FUZZY LABELED PRIVATE SET INTERSECTION WITH APPLICATIONS TO PRIVATE REAL-TIME BIOMETRIC SEARCH

Erkam Uzun, Simon Chung, Vladimir Kolesnikov, Alexandra Boldyreva and Wenke Lee
(Georgia Institute of Technology)

CREATING THE NEXT®

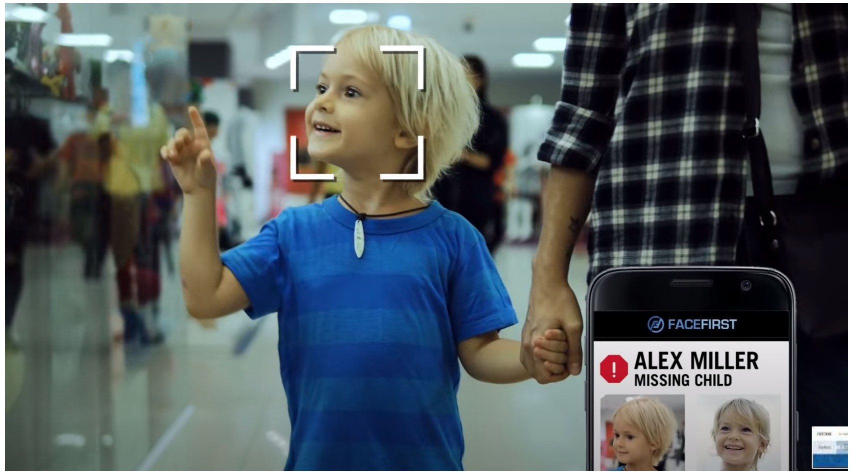
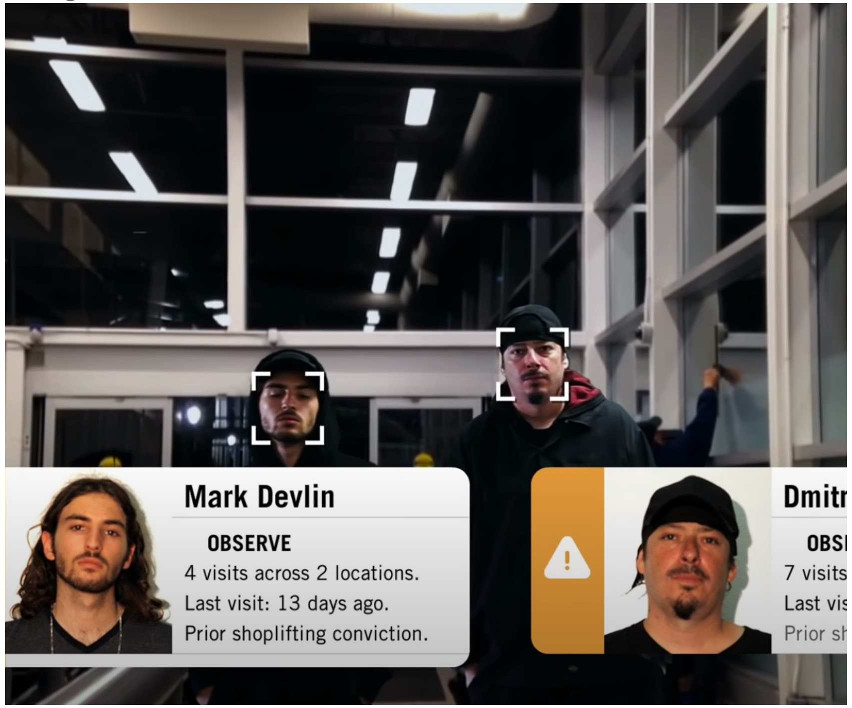


