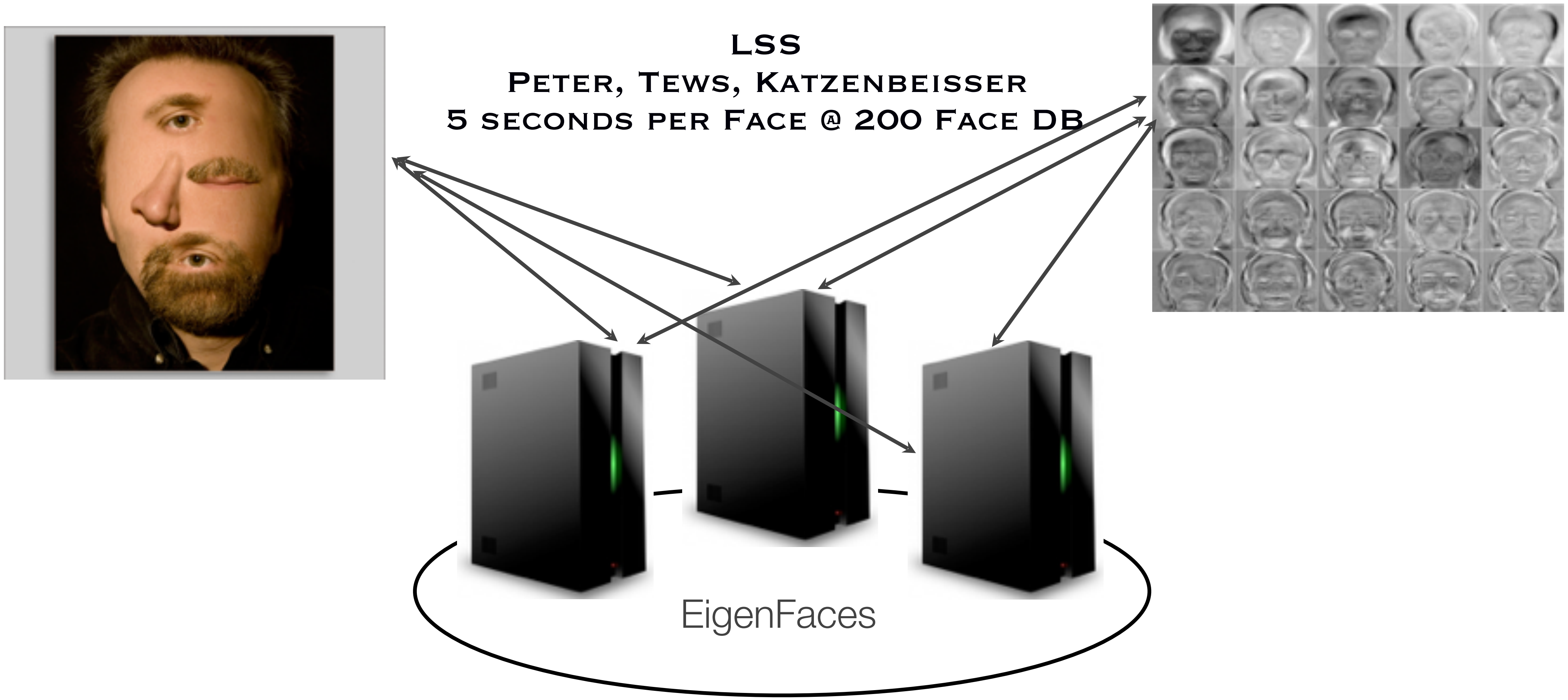
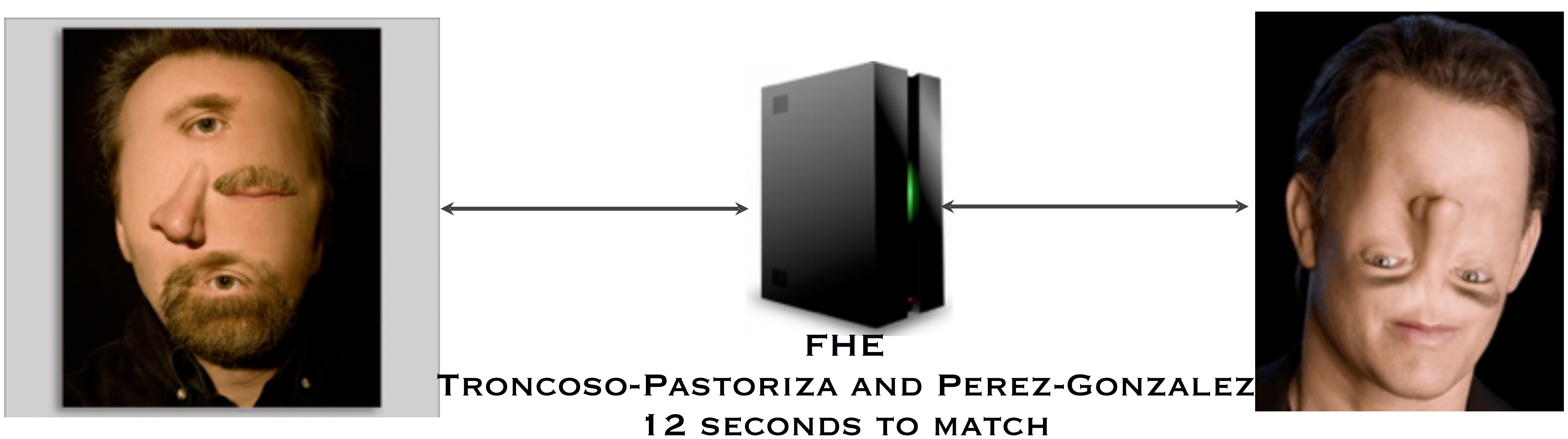
Practical Dynamic Composition in Secure Computation: VoIP and SQL

