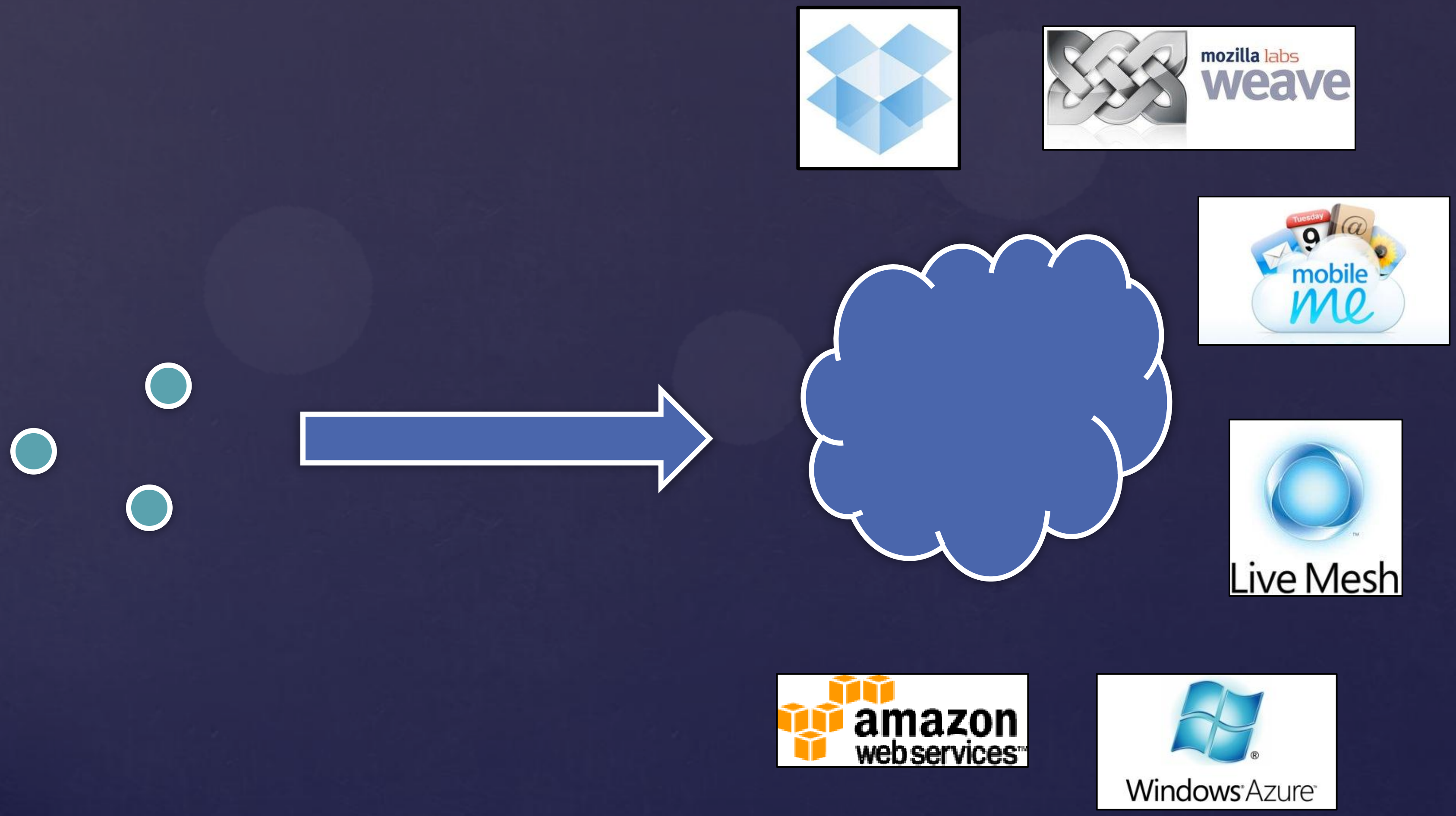
Structured Encryption and Controlled Disclosure

Melissa Chase
Seny Kamara

Microsoft Research

Cloud Storage



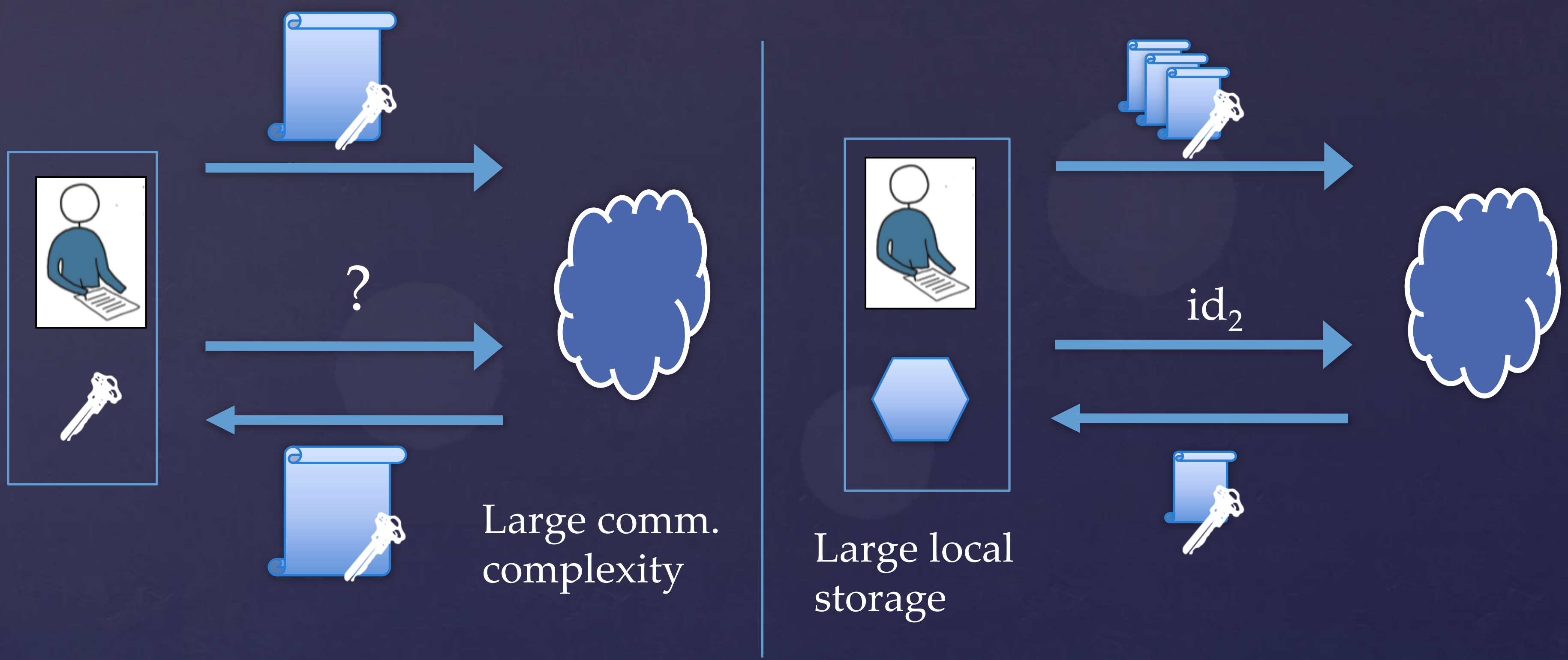
Security for Cloud Storage

- **Main concern:** *will my data be safe?*
  - it will be encrypted
  - it will be authenticated
  - it will be backed up
  - access will be controlled
  - …
- Security only vs.
  - outsiders
  - other tenants
- Q: can we provide security against the *cloud operator*?

