Kaires: Fully Decentralized Privacy-Preserving Machine Learning Framework

Islam Faisal* [1]
@islamfaisalm

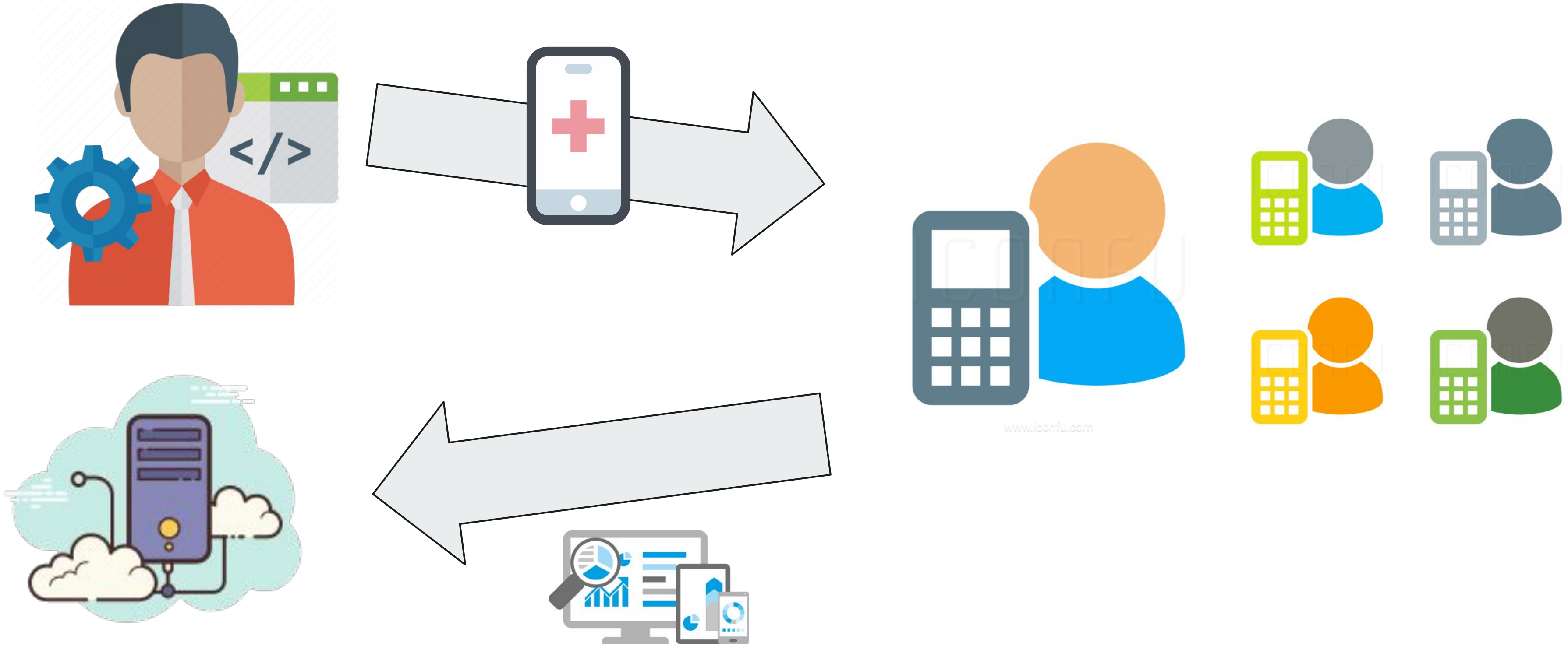
Eleftherios Kokoris-Kogias [2]
@LefKok

Bryan Ford [2]
@brynosaurus