Mind the Gap!

Dave Archer
Galois, Inc.

Face Matching (LSS, FHE)





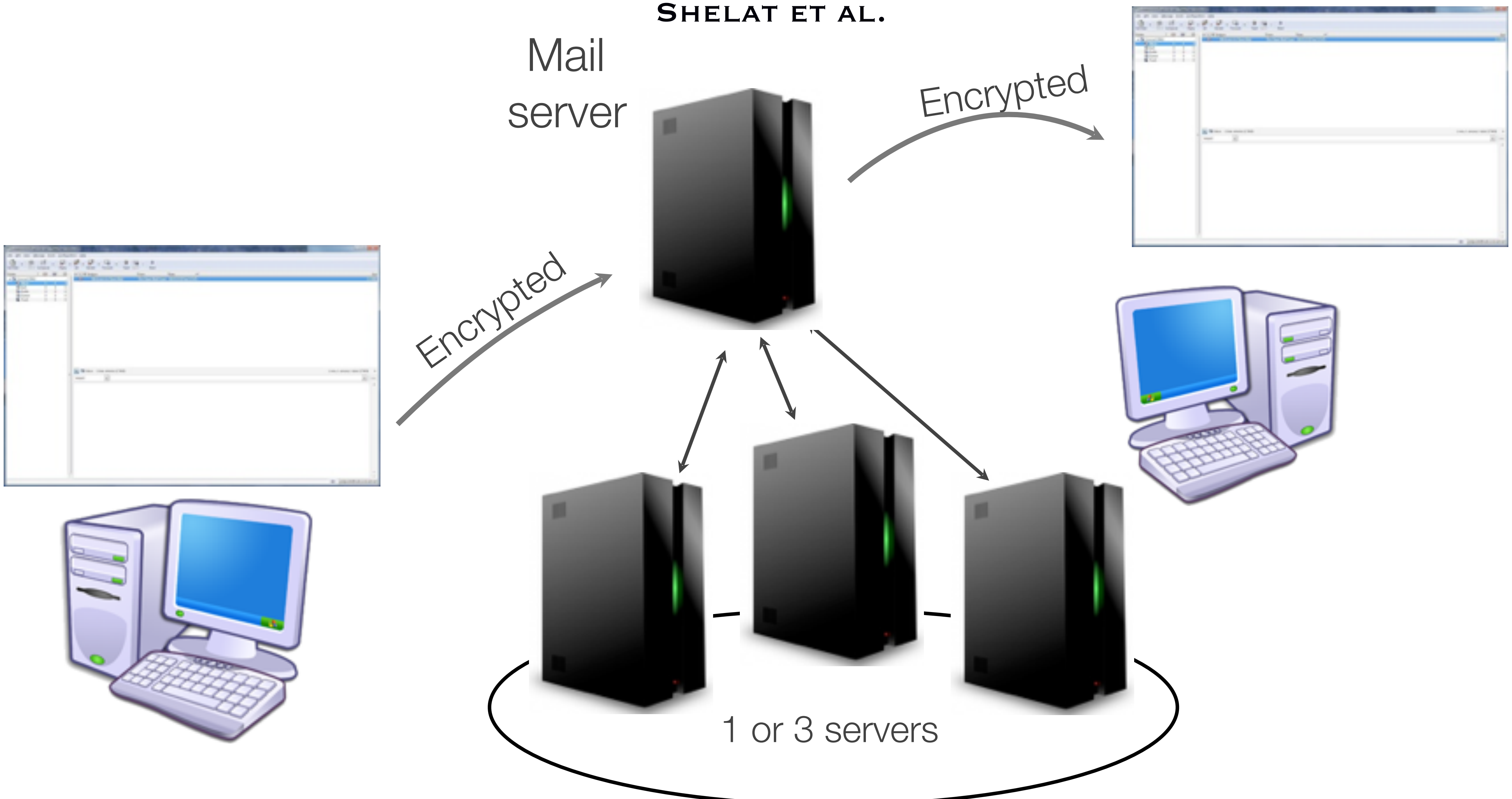
Scanning Encrypted e-Mail

TRAINABLE BAYESIAN SPAM FILTER @ ~256B / < 1 MINUTE
BOGDANOV ET AL.

FHE STRING MATCH DETECTION @ 256 BYTES / 10 SEC
ROHLOFF, COUSINS, ET AL.

LSS REGEXP MATCH DETECTION @ KBYTE/ 30 SEC
GALOIS

GC REGEXP MATCH DETECTION
SHELAT ET AL.



