

Bringing Decentralized Search to Decentralized Services

Mingyu Li¹, Jinhao Zhu¹, Tianxu Zhang¹, Cheng Tan²,
Yubin Xia¹, Sebastian Angel³, Haibo Chen¹



IPADS
INSTITUTE OF PARALLEL
AND DISTRIBUTED SYSTEMS



Northeastern
University



Penn
UNIVERSITY OF PENNSYLVANIA



Outline

- **Problem Statement**
- **Design: decentralized search**
- **Design: verifiable search**
- **Design: private search**
- **Evaluation**

Status Quo of DApps

- Decentralized applications are booming

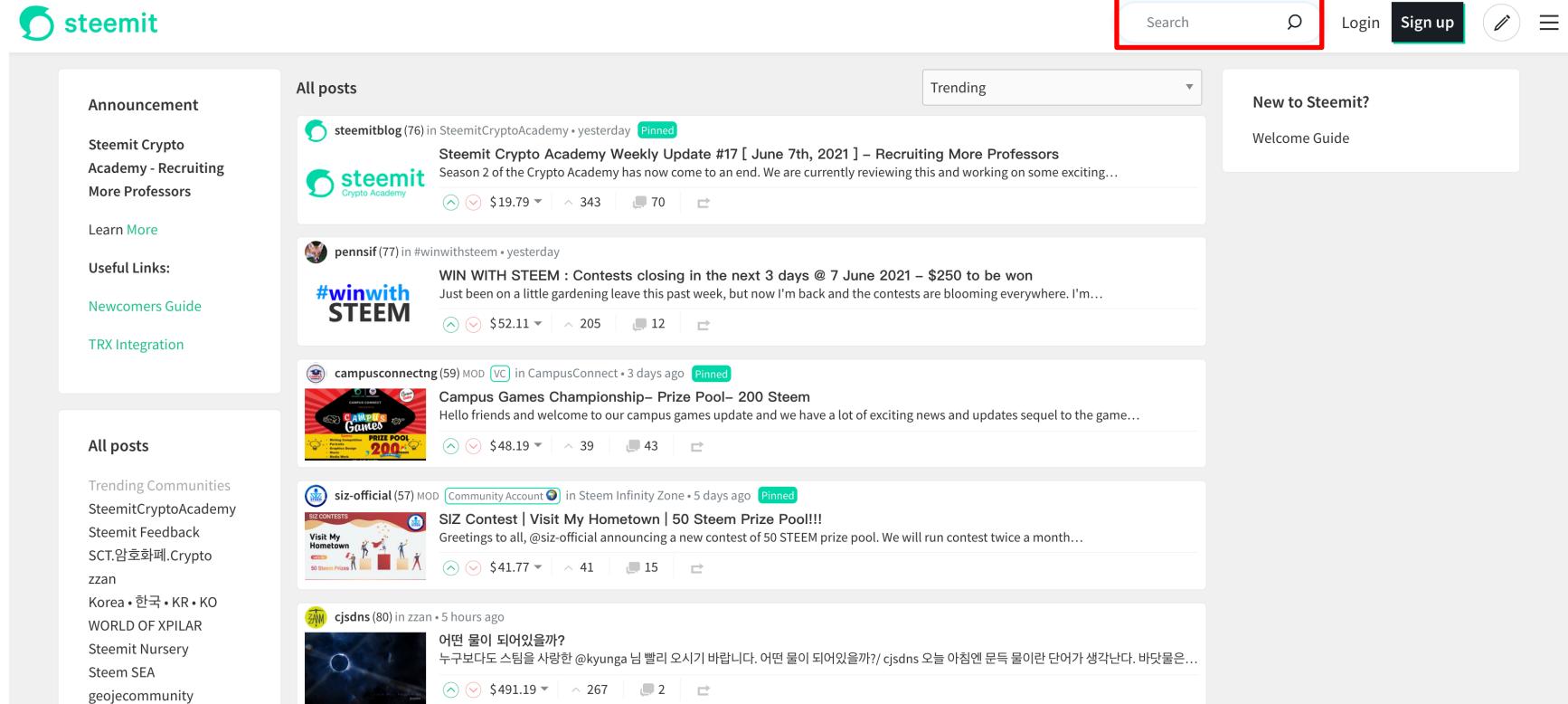
Service	Centralized	Decentralized
E-commerce	eBay	OpenBazaar
Social Media	Twitter	Steemit
Video Sharing	Youtube	DTube
Public Storage	DropBox	IPFS
Messaging	Slack	Matrix
Video Conference	Zoom	Zipcall
...

Search as DApps' Entrance

The screenshot shows the OpenBazaar web interface. At the top, there's a navigation bar with icons for OpenBazaar Discover, OB1, and a plus sign, followed by 'Transactions' and 'My Page'. A red box highlights the 'Search' button on the right side of the main search bar. A modal dialog titled 'Add search provider' is open, containing a text input for 'URL for the search provider' and a 'Cancel' or 'Add' button. Below the modal, the search results are displayed under the heading '7546 listings'. The results include three items: 1) 'Video Clase Pensamiento...' with a price of \$2.41, 2) 'Reddit PREMIUM Accou...' with a price of \$15.00, and 3) 'Abstract Geo Print Halte...' with a price of \$47.99. Each listing includes a small image, a title, a rating (all 0.0), and a 'FREE SHIPPING' badge.

OpenBazaar users need to search shopping items

Search as DApps' Entrance



The screenshot shows the Steemit homepage. At the top right, there is a search bar with a magnifying glass icon, which is highlighted with a red box. To the right of the search bar are buttons for "Login" and "Sign up". Below the search bar is a navigation menu with icons for edit and more.

The main content area displays a feed of posts. A "Trending" dropdown menu is visible above the first post. The posts include:

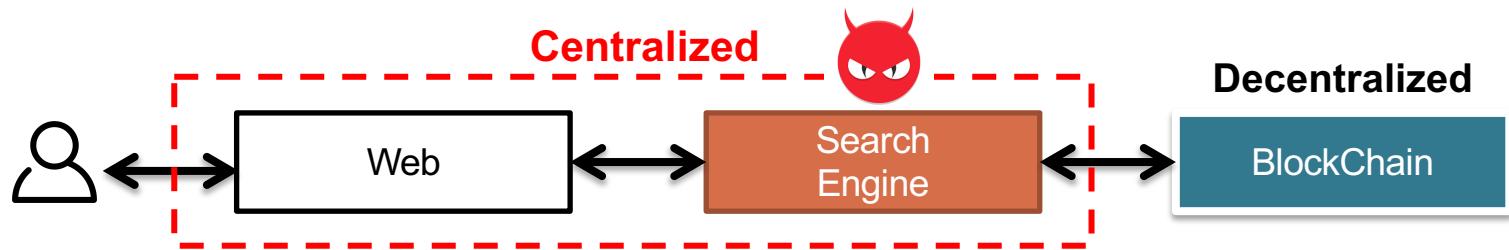
- steemitblog (76)** in SteemitCryptoAcademy • yesterday Pinned
Steemit Crypto Academy Weekly Update #17 [June 7th, 2021] – Recruiting More Professors
Season 2 of the Crypto Academy has now come to an end. We are currently reviewing this and working on some exciting...
 \$19.79 ▾ 343 70
- pennsif (77)** in #winwithsteem • yesterday
WIN WITH STEEM : Contests closing in the next 3 days @ 7 June 2021 – \$250 to be won
Just been on a little gardening leave this past week, but now I'm back and the contests are blooming everywhere. I'm...
 \$52.11 ▾ 205 12
- campusconnectng (59) MOD** in CampusConnect • 3 days ago Pinned
Campus Games Championship– Prize Pool- 200 Steem
Hello friends and welcome to our campus games update and we have a lot of exciting news and updates sequel to the game...
 \$48.19 ▾ 39 43
- siz-official (57) MOD** in Steem Infinity Zone • 5 days ago Pinned
SIZ Contest | Visit My Hometown | 50 Steem Prize Pool!!!
Greetings to all, @siz-official announcing a new contest of 50 STEEM prize pool. We will run contest twice a month...
 \$41.77 ▾ 41 15
- cjsdns (80)** in zzan • 5 hours ago
어떤 물이 되어있을까?
누구보다도 스팀을 사랑한 @kyunga 님 빨리 오시기 바랍니다. 어떤 물이 되어있을까?/ cjsdns 오늘 아침엔 문득 물이란 단어가 생각난다. 바닷물은...
 \$491.19 ▾ 267 2

Steemit users need to search blogs of interest

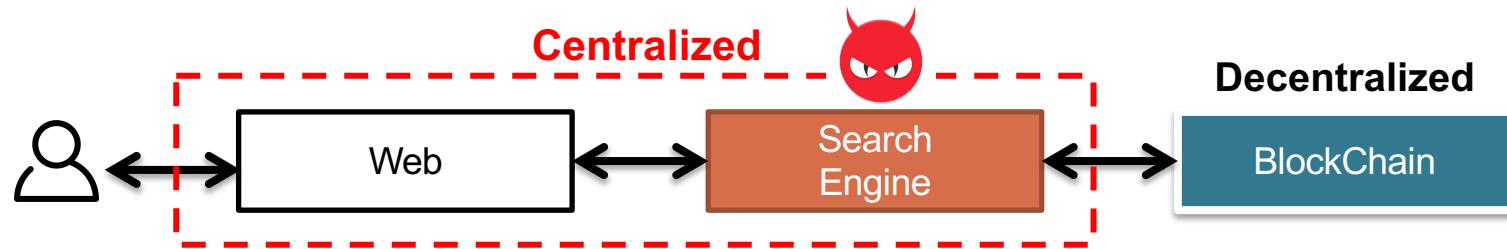
Why centralized search not fit?



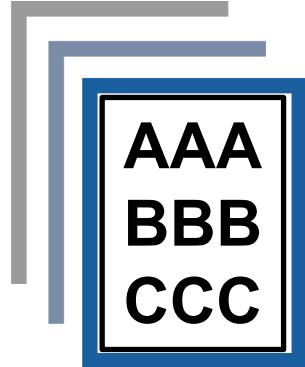
Why centralized search not fit?



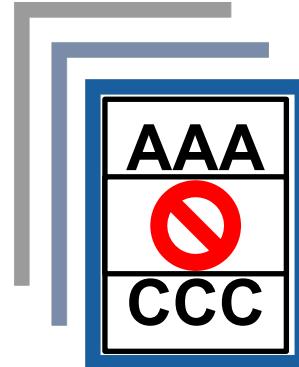
Why centralized search not fit?



Expected

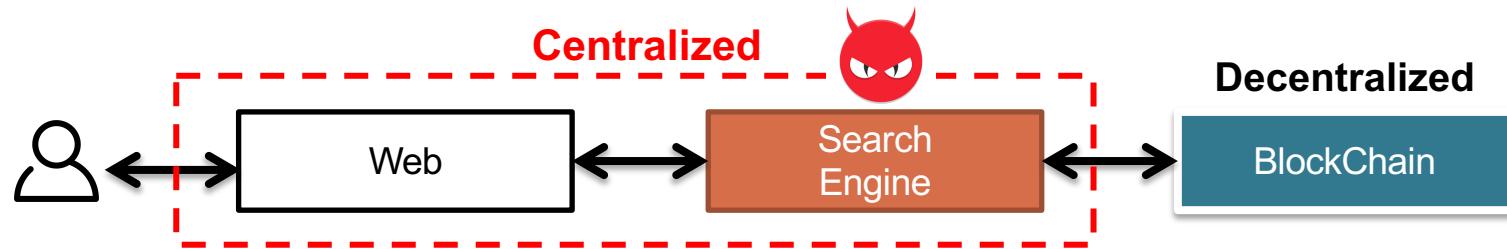


Opaque Censorship^[1]

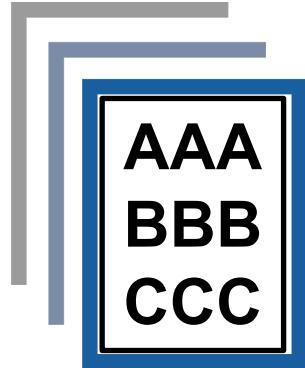


[1] Steemit censoring users on immutable social media blockchain's front-end.
<https://cryptoslate.com/steemit-censoring-users-immutable-blockchainsocial-media/>

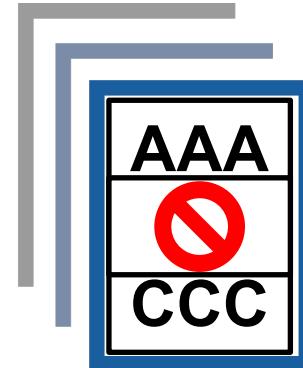
Why centralized search not fit?



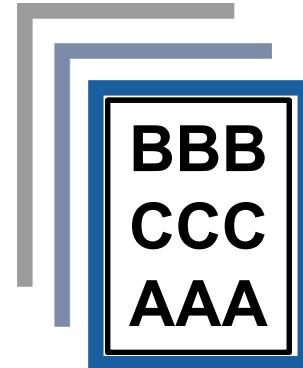
Expected



Opaque Censorship^[1]



Paid/Bias Listings



[1] Steemit censoring users on immutable social media blockchain's front-end.
<https://cryptoslate.com/steemit-censoring-users-immutable-blockchainsocial-media/>

Why not build a decentralized search engine for decentralized apps?

Why not build a decentralized search engine for decentralized apps?



DeSearch

Search Pipeline

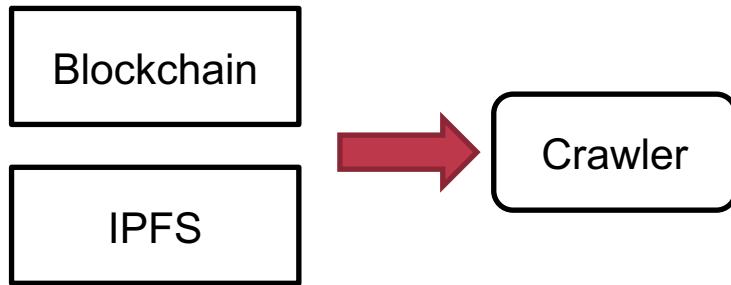
Public data source

Blockchain

IPFS

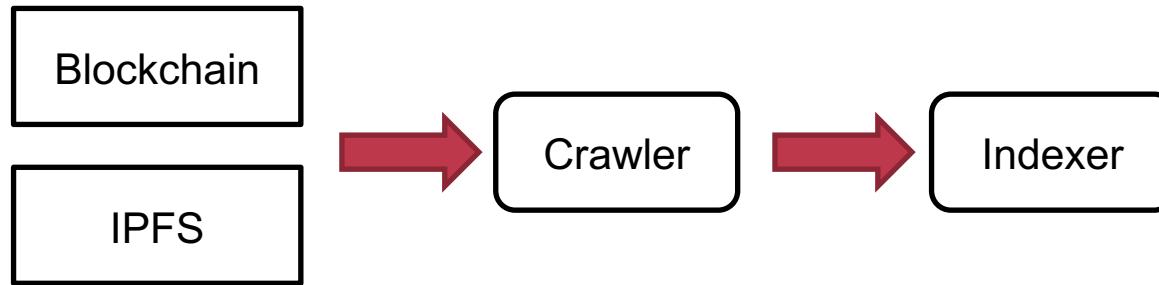
Search Pipeline

Public data source



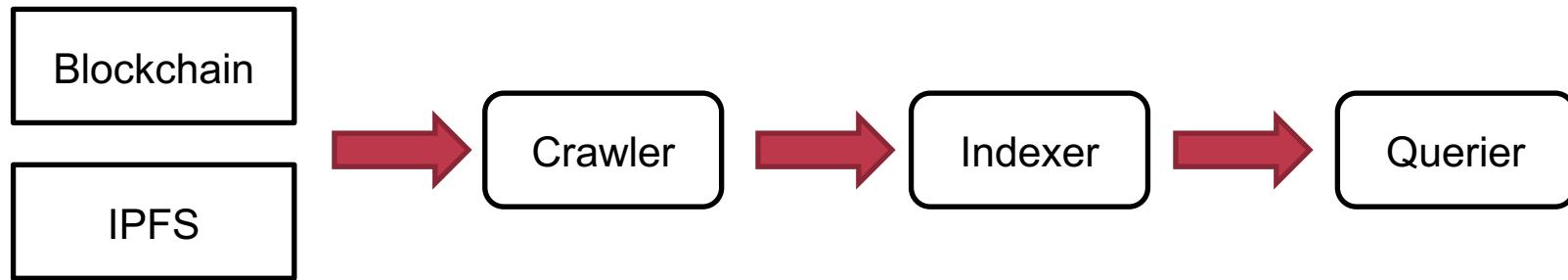
Search Pipeline

Public data source



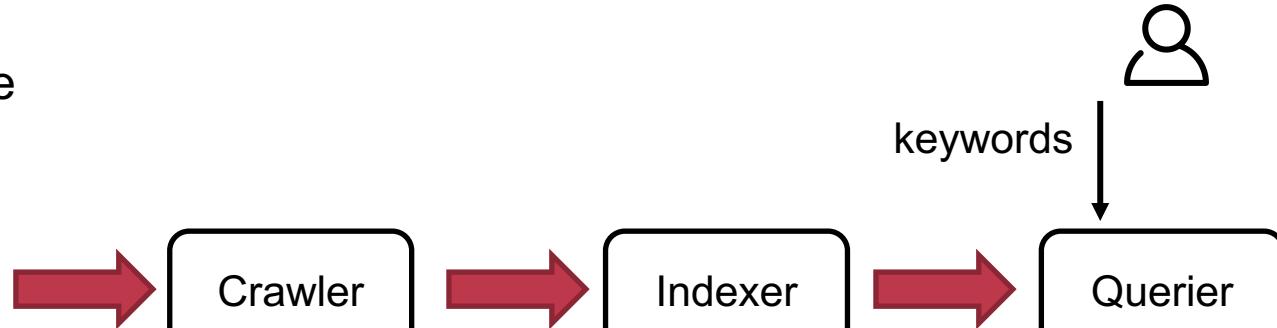
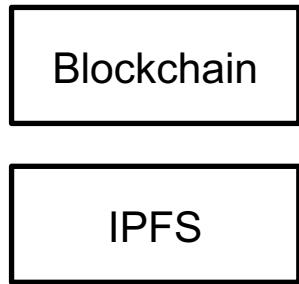
Search Pipeline

Public data source



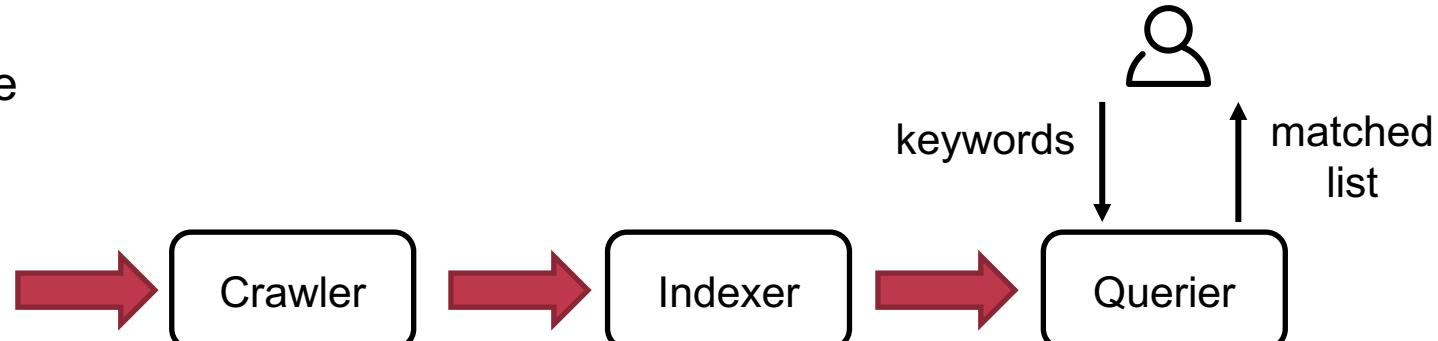
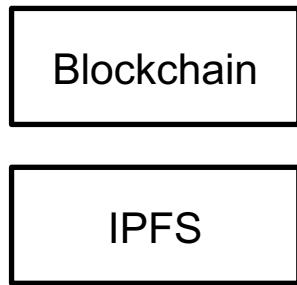
Search Pipeline

