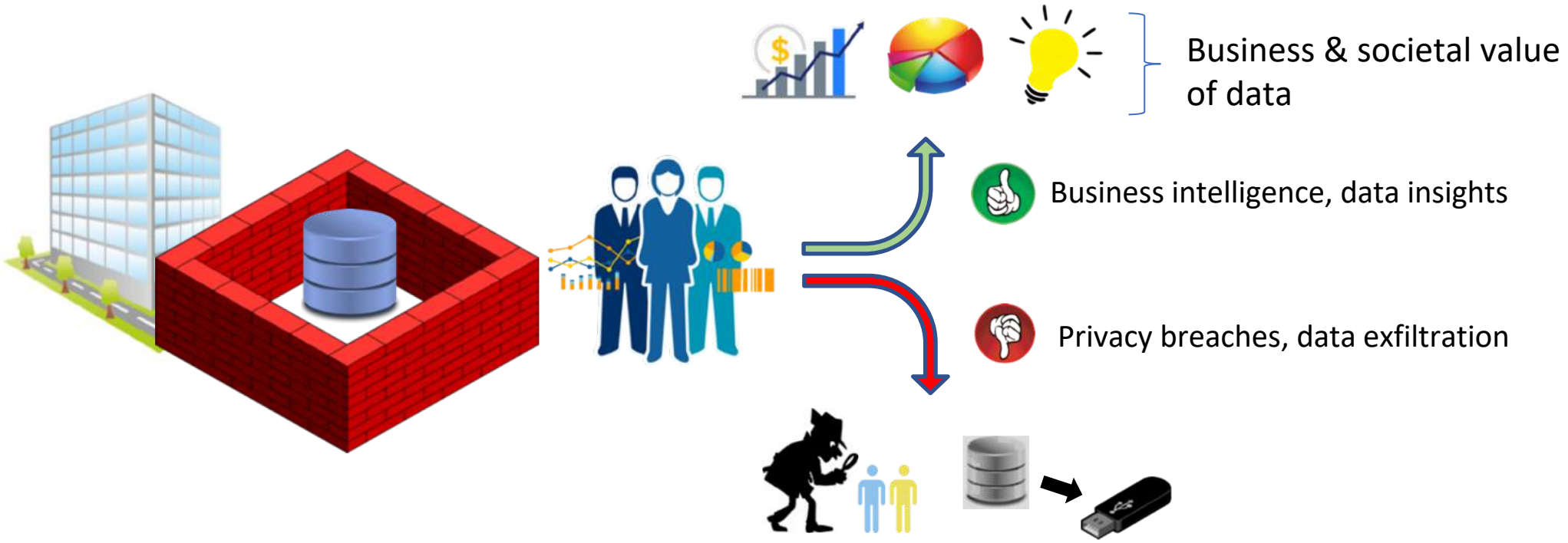


BBC Sign in News Sport Weather Shop Earth Travel More

NEWS

Massive Equifax data breach hits 143 million

Most Data Is Siloed



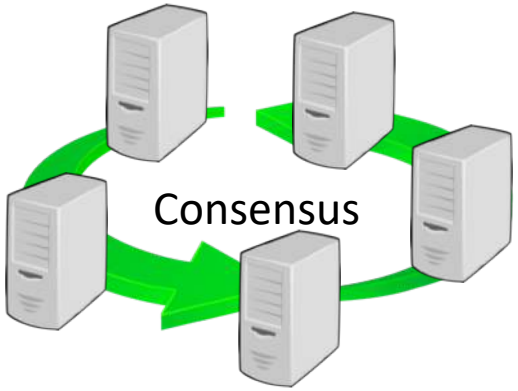
Users Are Losing Control of Their Data

Technology Intelligence

Millions of private Gmail messages read by third parties

Blockchain: a Transformative Technology

Openness & transparency



No reliance on a central party

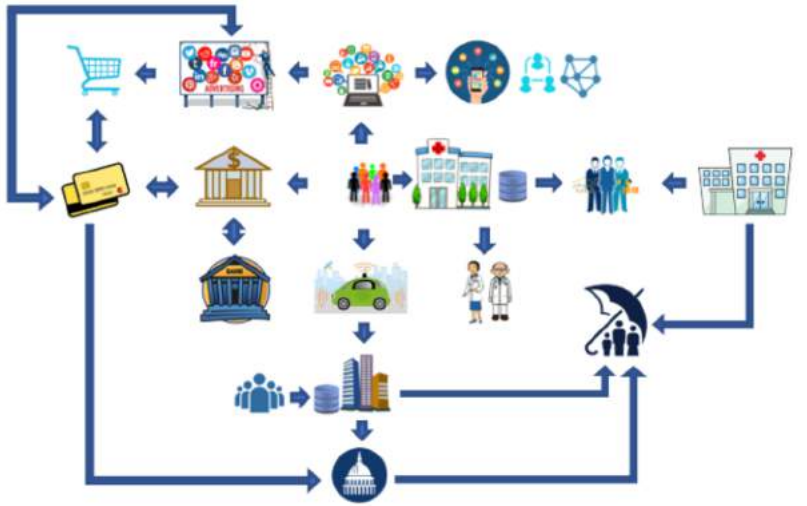
Automatic enforcement of agreements

The future of blockchain

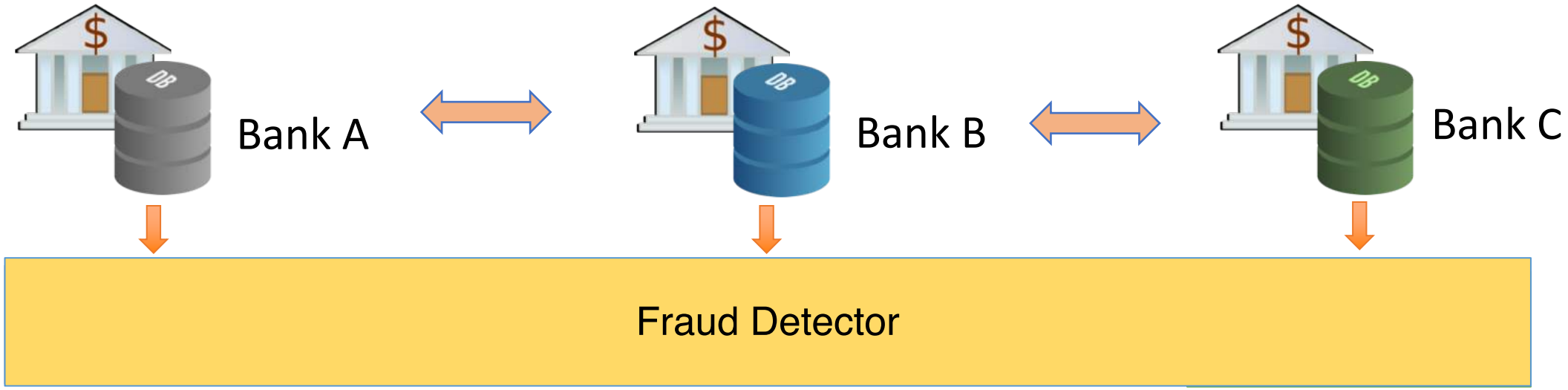
- Fraud detection
- Credit scoring
- Decentralized exchange
- Decentralized hedge fund
- Medical diagnostics
- Personalized medicine
- Private auctions
- Internet of Things applications