Image credits **FaceFirst**



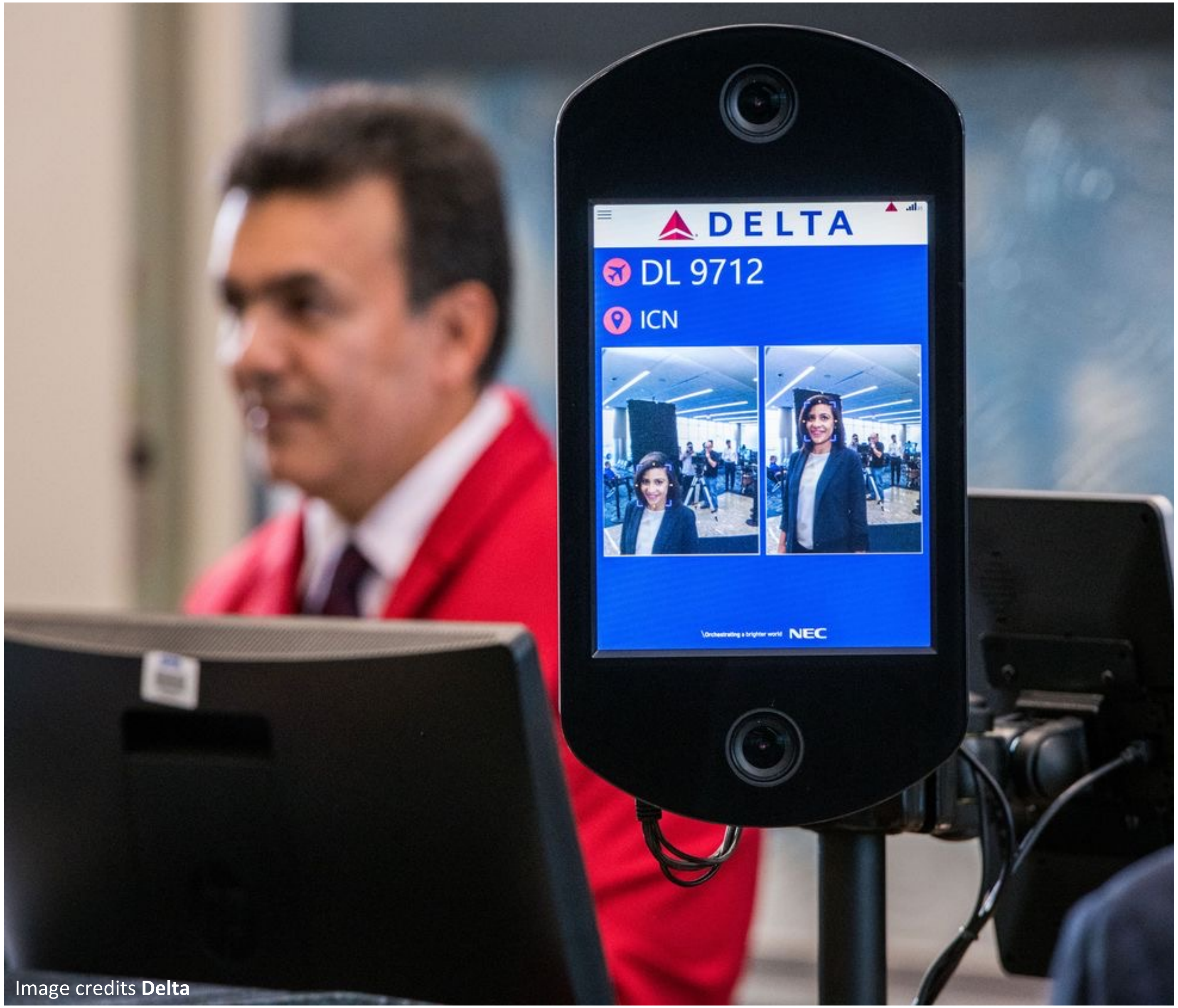


Image credits **Delta**

Current practice: privacy risk

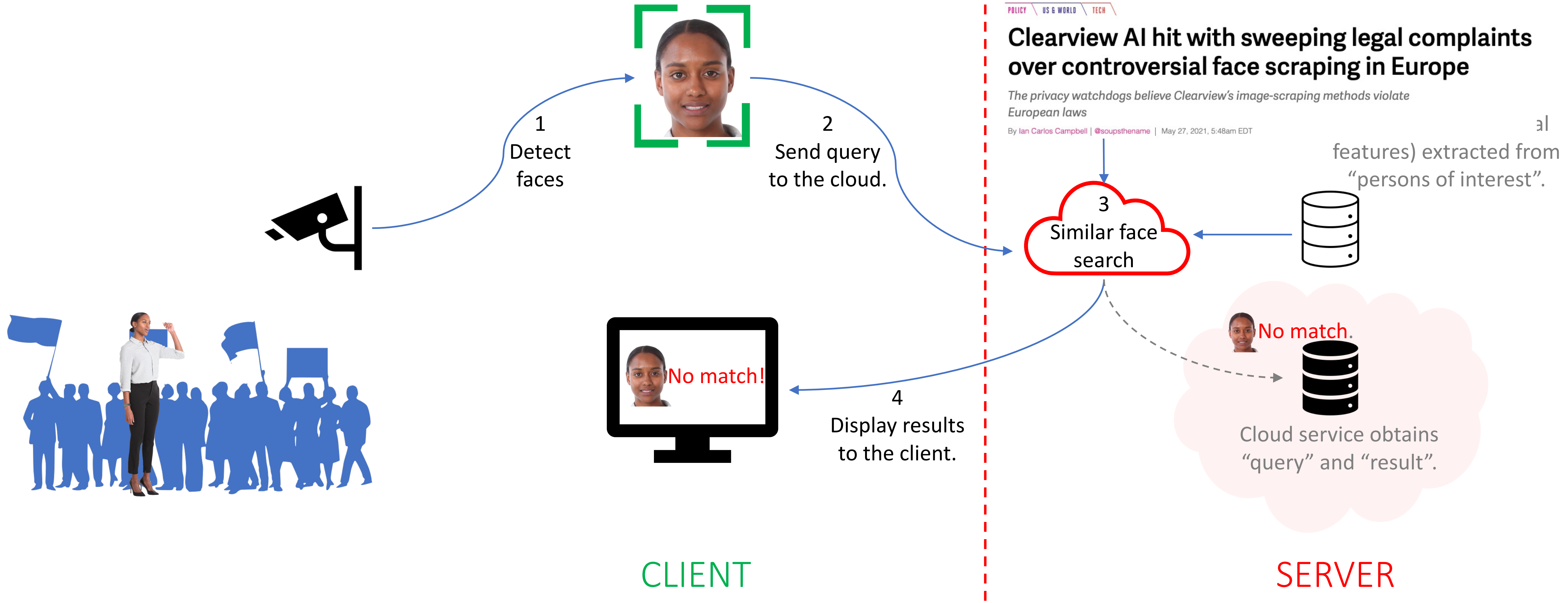
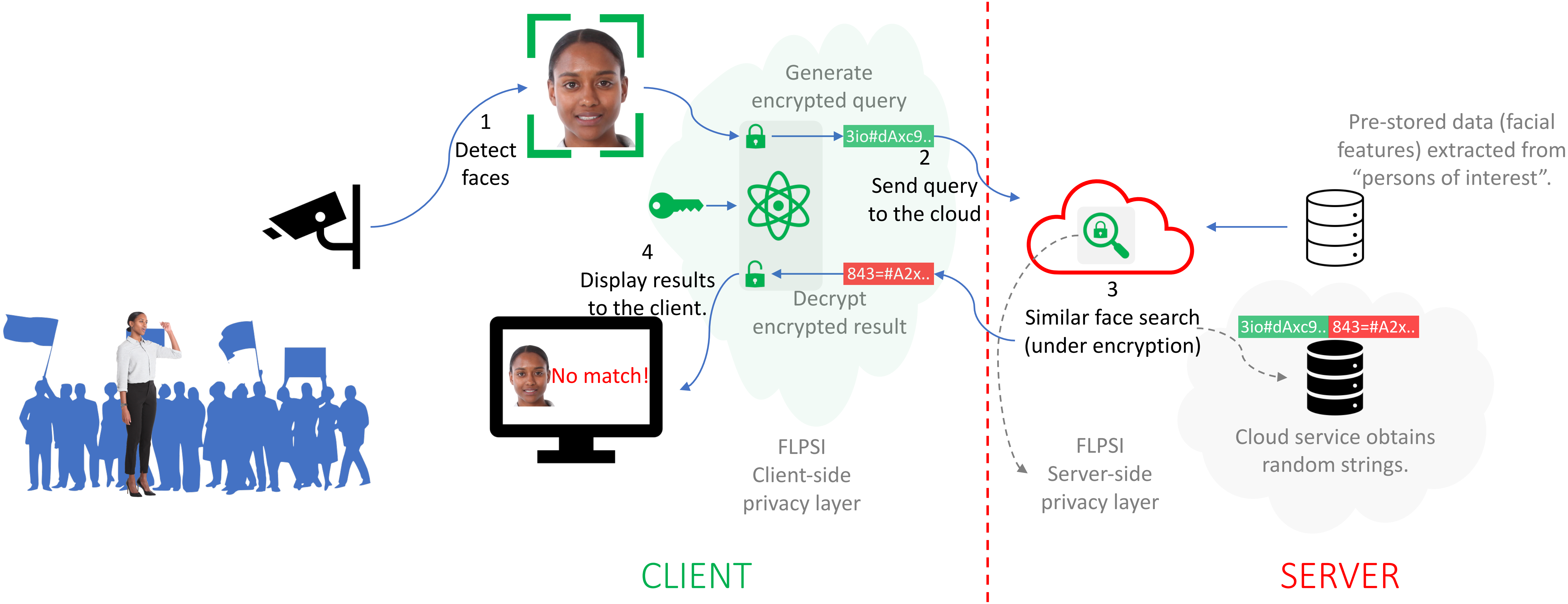