Confidentiality in Cloud Storage

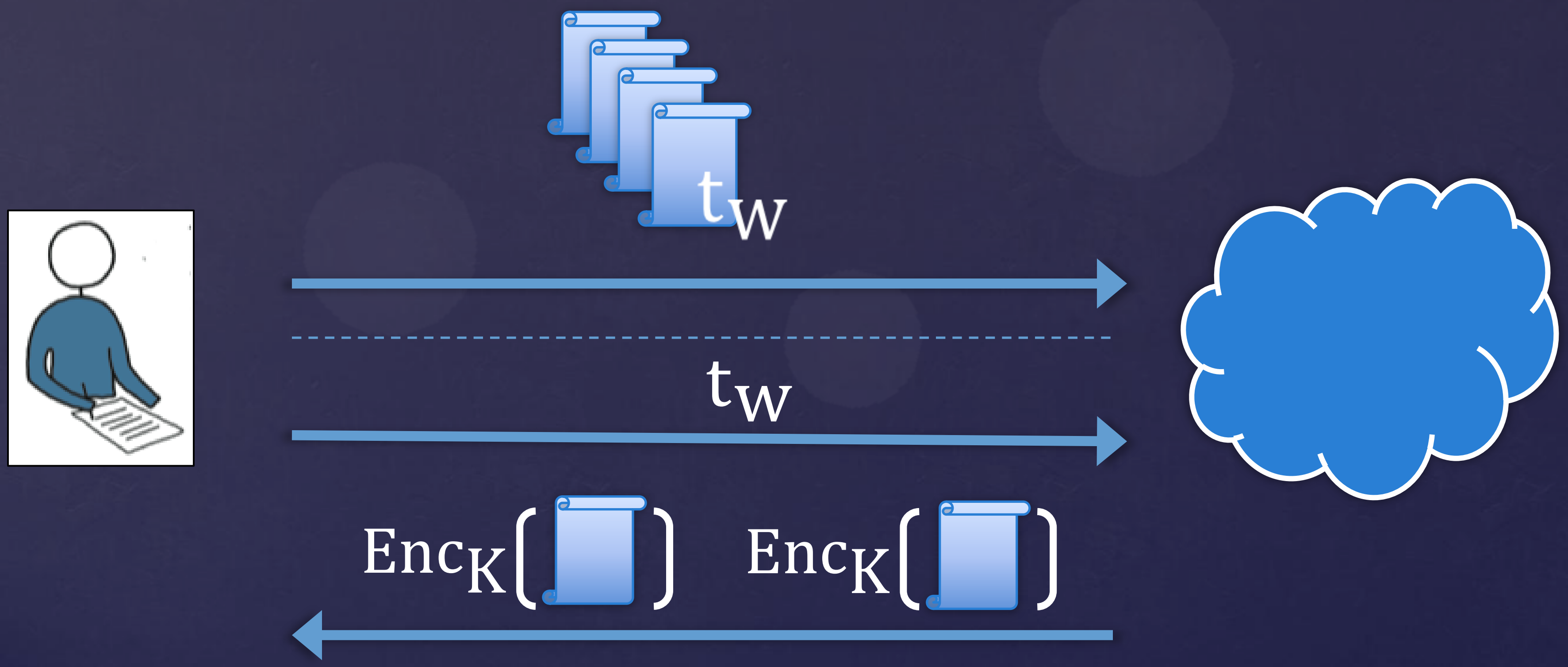
- How do we preserve confidentiality of data in the cloud?
  - Encryption!
  - What happens when I need to retrieve my data?
  - e.g., search over emails or pictures

Two Simple Solutions



Q: can we achieve O(1) storage at client and ``small" comm. complexity?

Searchable Symmetric Encryption
[Song-Wagner-Perrig01]



Related Work

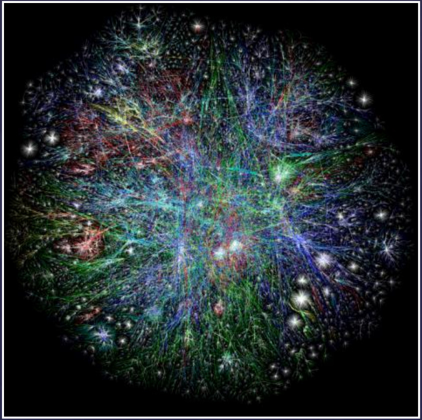
- Two-party computation [Yao82]
  - O(|data|) OTs & poly(|data|) server computation
- Oblivious RAMs [Goldreich-Ostrovsky96]
  - O(log n) rounds & polylog(n) server computation
- Fully-homomorphic encryption [Gentry09]
  - 1 round & poly(|data|) server computation
- Searchable encryption
  - [SWP01,Goh03,Chang-Mitzenmacher05,Boneh-diCrescenzo-Ostrovsky-Persiano04,…]: 1 round & O(n) server computation
  - [Curtmola-Garay-K-Ostrovsky06]: 1 round & O(# of docs w/ word) server computation

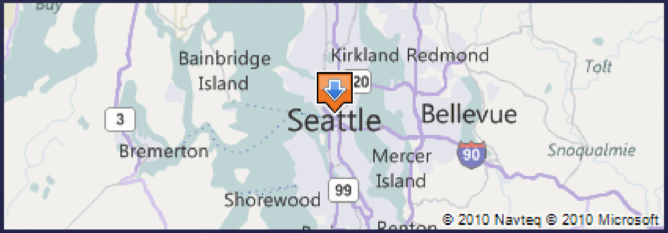
Limits of Searchable Encryption

- Private *keyword search* over encrypted *text* data
- Q: can we privately query other types of encrypted data?
  - maps
  - image collections
  - social networks
  - web page archives