Payments Tokens Cryptokitties

ethereum



Motivating example: Fraud detection

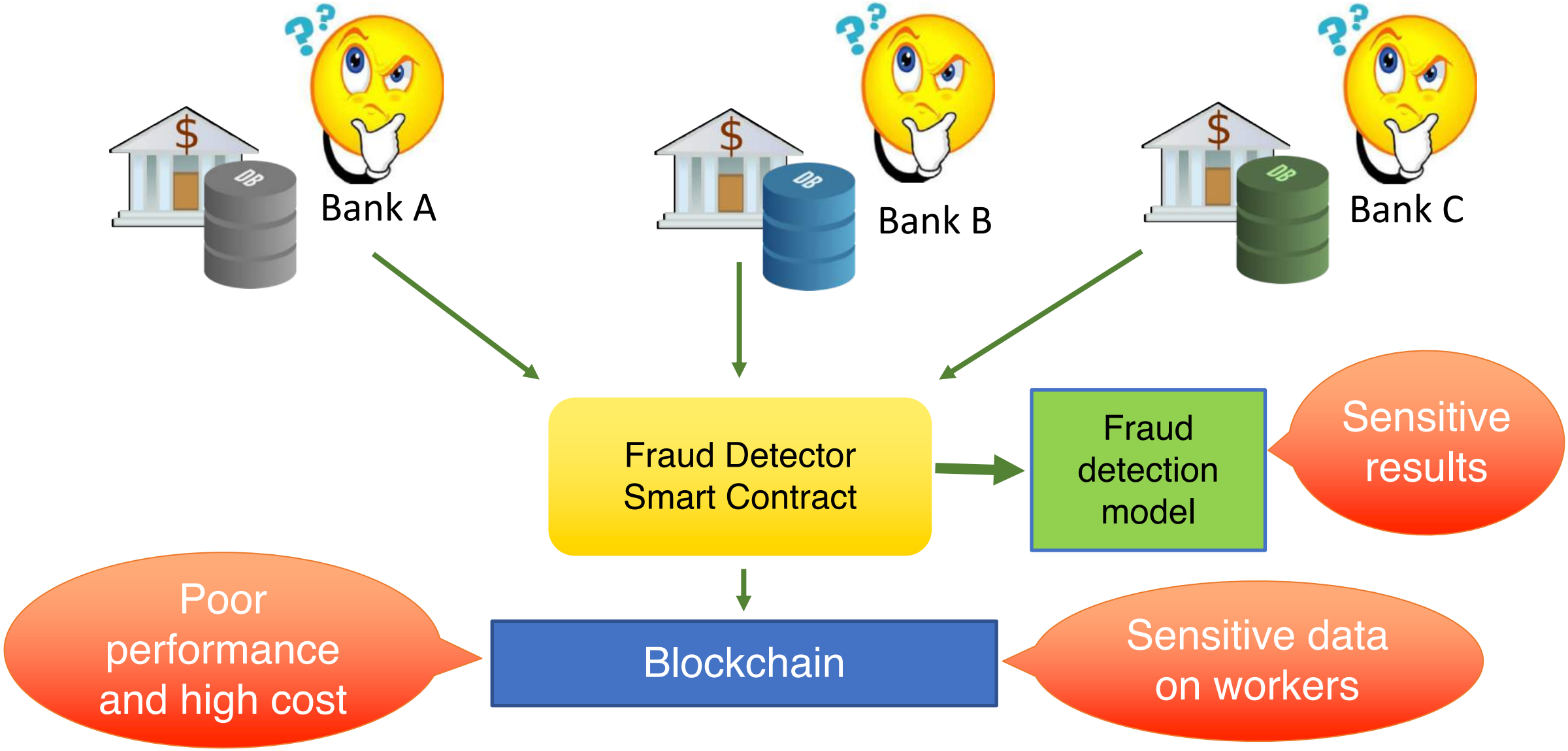


Banks would all benefit by combining data to train better model

Can’t do this today because:

- Privacy concerns
- Regulatory risk
- Misaligned incentives

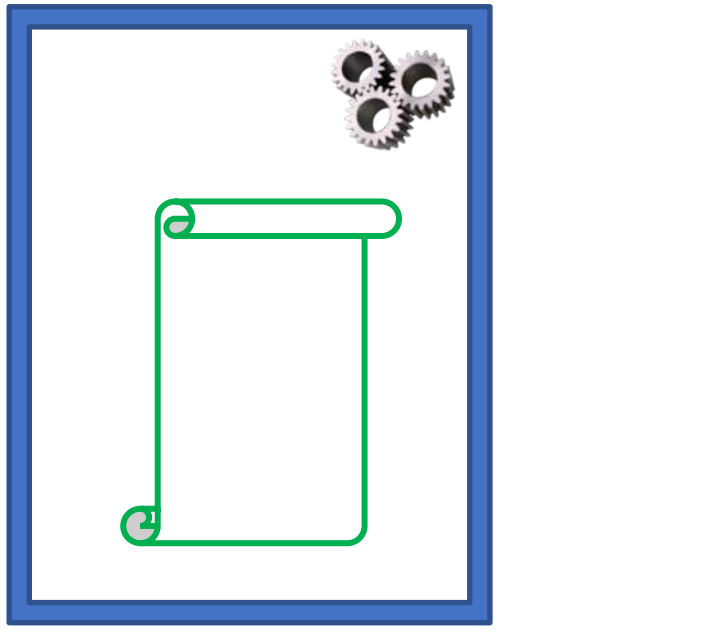
Motivating example: Fraud detection



Oasis: Privacy-preserving Smart Contracts at Scale

Our Solution

Privacy-preserving Smart contract

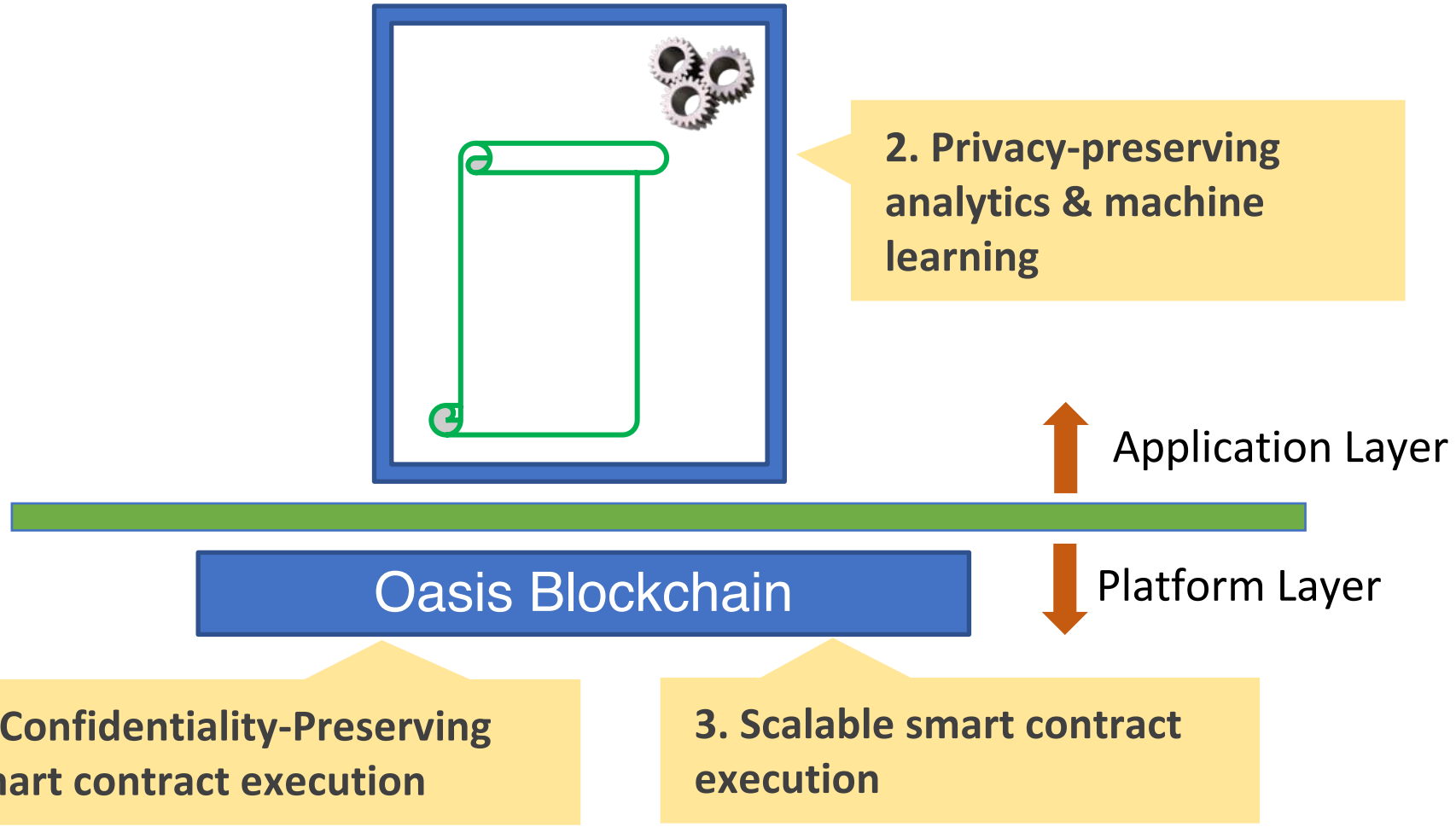


Oasis Blockchain

Properties of Our Solution

- Automatic enforcement of codified privacy requirements
- Without relying on any central party
- Scale to real-world applications including machine learning
- Easy to use for developers without privacy expertise

Privacy-Preserving Smart Contracts At Scale



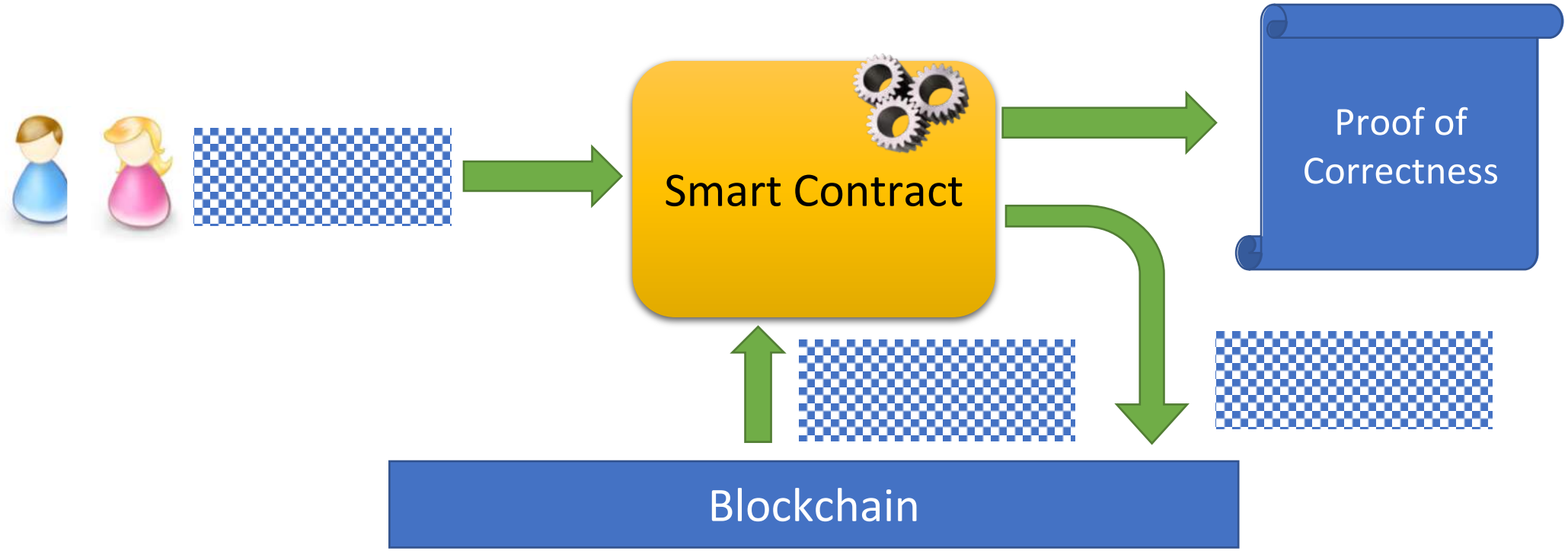
Outline

1. **Confidentiality-Preserving smart contract execution**
2. **Privacy-preserving analytics & machine learning**
3. **Scalable smart contract execution**

Outline

1. **Confidentiality-Preserving smart contract execution**
2. **Privacy-preserving analytics & machine learning**
3. **Scalable smart contract execution**

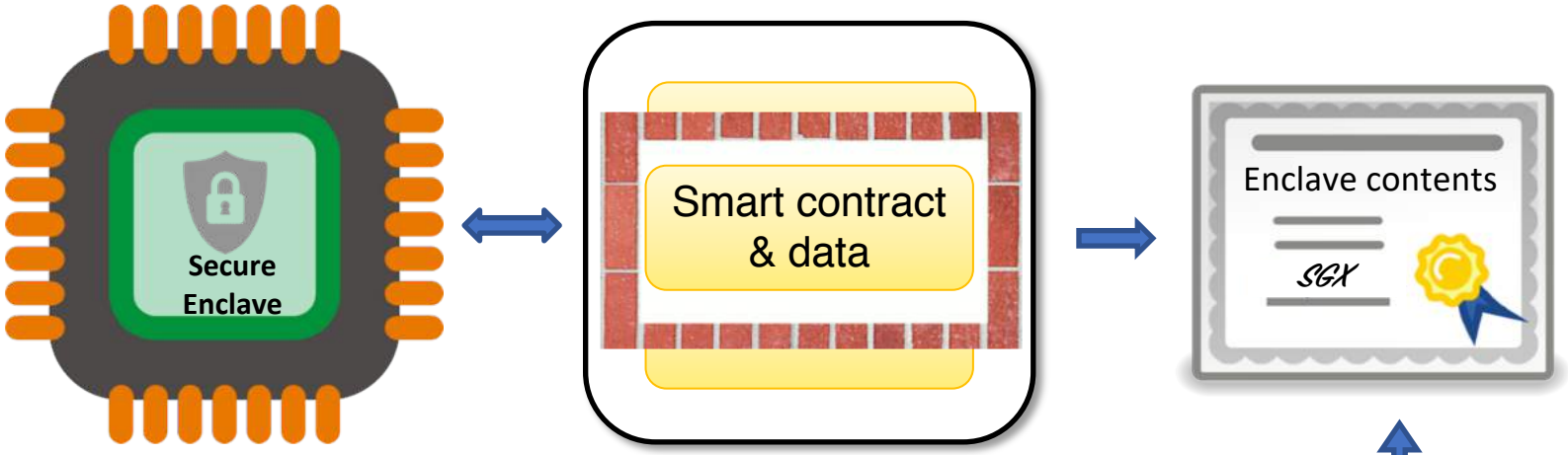
Confidentiality-preserving Smart Contract Execution



Secure computation techniques

	Performance	Support for general-purpose computation	Security mechanisms
Trusted hardware			Secure hardware
Secure multi-party computation			Cryptography, distributed trust
Zero-knowledge proof			Cryptography, local computation
Fully homomorphic encryption			Cryptography