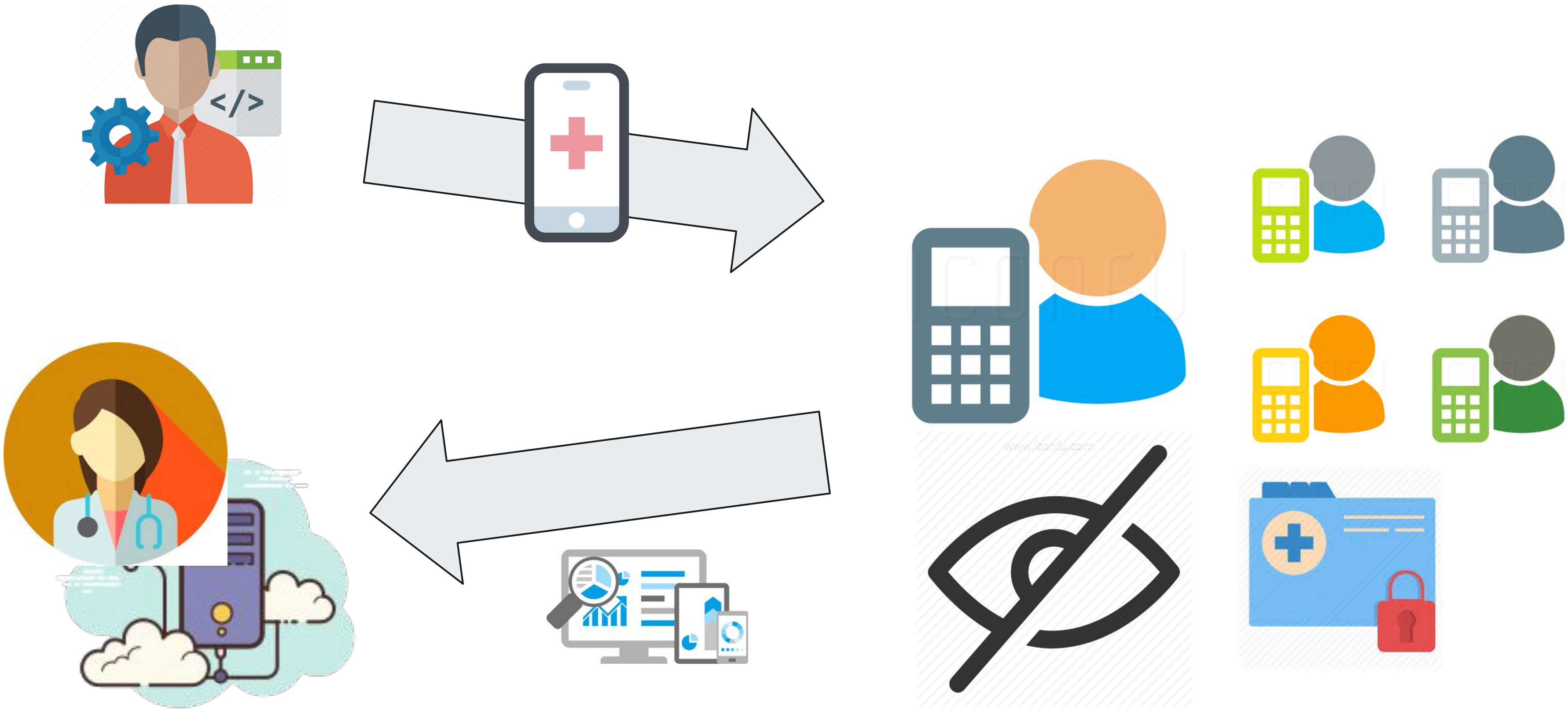
Sherif El-Kassas [1]
@kassas0

[1] The American University in Cairo (AUC)
[2] Ecole polytechnique fédérale de Lausanne (EPFL)
* Work mostly done while visiting EPFL.