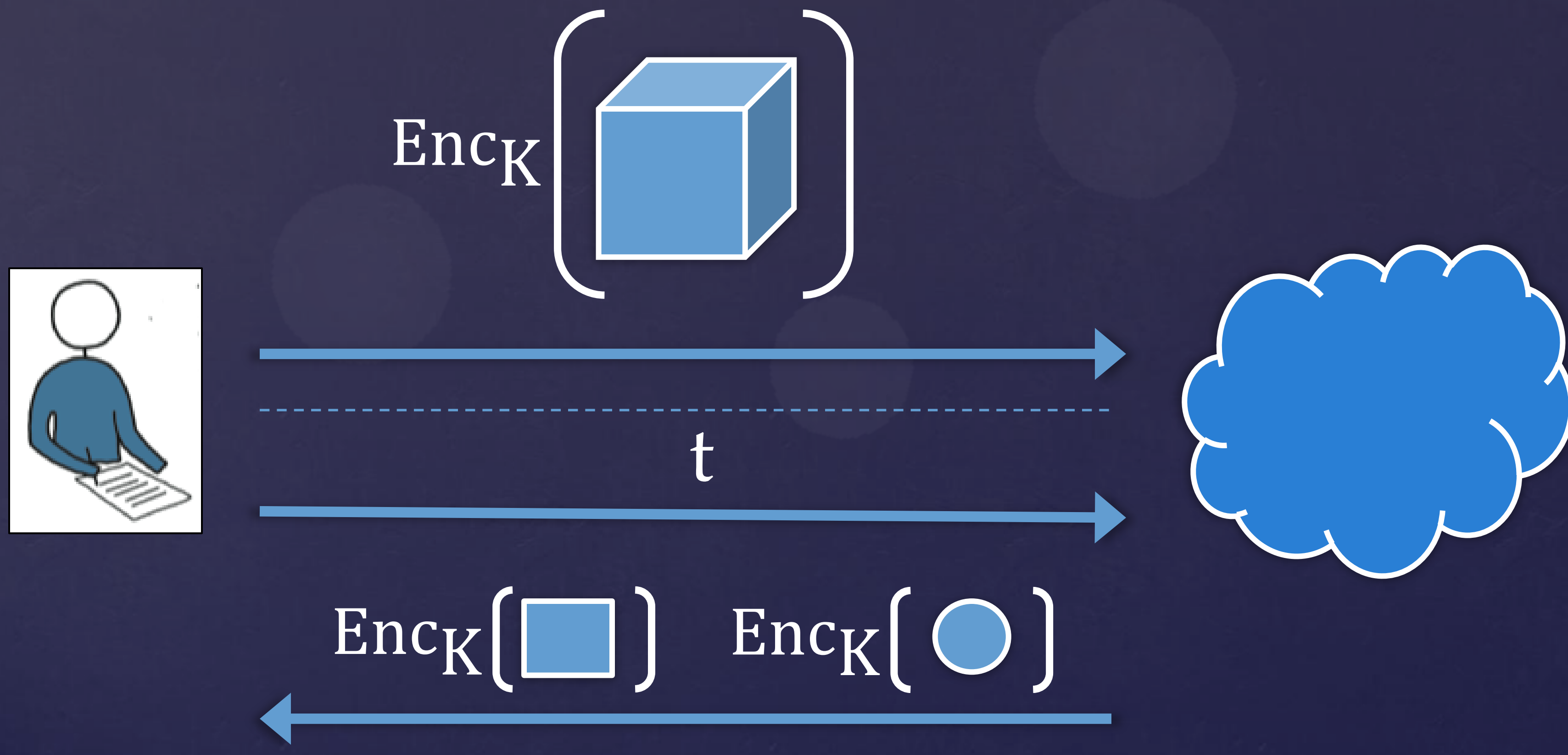
Graph Data

- Communications
  - email headers, phone logs
- Networks
- Social networks
- Web crawlers
- Maps





Structured Encryption



Our Results

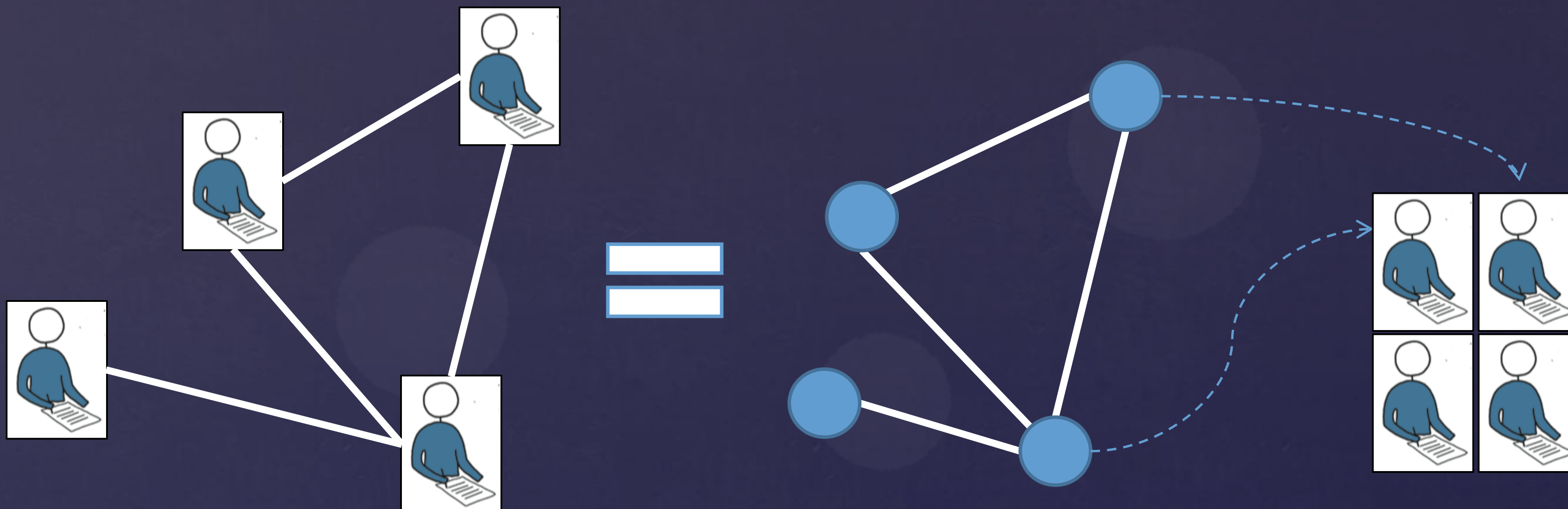
- Structured Encryption
- Formal security definition
  - simulation-based
- Constructions
  - *Adjacency* queries on encrypted graphs
  - *Neighbor* queries on encrypted graphs
  - *Focused subgraph* queries on encrypted web graphs
- Controlled disclosure
  - Application to cloud-based data brokering

Structured Encryption

Structured Data

- Email archive = Index + Email text

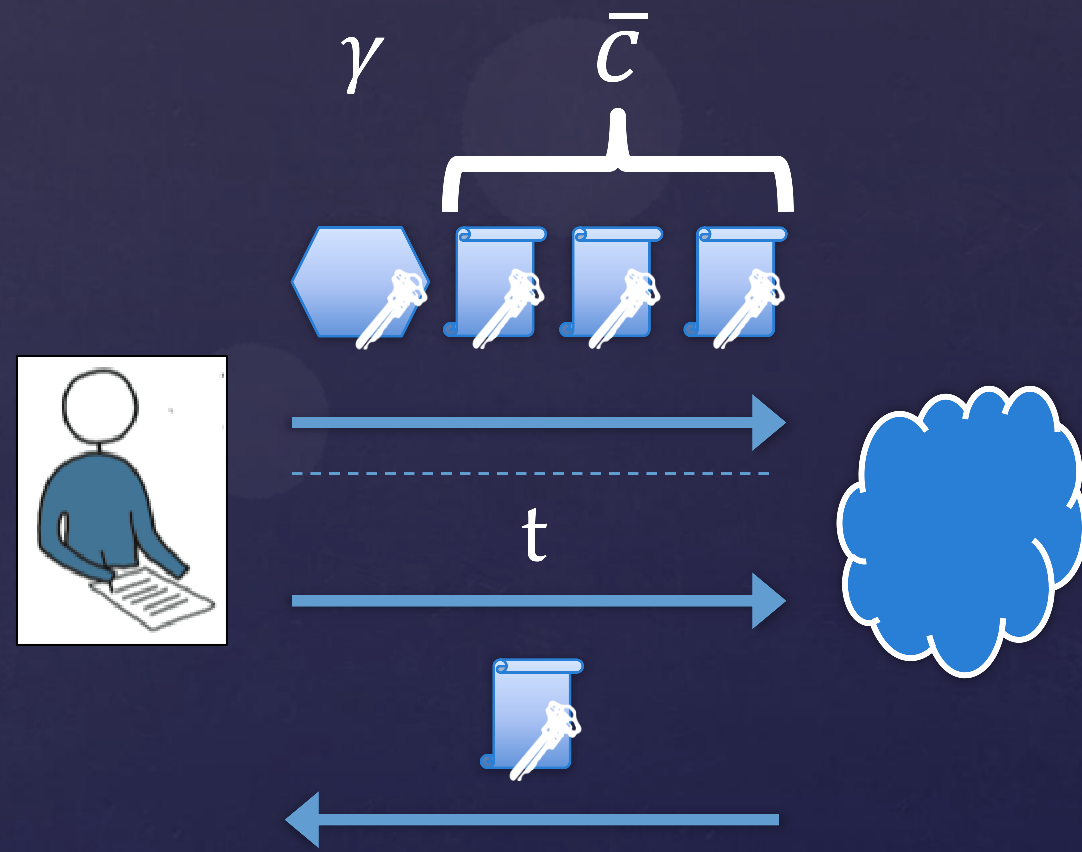
Structured Data



- Social network = Graph + Profiles

Structured Encryption

- $\mathrm{Gen}(1^k) \Rightarrow \mathrm{K}$
- $\mathrm{Enc}_K(\delta, \overline{m}) \Rightarrow (\gamma, \bar{c})$
- $\mathrm{Token}_K(q) \Rightarrow t$
- $\mathrm{Query}(\gamma, t) \Rightarrow I$
- $\mathrm{Dec}_K(c_i) \Rightarrow m_i$



CQA2-Security

- Security against *adaptive* chosen query attacks
  - generalizes CKA2-security from [Curtmola-Garay-K-Ostrovsky06]