Public data source



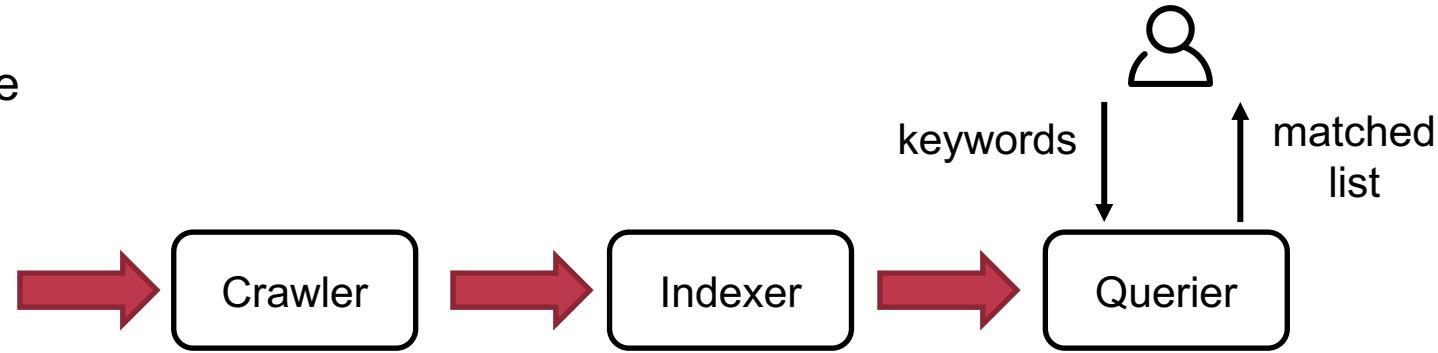
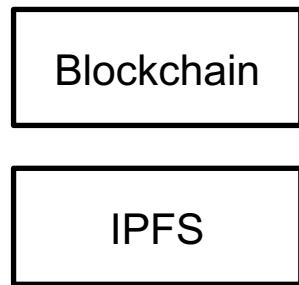
Search Pipeline

Public data source



Search Pipeline

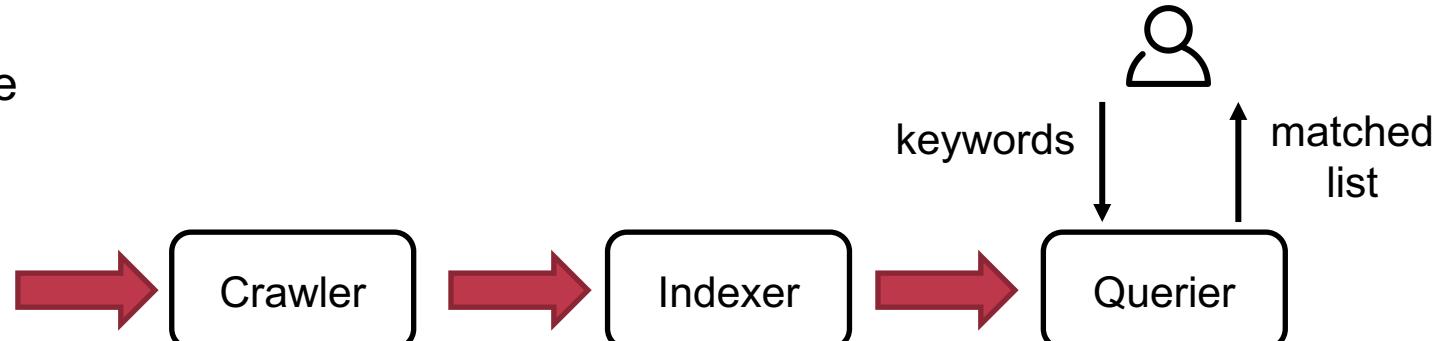
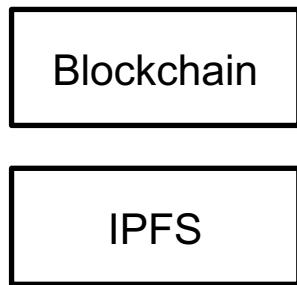
Public data source



Can we use smart contracts?

Search Pipeline

Public data source

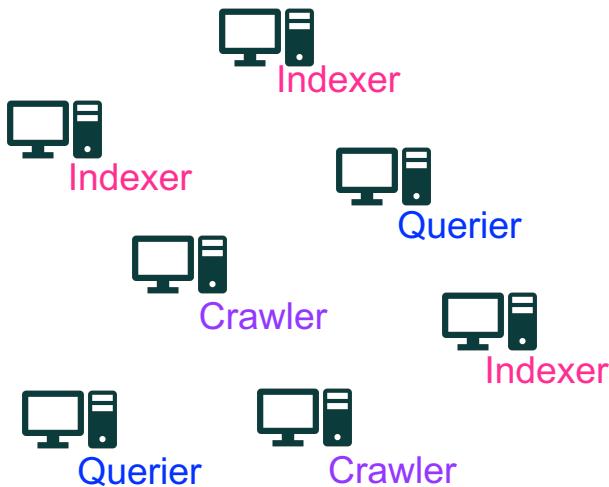


Can we use smart contracts?

Not practical <= long search latency

DeSearch Challenges

Decentralized P2P Network



How to ensure state availability?

DeSearch Challenges

Decentralized P2P Network



Querier



Querier

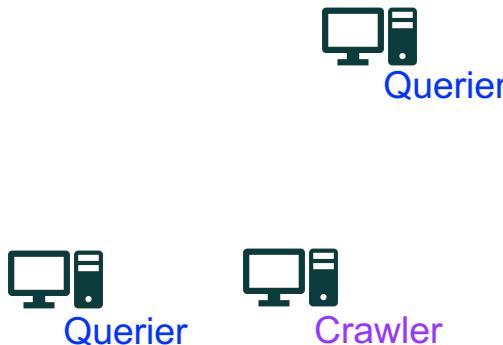


Crawler

How to ensure state availability?

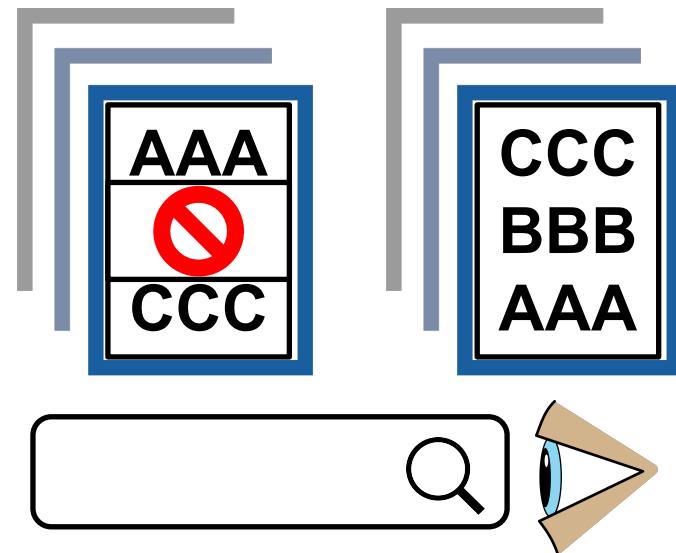
DeSearch Challenges

Decentralized P2P Network



How to ensure state availability?

Untrusted Environments



How to detect dishonest behaviors?

Outline

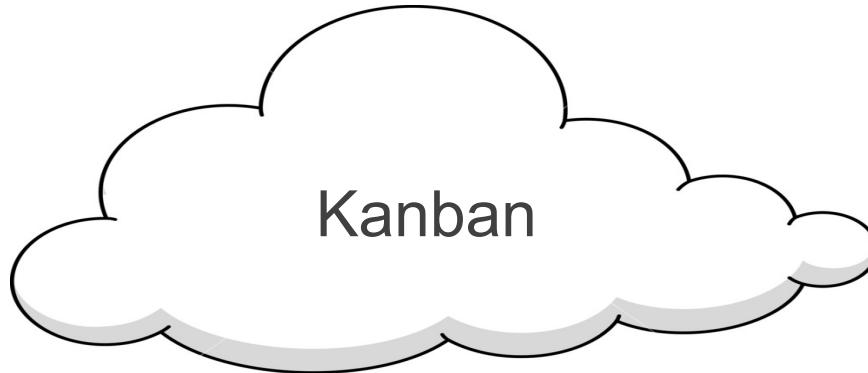
- **Problem Statement**
- **DeSearch Design: decentralized search**
 - how to ensure state availability?
- **Design: verifiable search**
- **Design: private search**
- **Evaluation**

Decouple states from computation

Kanban

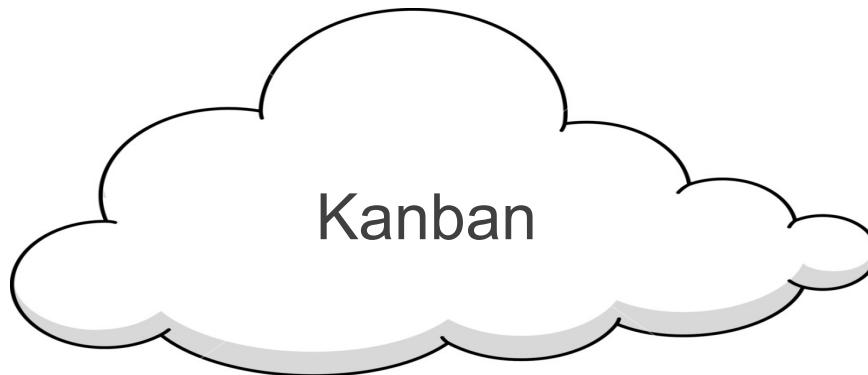
States with high availability

Decouple states from computation

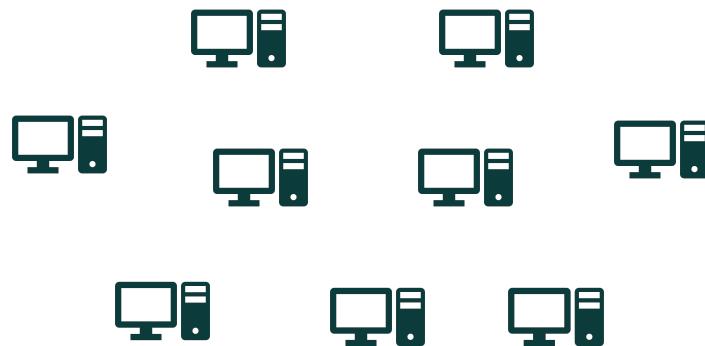


States with high availability

Decouple states from computation



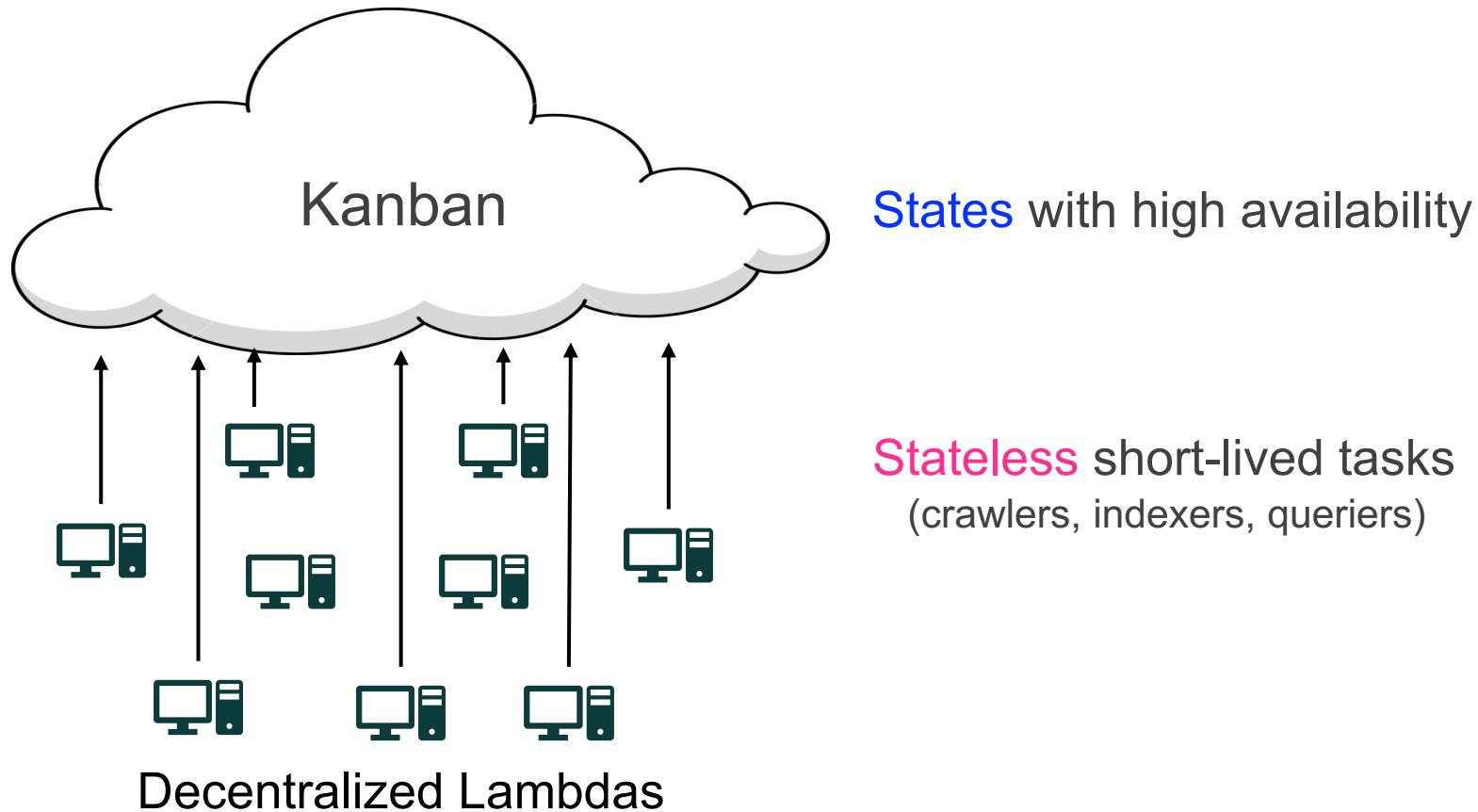
States with high availability



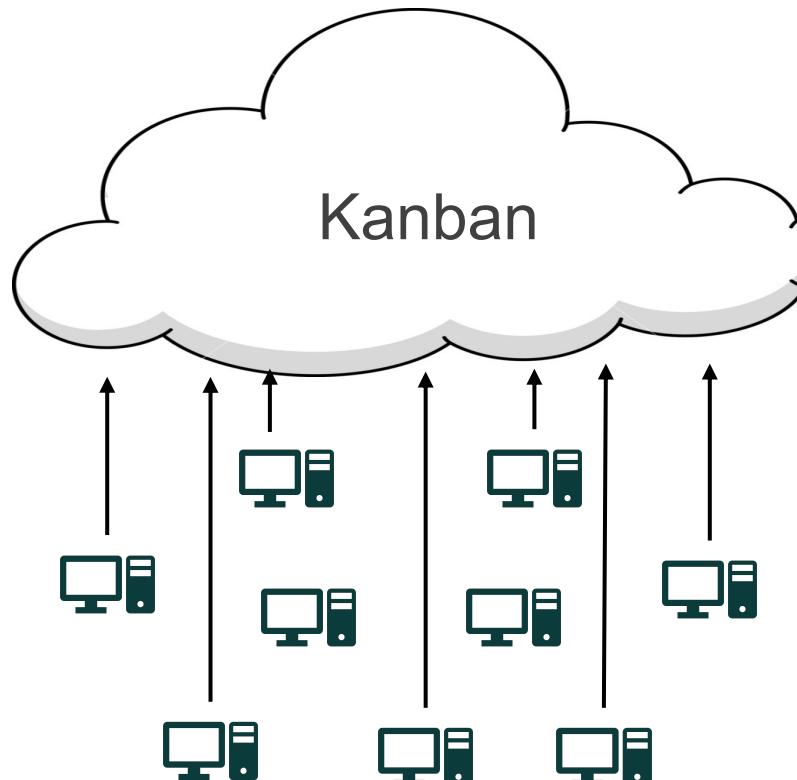
Stateless short-lived tasks
(crawlers, indexers, queriers)

Decentralized Lambdas

Decouple states from computation



Decouple states from computation



States with high availability

Stateless short-lived tasks
(crawlers, indexers, queriers)



good scalability



good fault tolerance

Decentralized Lambdas

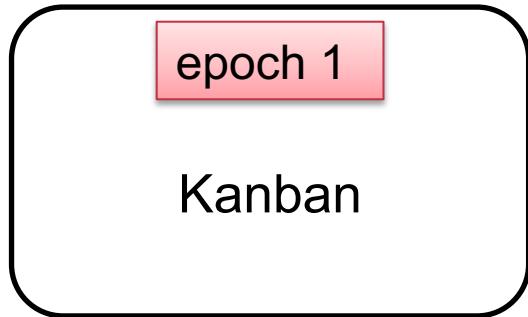
Outline

- Problem Statement
- DeSearch Design: decentralized search
- DeSearch Design: verifiable search
 - Verifiable Kanban
 - Verifiable lambda
- Evaluation