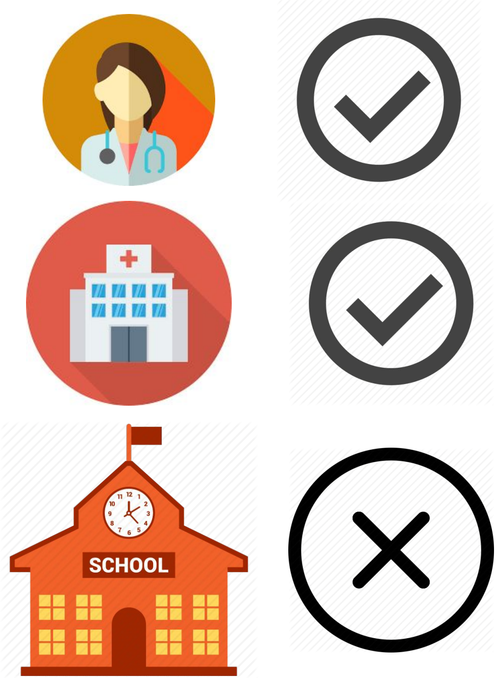
Data Analytics in the Wild



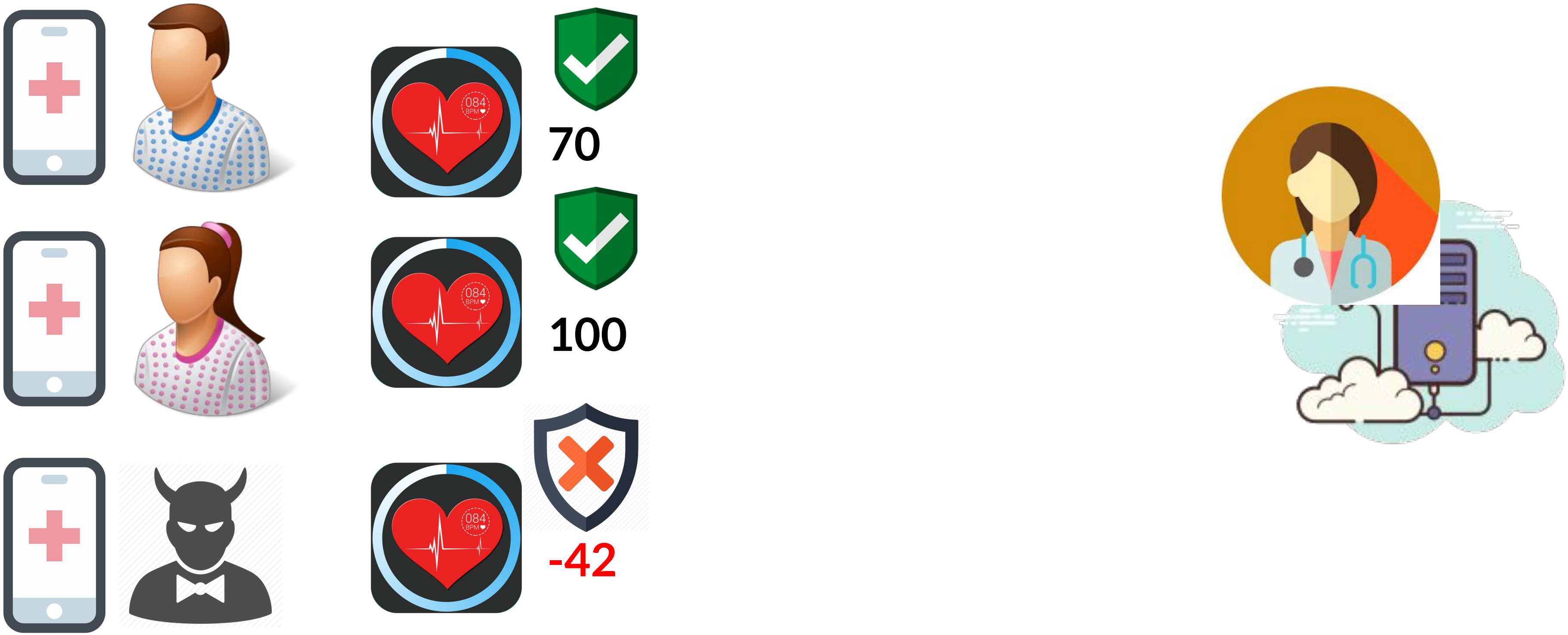
Privacy Vs. Functionality



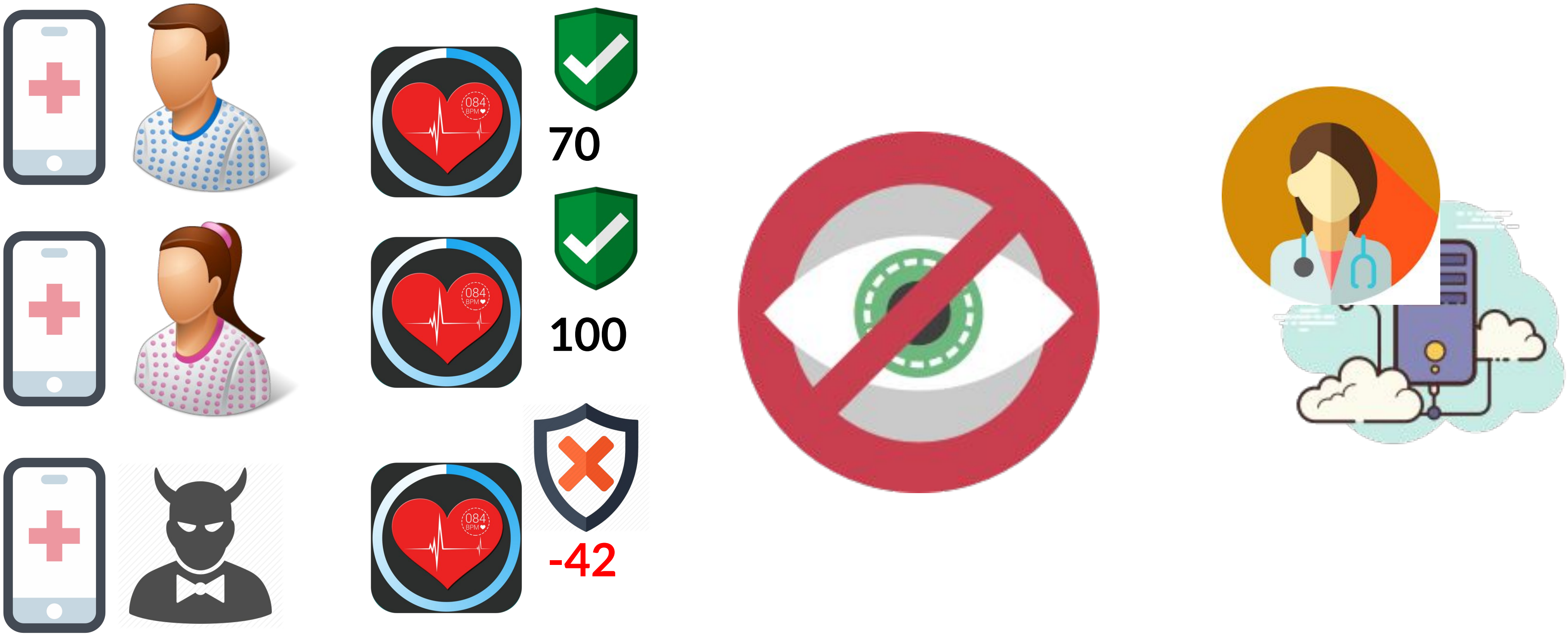
Access-Control



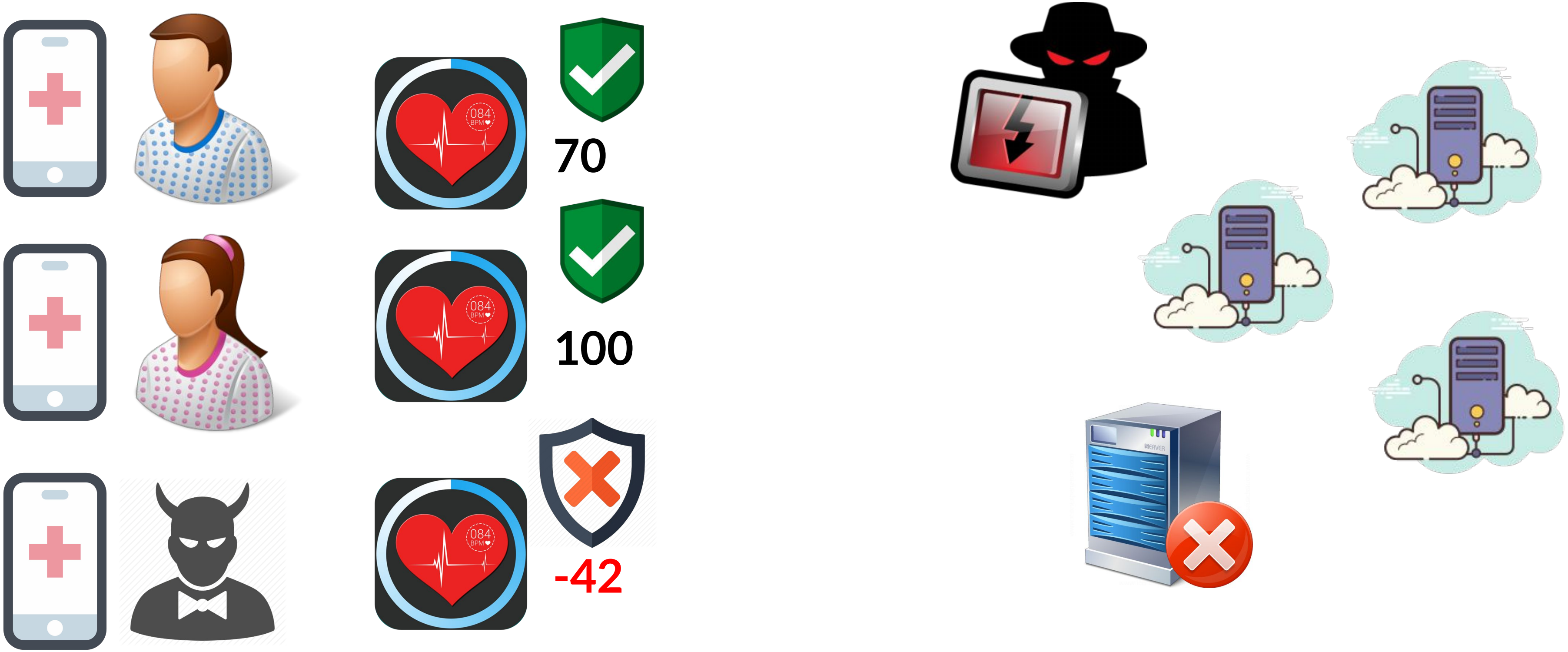
Robustness against “bad” clients



Robustness against “bad” clients



Robustness against “bad” servers



System Goals

Access-Control

Users have revokable fine-grained access control over their data.

Privacy

Privacy of confidential data points as well as machine learning models.

Fairness

Fair exchange of data points and in-return value.

Robustness

Model can be built even with a failing (or dishonest) minority.

Decentralization

No single point of compromise or failure.