Secure Hardware

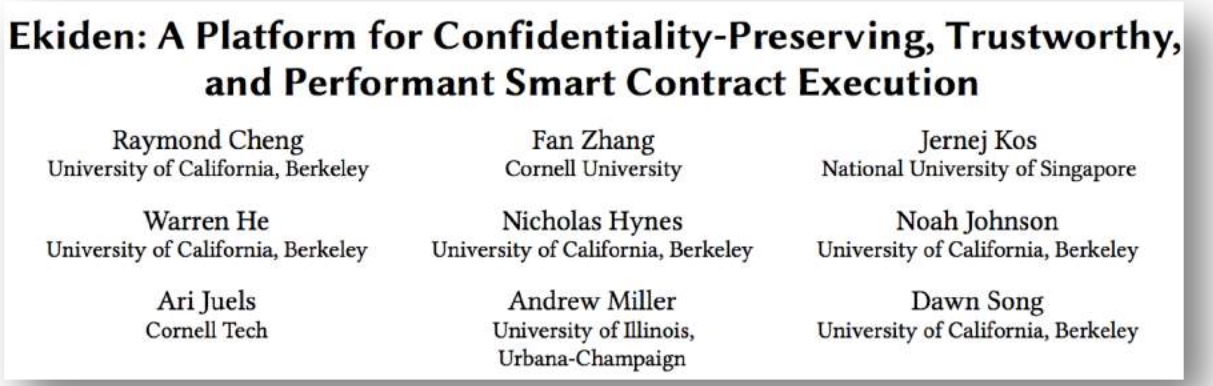


Integrity

Confidentiality

Remote Attestation

Ekiden: Confidentiality-preserving Smart Contracts



Ekiden: A Platform for Confidentiality-Preserving, Trustworthy, and Performant Smart Contract Execution

Raymond Cheng
University of California, Berkeley

Fan Zhang
Cornell University

Jernej Kos
National University of Singapore

Warren He
University of California, Berkeley

Nicholas Hynes
University of California, Berkeley

Noah Johnson
University of California, Berkeley

Ari Juels
Cornell Tech

Andrew Miller
University of Illinois, Urbana-Champaign

Dawn Song
University of California, Berkeley

https://arxiv.org/abs/1804.05141

- Smart contract execution using **secure computation**:
  - Secure Enclave (e.g. Intel SGX)
  - Cryptographic protocols: secure MPC or Zero-knowledge Proofs
- **Security proof**: Universal Composability

Ekiden: Sample Applications

Application	**Secret Input/Output**	**Secret State**
Machine Learning	Training data, predictions	Model
Thermal Modeling	Sensor data, temperature	Building model
Token (Rust)	Transfer(from, to, amount)	Account balances
Poker	Players' cards	Shuffled deck
Cryptokitties	Random mutations	Breeding algorithm
Ethereum VM	Input and output	Contract state

Secure Enclave as a Cornerstone Security Primitive

- Strong security capabilities
  - Authenticate itself (device)
  - Authenticate software
  - Guarantee the integrity and privacy of execution
- Platform for building new security applications
  - Couldn’t be built otherwise for the same practical performance
  - Many examples
    - Haven [OSDI’14], VC3 [S&P’15], M2R[USENIX Security‘15], Ryoan [OSDI’16], Opaque [NSDI’17]

Trusted hardware timeline

Closed source

2014

ARM TrustZone

Hardware-based isolation for embedded devices

2015

SGX: Software Guard Extensions

Built in to all Core™ processors (6th-generation and later)

2016

Trusted Execution Environment

- Hardware-based isolation
- TLK: open-source stack for TEE

2017

SEV: Secure Encrypted Virtualization

- Introduced in EYPC server processor line
- Provides confidentiality but not integrity

2018

Intel SGX version 2

- In pipeline
- Drivers already available

Open source

Keystone: Open-source secure enclave

https://keystone-enclave.github.io

- Collaboration between Berkeley & MIT
- Remedies issues in previous secure hardware
- Can be publicly analyzed and verified
- Can be manufactured by any manufacturer
- First release: Fall 2018

Challenges in Secure Hardware

- How secure can it be? Under what threat models?
- What would you entrust with secure hardware?
  - Your bitcoin keys
  - Financial data
  - Health data
- Can we create trustworthy secure enclave as a cornerstone security primitive?
  - Widely deployed, enable secure systems on top
  - A new secure computation era

Path to Trustworthy Secure Enclave

- Open source design
  - Provides transparency & enables high assurance
  - Builds a community
- Formal verification
- Secure supply-chain management