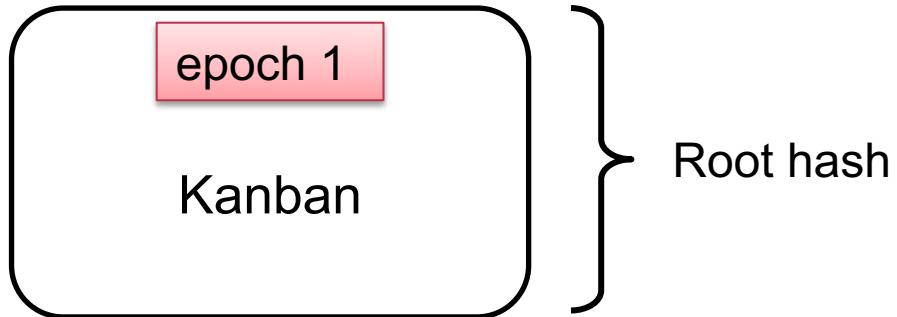
Verifiable Kanban

Kanban

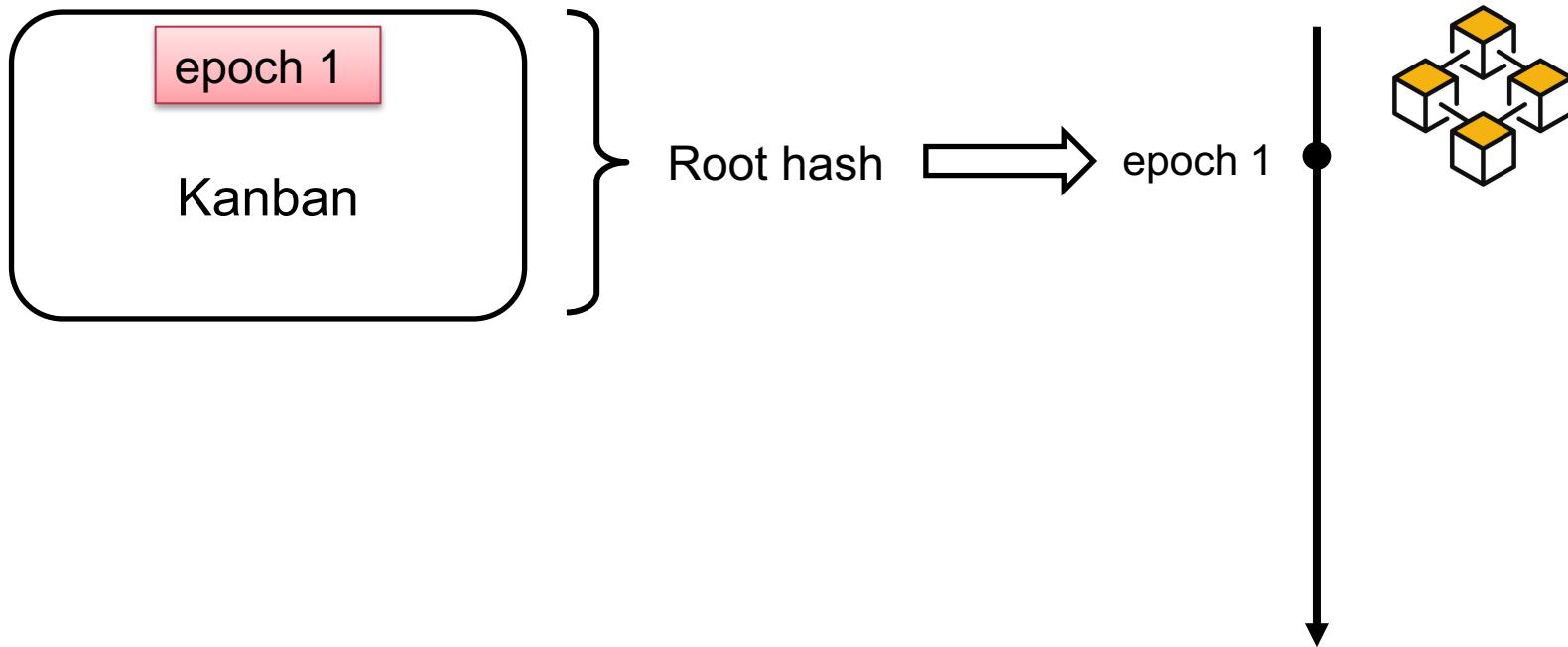
Verifiable Kanban



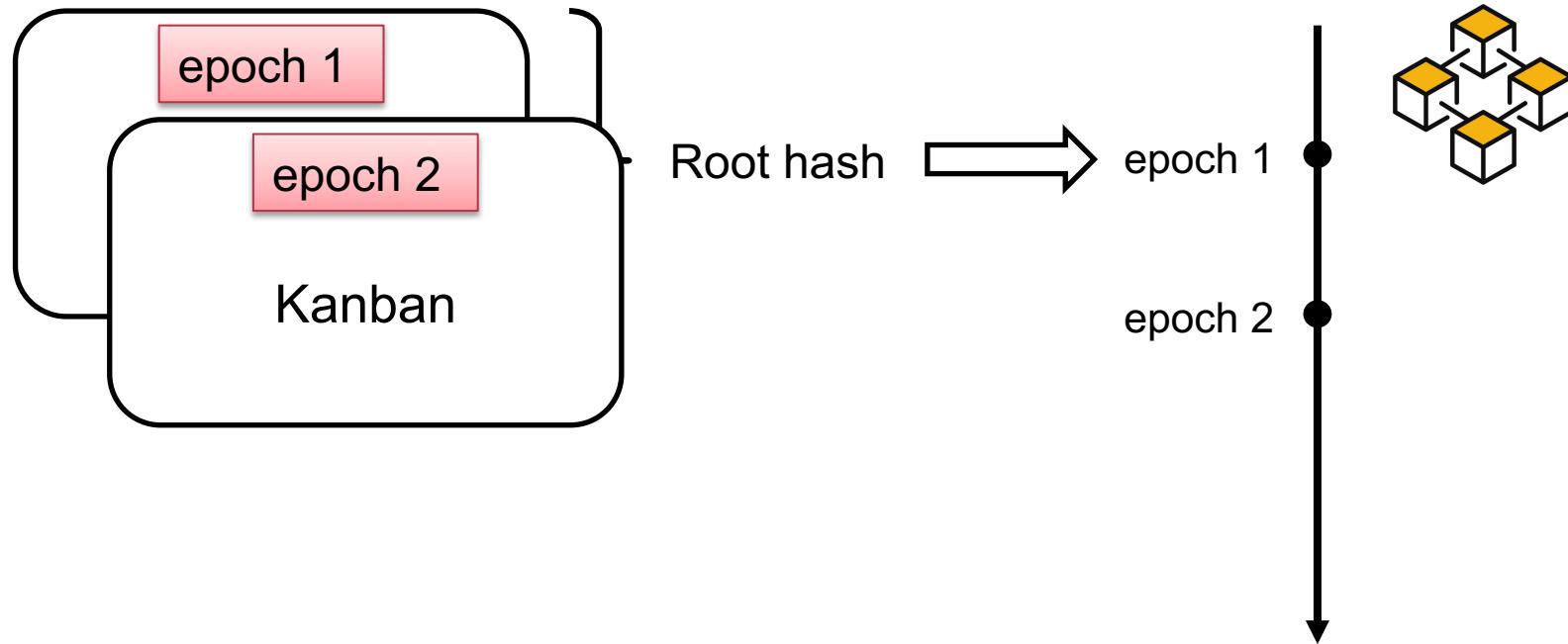
Verifiable Kanban



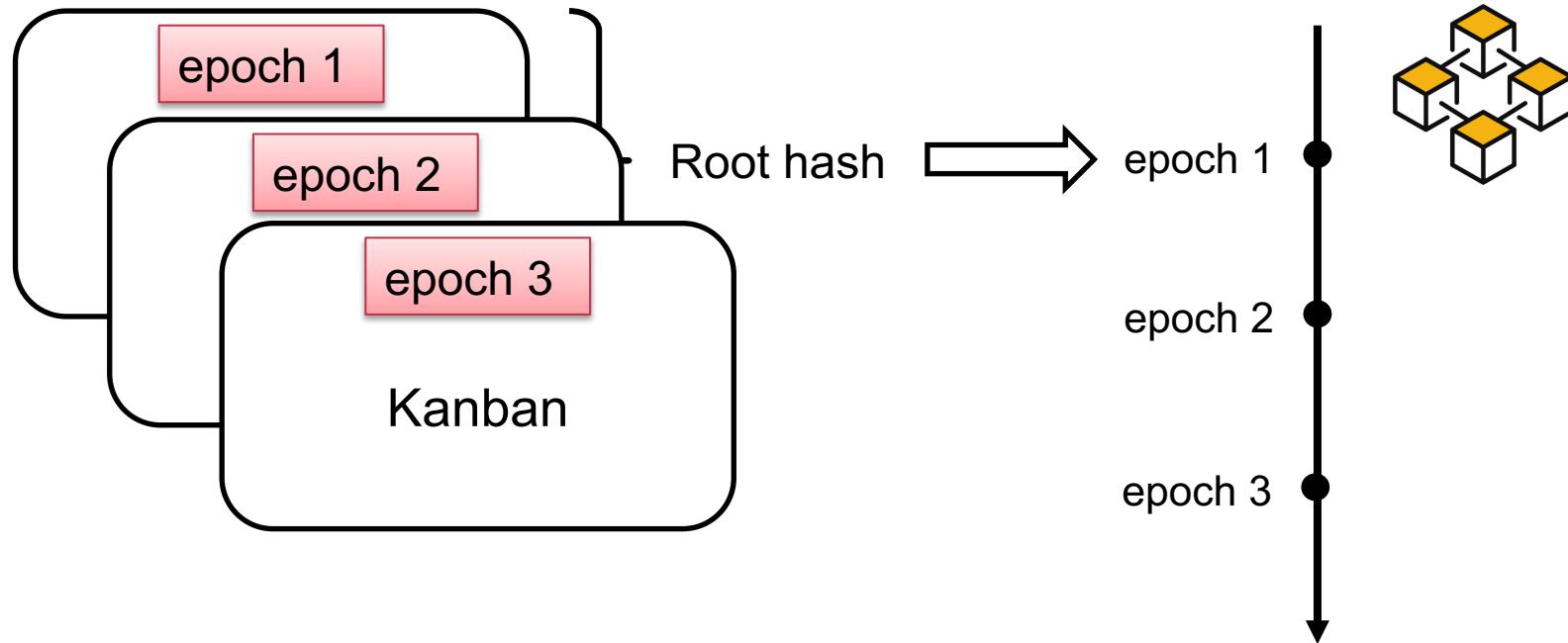
Verifiable Kanban



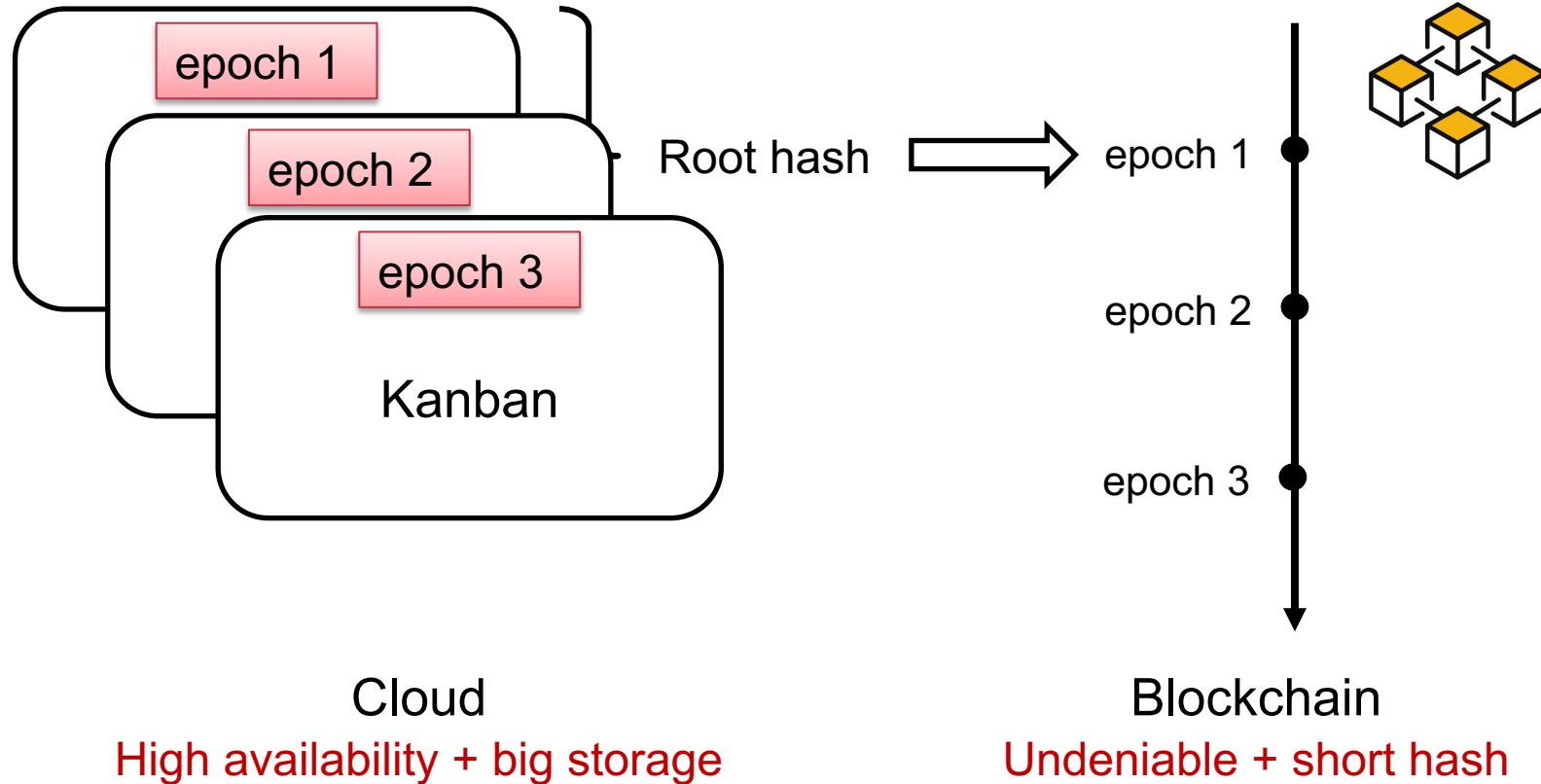
Verifiable Kanban



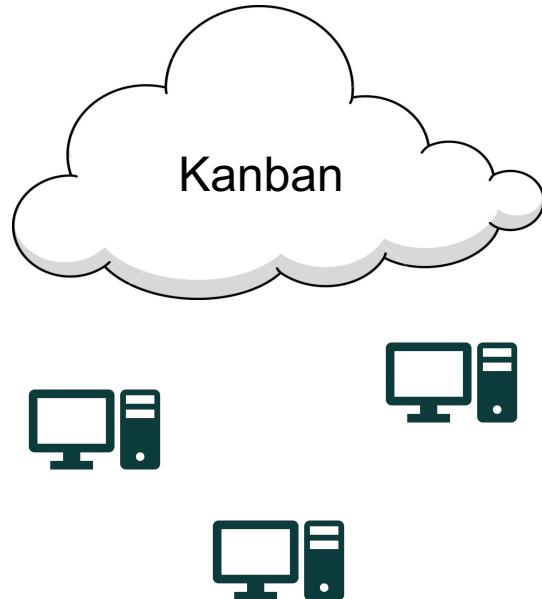
Verifiable Kanban



Verifiable Kanban

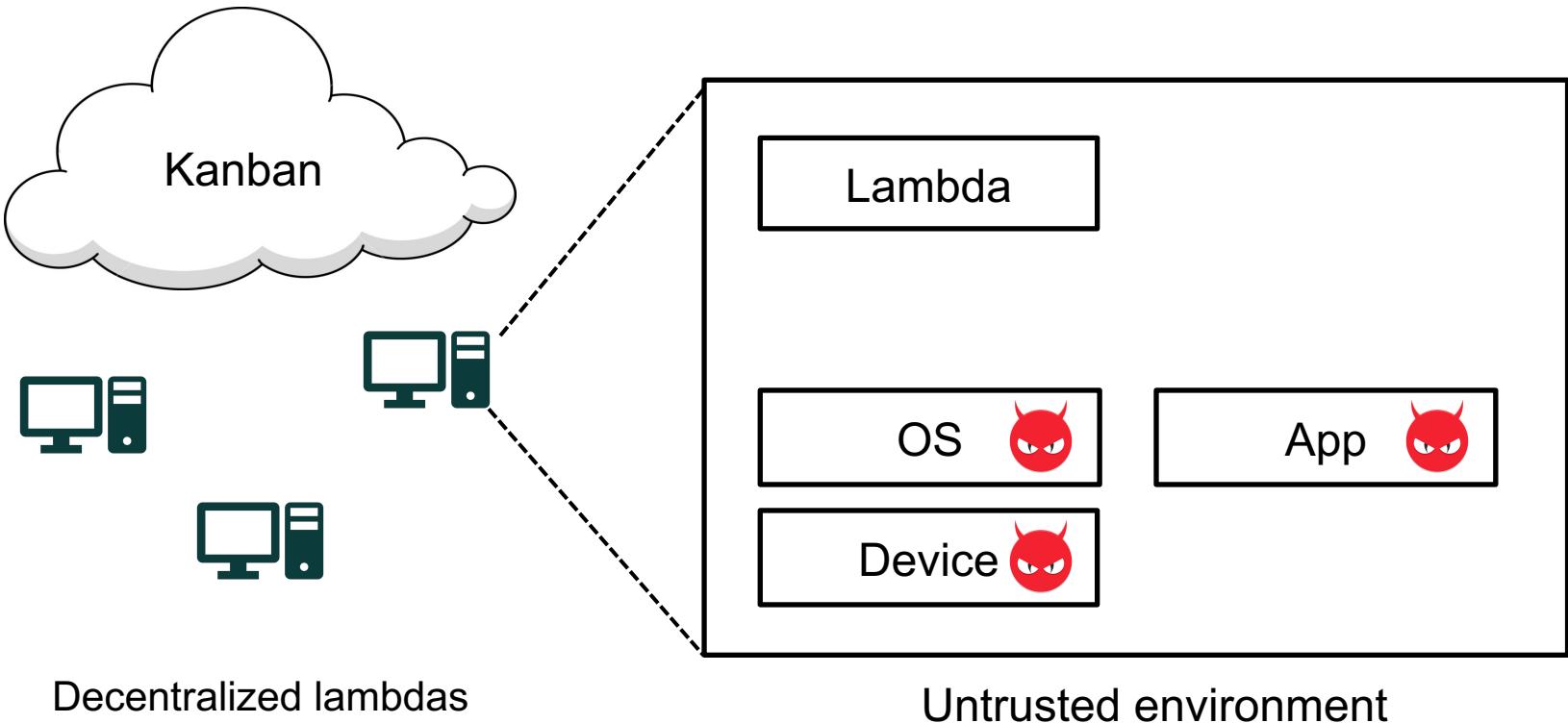


TEE-based Lambda

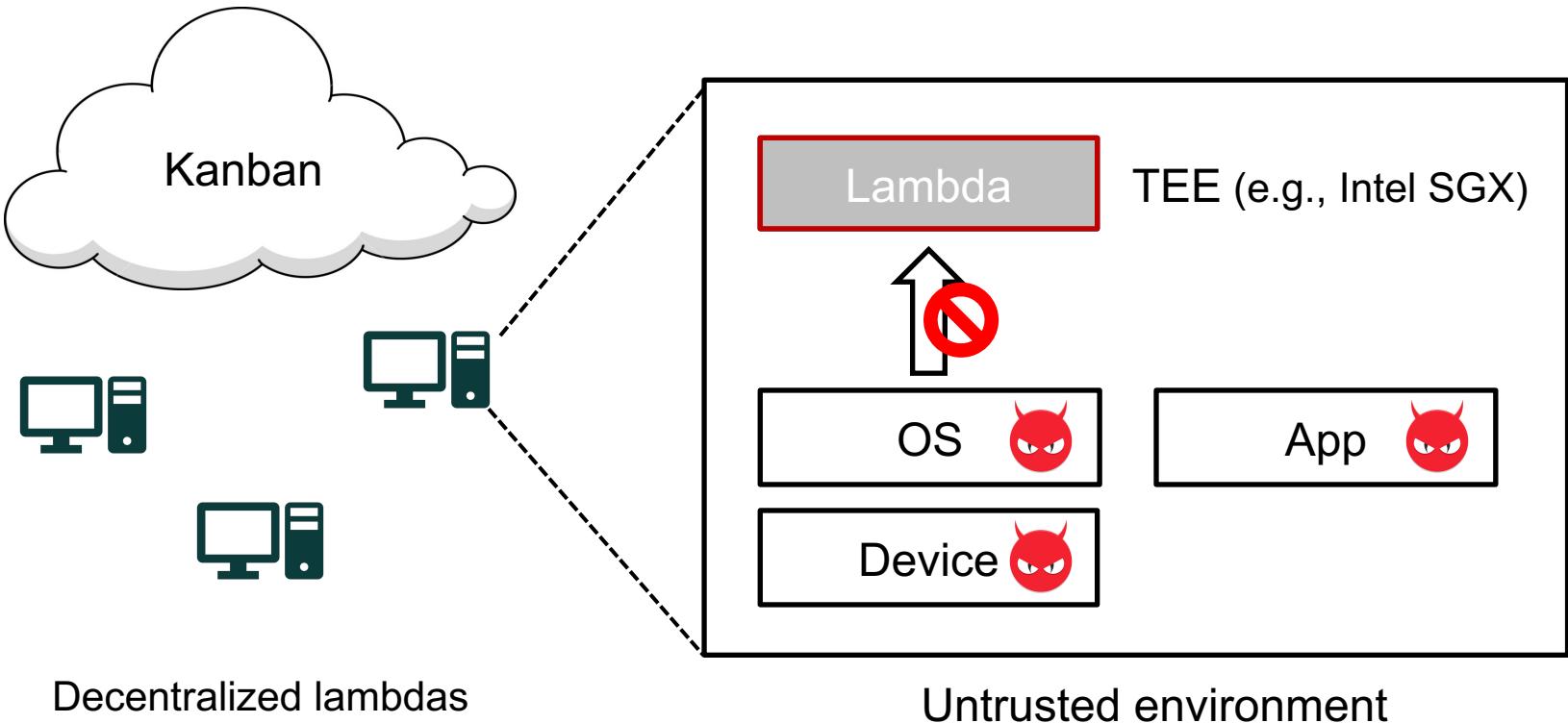


Decentralized lambdas

TEE-based Lambda

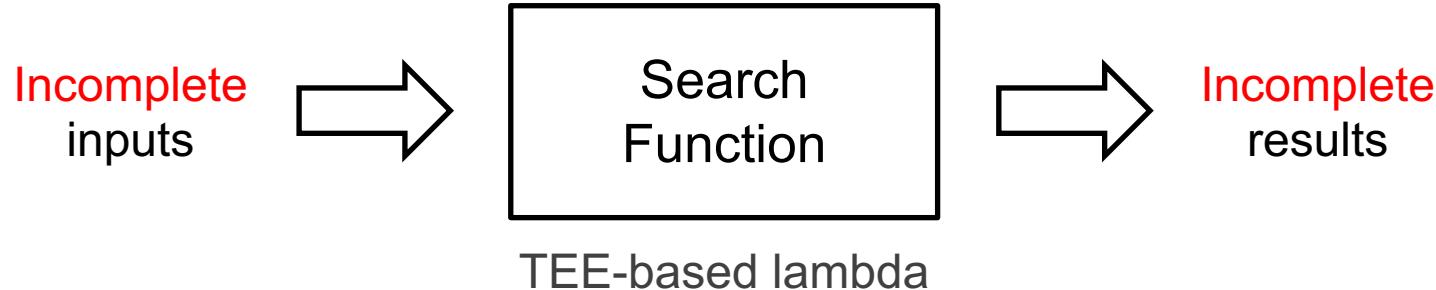


TEE-based Lambda



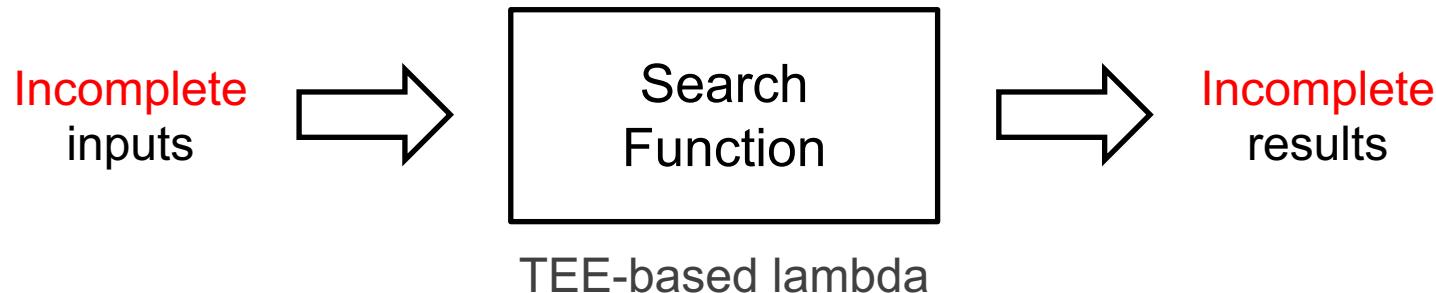
Why TEE is not enough?