- *Simulation*-based definition
  - ``given the ciphertext and the tokens no adversary can learn any information about the data and the queries, even if the queries are made adaptively"

- Too strong
  - e.g., SSE constructions leak some information
  - access pattern: pointers to documents that contain keyword
  - search pattern: whether two queries were for the same keyword

CQA2-Security

- Security is *parameterized* by 2 stateful leakage functions

- *Simulation*-based definition
  - ``given the ciphertext and the tokens no adversary can learn any information about the data and the queries other than what can be deduced from the $\mathcal{L}_1$ and $\mathcal{L}_2$ leakages…"
  - "…even if queries are made adaptively"

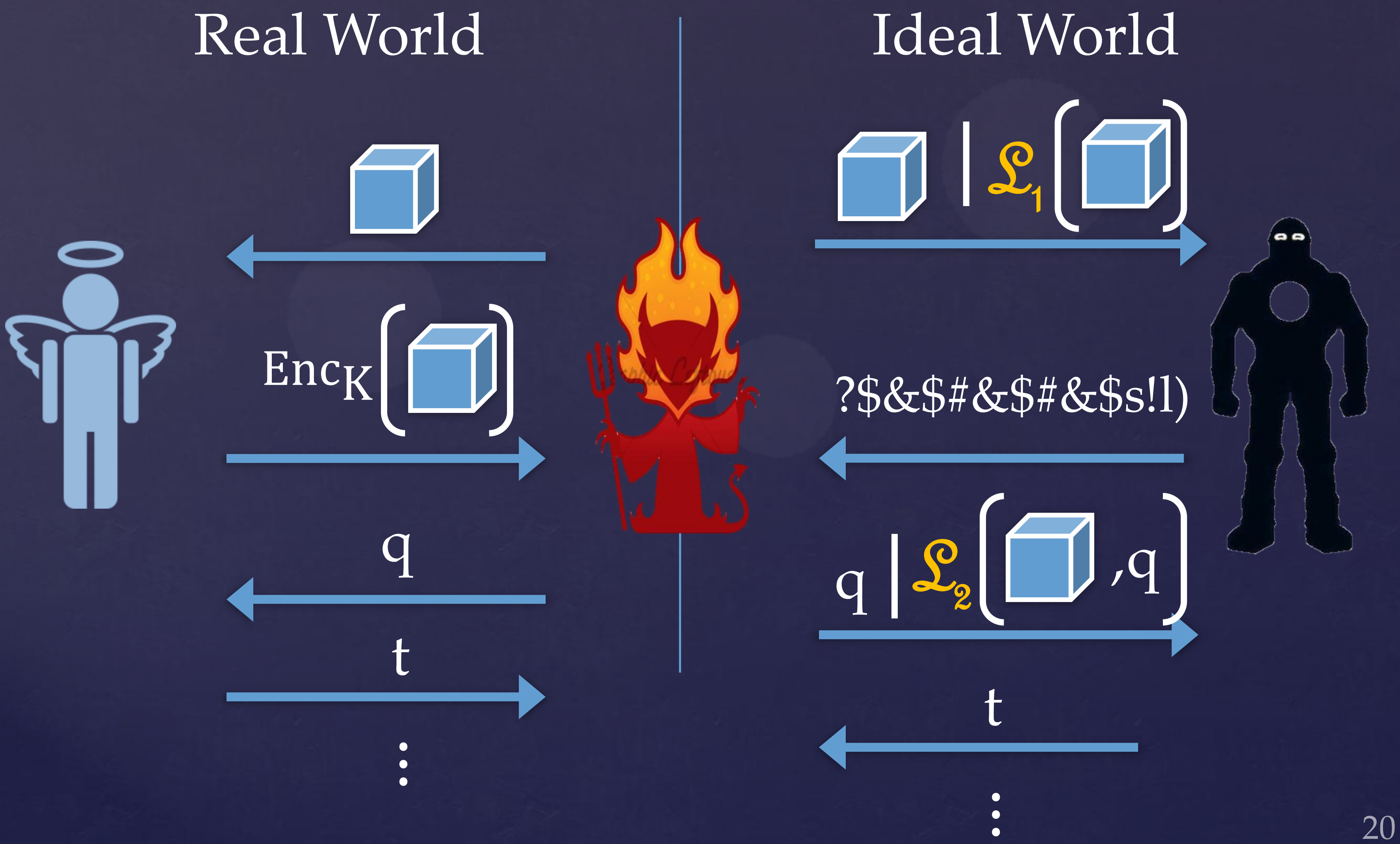
Leakage Functions

- 2 leakage functions
  - L1: leakage about data items
  - L2: leakage about data items and queries
- Previous work on SSE -- except [Goldreich-Ostrovsky96]
  - L1: number of items and length of each item
  - L2: access pattern and search pattern
- This work:
  - L1: number of items and length of each item
  - L2: *intersection* pattern and query pattern
- intersection pattern $\ll$ access pattern

Access vs. Intersection Patterns

- Access pattern
  - Pointers to relevant data items (i.e., result of query)
- Intersection pattern
  - Replace each pointer in access pattern with random value in [1,n]
- Note:
  - access pattern could reveal information about query

CQA2-Security



Adaptiveness

- Simulator "commits" to encryptions before queries are made
  - requires equivocation and some form of non-committing encryption
- Lower bound on token length ≈ [Nielsen02]
  - $\Omega\left(\log\binom{n}{\lambda}\right)$ (w/o ROs)
    - n: # of data items
    - λ: # of relevant items
- All our constructions achieve lower bound

vs. Functional Encryption
[Boneh-Sahai-Waters10]

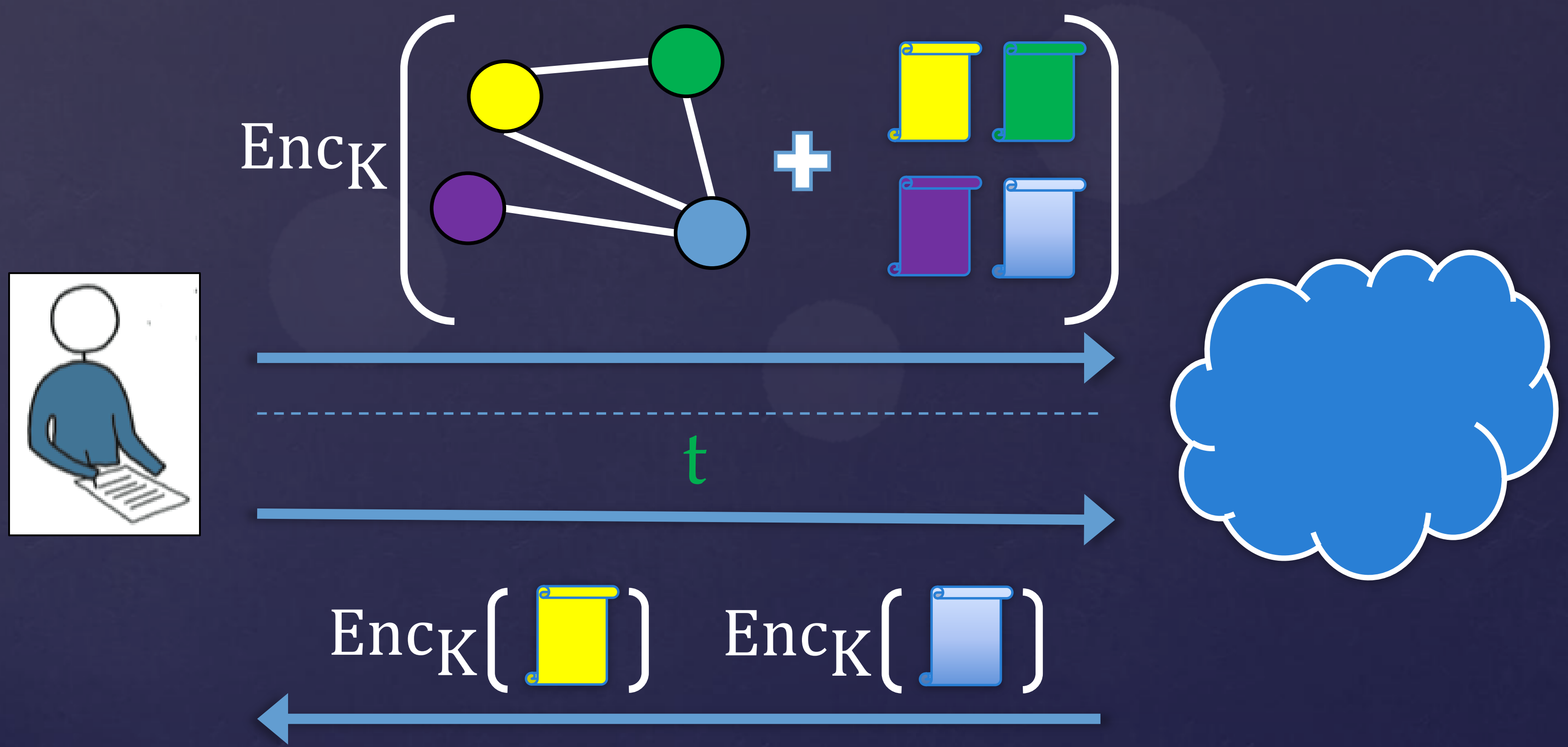
- Functional encryption
  - token can be used on *multiple* ciphertexts
  - Indistinguishability-based definitions
    - Simulation-based definitions are impossible (w/o ROs)
  - Currently can handle: inner products (i.e., polynomial predicates, AND, OR, boolean DNF & CNF)
- Structured encryption
  - token can be used on a *single* ciphertext
  - Simulation-based definition
  - Currently can handle: keyword search on text data; neighbor & adjacency queries on graphs; focused subgraph queries on web graphs; …