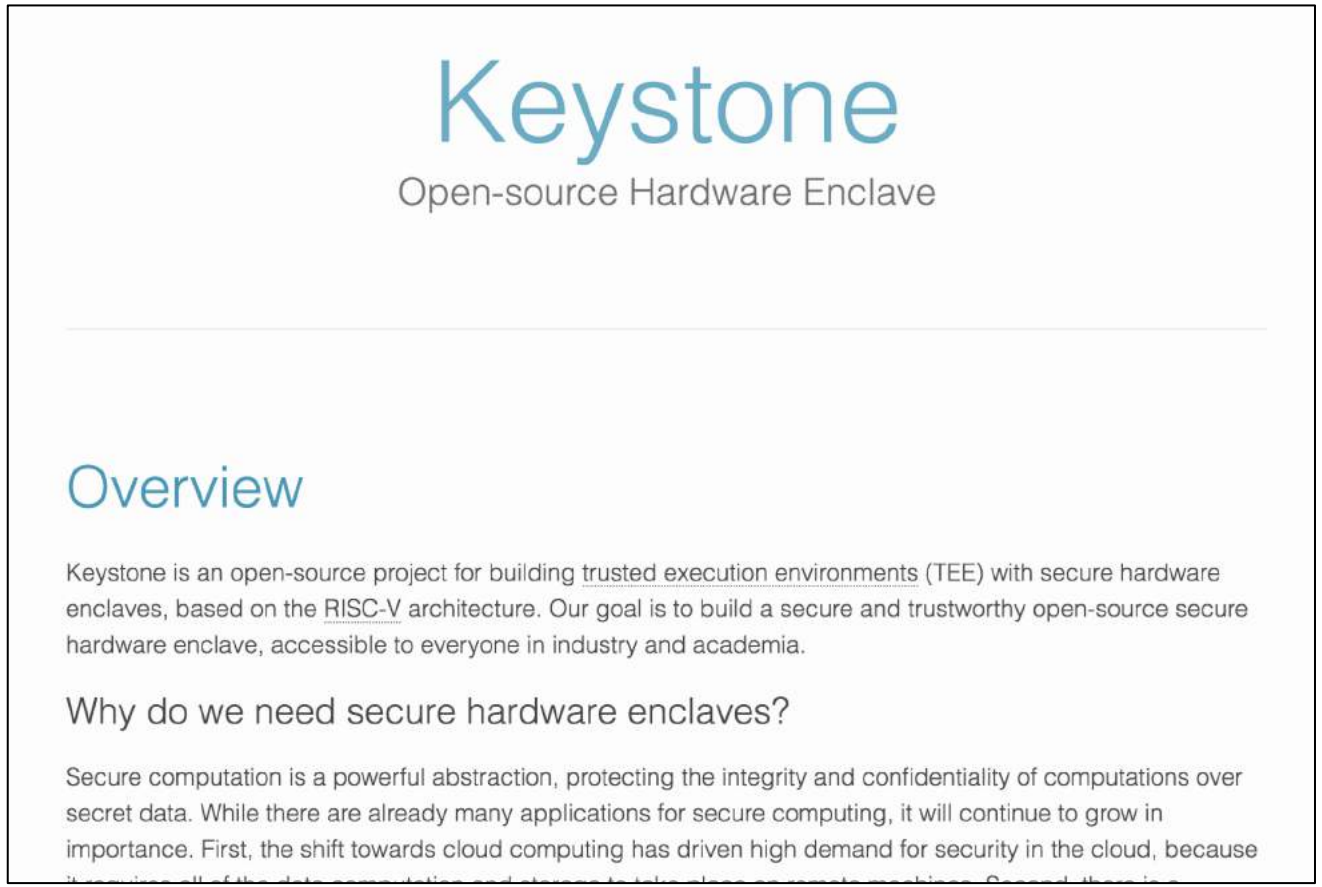
Keystone Enclave

- What is the Keystone Enclave?
  - Open-source Trusted Execution Environment (TEE) based on RISC-V
- Strong Memory Isolation
  - ISA-enforced memory access management
  - Separate virtual memory management without relying on the OS
- Simple and Portable
  - Exploits standard RISC-V ISA primitives: PMP, TVM
- Remote Attestation
  - Extends MIT Sanctum's remote attestation
- Open Source
  - Full software/hardware stack will be released
  - Run on many platforms: QEMU, Amazon AWS FPGA (FireSim), HiFive Unleashed, ...

Foundation: 100+ Members

Keystone Goals and Roadmap

Website: https://keystone-enclave.org



Keystone

Open-source Hardware Enclave

Overview

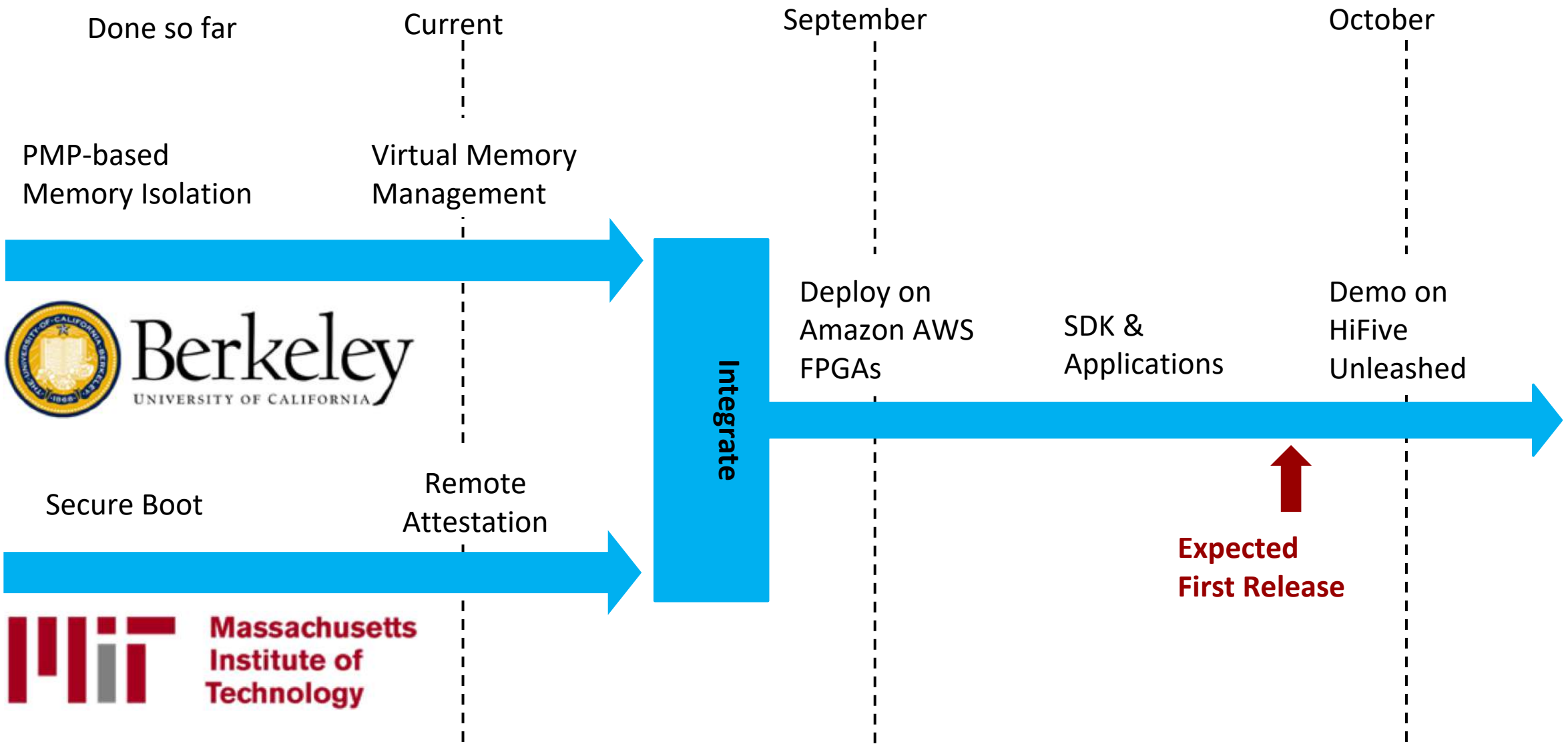
Keystone is an open-source project for building trusted execution environments (TEE) with secure hardware enclaves, based on the RISC-V architecture. Our goal is to build a secure and trustworthy open-source secure hardware enclave, accessible to everyone in industry and academia.

Why do we need secure hardware enclaves?

Secure computation is a powerful abstraction, protecting the integrity and confidentiality of computations over secret data. While there are already many applications for secure computing, it will continue to grow in importance. First, the shift towards cloud computing has driven high demand for security in the cloud, because

1. Chain of Trust
   - Secure boot
   - Remote attestation
   - Secure key provisioning (PUF)
2. Memory Isolation
   - Physical memory protection
   - Page table isolation
3. Defense against Physical Attack
   - Memory encryption
   - Memory address bus encryption
4. Defense against Side-channel Attack
   - Isolated architecture
5. Formal Verification
6. Deployment
   - RISC-V QEMU
   - Amazon AWS FPGAs (FireSim)
   - HiFive Unleashed
7. Tape Out to Chip
8. Secure supply-chain management

Timeline


Done so far
Current
September
October
PMP-based
Memory Isolation
Virtual Memory
Management
Berkeley
UNIVERSITY OF CALIFORNIA
Integrate
Deploy on
Amazon AWS
FPGAs
SDK &
Applications
Demo on
HiFive
Unleashed
Secure Boot
Remote
Attestation
Expected
First Release
Massachusetts
Institute of
Technology

Outline

1. Confidentiality-Preserving smart contract execution

2. Privacy-preserving analytics & machine learning

3. Scalable smart contract execution

Privacy Risks in Analytics

How many trips were taken in New York last year?