- Execution integrity \neq Data integrity



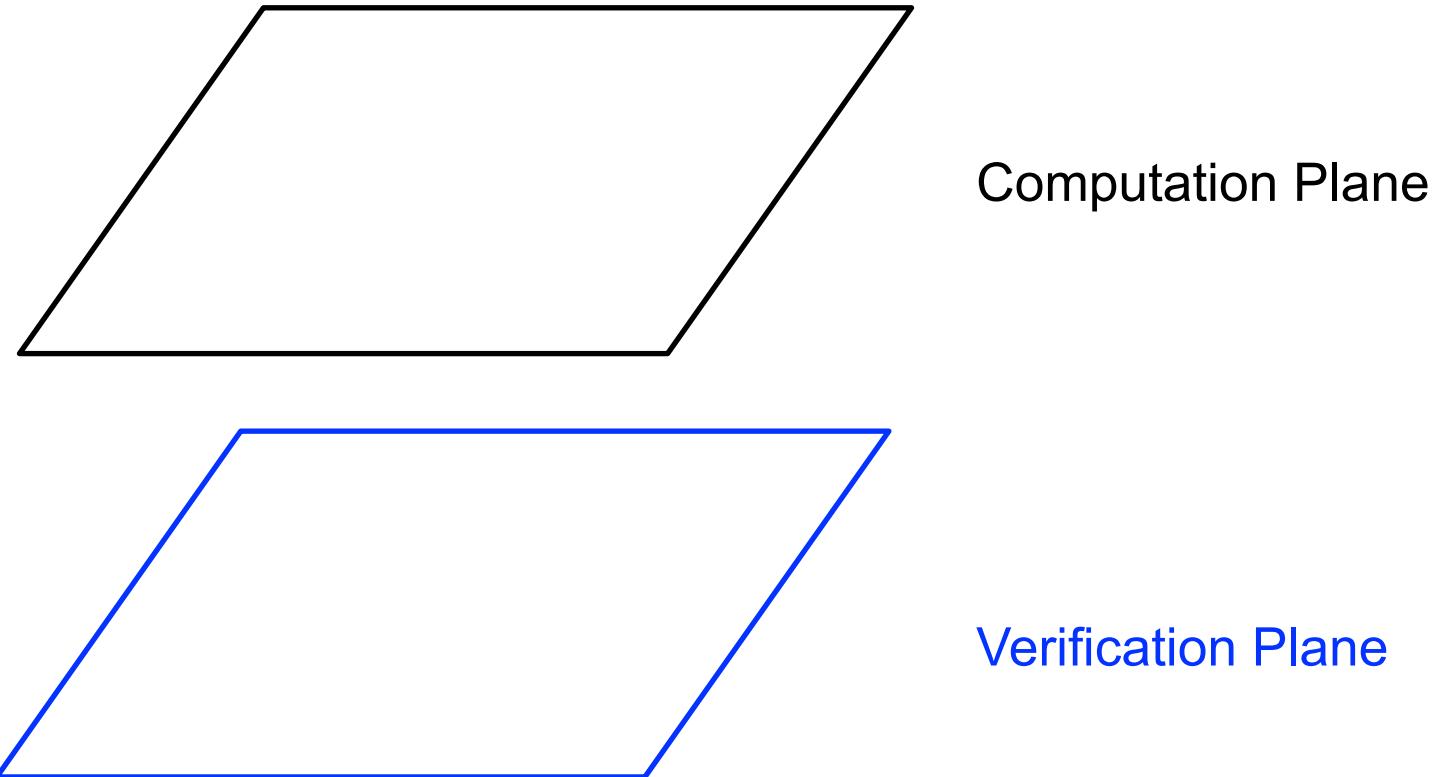
Why TEE is not enough?

- Execution integrity \neq Data integrity

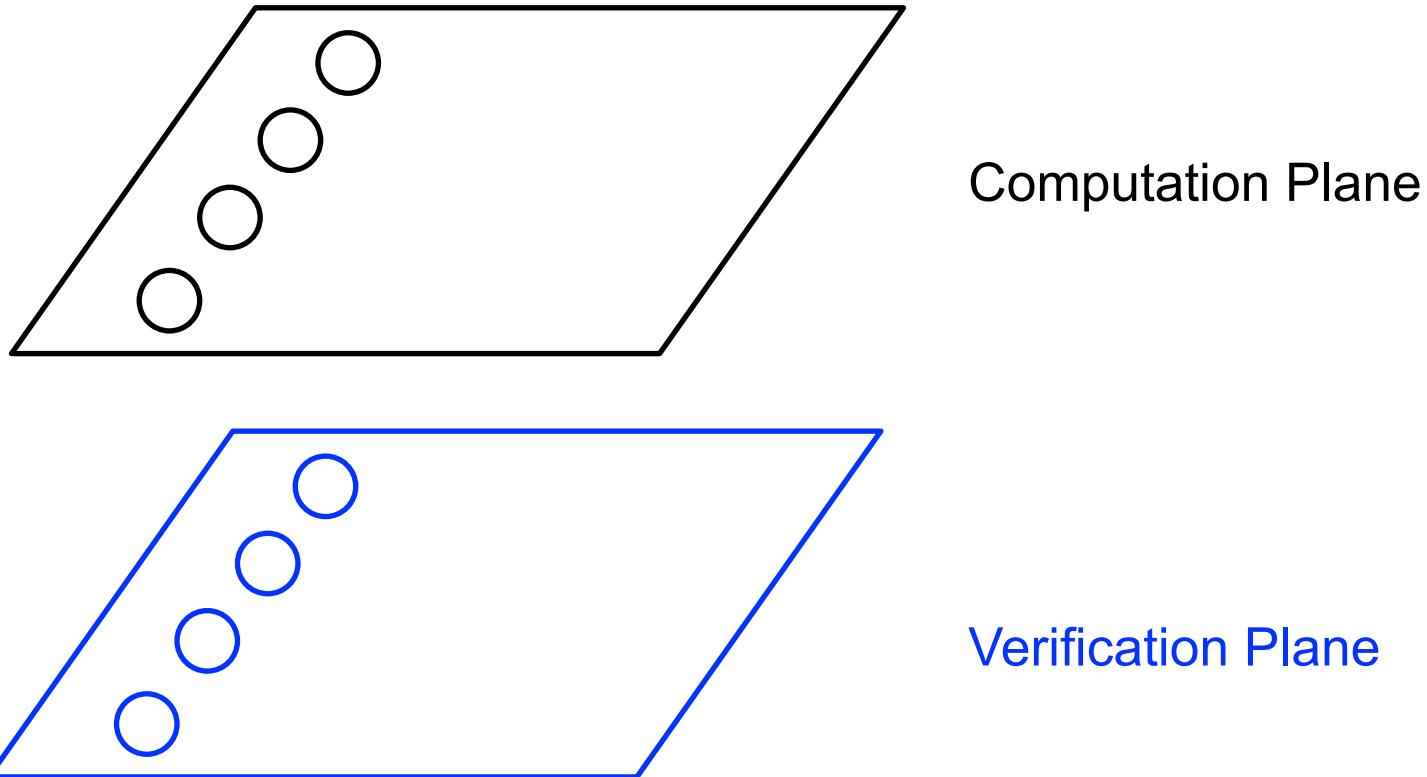


- Put verification code into lambdas?
 - Lambdas are **ephemeral** (verification time >> execution time)
 - Lambdas are **scattered** (lacking a global view)