Constructions

Constructions

- Adjacency queries on encrypted graphs
  - from lookup queries on encrypted matrices
- *Neighbor queries on encrypted graphs*
  - from keyword search on encrypted text (i.e., SSE)
- *Focused subgraph queries on encrypted web graphs*
  - from keyword search on encrypted text
  - from neighbor queries on encrypted graphs

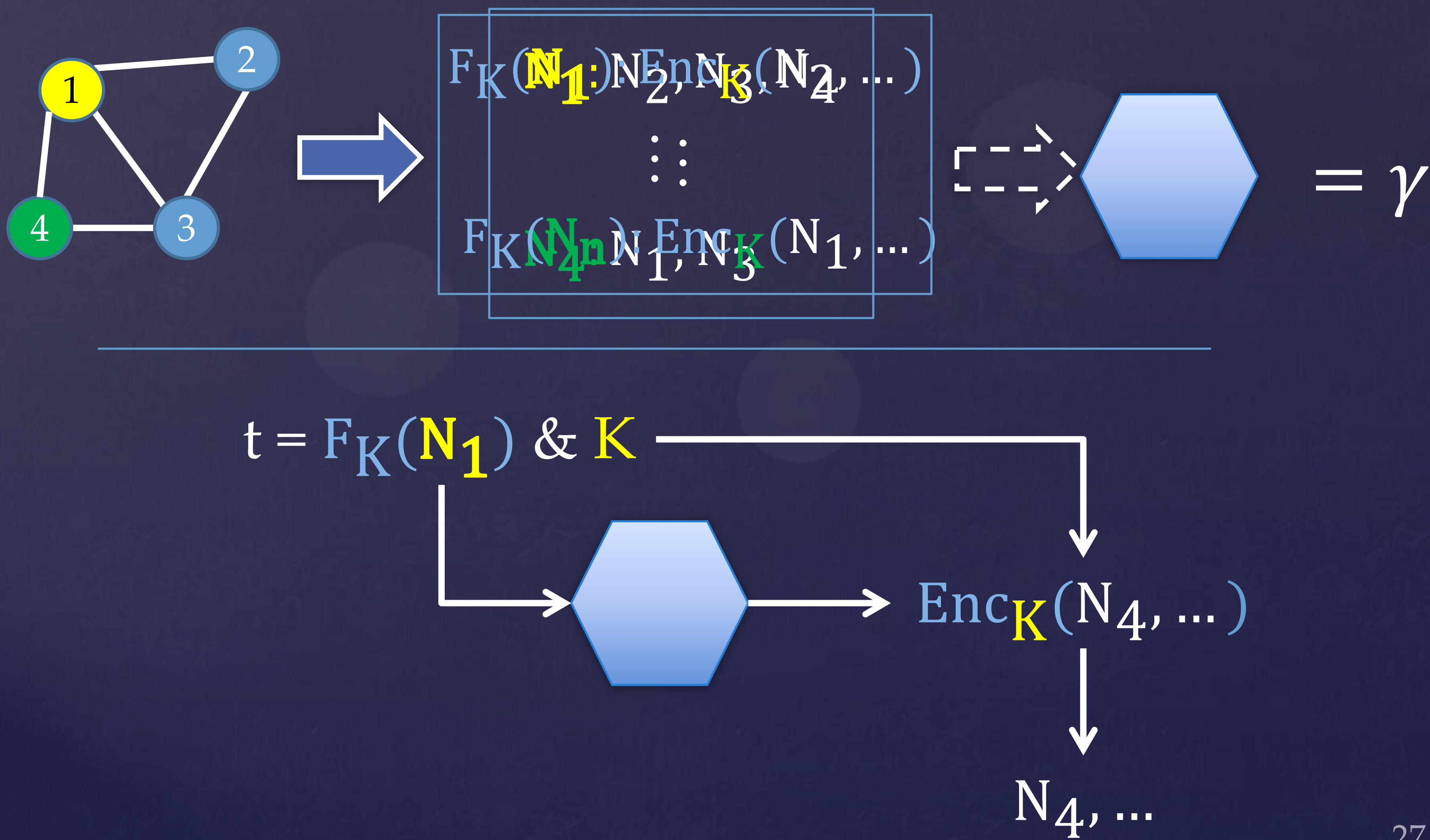
Neighbor Queries on Graphs



Neighbor Queries on Graphs

- Building blocks
  - Dictionary (i.e., key-value store)
  - Pseudo-random function
  - Non-committing symmetric encryption
    - PRF + XOR $\Longrightarrow$ tokens are as long as query answer
    - RO + XOR $\Longrightarrow$ tokens are as long as security parameter

Neighbor Queries on Graphs



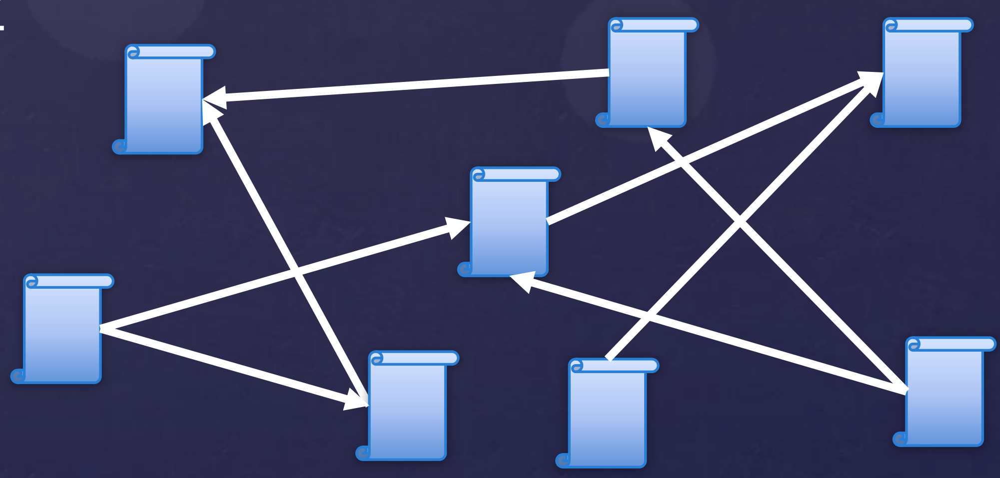
FSQ on Web Graphs

- Web graphs
  - Text data -- pages
  - Graph data --- hyperlinks
- Simple queries on web graphs
  - All pages linked from P
  - All pages that link to P
- Complex queries on web graphs
  - ``mix" both text and graph structure
  - search engine algorithms based on link-analysis
    - Kleinberg's HITS [Kleinberg99]
    - SALSA [LM01]
    - …