Reflects a **trend**

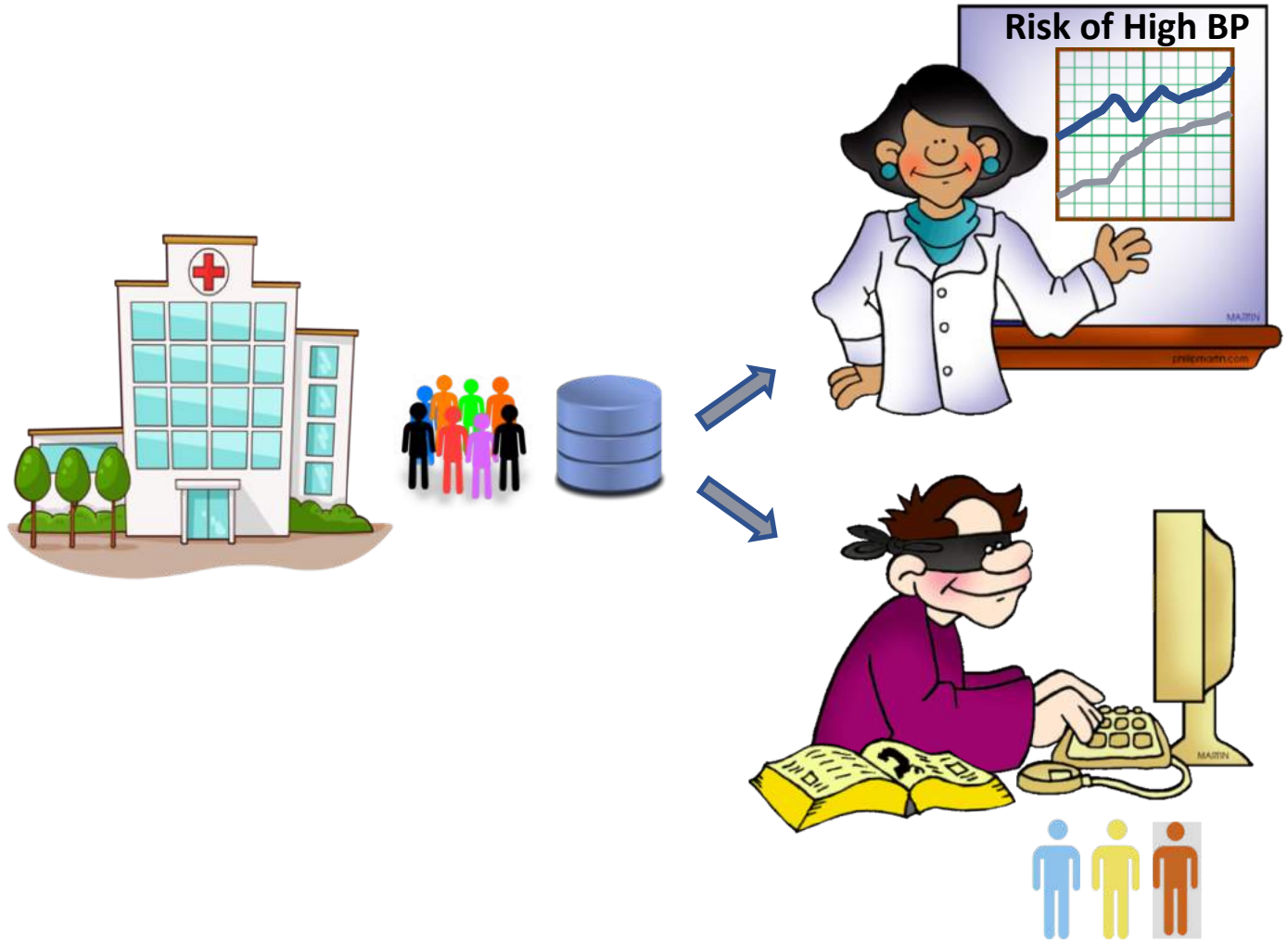
How many trips did Joe take last week?

Reflects an **individual**

Access control policies cannot enable the use of data while protecting the privacy of individuals

Data Anonymization

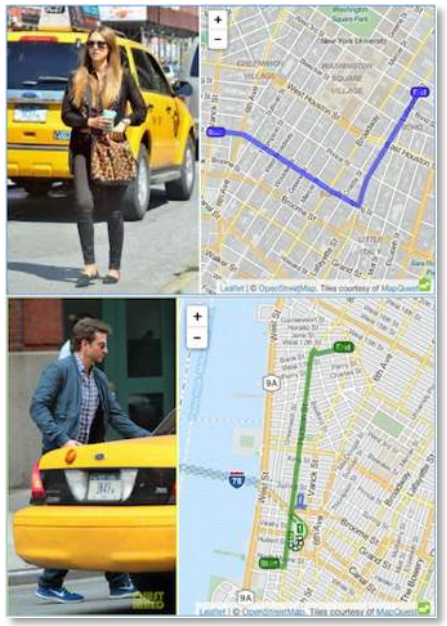
Data Anonymization



Reidentification attacks

Netflix prize (Narayanan et al.)

NYC taxi data (Anthony Tockar)



Do Neural Networks Remember Training Data?

Can Attackers Extract Secrets (in Training Data) from (Querying) Learned Models?

N Carlini, C Liu, J Kos, Ú Erlingsson, and D Song.
"The Secret Sharer: Measuring Unintended Neural Network Memorization & Extracting Secrets". 2018.

Extracting Social Security Number from Language Model

- Learning task: train a language model on Enron Email dataset
  - Containing actual people's credit card and social security numbers
- New attacks: can extract 3 of the 10 secrets completely by querying trained models
- New measure "Exposure" for memorization
  - Used in Google Smart Compose

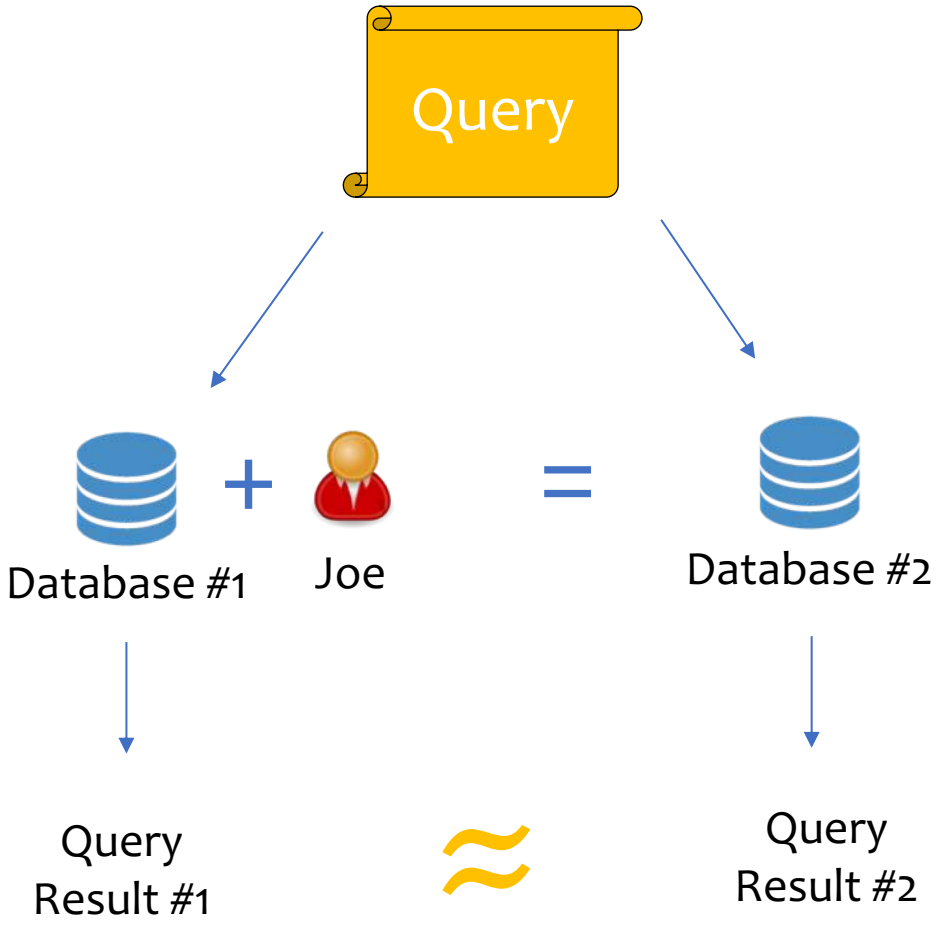
User	Secret Type	Exposure	Extracted?
A	CCN	52	✓
B	SSN	13	
C	SSN	16	
	SSN	10	
	SSN	22	
D	SSN	32	✓
F	SSN	13	
G	CCN	36	
	CCN	29	
	CCN	48	✓