Image credits European Digital Rights

Image credits Financial Times

Ban vs Keep Using

Solution: Fuzzy Labeled PSI (FLPSI)



State-of-the-art

Exact private match: CHLR18

- (Labeled) Private Set Intersect.
  - E.g., contact list discovery
- Chen et al. (CCS’17, CCS’18)

  Sublinear communication.

  Efficient computation.

  Not directly be applied to fuzzy (e.g., biometrics) match.

Fuzzy private match: SANNS

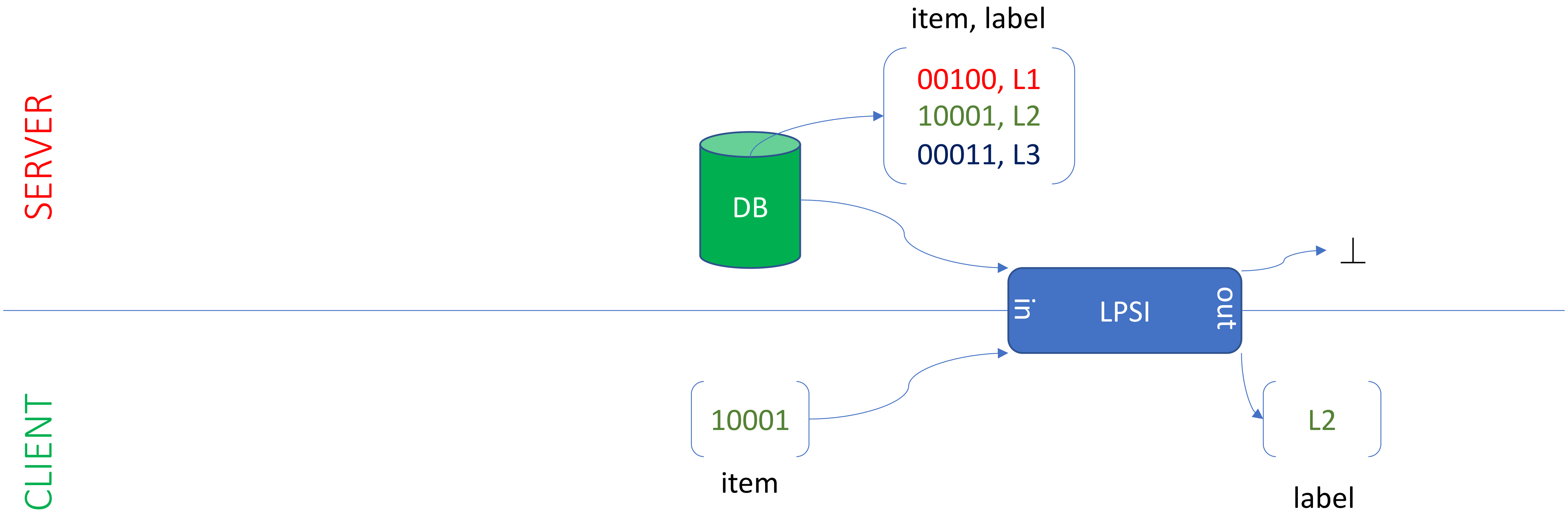
- Secure Approximate NNS
  - E.g., top-k closest embedding vector search
- Chen et al. (Usenix’20)

  Accommodate fuzzy matching.