Auditability

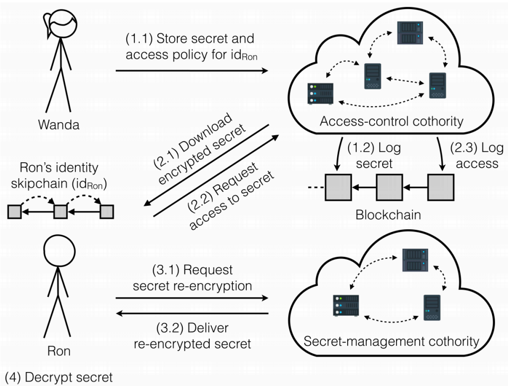
Publicly verifiable tamper-proof access logs for data accesses.


TRUSTED THIRD PARTY
TRUSTED HARDWARE
APPLY CRYPTOGRAPHY
imgflip.com

CALYPSO: Auditable Sharing of Private Data (Kokoris-Kogias et al.)

System Properties

- Confidentiality
- Auditability
- Atomic Data Delivery
- Dynamic Identity Management
- Decentralization



Machine Learning with CALYPSO

- Use CALYPSO to store and retrieve the data points
- Central machine learning node
- Data points are collected in a publicly auditable access-controlled system
- Data consumers have access to plain-text data points

Property	CALYPSO
Access-Control	✓
Fairness	✓
Auditability	✓
Decentralization	?[1]
Privacy	✗
Robustness	✗

[1] Access-control and secret-sharing are decentralized, but learning is centralized.

Prio (Corrigan-Gibbs–Boneh)

- A decentralized system for computing aggregate statistics
- Provides client-privacy as long as there is at least one honest server
- Servers learn about the data no more than they can learn from statistics

Prio (Corrigan-Gibbs–Boneh)

Secret-Shared Non-Interactive Proofs (SNIPs)

- Distributed zero-knowledge proofs that can prove whether a certain point x satisfies a boolean circuit Valid(x)
- Provides robustness against adverserial clients

Multi-Party Computation

- Local aggregators compute local values from shares
- A global aggregator combines all the local aggregators to obtain the model