Preventing Memorization

- **Differential Privacy:** a formal notion of privacy to protect sensitive inputs
- Solution: train a differentially-private neural network
  - Exposure is lower empirically
  - Attack unable to extract secrets

	Optimizer	ε	Testing Loss	Estimated Exposure
With DP	RMSProp	0.65	1.69	1.1
	RMSProp	1.21	1.59	2.3
	RMSProp	5.26	1.41	1.8
	RMSProp	89	1.34	2.1
	RMSProp	2×10^8	1.32	3.2
	RMSProp	1×10^9	1.26	2.8
	SGD	∞	2.11	3.6
No DP	SGD	N/A	1.86	9.5
	RMSProp	N/A	1.17	31.0

Differential Privacy: a Formal Privacy Definition



- Outcome is the same **with or without** Joe's data
  - Holds for *every* user and *every* database
- Immune to re-identification attacks
- Parameterized by ε (the *privacy budget*)

Real-world Use of Differential Privacy

- Previous work on differential privacy is either:
  - Theoretical
  - Targeted for specialized applications
    - Google: top websites visited
    - Apple: top emojis used
- No previous real-world deployments of differential privacy for general-purpose analytics

Challenges for Practical General-purpose Differential Privacy for SQL Queries

- Usability for non-experts
- Broad support for analytics queries
- Easy integration with existing data environments

No existing system addresses these issues

Collaboration with Uber: address practical deployment challenges

Chorus: a Framework for Privacy-preserving Analytics

- **Usable by non-experts**
  - Analyst does not need to understand differential privacy
  - Chorus automatically enforces differential privacy for SQL queries

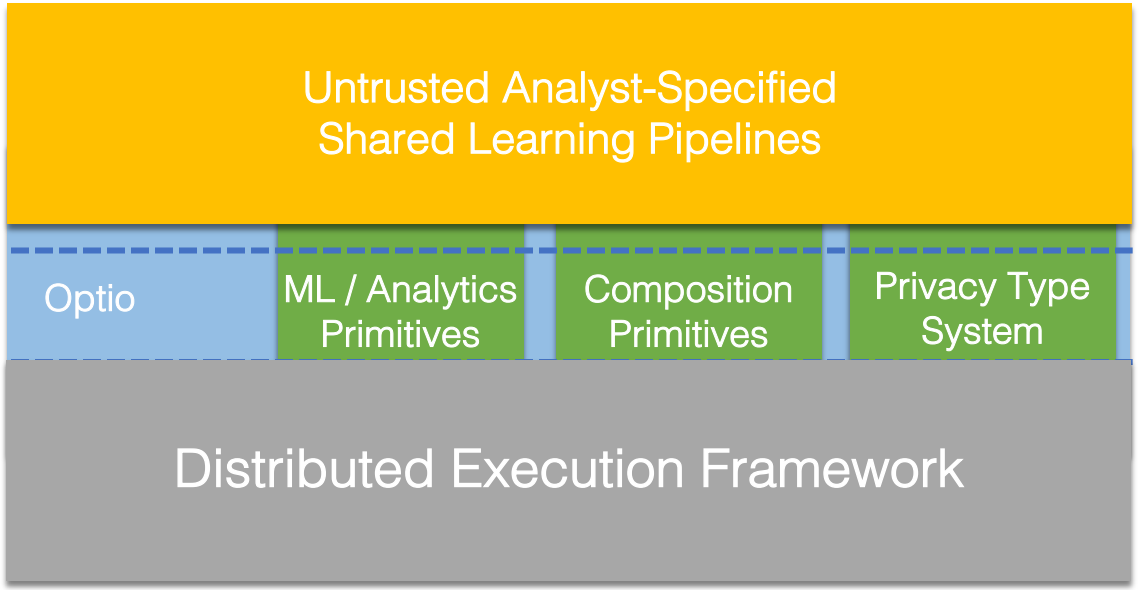
- **Broad support for analytics queries**
  - Modular design to support wide variety of mechanisms
  - Implemented mechanisms support 93% of queries in our workload

- **Easy integration with existing data environments**
  - Chorus works with standard SQL databases

- **Designed for real-world use**
  - Deployment underway at Uber

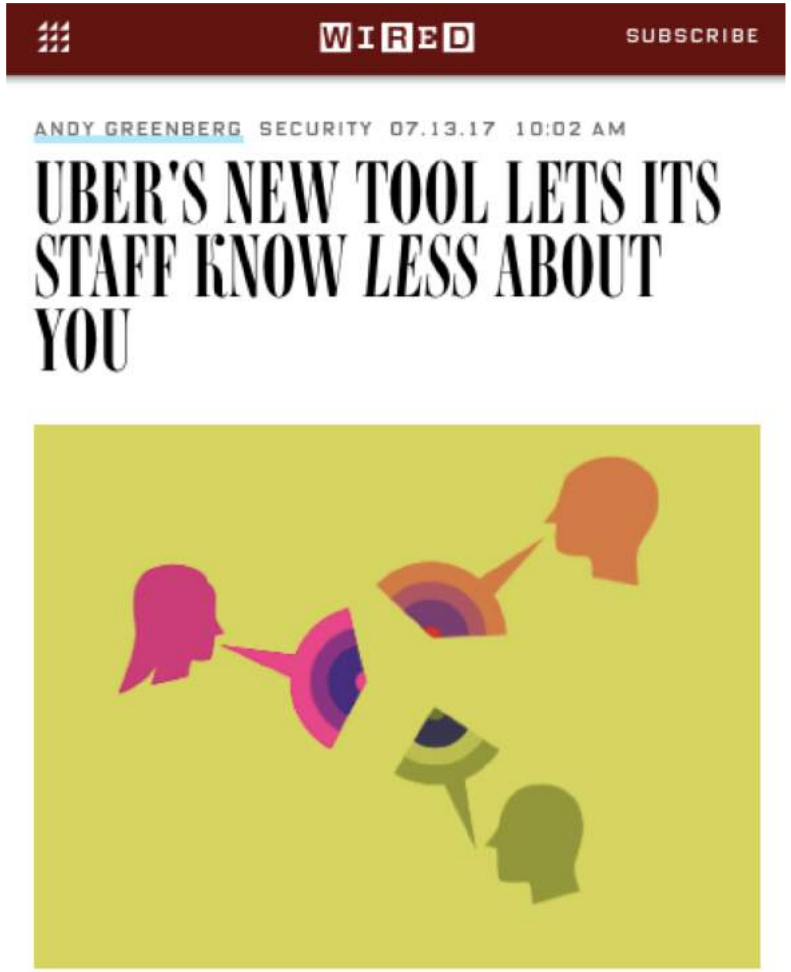
Optio: Privacy-Preserving Machine Learning

- Optio provides automatic differential privacy guarantees
  - Rewriting and verifying analytics and ML pipelines
  - Type system to enforce privacy policies



Real-world Deployment at Uber

- Ongoing **deployment** for analytics
  - Differential privacy
  - GDPR
- Plans for public-facing systems
- Open-source release: https://github.com/uber/sql-differential-privacy