Focused Subgraph Queries

- HITS algorithm
  - Step 1: compute *focused subgraph*
  - Step 2: run iterative algorithm on focused subgraph

Singapore



FSQ on Encrypted Graphs

- Encrypt
  - pages with SE-KW
  - graph with SE-NQ
  - does not work!
- *Chaining* technique
  - combine SE schemes (e.g., SE-KW with SE-NQ)
  - preserves token size of first SE scheme
- Requires *associative* SE
  - message space: private data items and semi-private information
  - answer: pointers to data items + associated semi-private information
  - [Curtmola-Garay-K-Ostrovsky06]: associative SE-KW but not CQA2-secure!

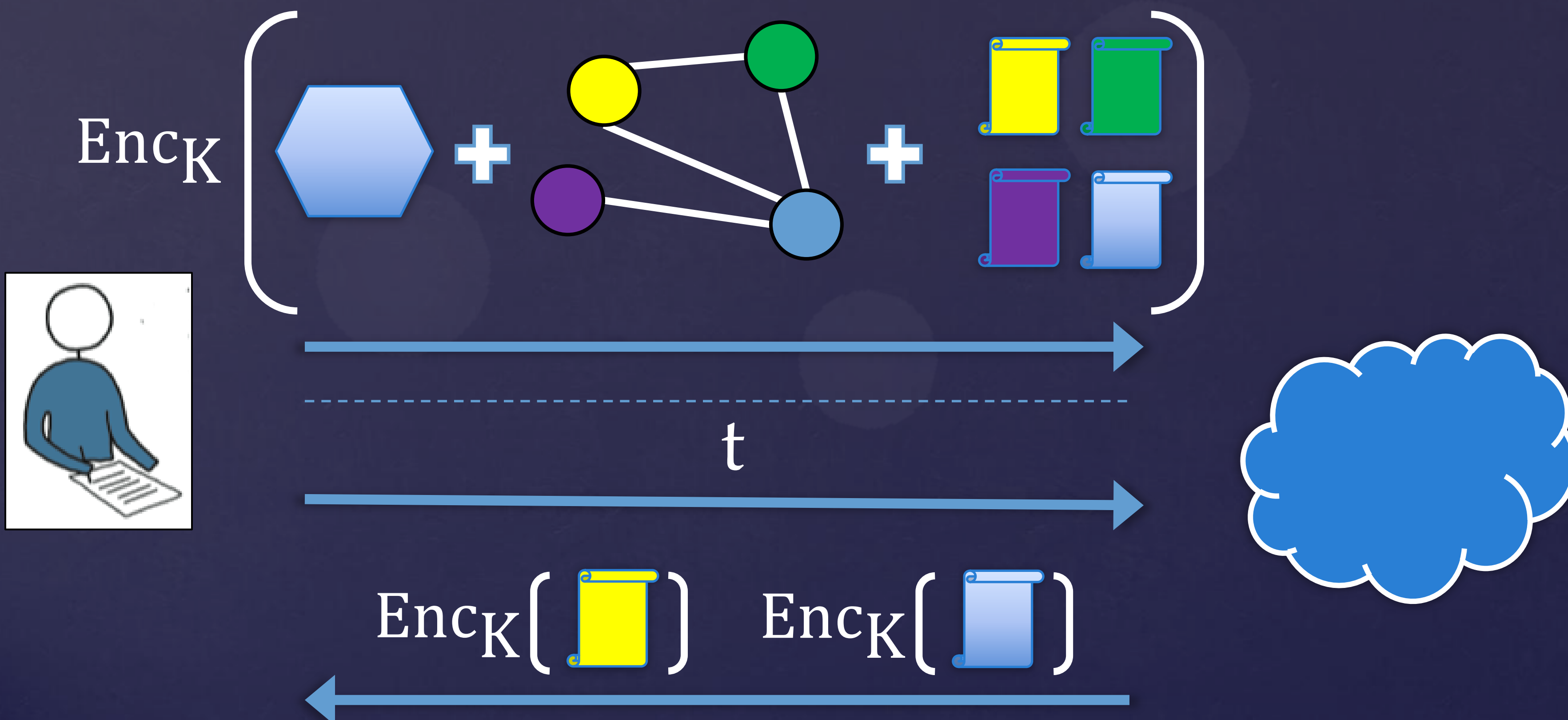
Associativity

- $\text{Gen}(1^k) \Rightarrow \text{K}$
- $\text{Enc}_K(\delta, \overline{m}) \Rightarrow (\gamma, \bar{c})$
- $\text{Token}_K(q) \Rightarrow t$
- $\text{Query}(\gamma, t) \Rightarrow I$
- $\text{Dec}_K(c_i) \Rightarrow m_i$