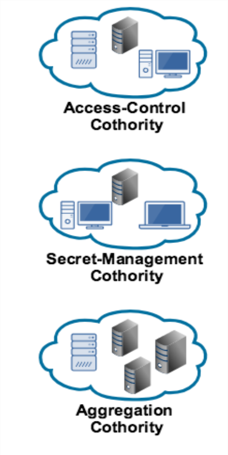
System Design

- Combine CALYPSO with Prio to get a decentralized design
- Neither data consumers nor aggregation servers see the data in plain-text
- Extend Prio so that only the data consumer has access to the model

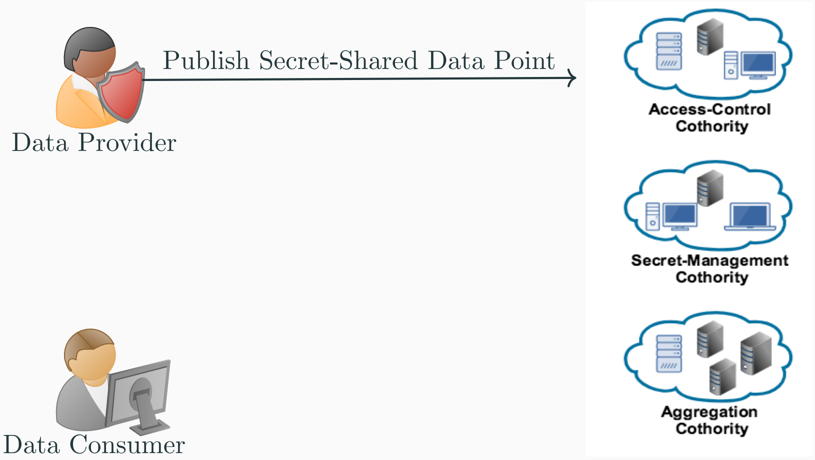
System Design

Data Provider

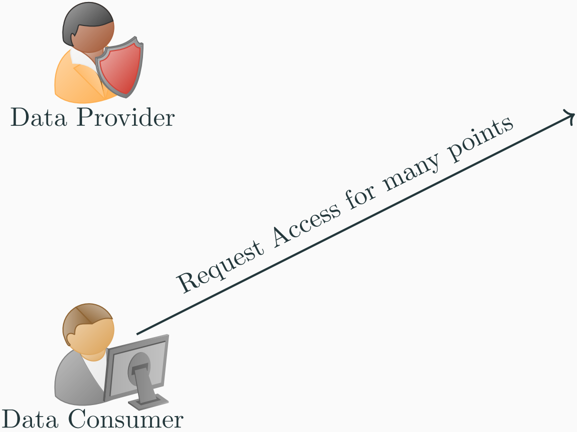
Data Consumer

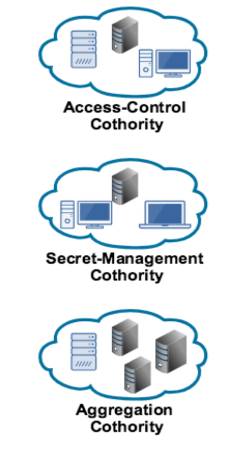


System Design

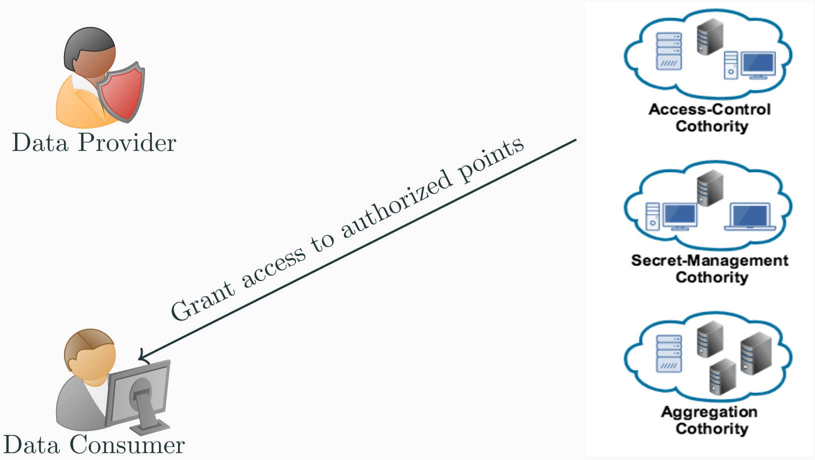


System Design

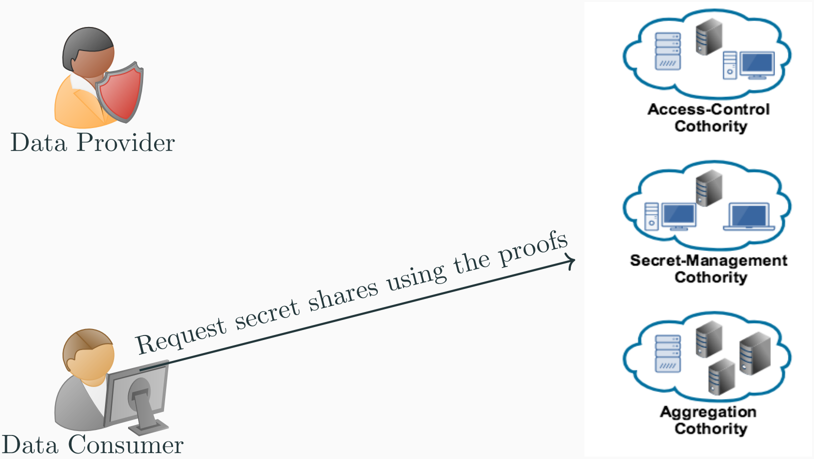




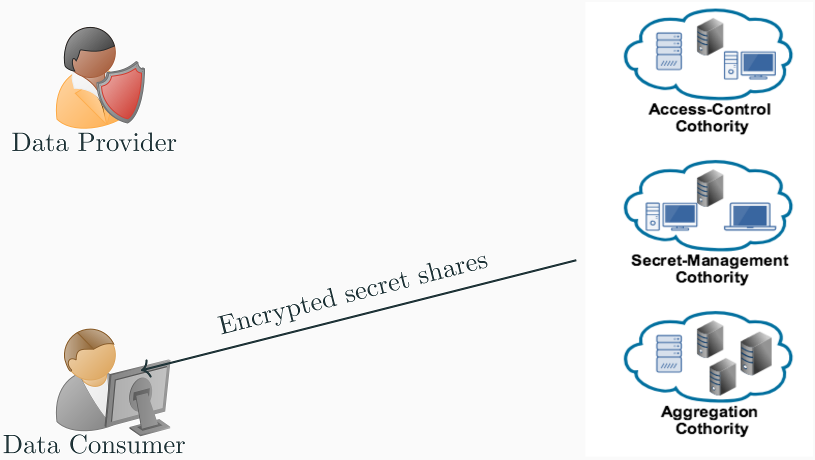
System Design



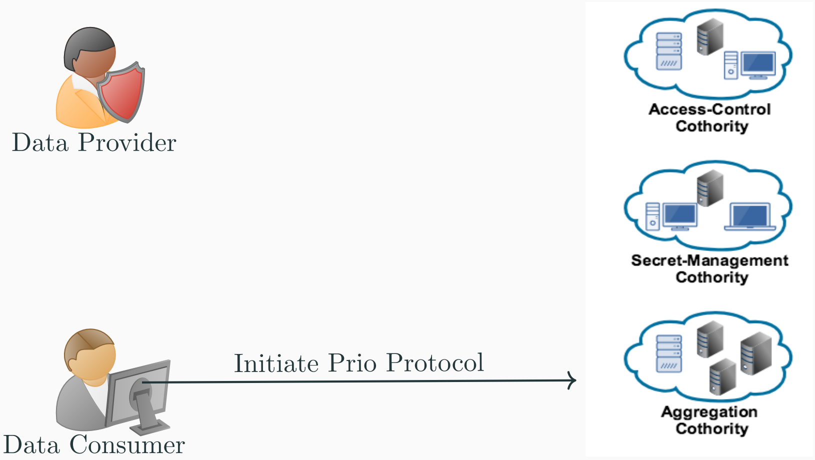
System Design



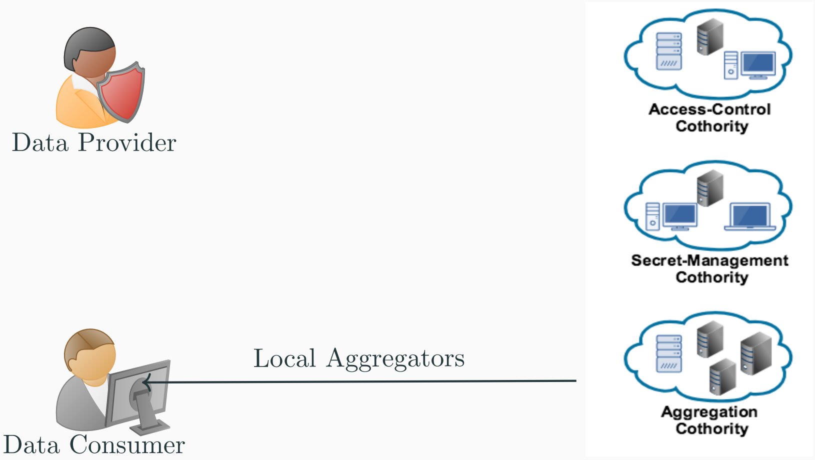
System Design



System Design



System Design

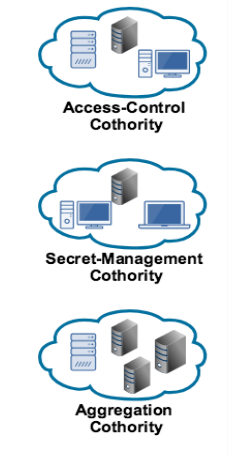


System Design

Data Provider

Combine all local aggregators

Data Consumer



System Design

Property	CALYPSO	+Prio
Access-Control	✓	✓
Fairness	✓	✓
Auditability	✓	✓
Decentralization	?[1]	✓
Privacy	✗	✓
Robustness	✗	✗[2]

[1] Access-control and secret-sharing are decentralized, but learning is centralized.