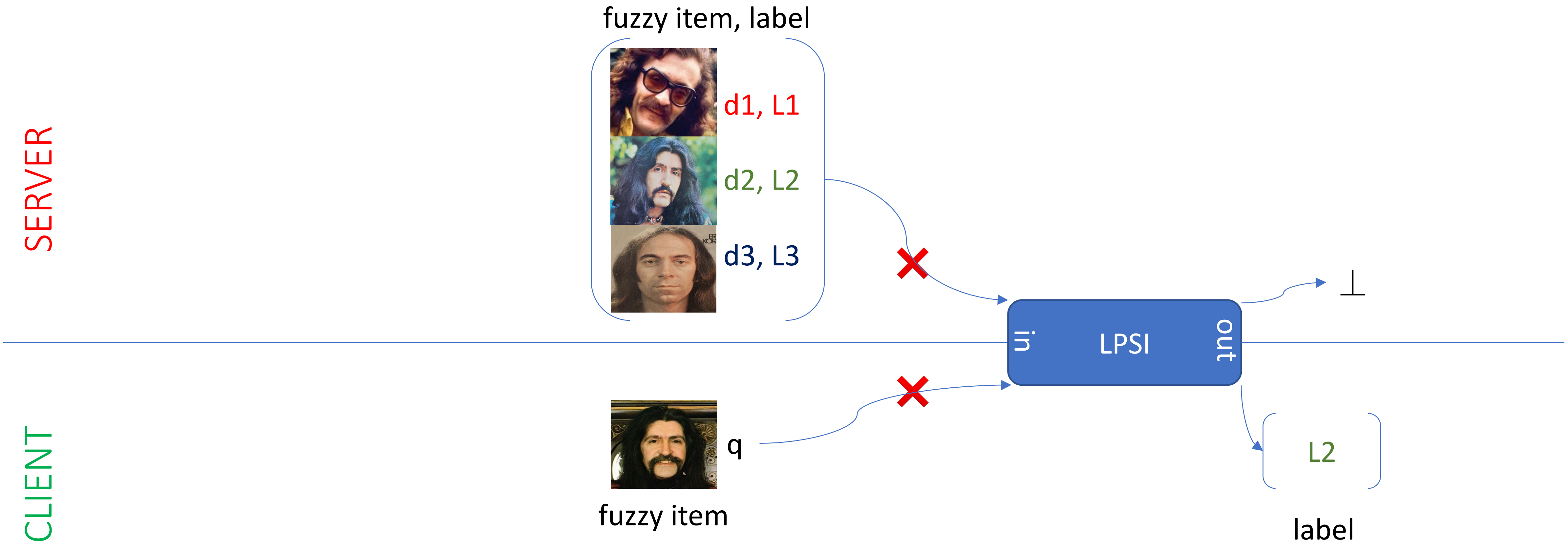
  High bandwidth requirement.
  - 1.7-5.4 GB communication to search a face over 1M-row DB.