Linear Regression

3D Linear Regression
10k points in ~5 seconds
Galois

SHE linear regression
10k points in 280 seconds
Boneh at al.

HE to aggregate inputs, GC to solve CHOLESKY DECOMPOSITION
100M 20D ridge regression in 10 hours
Nikolaenko, Weinsberg, et al.

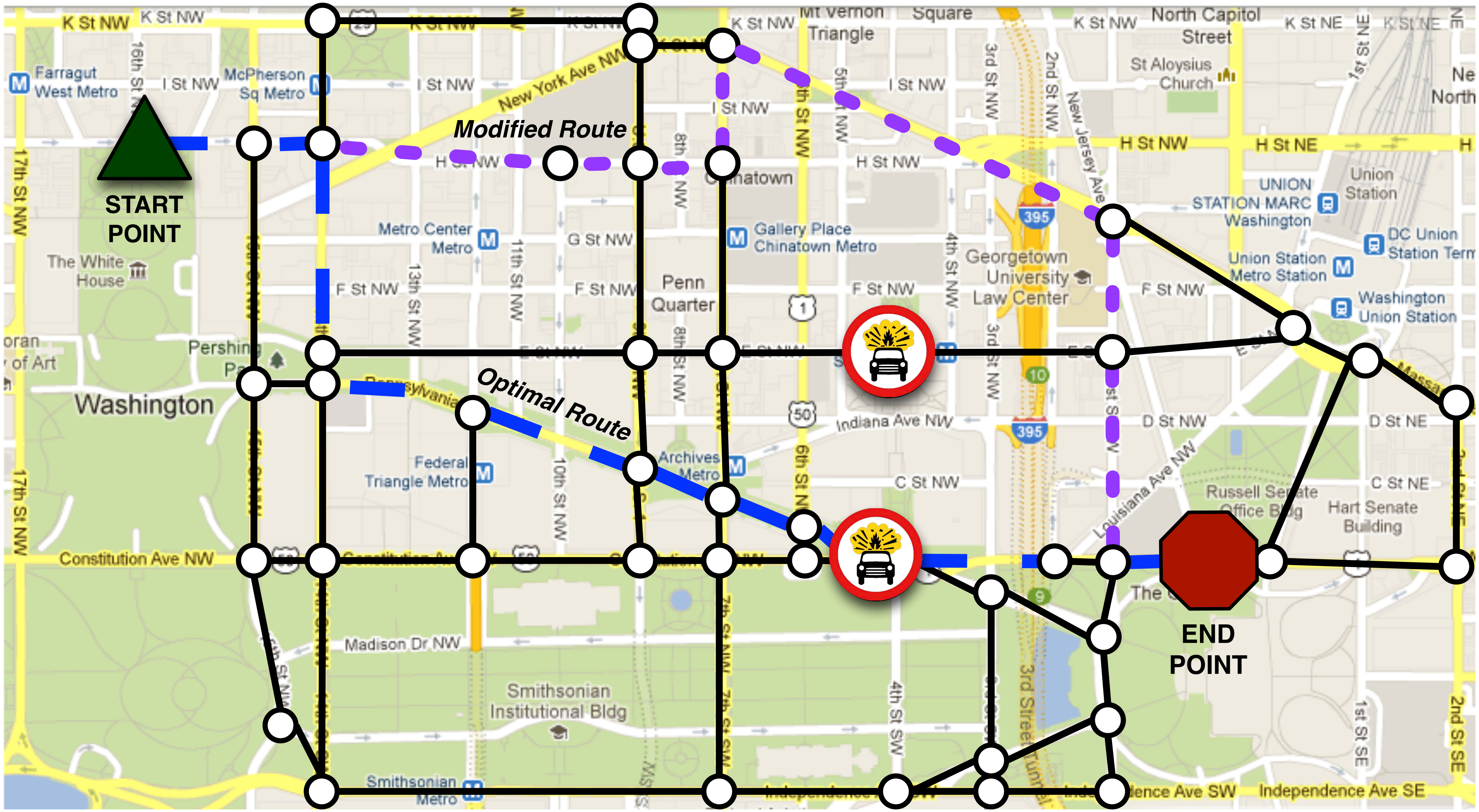
$$m = \frac{(n\sum xy) - (\sum x \sum y)}{(n\sum x^2) - (\sum x)^2}$$