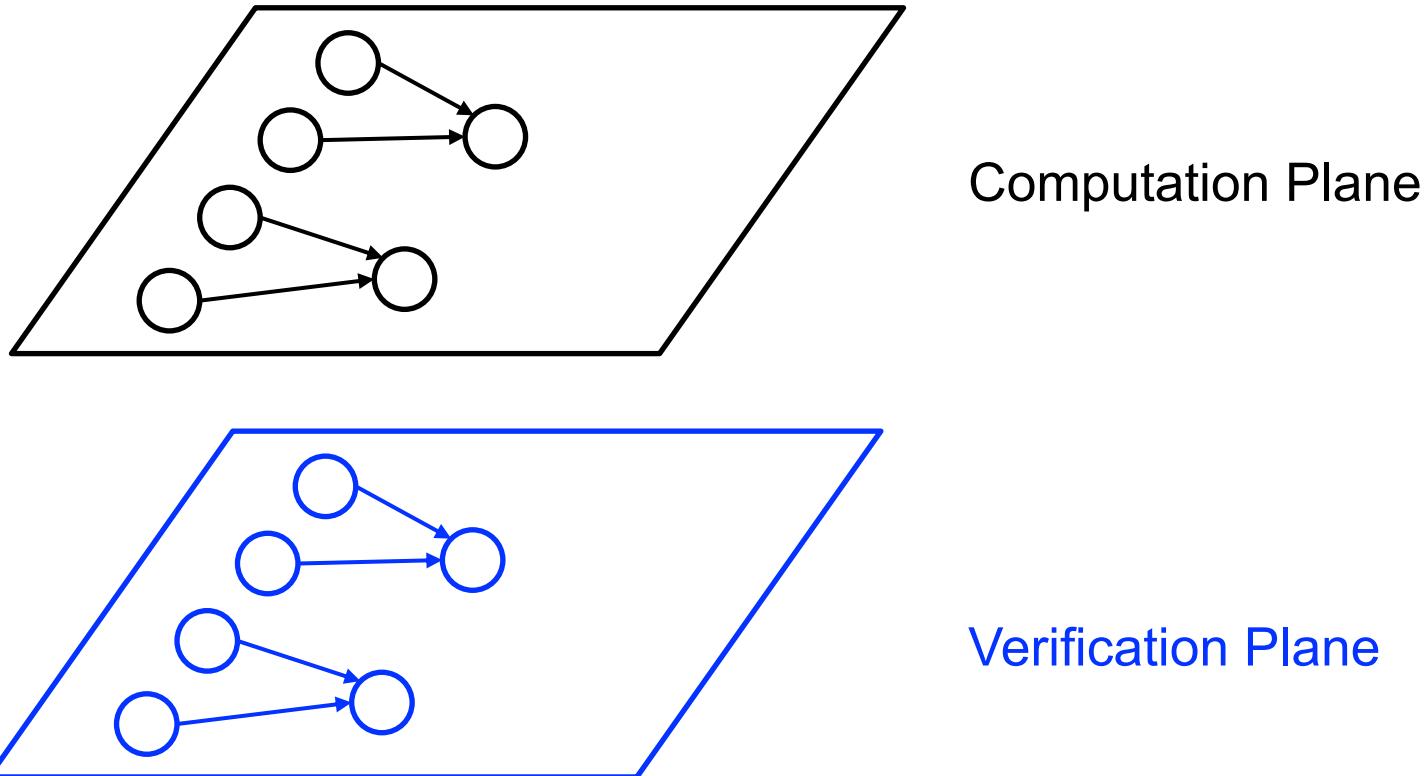
Decouple verification from computation



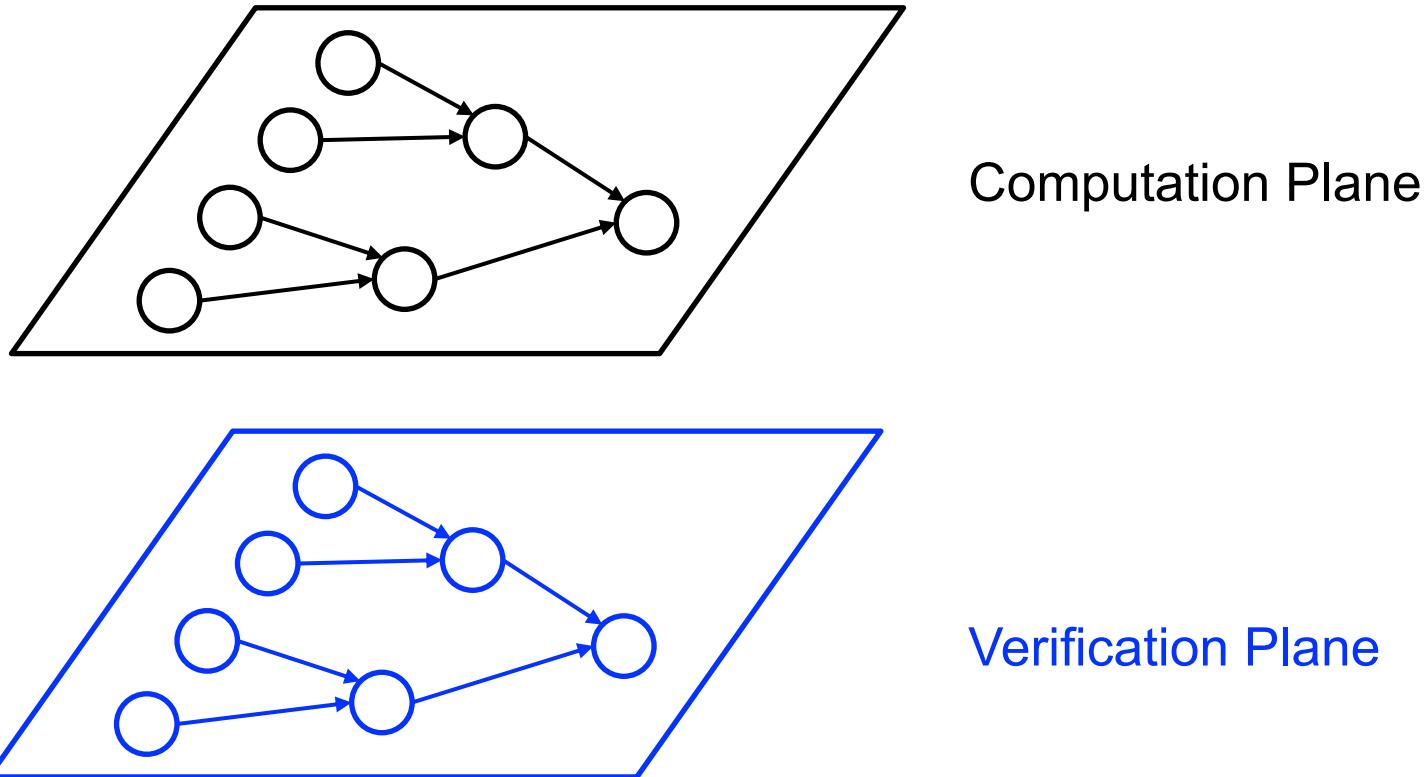
Decouple verification from computation



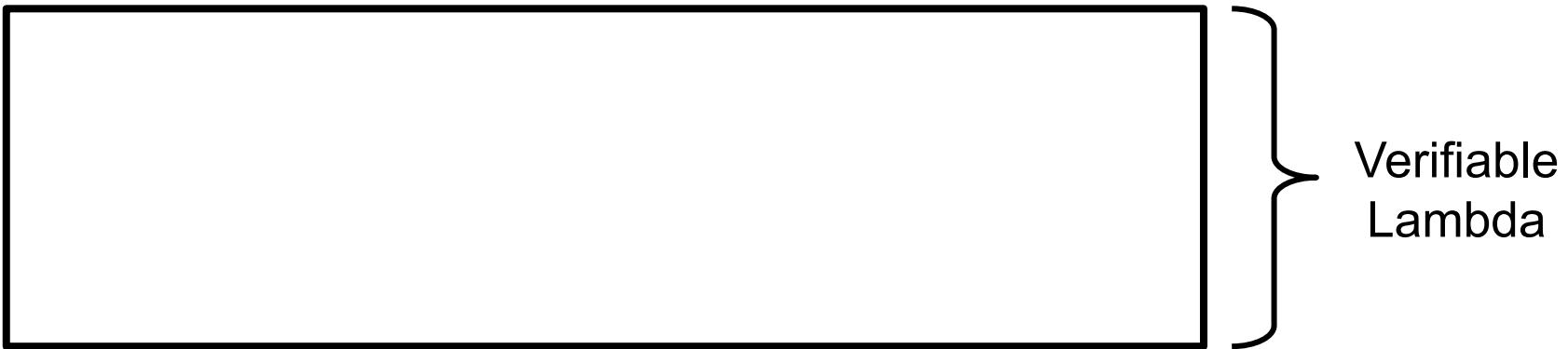
Decouple verification from computation



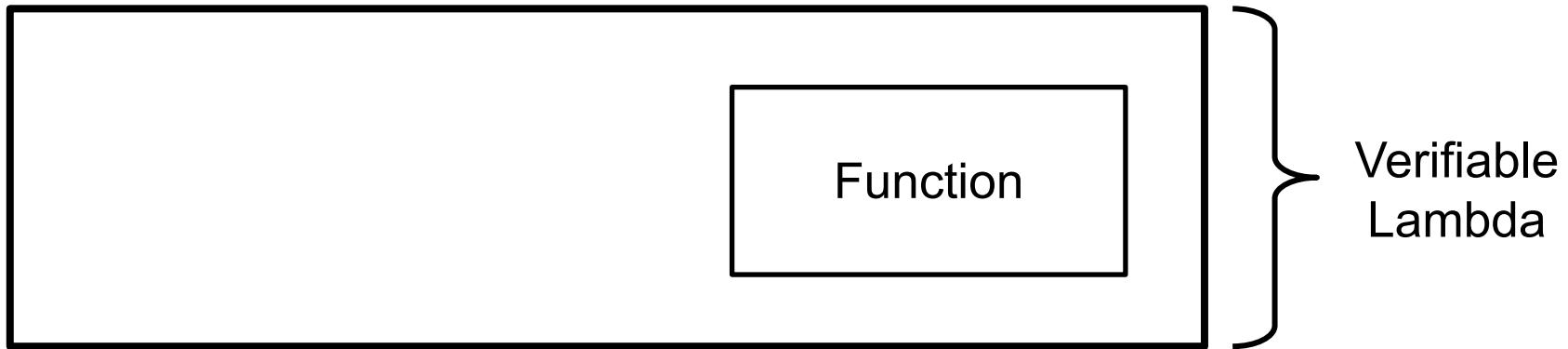
Decouple verification from computation



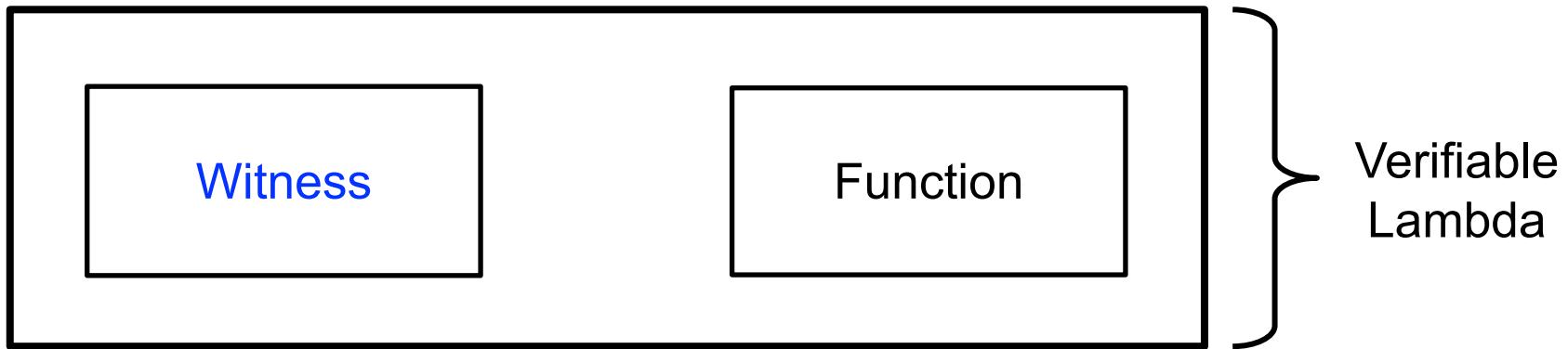
Verifiable Lambda and Witness



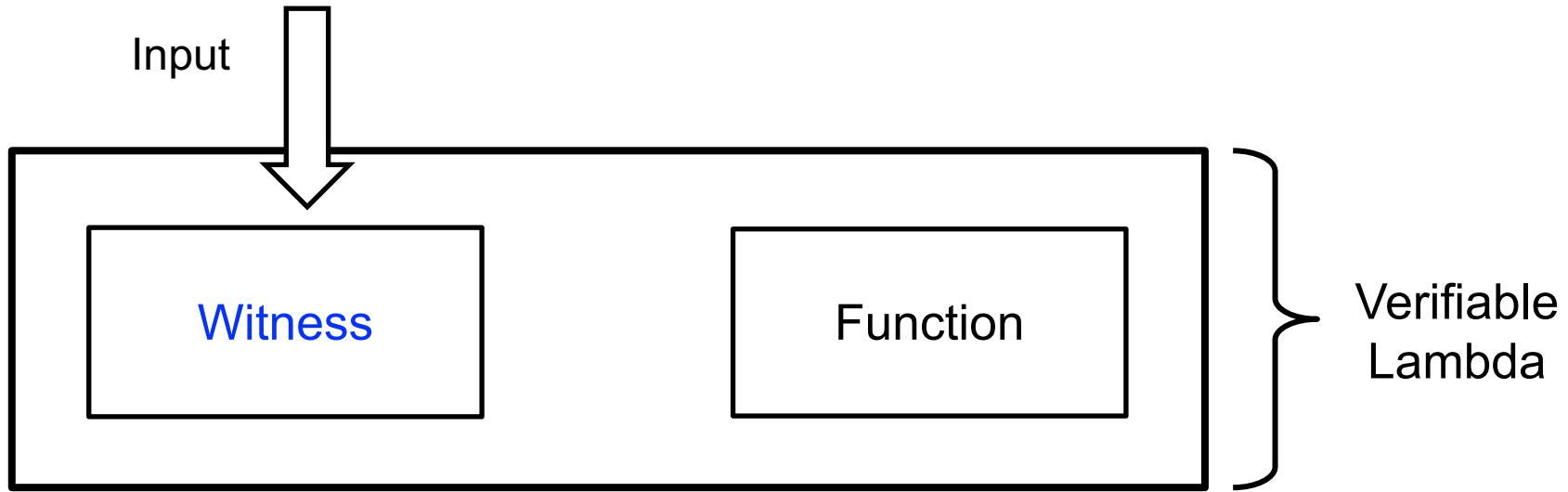
Verifiable Lambda and Witness



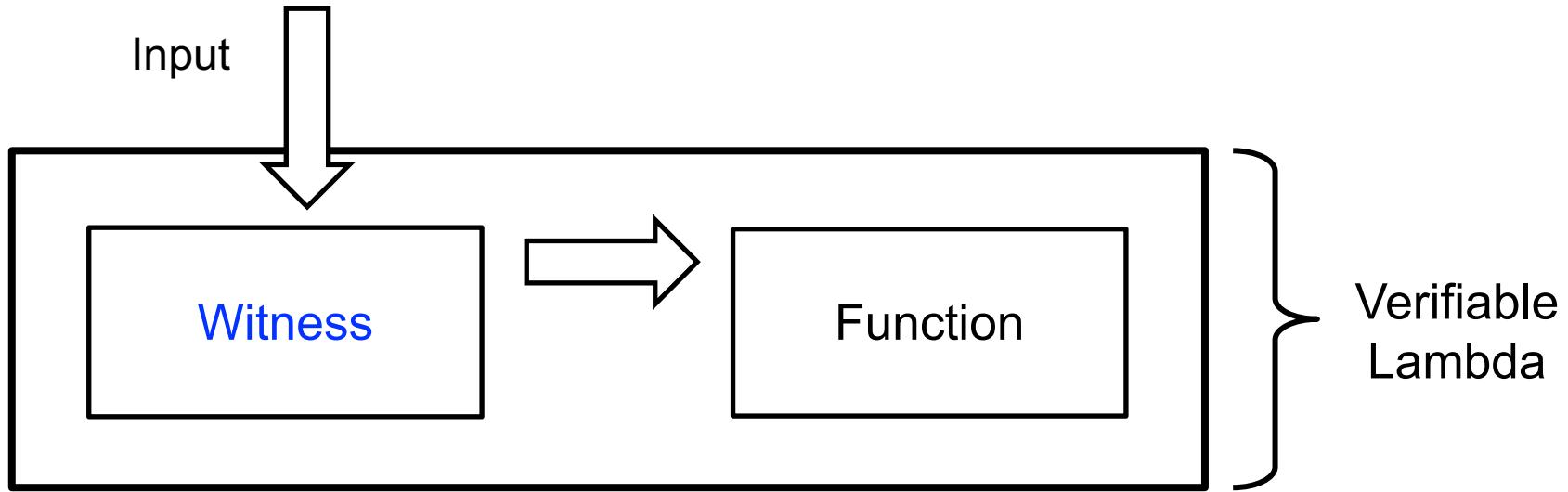
Verifiable Lambda and Witness



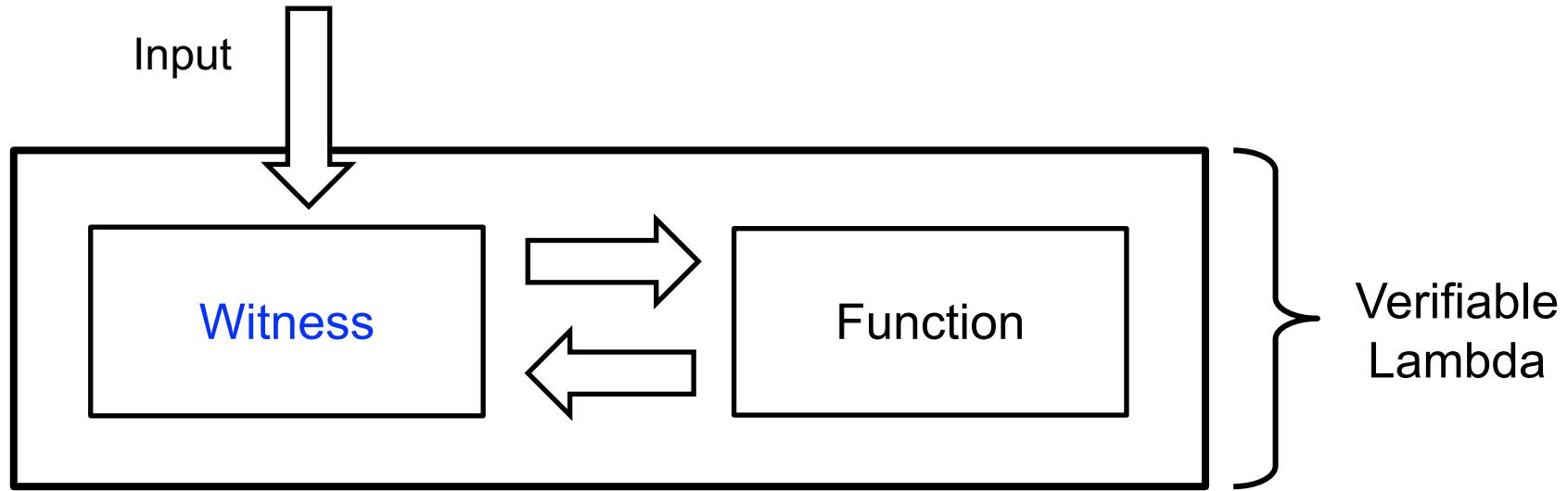
Verifiable Lambda and Witness



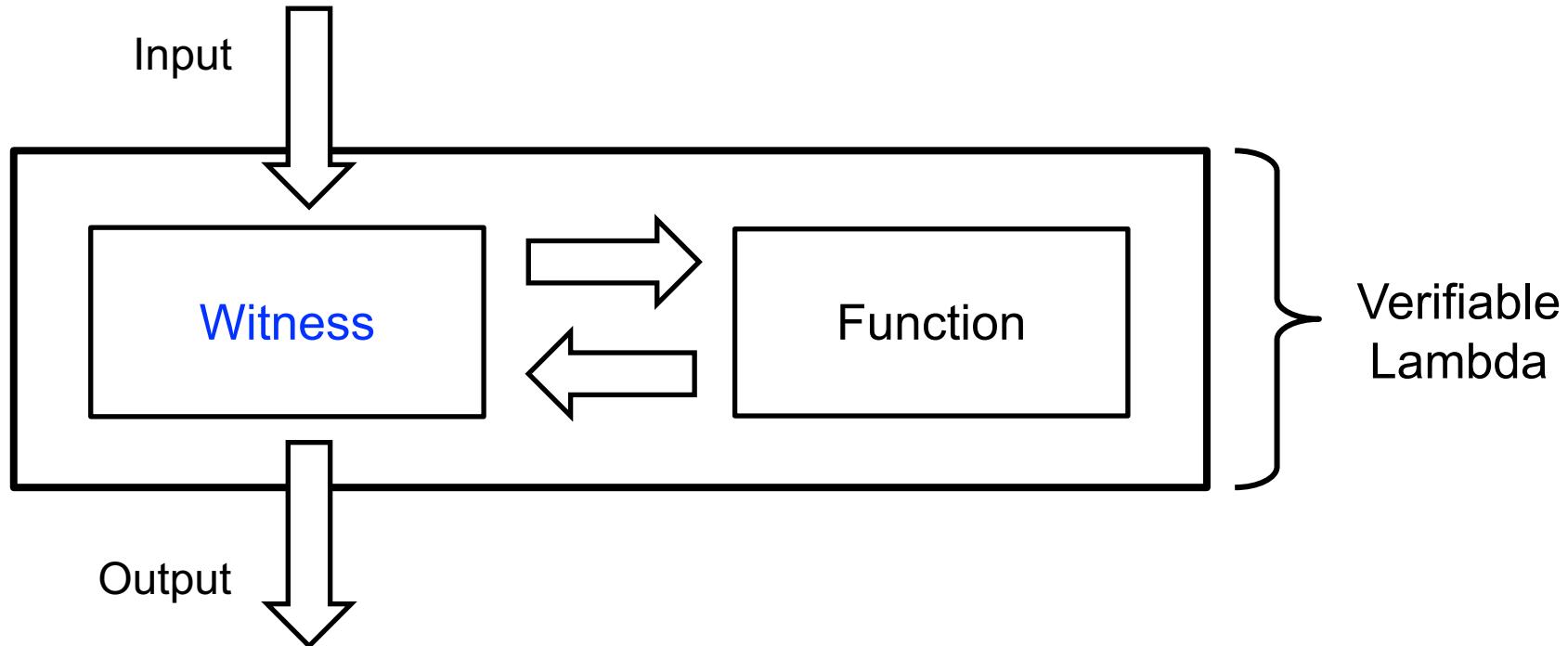
Verifiable Lambda and Witness



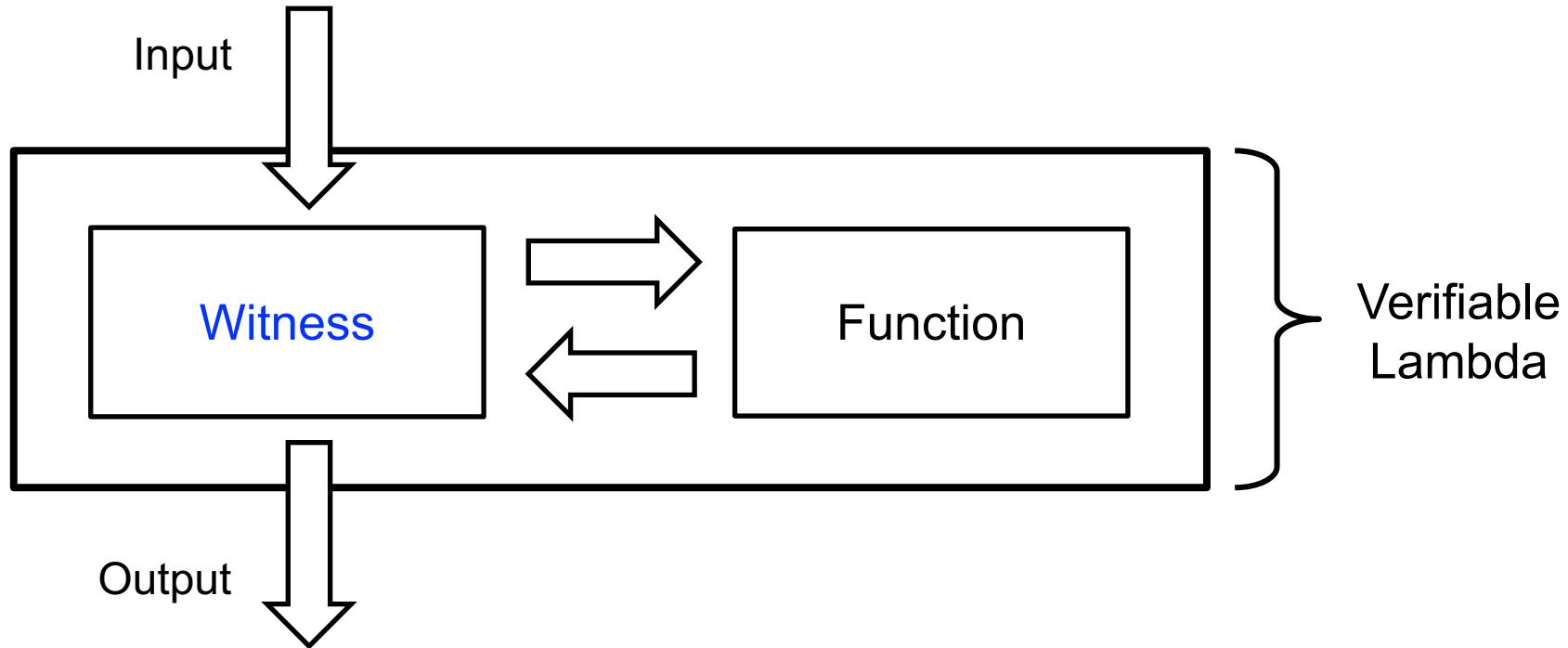
Verifiable Lambda and Witness



Verifiable Lambda and Witness

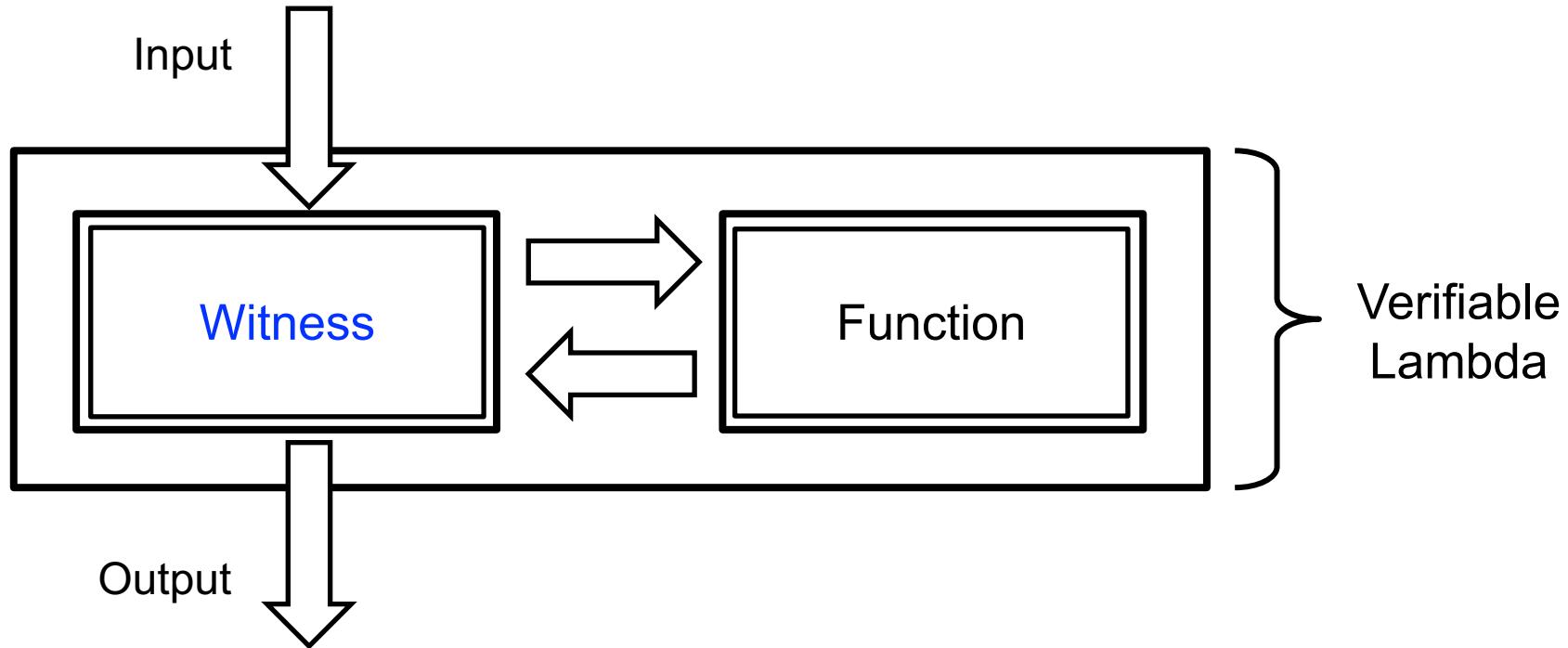


Verifiable Lambda and Witness



$\langle \text{Hash}(\text{Input}), \text{Hash}(\text{Function}), \text{Hash}(\text{Output}) \rangle_{\text{signed}}$

Verifiable Lambda and Witness



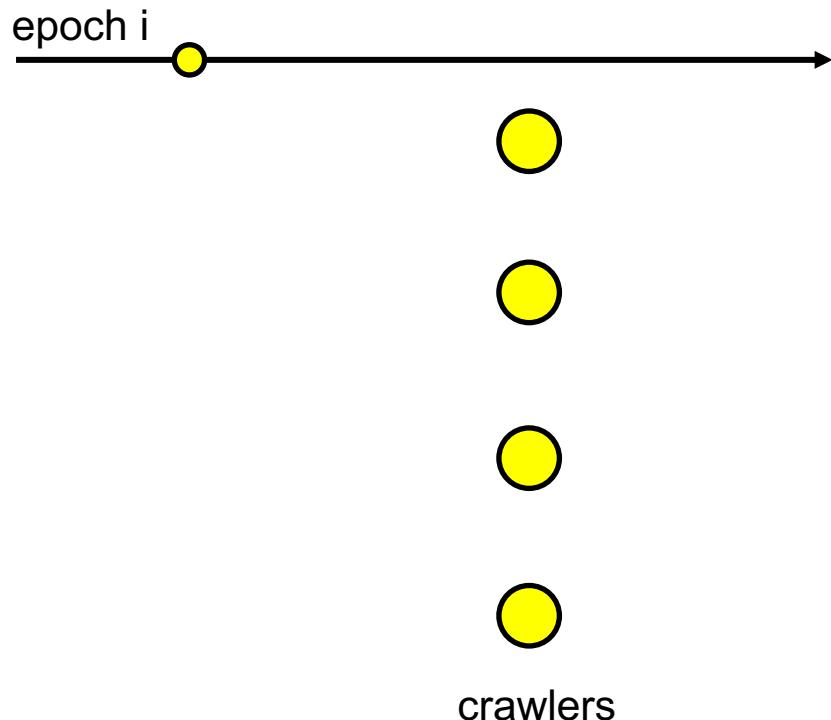
$\langle \text{Hash}(\text{Input}), \text{Hash}(\text{Function}), \text{Hash}(\text{Output}) \rangle_{\text{signed}}$