Kara

A Privacy-Preserving Tokenized Data Market for Medical Data

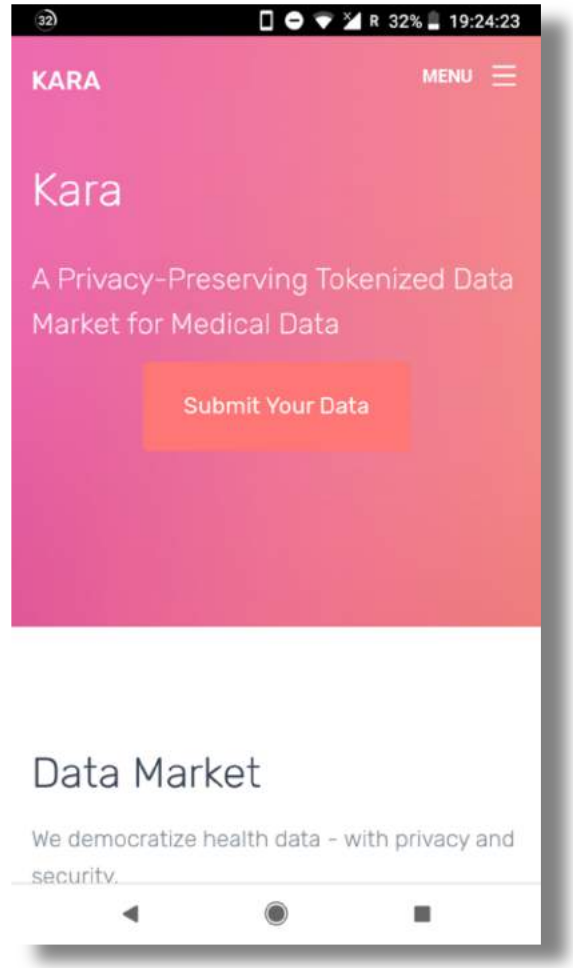
Oasis Labs Berkeley UNIVERSITY OF CALIFORNIA ETH zürich Stanford HEALTH CARE

Medical data *is locked in “Data Silos”.*

Goal: *Incentivize* *doctors and patients to share data and* *improve medical research*!

Meet Kara!

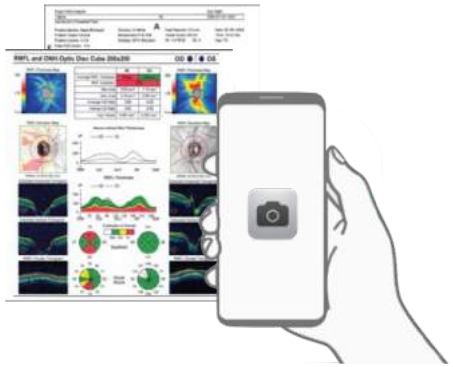
- **Kara** is a privacy-preserving tokenized data market
- **Easy, fast and secure way** for doctors and patients to **earn tokens by sharing data**
- Data is stored **securely and privately** in Oasis Blockchain Platform
- Researchers, doctors, industry can look for certain diseases / categories and **pay to train their models with privacy-preserving machine learning**



How it works

Doctors / Patients

Researchers

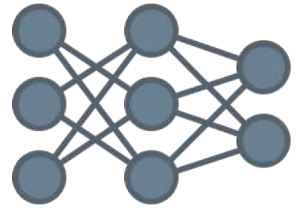


1. Patients take picture of their medical data

Kara App

2. App automatically processes data and stores it to Oasis

Private Training

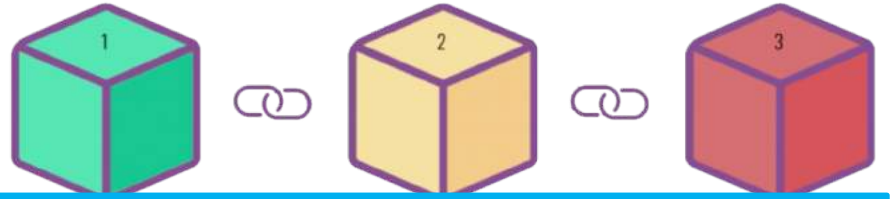


3. Researchers pay for model training

Untrained Model

Data Market

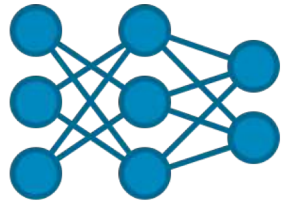
Token



Oasis Blockchain

4. Receive **payments**

Privacy-preserving smart contracts serve as business logic



4. Receive **trained** model

Nick Hynes, Raymond Cheng, Noah Johnson, David Dao, Dawn Song. "**A Demonstration of Sterling: A Privacy-Preserving Data Marketplace"** in VLDB'18 (Demo Track)

David Dao, Dan Alistarh, Claudiu Musat, Ce Zhang. "**DataBright: Towards a Global Exchange for Decentralized Data Ownership and Trusted Computation**"

Oasis: Example use cases

Private escrow

Tokens

Prediction market

Regulatory compliance (enterprise)

Privacy-preserving machine learning

Personalized medicine

Decentralized exchange

Collaborative analytics

Blind auction

Portfolio manager

Credit scoring

Blockchain games (e.g. Poker, Cryptokitties)

Oasis Labs Just Launched!


MIT Technology Review

Login / Create an account Search

Topics+ The Download Magazine Events More+

Meet Oasis Labs, the blockchain startup Silicon Valley is buzzing about


Forbes

Big Hitter Crypto Funds Pile Into Privacy-Enhanced Smart Contract Startup Oasis Labs


THE WALL STREET JOURNAL.

U.S. Edition | July 10, 2018 | Today's Paper | Video

Oasis Labs Building Cloud Computing on Blockchain With $45 Million

Backers include a16zcrypto, Accel Partners, Binance, Polychain, Metastable


WIRED

BUSINESS CULTURE

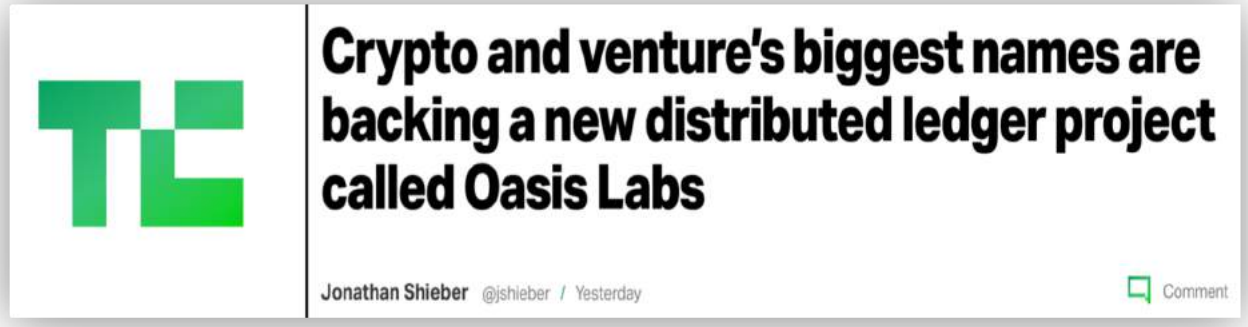
TOM SIMONITE BUSINESS 07.11.18 08:00 AM

HOW A STARTUP IS USING THE BLOCKCHAIN TO PROTECT YOUR PRIVACY


VB CHANNELS EVENTS NEWSLETTERS

Oasis Labs raises $45 million for 'privacy first' cloud on blockchain

DEAN TAKAHASHI @DEANTAK JULY 9, 2018 3:00 AM


Crypto and venture's biggest names are backing a new distributed ledger project called Oasis Labs

Jonathan Shieber @jshieber / Yesterday

Comment

Oasis Testnet

Interested in building an application on Oasis?

Join our private testnet!

https://www.oasislabs.com/developers

Oasis Labs

Building a privacy-first, high performance cloud computing platform on blockchain.

We're hiring!

www.oasislabs.com

Thank you!

Microsoft