$$b = \frac{(\sum y \sum x^2) - (\sum x \sum xy)}{(n\sum x^2) - (\sum x)^2}$$

$$y = mx + b$$

Route Mapping

Garbled circuits with outsourcing to Cloud servers
Butler, Traynor et al.

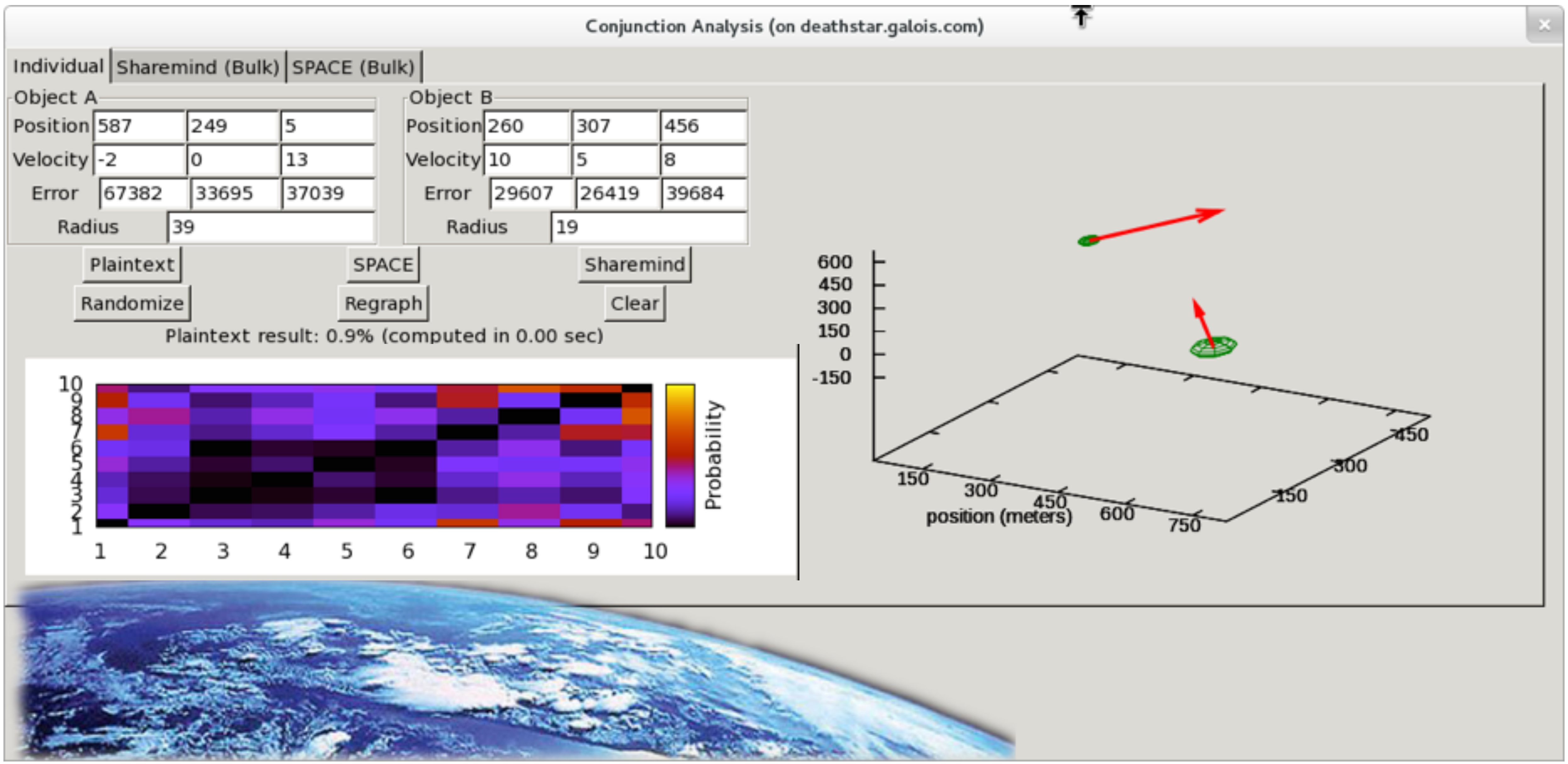


Satellite Conjunction Analysis

2-party
Ostrovsky and Lu

LSS - bulk analysis & Heat map
Bogdanov et al.

$$P = \frac{1}{\sqrt{8\pi}\sigma_x} \int_{-R}^{R} \left[\mathrm{erf}\left(\frac{y_m + \sqrt{R^2 - x^2}}{\sqrt{2}\sigma_y} \right) + \mathrm{erf}\left(\frac{-y_m + \sqrt{R^2 - x^2}}{\sqrt{2}\sigma_y} \right) \right] \exp\left(\frac{-(x + x_m)^2}{2\sigma_x^2} \right) dx$$