Building FLPSI

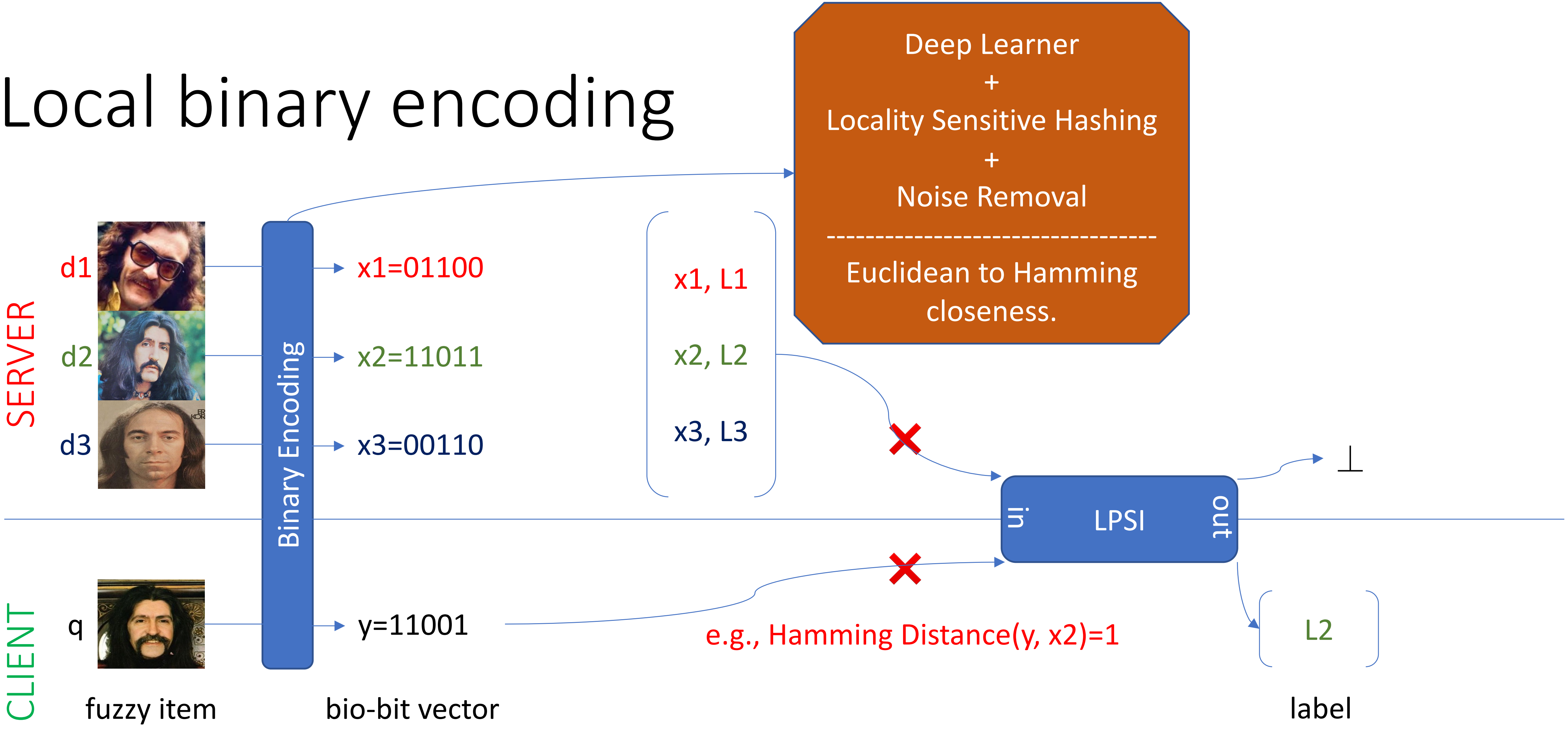
Accommodating exact matching



Accommodating exact matching



Local binary encoding



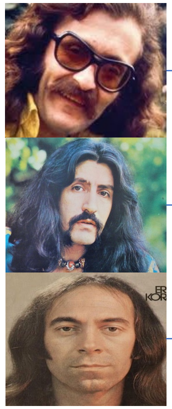
Subsampling

SERVER

d1 x1=01100

d2 x2=11011

d3 x3=00110



CLIENT

q y=11001

fuzzy item

Binary Encoding

bio-bit vector

Subsampling

S samples an AES key *k* and masks.

mask1=10101

mask2=11001

mask3=01010

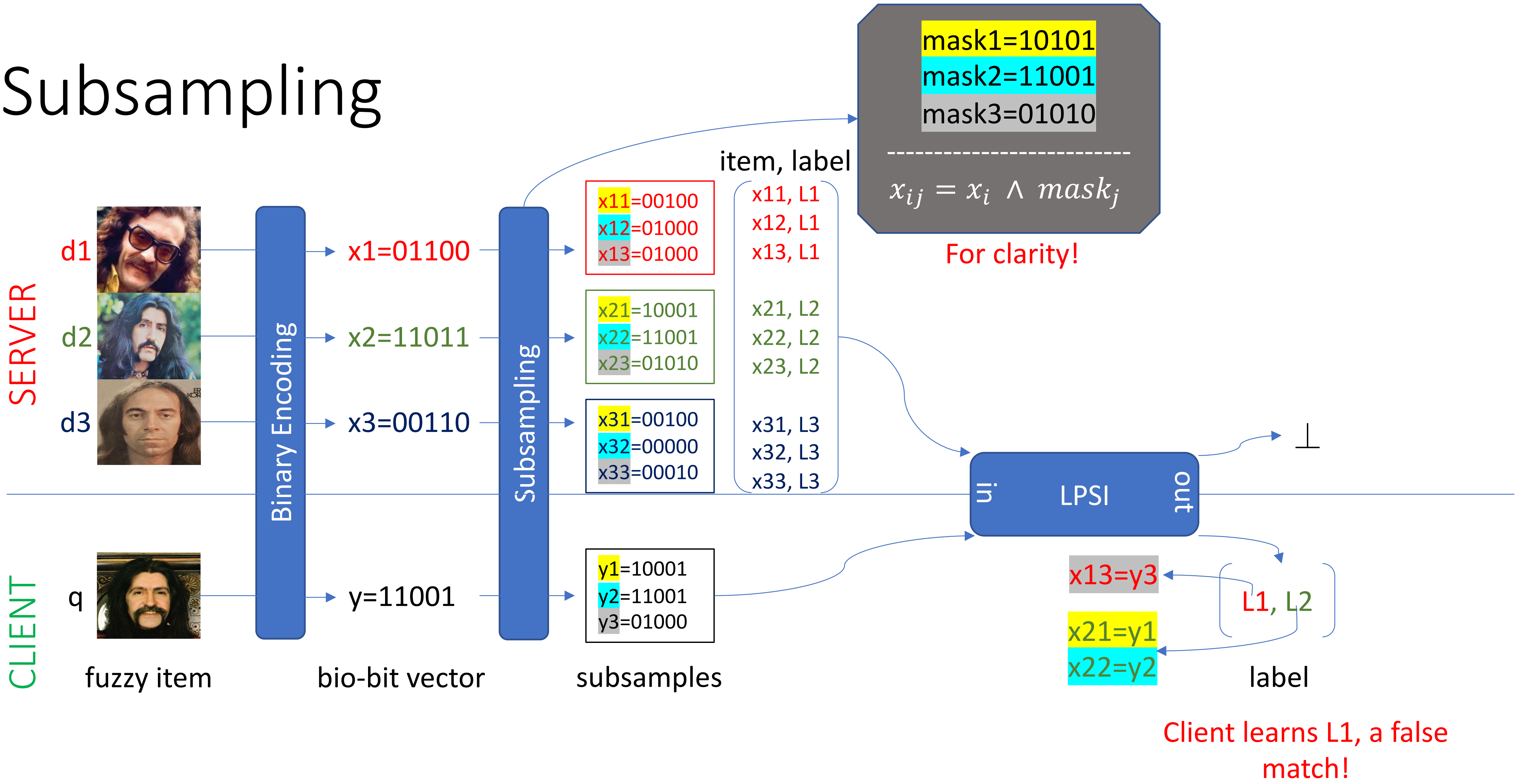
S locally computes:

$$x_{ij} = AES_k(x_i \wedge mask_j)$$

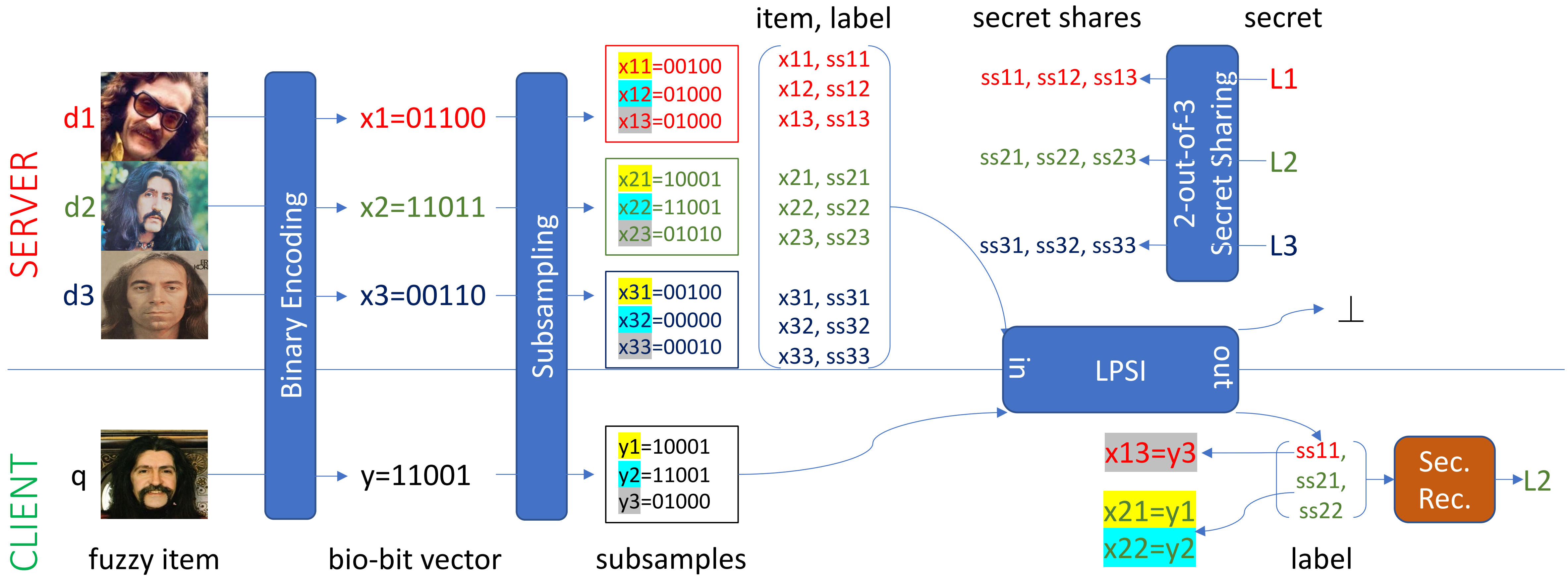
C and *S* computes via 2PC:

$$y_j = AES_k(y \wedge mask_j)$$

Subsampling

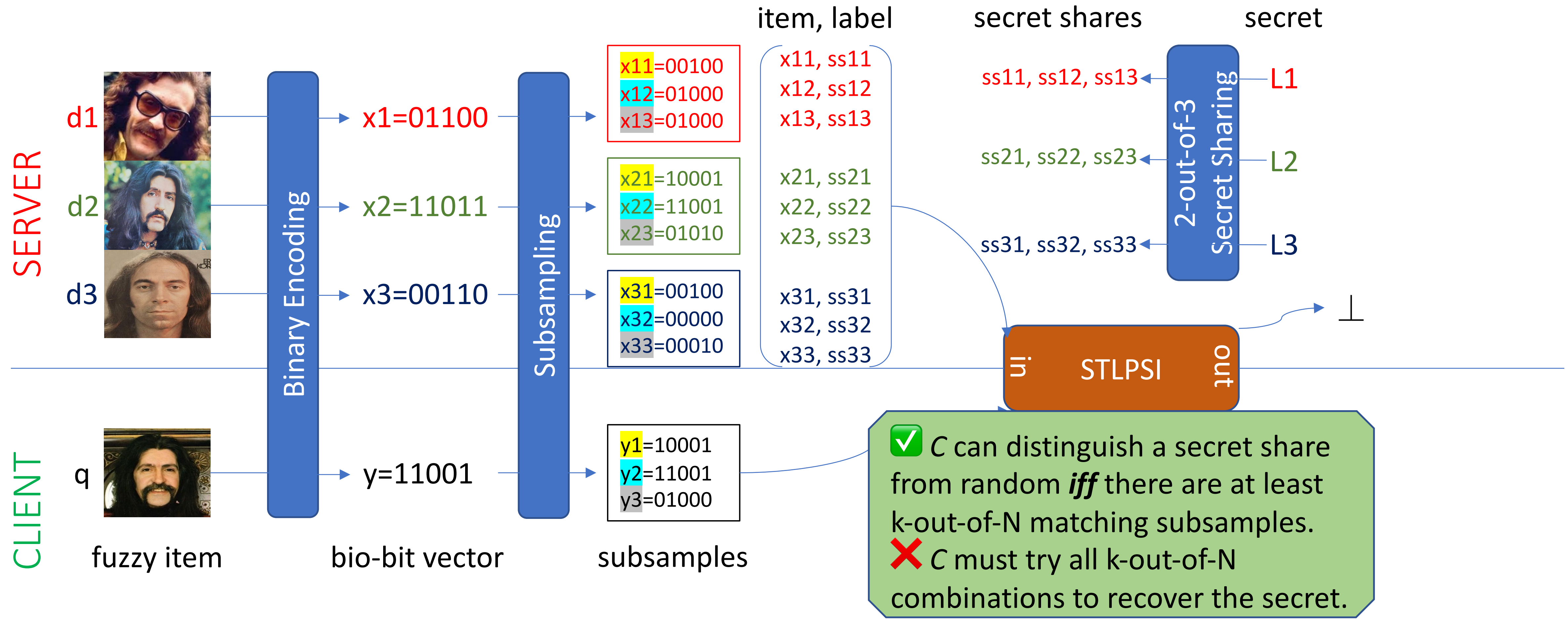


k-out-of-N Secret Sharing

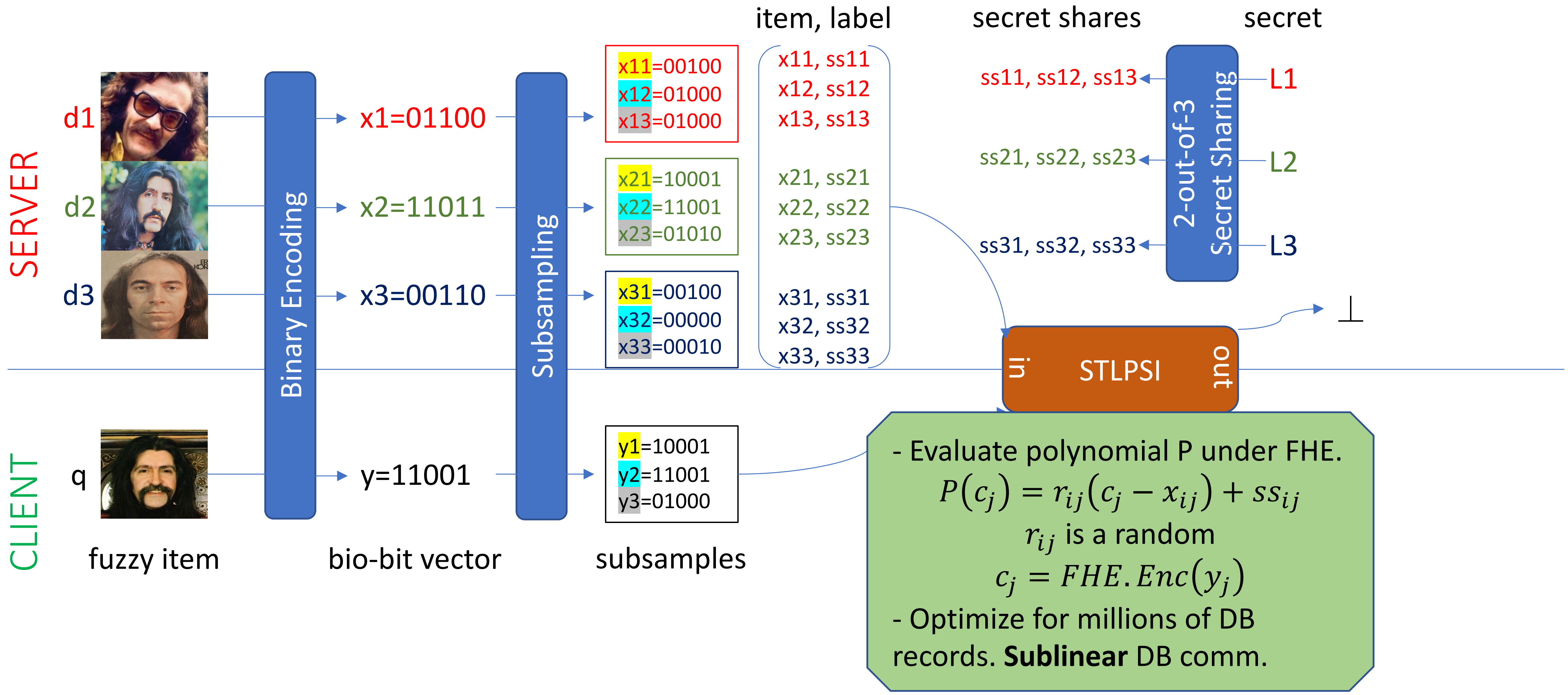


Client does not learn L1, but partial matches are **still** leaked!

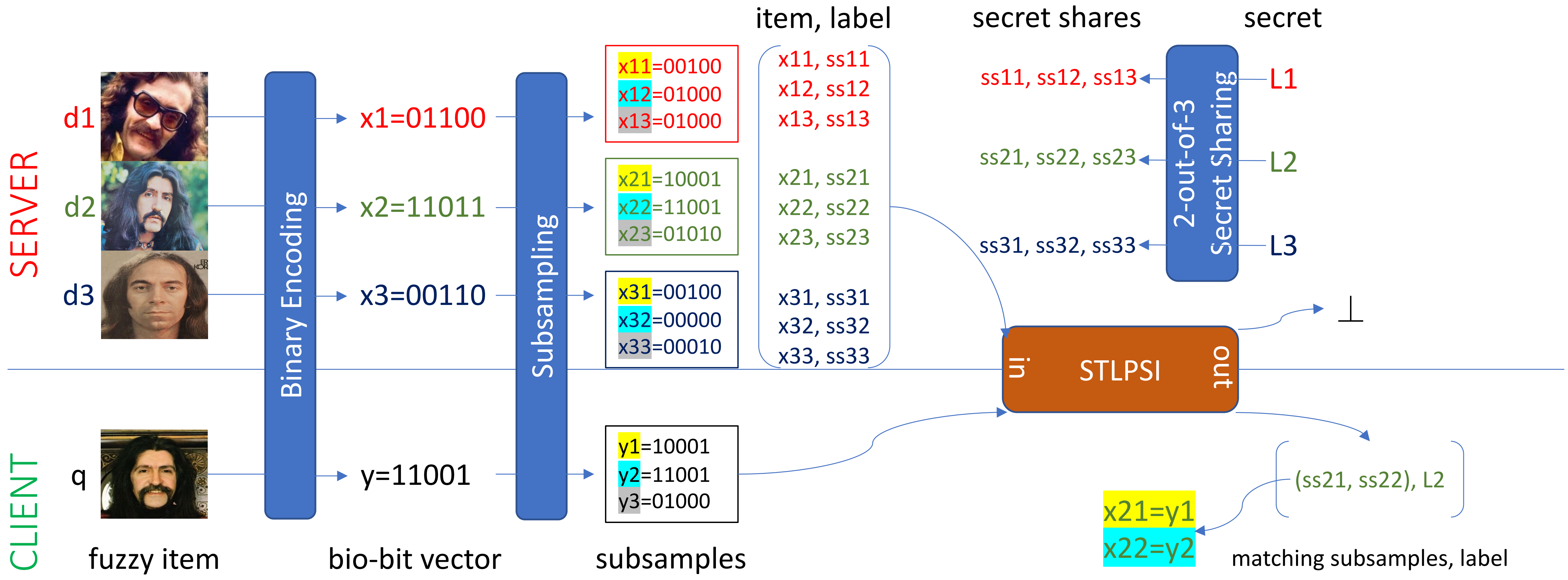
Set Threshold LPSI: k-out-of-N private match