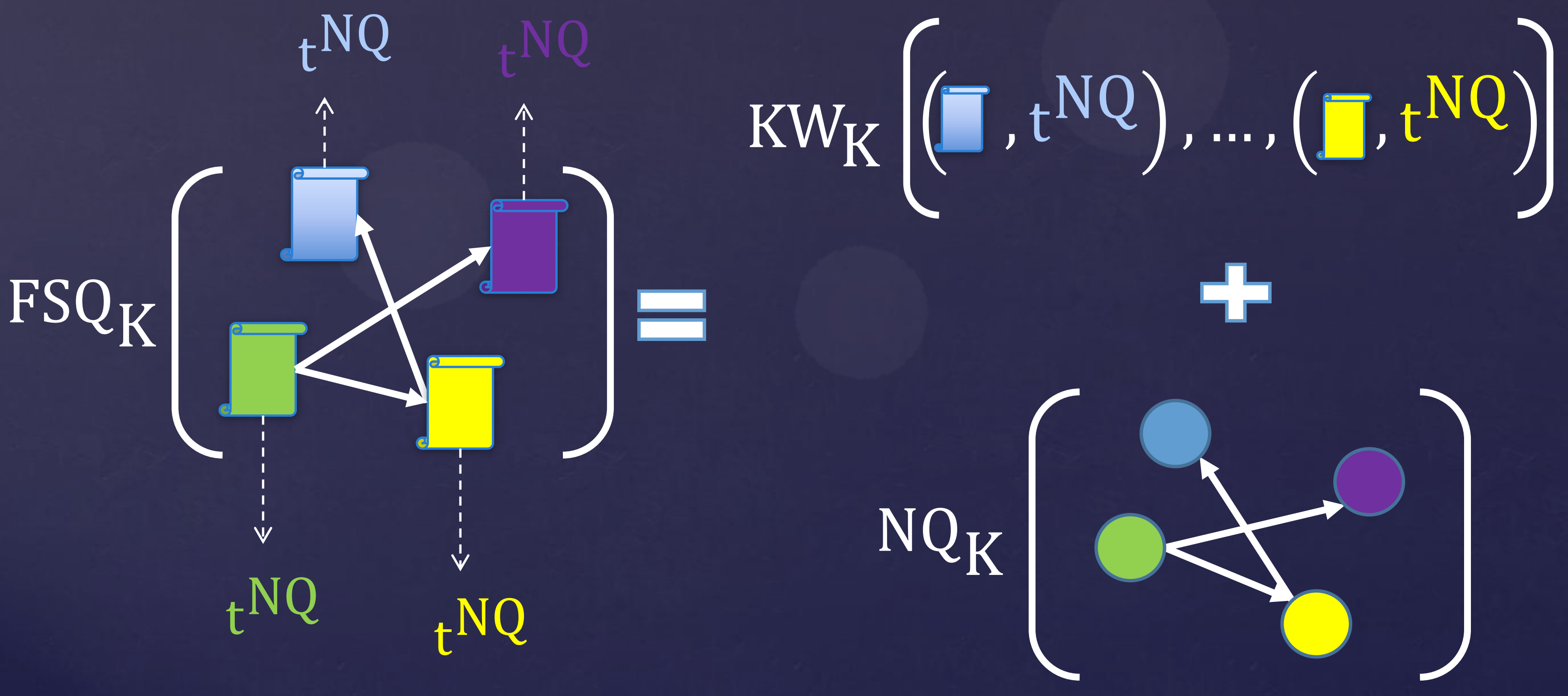
Associativity

- $\text{Gen}(1^k) \Rightarrow \text{K}$
- $\text{Enc}_K(\delta, \bar{m}, \bar{v}) \Rightarrow (\gamma, \bar{c})$
- $\text{Token}_K(q) \Rightarrow t$
- $\text{Query}(\gamma, t) \Rightarrow (I, \{v_i : i \in I\})$
- $\text{Dec}_K(c_i) \Rightarrow m_i$