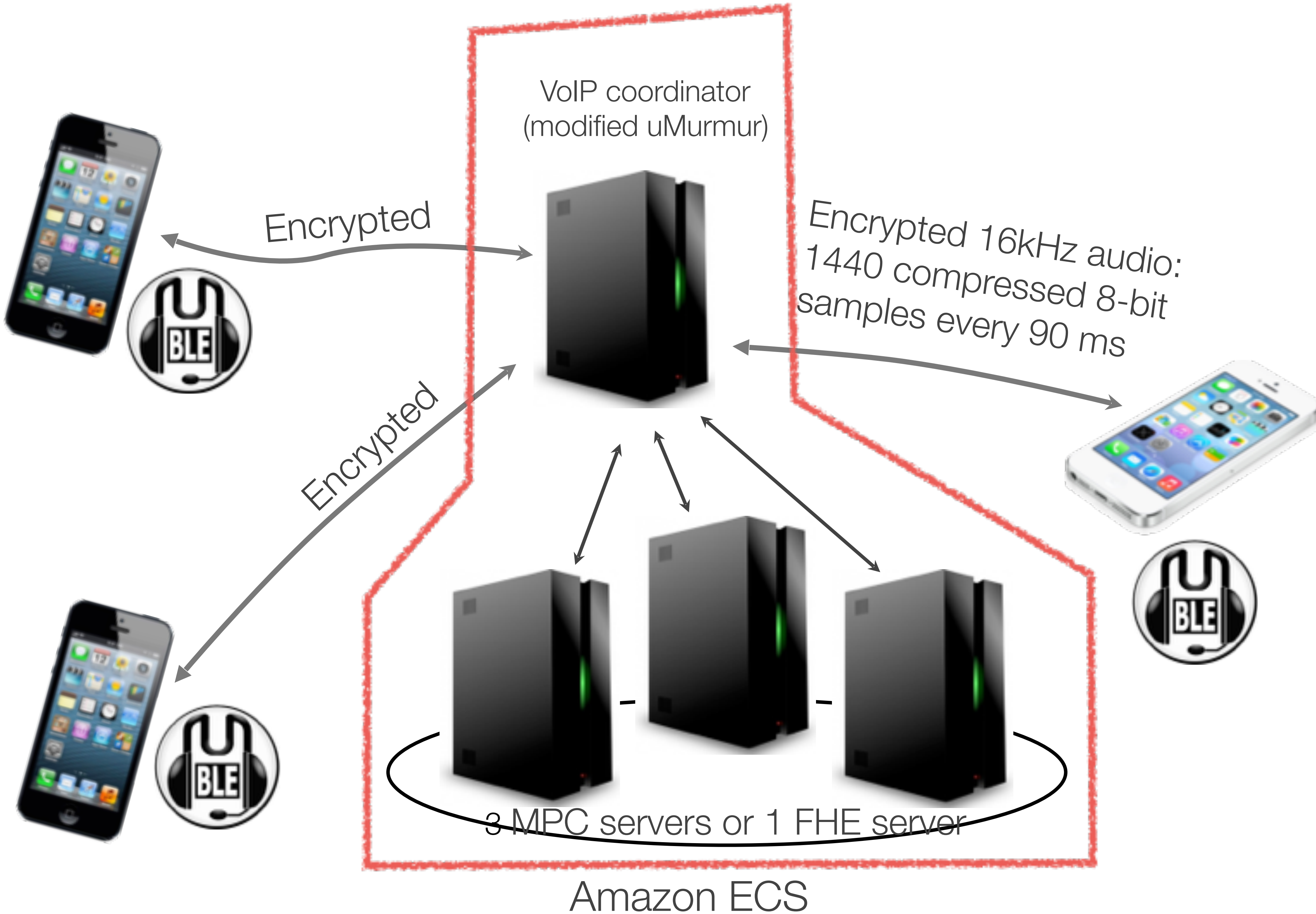
And so on…

- Distributed Shared HMAC verification
- Distributed Shared Digital Signatures
- Secure statistical