[2] Robustness against adverserial clients only.

How can we make the system tolerate a faulty minority?

Tolerating a faulty minority

- SNIPs are essentially a multi-party computation of a certain arithmetic circuit C_f
- If only we could replace the circuit evaluation protocol by one that is fault-tolerant ...

Multi-party computation against a faulty minority [1]

Creates a multi-party evaluation protocol based on a verifiable secret-sharing scheme (VSS)

Verifiable Secret-Sharing (VSS)

- VSS Share: Allows a dealer to share a certain with n nodes.
- VSS Reconstruct: Reconstruction protocol to be run by the nodes.

[2] Ronald Cramer et al. "Efficient Multiparty Computations Secure Against an Adaptive Adversary". In: EUROCRYPT '99. 1999, p. 1.

Multi-party computation against a faulty minority [1]

Creates a multi-party evaluation protocol based on a verifiable secret-sharing scheme (VSS)

Verifiable Secret-Sharing (VSS)

- VSS Share: Allows a dealer to share a certain with n nodes.
- VSS Reconstruct: Reconstruction protocol to be run by the nodes.

Properties

- Secrets in VSS can be reconstructed with up to n/2 failing or dishonest nodes
- Computations in the MPC protocol based on VSS is robust against up to n/2 failing or dishonest nodes

[2] Cramer et al., "Efficient Multiparty Computations Secure Against an Adaptive Adversary", p. 1.

Properties Recap

Property	CALYPSO	+Prio	+BFT Prio
Access-Control	✓	✓	✓
Fairness	✓	✓	✓
Auditability	✓	✓	✓
Decentralization	?[3]	✓	✓
Privacy	✗	✓	✓
Robustness	✗	✗[4]	✓

[3] Access-control and secret-sharing are decentralized, but the learning is centralized.

[4] Robustness against adverserial clients only.

Conclusion

- Designed a fully-decentralized fault-tolerant machine learning framework that is private, fair, auditable, and robust.
- The first system to our knowledge that achieves these properties without reliance on extra assumptions such as trusted hardware.
- Integrated the ByzCoin distributed ledger and Calypso service with the Prio MPC primitives to implement the system in Go.

Feedback is welcome!

islam@bu.edu

Islam Faisal	@islamfaisalm
Eleftherios Kokoris-Kogias	@LefKok
Bryan Ford	@brynosaurus
Sherif El-Kassas	@kassas0