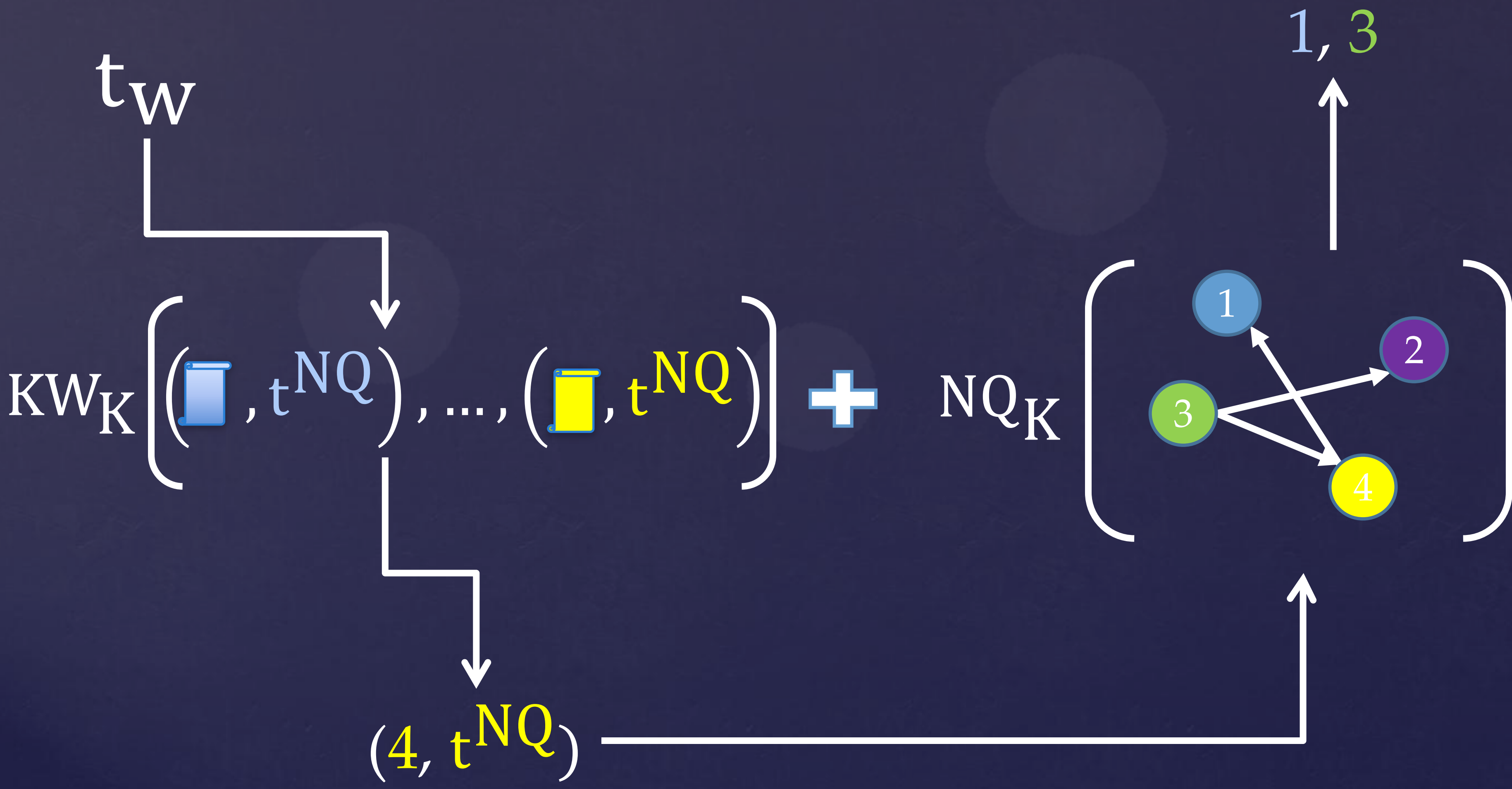
FSQ on Web Graphs



FSQ on Web Graphs



FSQ on Web Graphs



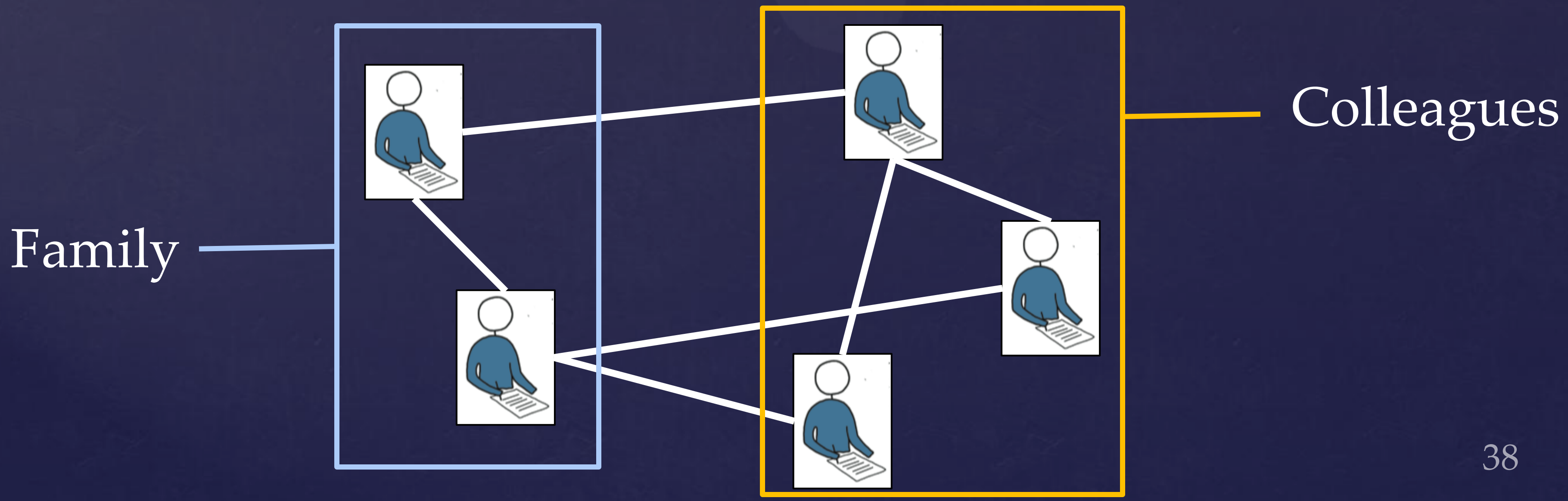
Controlled Disclosure

Limitations of Structured Encryption

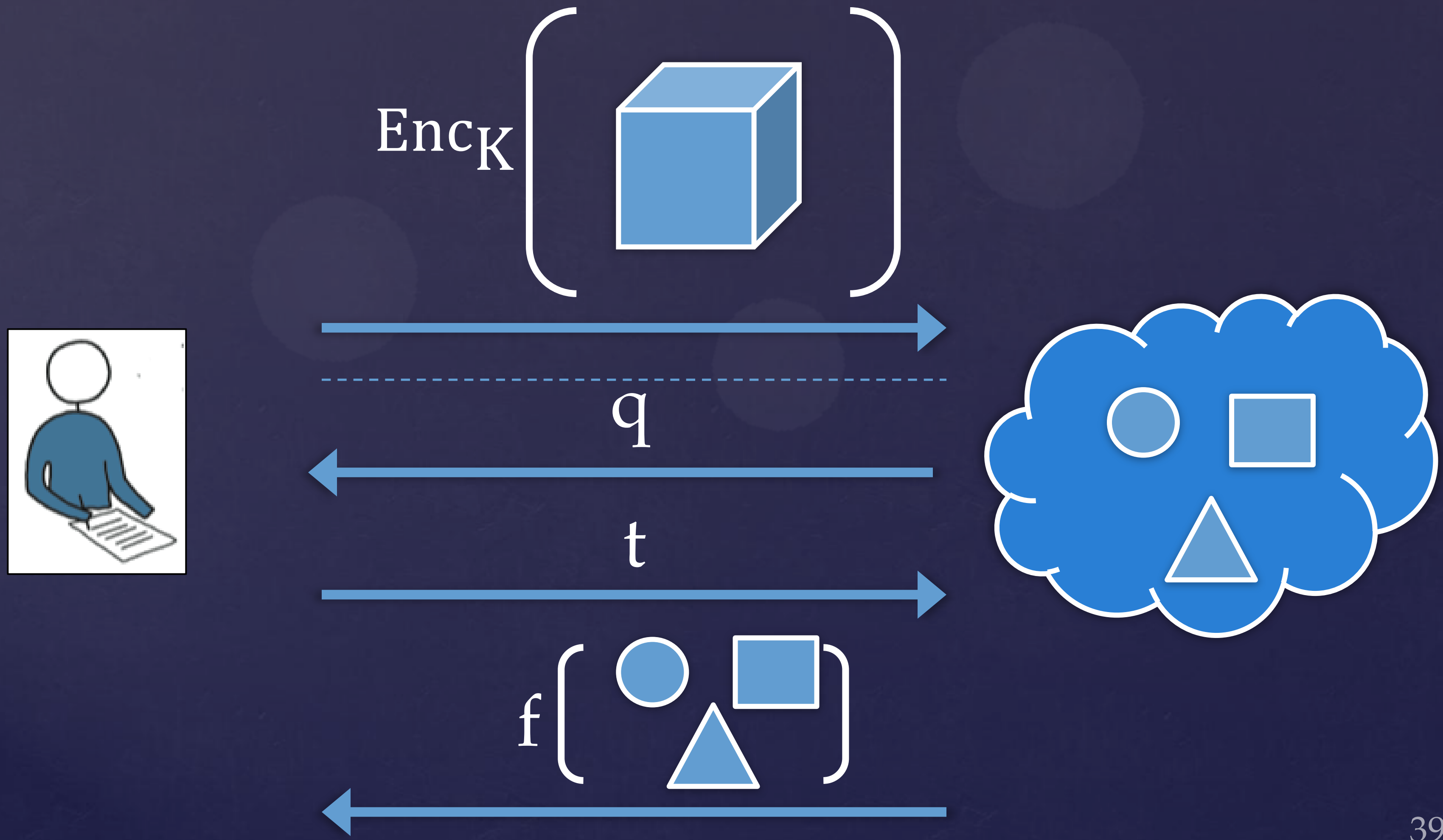
- Structured encryption
  - Private queries on encrypted data
- Q: what about computing on encrypted data?
  - Two-party computation
  - Fully-homomorphic encryption
- 2PC & FHE don't scale to massive datasets (e.g., Petabytes)
  - Do we give up security?

Controlled Disclosure

- Compromise
  - reveal only what is *necessary* for the computation
- Local algorithms
  - Don't need to ``see'' all their input
  - e.g., simulated annealing, hill climbing, genetic algorithms, graph algorithms, link-analysis algorithms, …



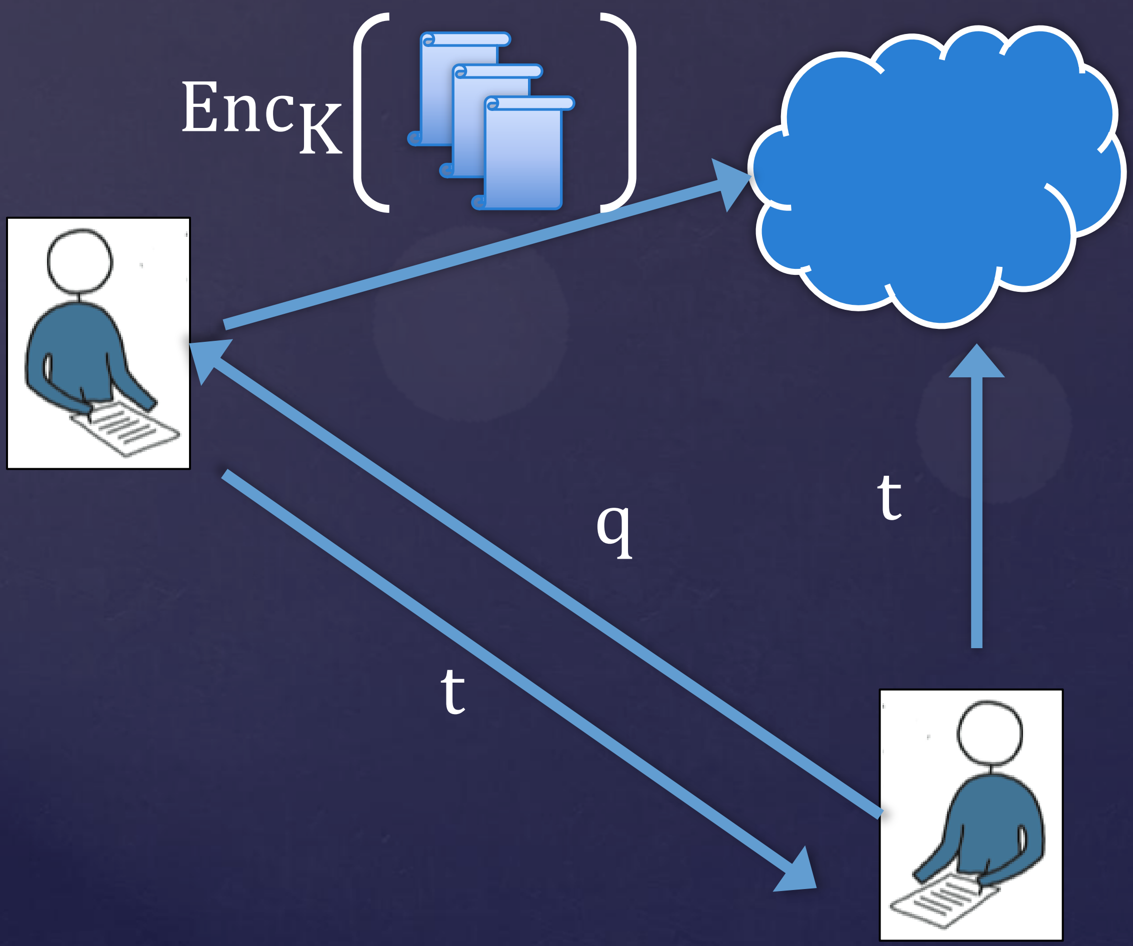
Controlled Disclosure



Cloud-based Data Brokerage

- Microsoft Azure Marketplace
- Infochimps

Secure Data Brokerage



- Producer
  - accurate count of data usage
- Collusions b/w
  - Cloud
  - Consumer

The End