analytics for Differential Privacy Computation
- …

End-to-End Encrypted VoIP Conferencing

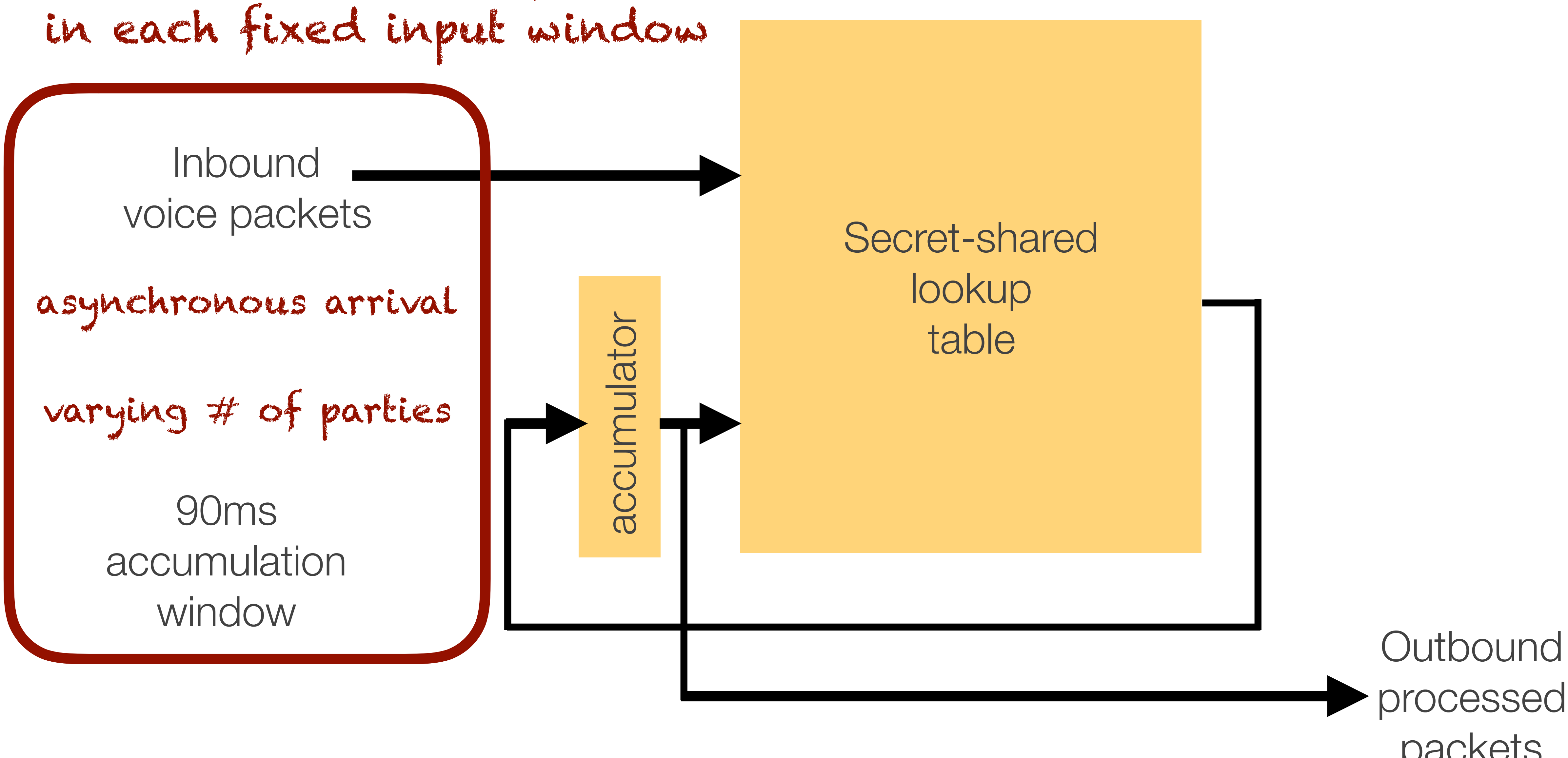
FHE
Rohloff, Cousins, et al.