Providing Verifiable Search

- Combing Lambda witnesses and Kanban epochs

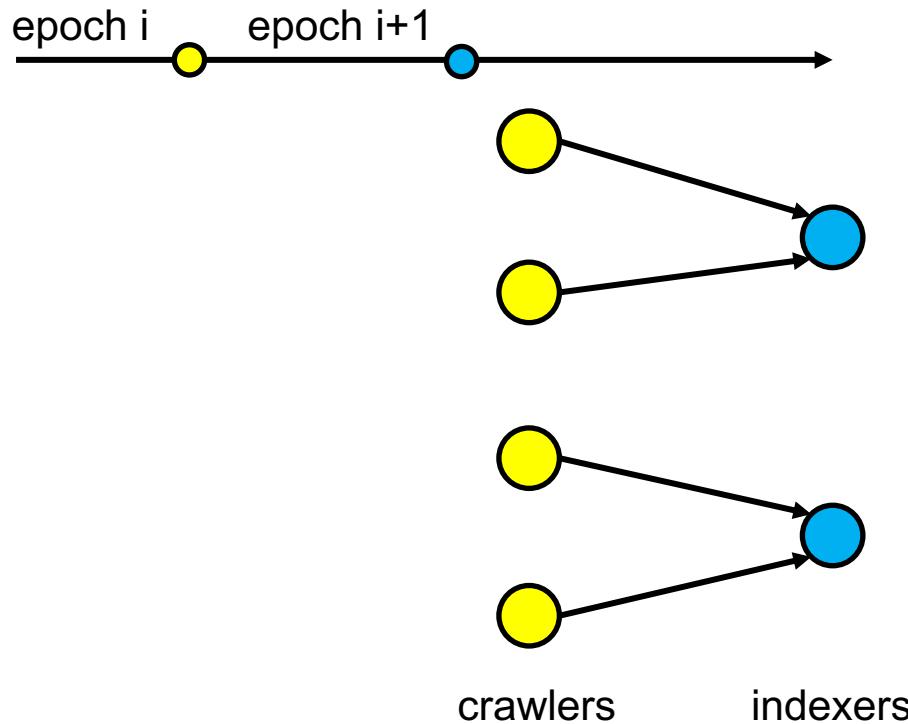
Providing Verifiable Search

- Combing Lambda witnesses and Kanban epochs



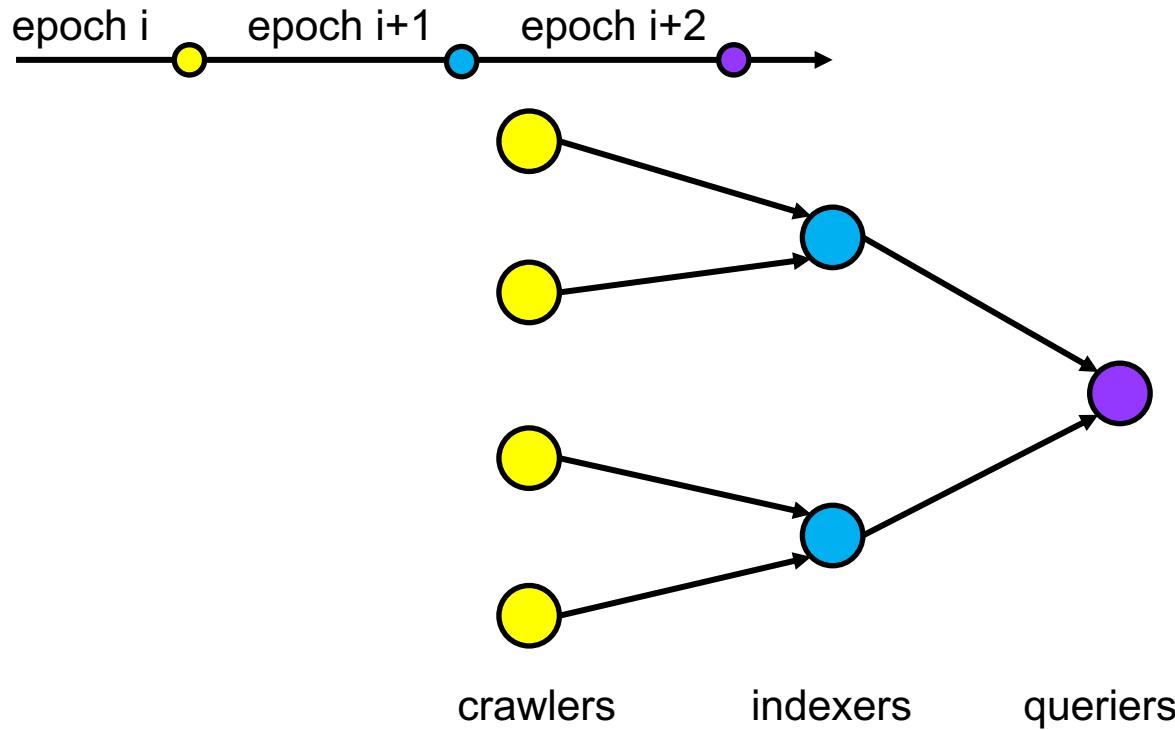
Providing Verifiable Search

- Combing Lambda witnesses and Kanban epochs



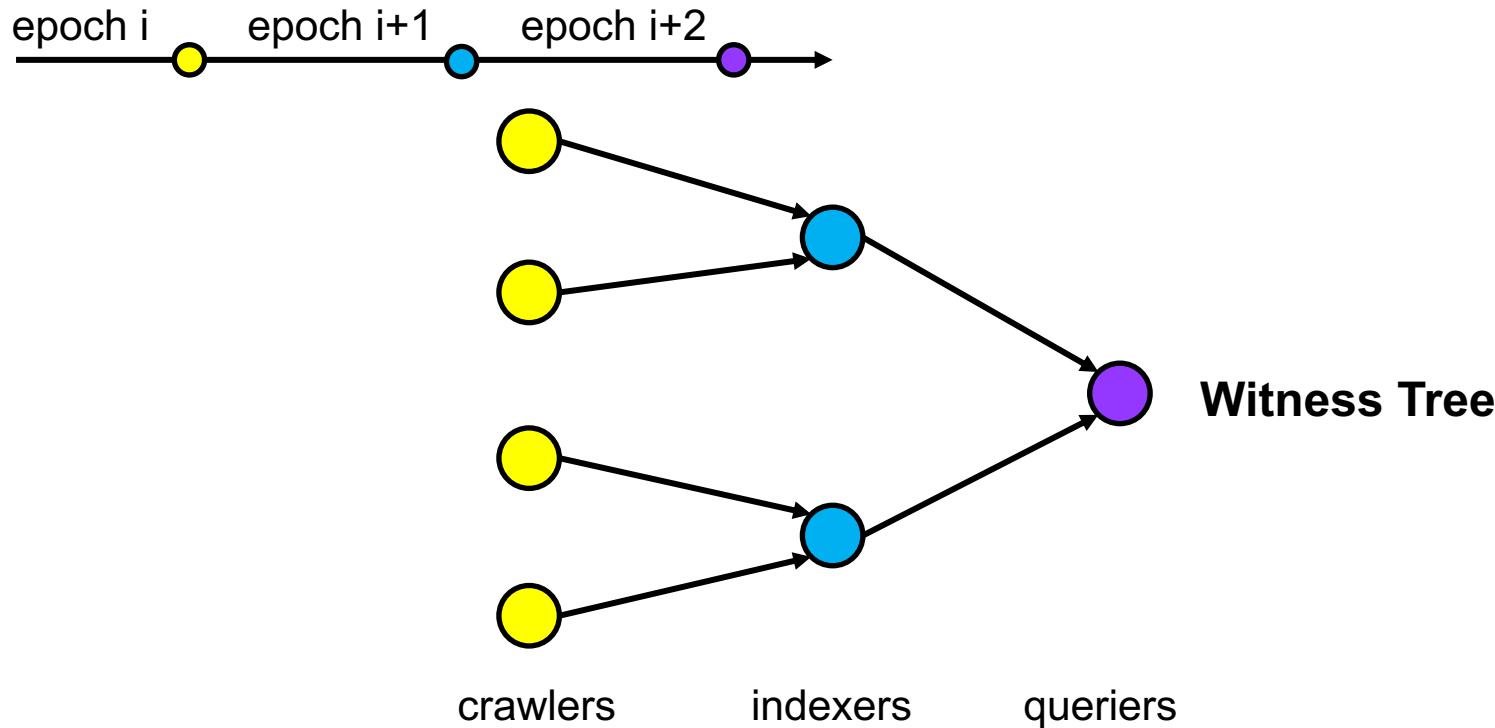
Providing Verifiable Search

- Combing Lambda witnesses and Kanban epochs



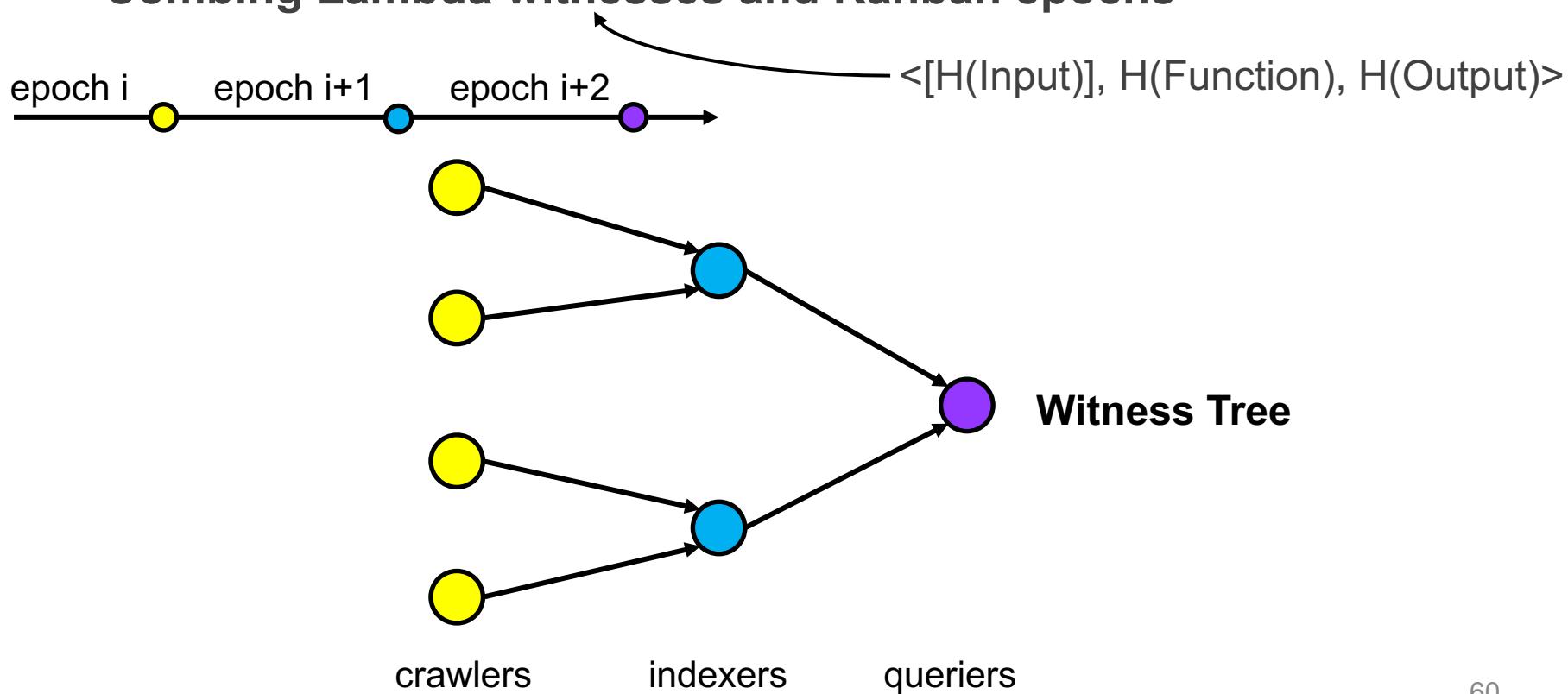
Providing Verifiable Search

- Combing Lambda witnesses and Kanban epochs



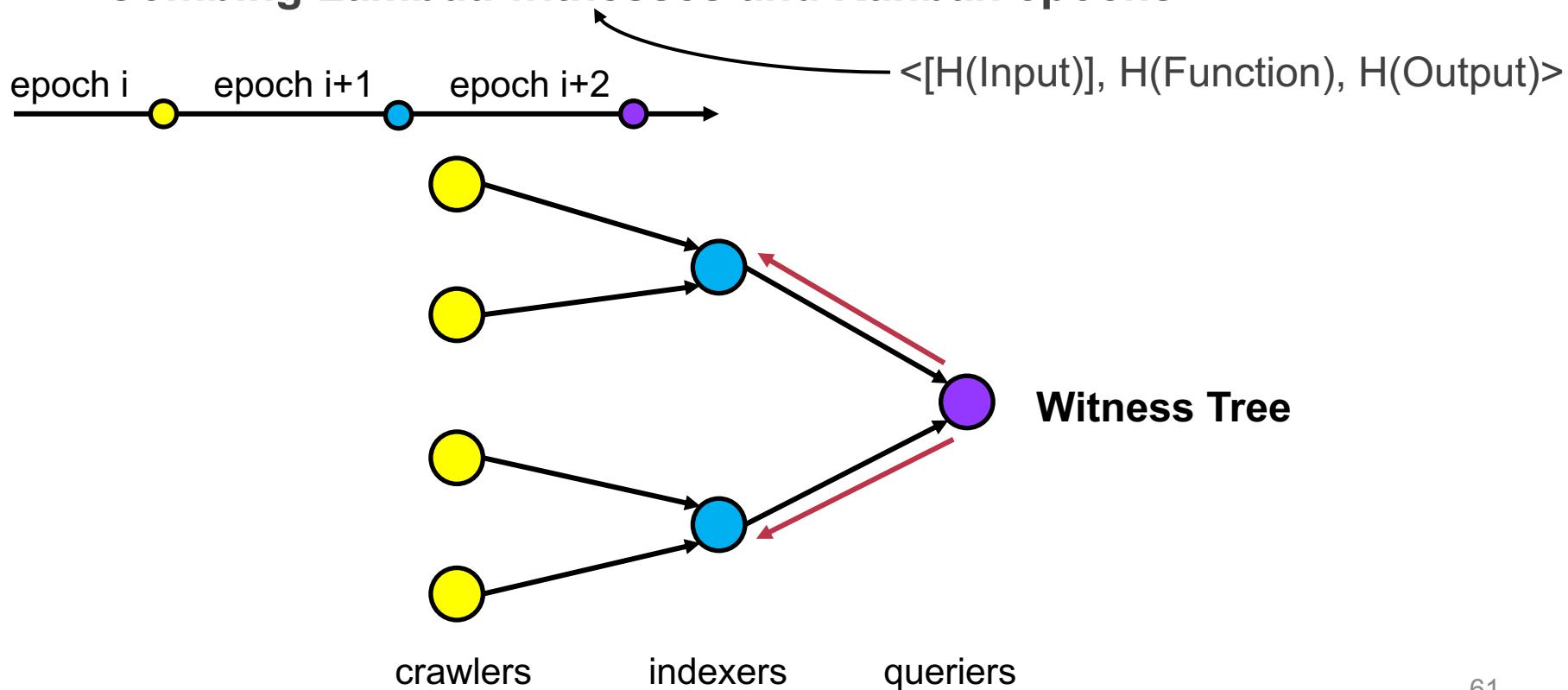
Providing Verifiable Search

- Combing Lambda witnesses and Kanban epochs



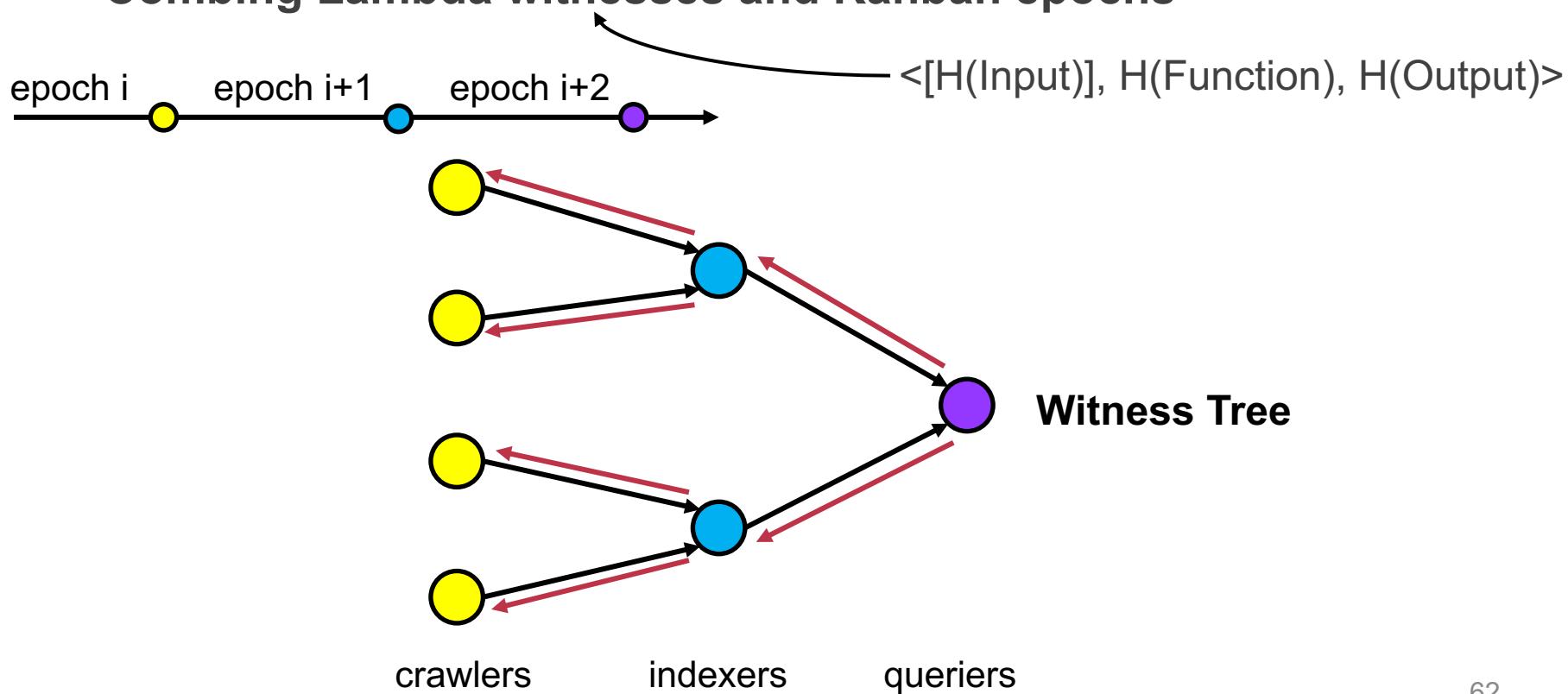
Providing Verifiable Search

- Combing Lambda witnesses and Kanban epochs



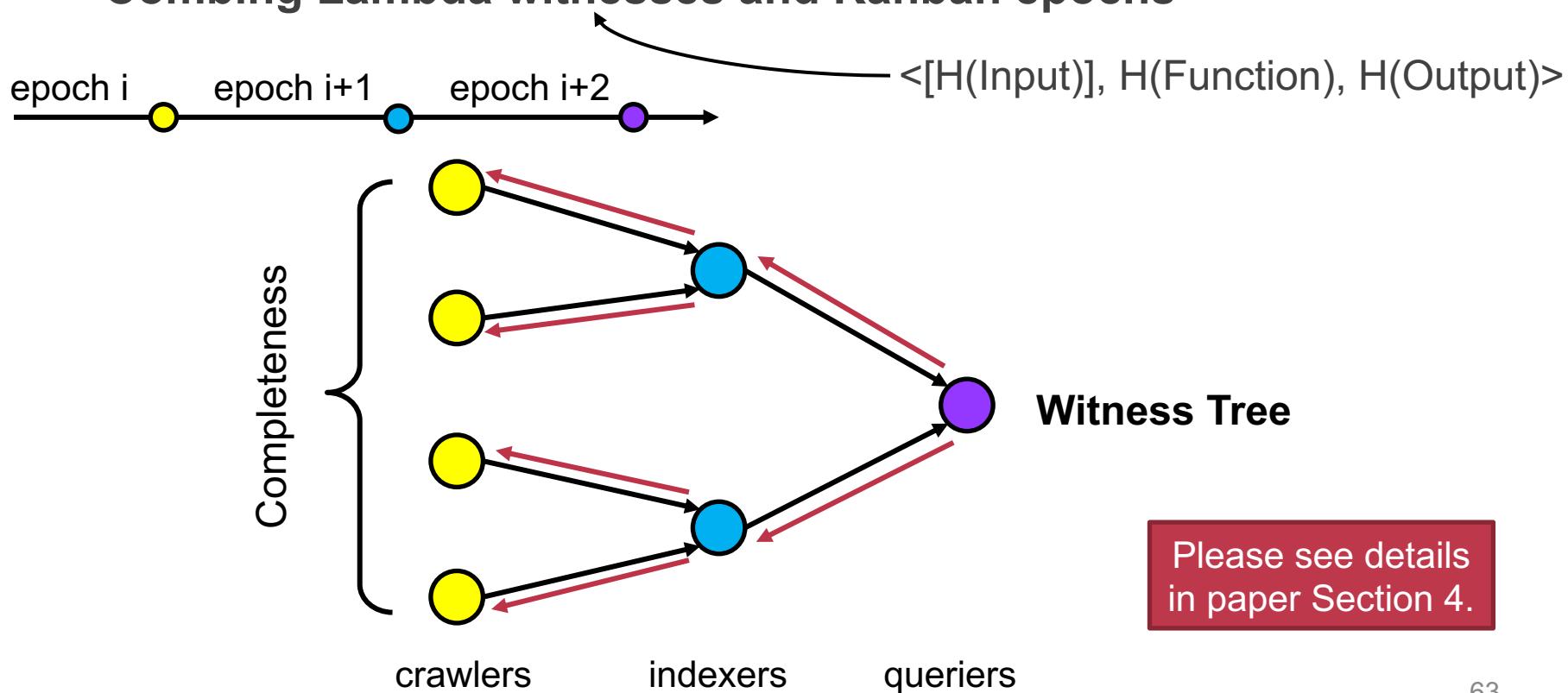
Providing Verifiable Search

- Combing Lambda witnesses and Kanban epochs



Providing Verifiable Search

- Combing Lambda witnesses and Kanban epochs

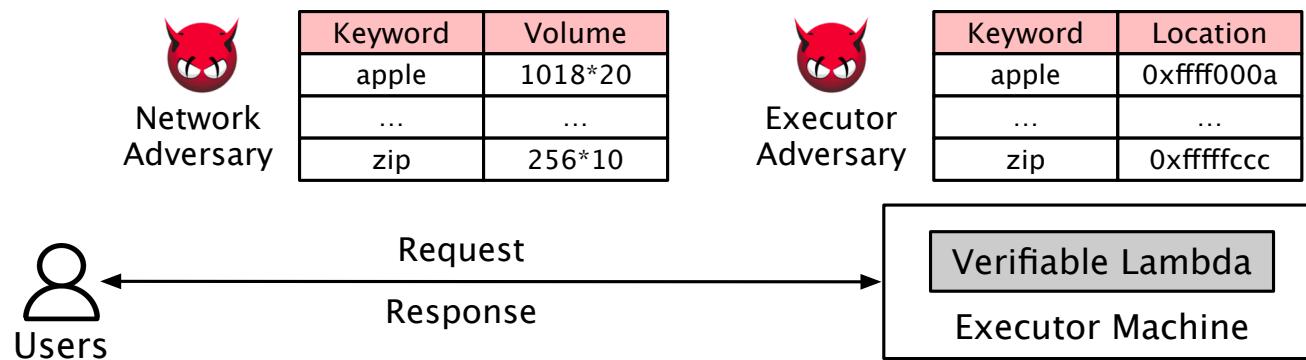


Outline

- **Problem Statement**
- **Design: decentralized search**
- **Design: verifiable search**
- **Design: private search**
- **Evaluation**

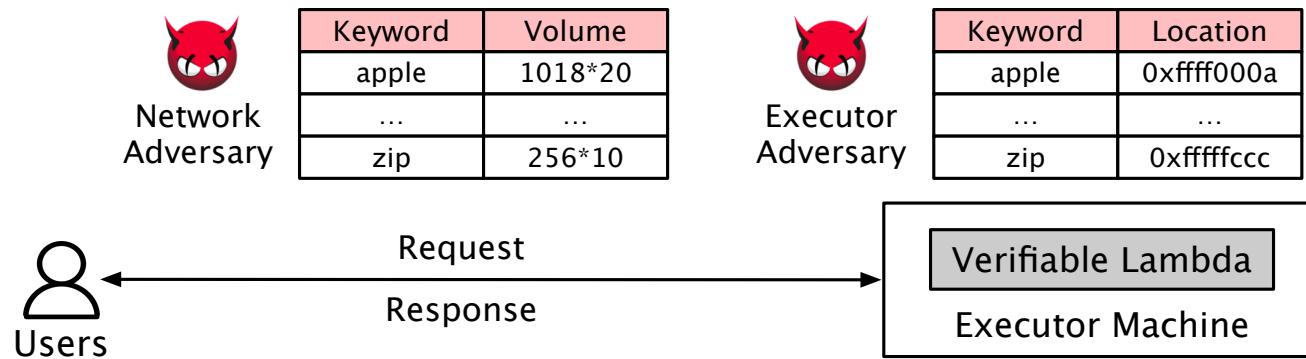
Private Search

- Two threats
 - Network adversary: infer search keywords via message lengths
 - Executor adversary: infer keywords via memory access pattern



Private Search

- Two threats
 - Network adversary: infer search keywords via message lengths
 - Executor adversary: infer keywords via memory access pattern

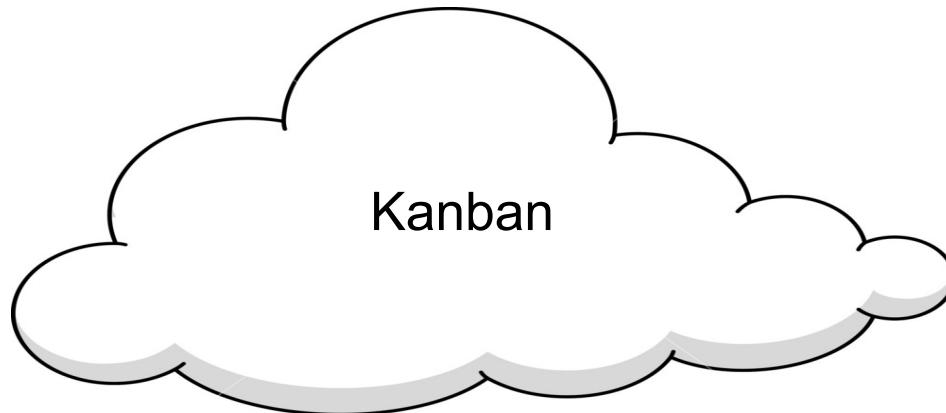


- Countermeasures
 - Network adversary: equalizing all message lengths
 - Executor adversary: SGX + Circuit-ORAM