Set Threshold LPSI: k-out-of-N private match

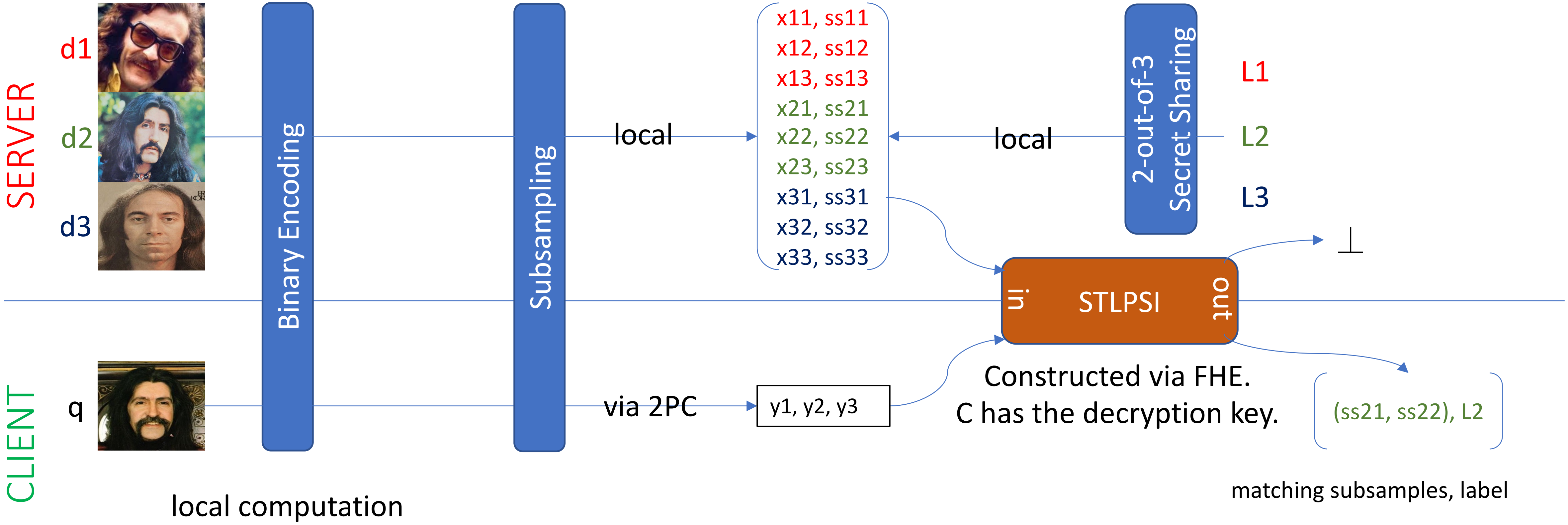


Set Threshold LPSI: k-out-of-N private match



Evaluating FLPSI

Security of FLPSI: in semi-honest model



Minor (allowed) leakage: C learns matching confidence.

Datasets

Used for	Query	Database
Face-1M	YouTube Face-YTF (1.6K)	YTF (1.6K) + StyleGAN (1M)
Deep1B-1M	10K image descriptors	1M
Deep1B-10M	10K	10M
AT&T	40 people	40

Environment and parameters

- Parameters are tuned to preserve plaintext accuracy.
  - 2-out-of-64 matching.
  - 0.67/0.75% of FRR for plaintext/FLPSI @10 false matches/query over Face-1M
- Same environment settings with SANNS (Chen et al. from Usenix'20).
  - Network settings: *fast* (500 MB/s) and *slow* (40 MB/s).
  - Azure F72s_v2 instance: 72 virtual cores, 144 GB of RAM

Performance results: Face-1M database

- Communication overhead : 40.8 MB
- Computation time:
  - @1 thread: 44 sec.
  - @72 threads: 1.36 sec.
- Best response time:
  - @fast network: 1.46 sec.
  - @slow network: 1.66 sec.

Comparison with threshold matching systems

Distance thresholding

- On AT&T dataset, single thread and same network speed (*fast*).
- Comparison with 7 systems.
  - 7.2x – 90x network save.
  - 121x – 7086x resp. time speed up.

k-out-of-N matching

- Asymptotic comparison with 3 systems.
- FLPSI is the *first* achieving communication sublinear to DB.

Comparison with kNN systems: SANNS

- The state-of-the-art: SANNS (Chen et al. from Usenix'20).
  - SANNS-linear: Searching over all DB items.
  - SANNS-approx: Searching over sub-DB items with slight accuracy penalty.

Database	**Protocol**	**Communication**		**Response time (fast/slow)**	
		Total	**Saving**	**Total (sec.)**	**Speed-up**
Deep1B-1M	FLPSI	40.8 MB	-	1.46/1.66	
	SANNS-linear	5.39 GB	132x	5.79/41.7	3.97/25.1x
	SANNS-approx	1.72 GB	42x	1.70/15.1	1.16/9.09x
Deep1B-10M	FLPSI	128 MB	-	12.7/13.5	-
	SANNS-linear	57.7 GB	452x	73.1/446	5.76/33x
	SANNS-approx	6.07 GB	48x	5.27/41.8	0.41/3.1x

Limitations

- Requires offline preprocessing before each query
  - 501 MB storage and 37.5 sec preprocessing for 1M-row DB.
- Client requires a public DL model.
- Not resilient against malicious attacks.
  - Server can return random outputs
  - Client can exploit allowed false matches to learn entire DB.
  - But prior systems are also semi-honest.

Questions?

Email:

euzun@gatech.edu

Project page:

https://sites.gatech.edu/euzun/projects/biometrics-surveillance