LSS - 4 voices @ streaming 12kb/s audio
Galois



Streaming Data, Real-time Constraints

Problem: variable sample count in each fixed input window



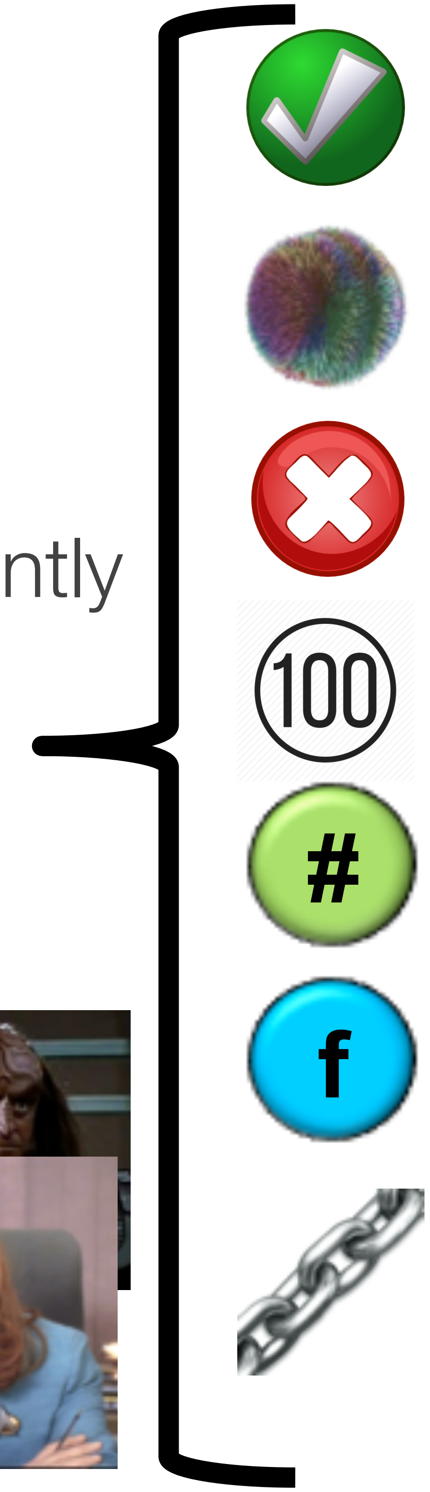
Solution: "Help out" the cryptographic solution:

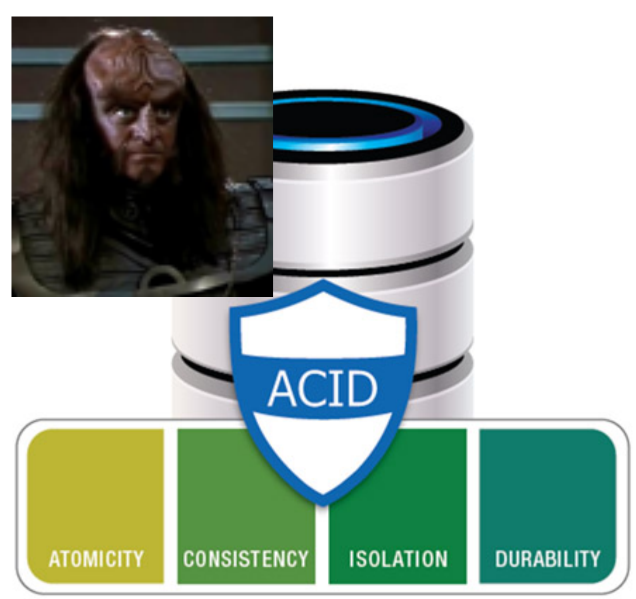
Impose external control loop and persistent state