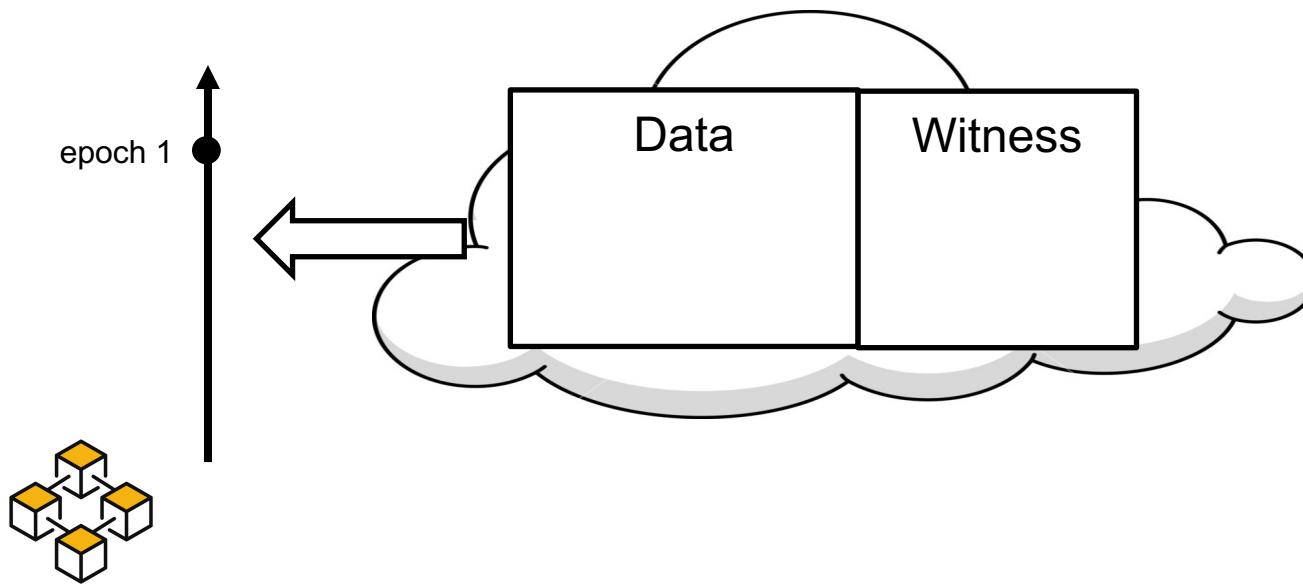


Putting all together

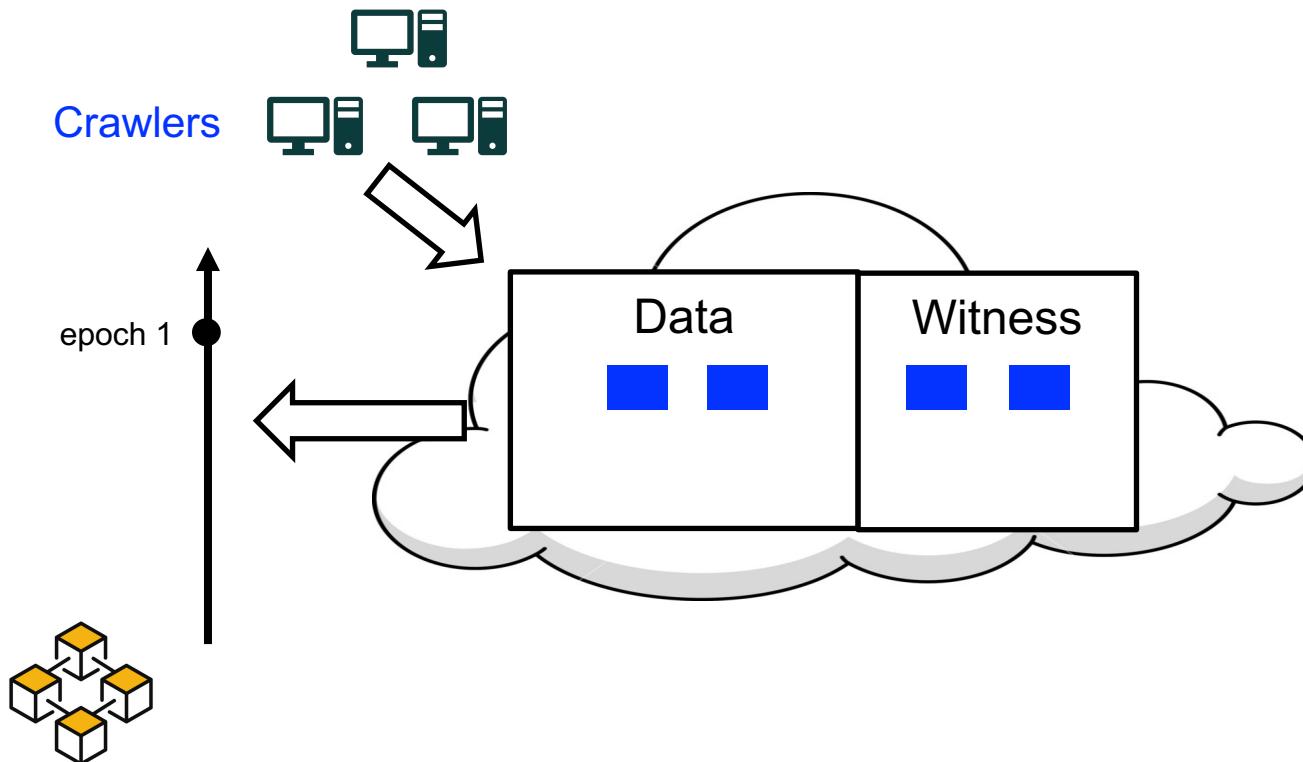
Putting all together



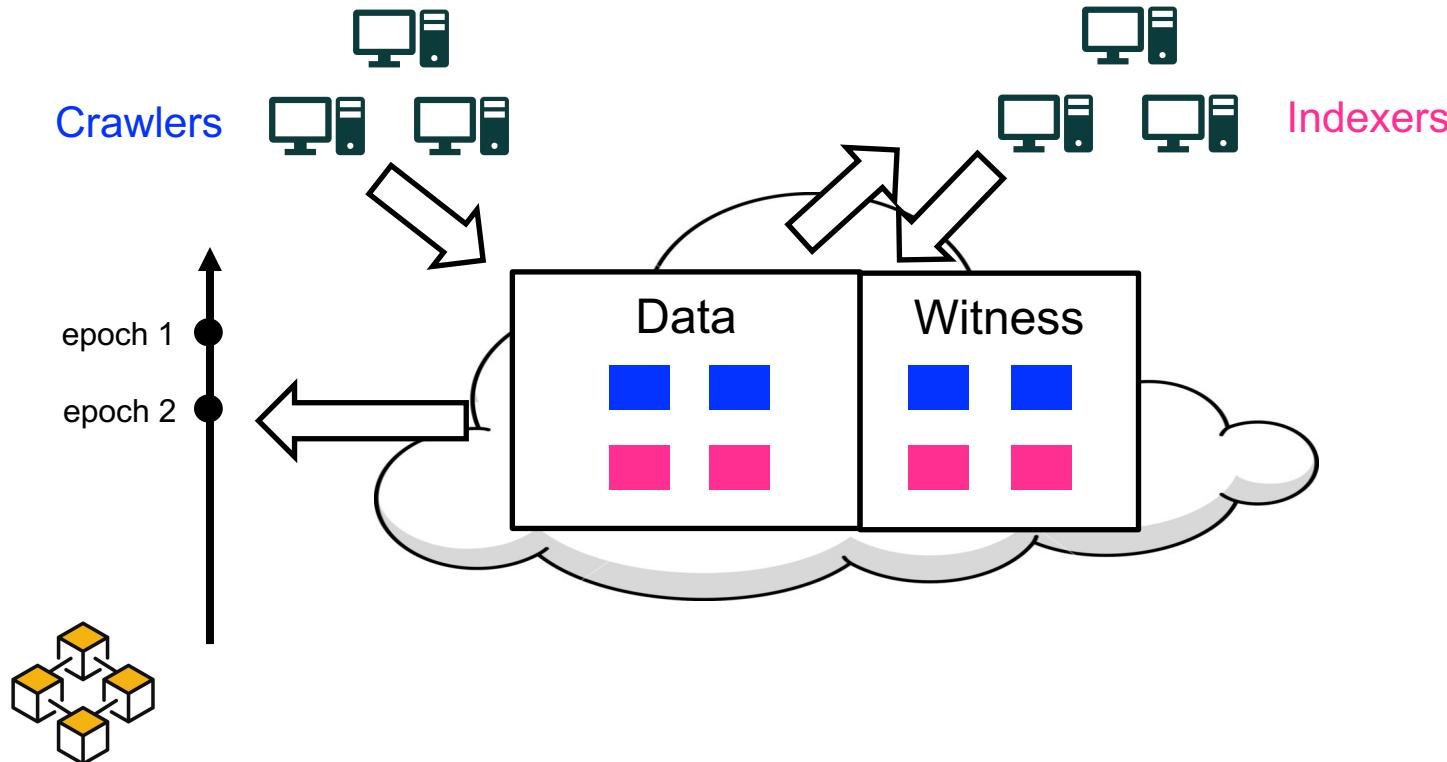
Putting all together



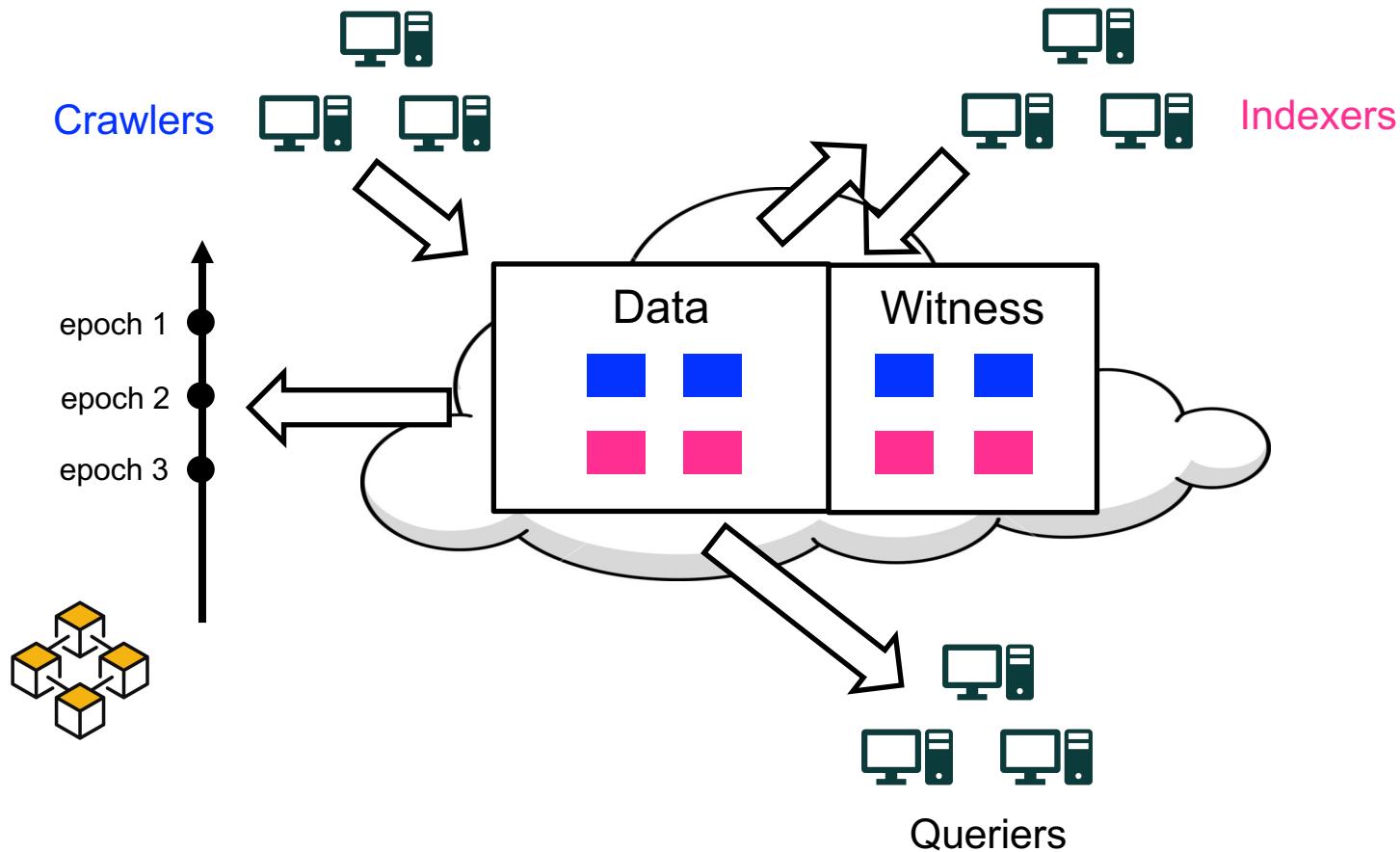
Putting all together



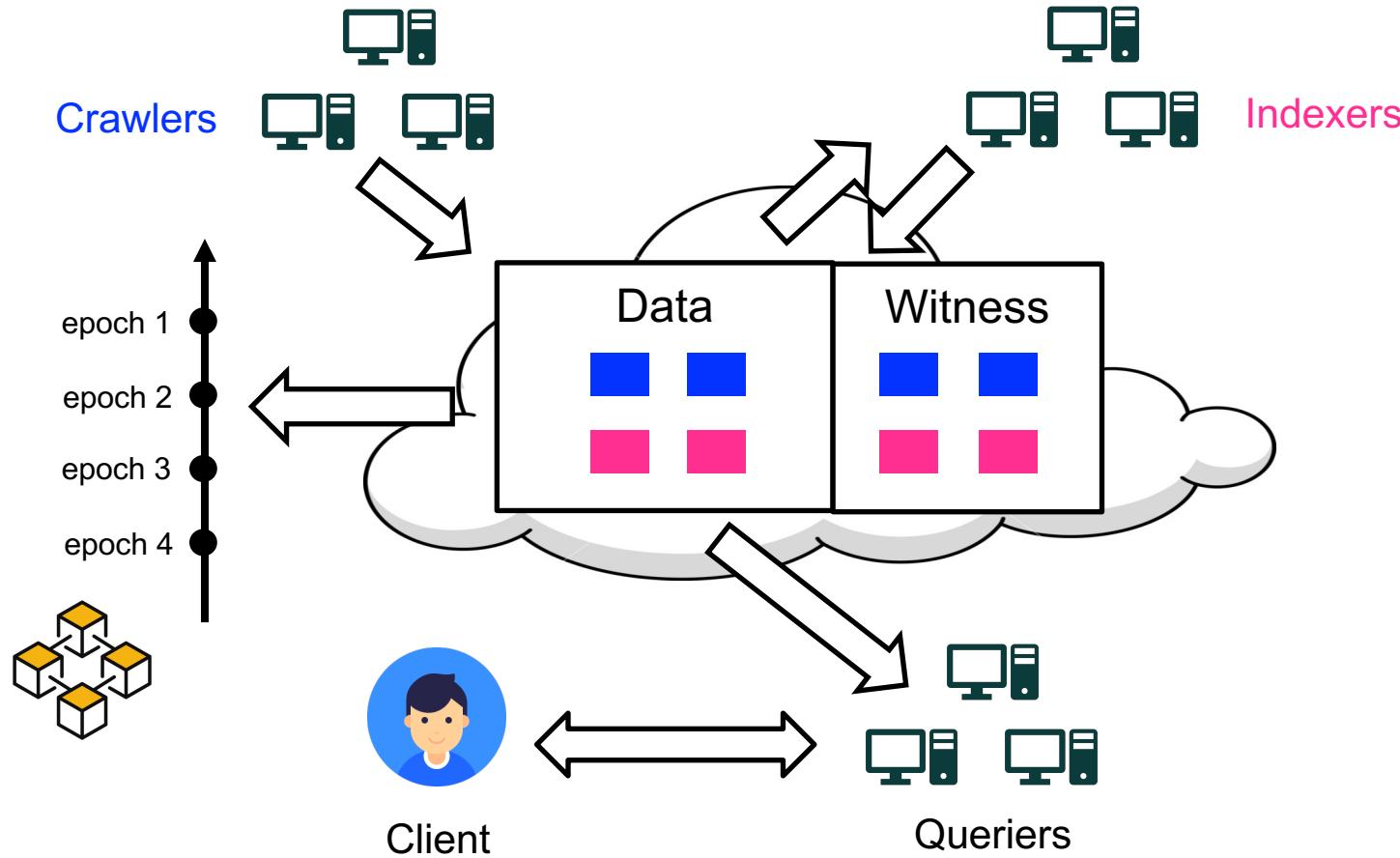
Putting all together



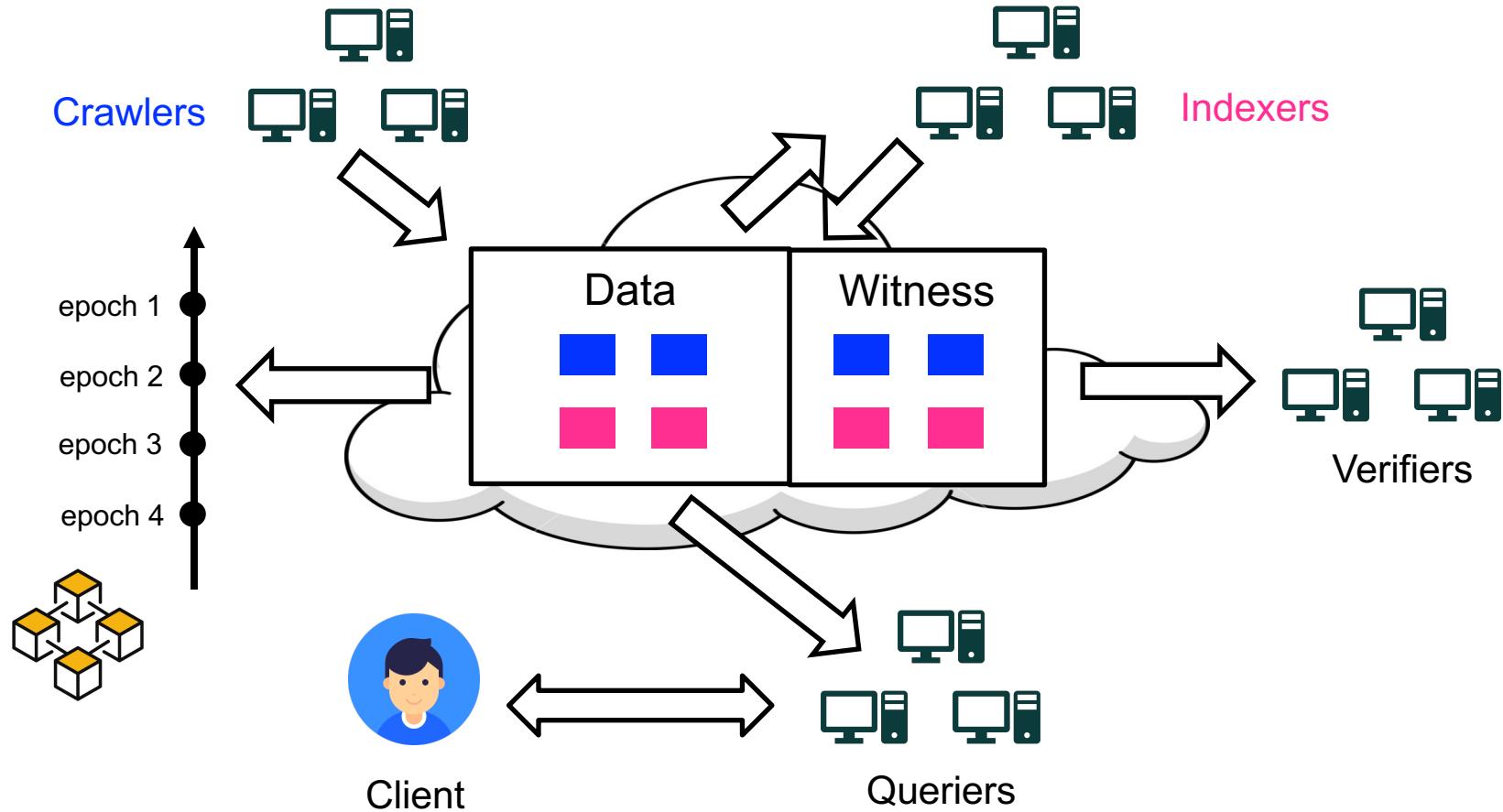
Putting all together



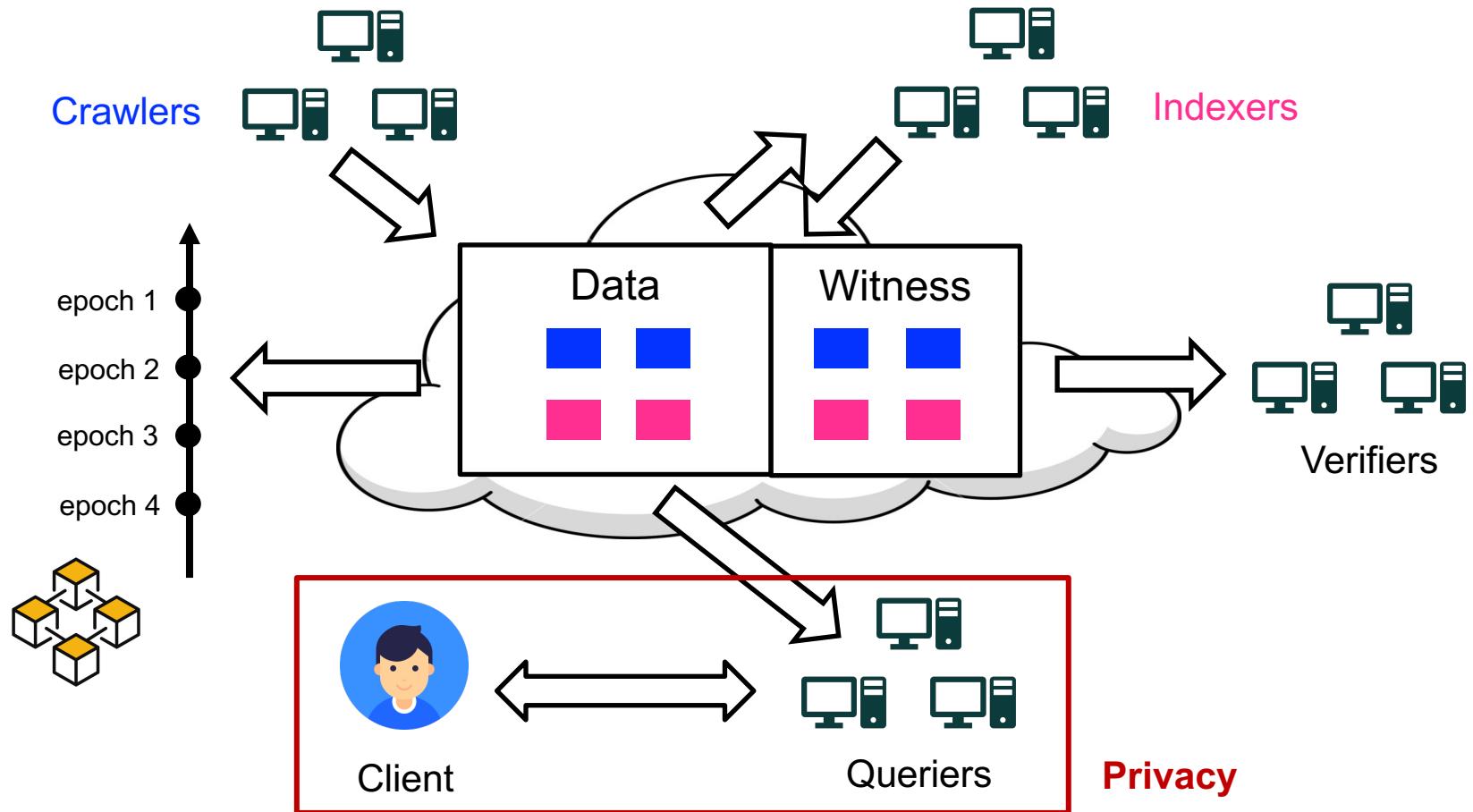
Putting all together



Putting all together



Putting all together





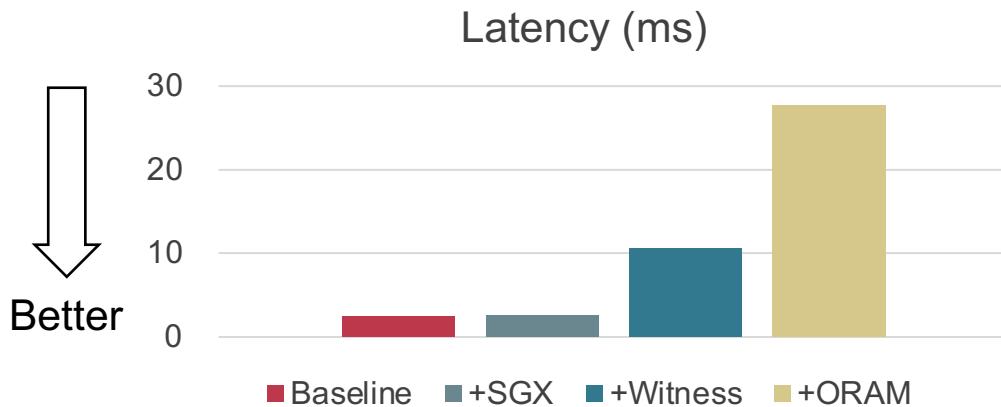
Outline

- **Problem Statement**
- **Design: decentralized search**
- **Design: verifiable search**
- **Design: private search**
- **Evaluation**

Evaluation Setup

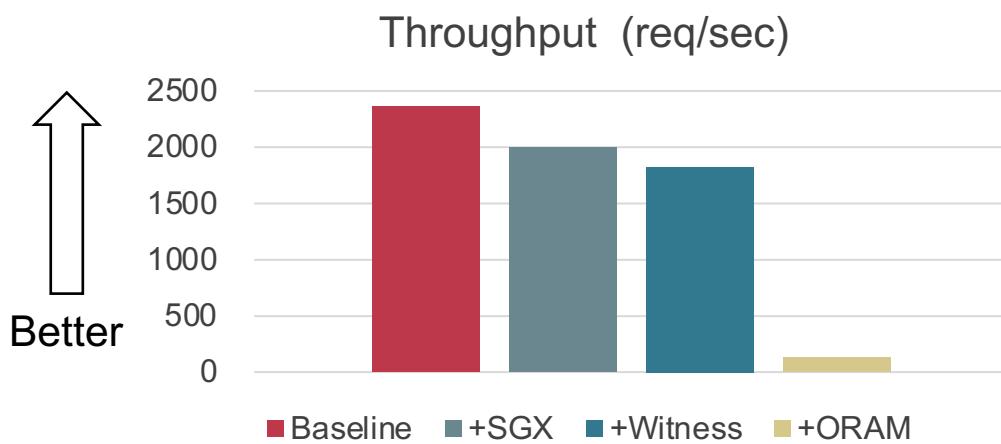
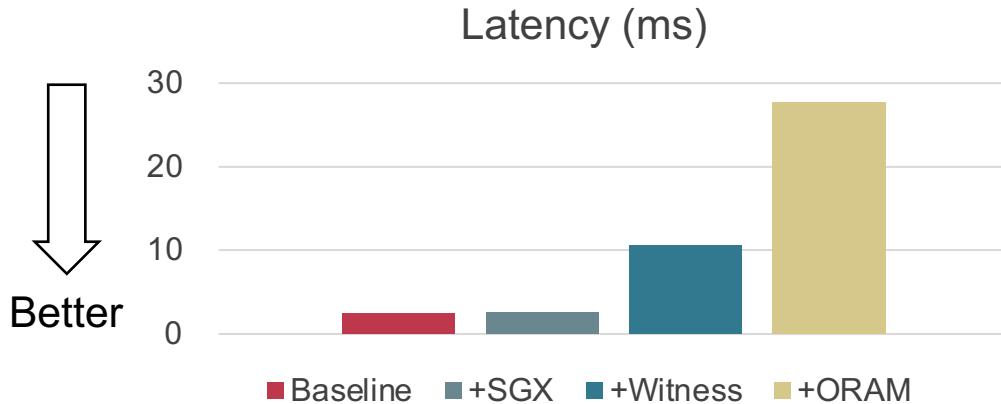
- **Datasets :**
 - Steemit: 234GB on-chain items
 - OpenBazaar: ~20K items per day
- **Lab testbed**
 - 9 SGX-enabled physical machines: at least 8-core, 8GB DRAM
- **AWS geo-distributed large-scale deployment**
 - 1300+ t2.medium EC2 instances: 2-core vCPU, 4GB DRAM

Microbenchmark



- **End-to-end latency**
 - < 50ms

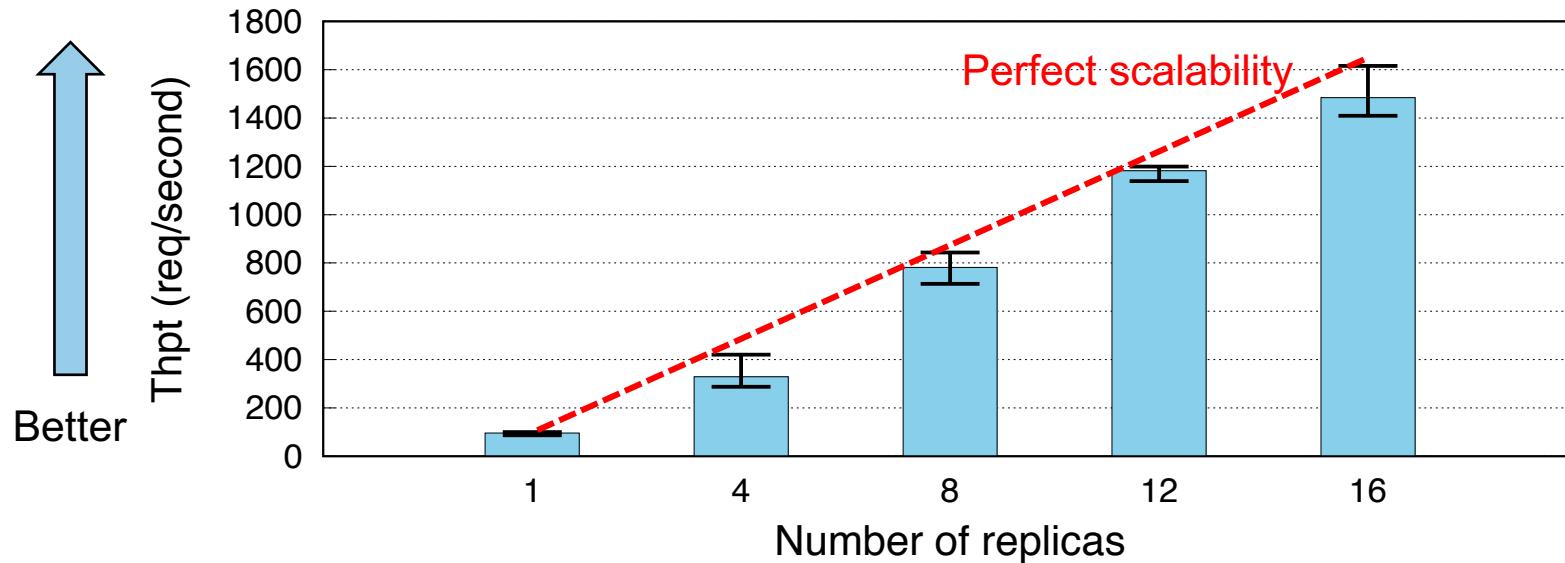
Microbenchmark



- **End-to-end latency**
 - < 50ms
- **Throughput :**
 - SGX: +15.2%
 - Witness: +7.6%
 - ORAM: +1260%
(**ORAM does not support concurrency**)

Large-scale Deployment

- Four geographic regions: Singapore (Asia), London (Europe), West Virginia (East America), California (West America)



From 1 to 16 replicas (each with 82 workers): nearly horizontally scalable

Verification and Speedup

	Native	Delegated
Execution Integrity		
Witness Download	517s	-
Signature Verification	4h25min	-
Data Integrity		
Witness Tree Verification	202s	-
Final-phase verify		
Verifier Interaction	0s	1.0s
Signature Verification	0s	0.2s
Total Time	4h33min	1.2s
		13681x

Witnesses can be reused to accelerate the verification

Summary

- **DeSearch is a decentralized search system for DApps**
- **Kanban: decouple states from computation**
 - cloud: avoid replicated storage and computation
 - blockchain: ensure Kanban contents verifiable
- **Witness: decouple verification from computation**
 - remove verification from lambda's fast paths
 - public dataflow proofs: enable data reuse and efficient verification
- **DeSearch project will be available at:**

<https://github.com/SJTU-IPADS/DeSearch>

Thanks!