Error-prone, awkward and complicated

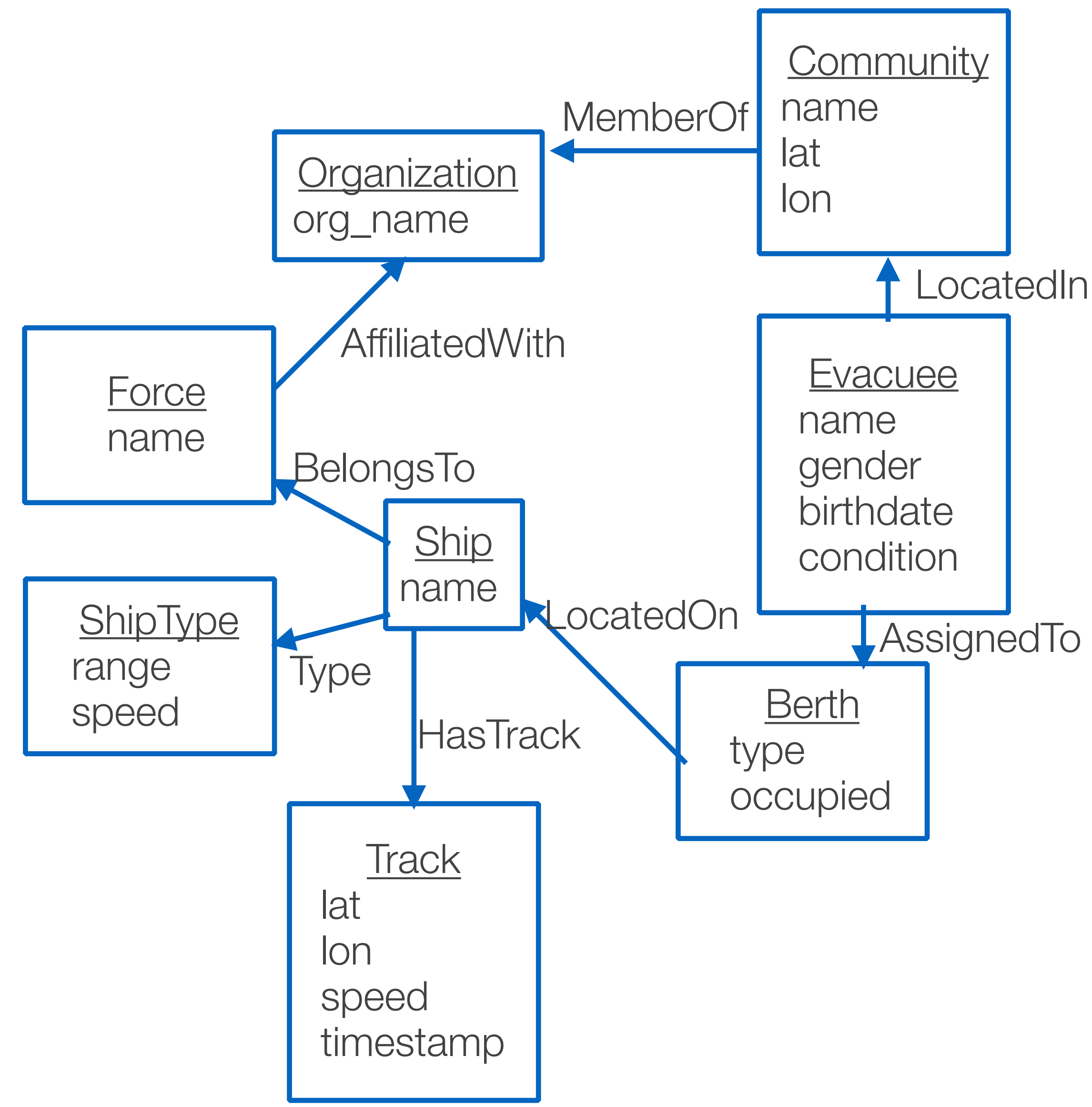
PDaaS

- only by people I authorize
- only under conditions I allow
- never for evil, and
- only when documented transparently



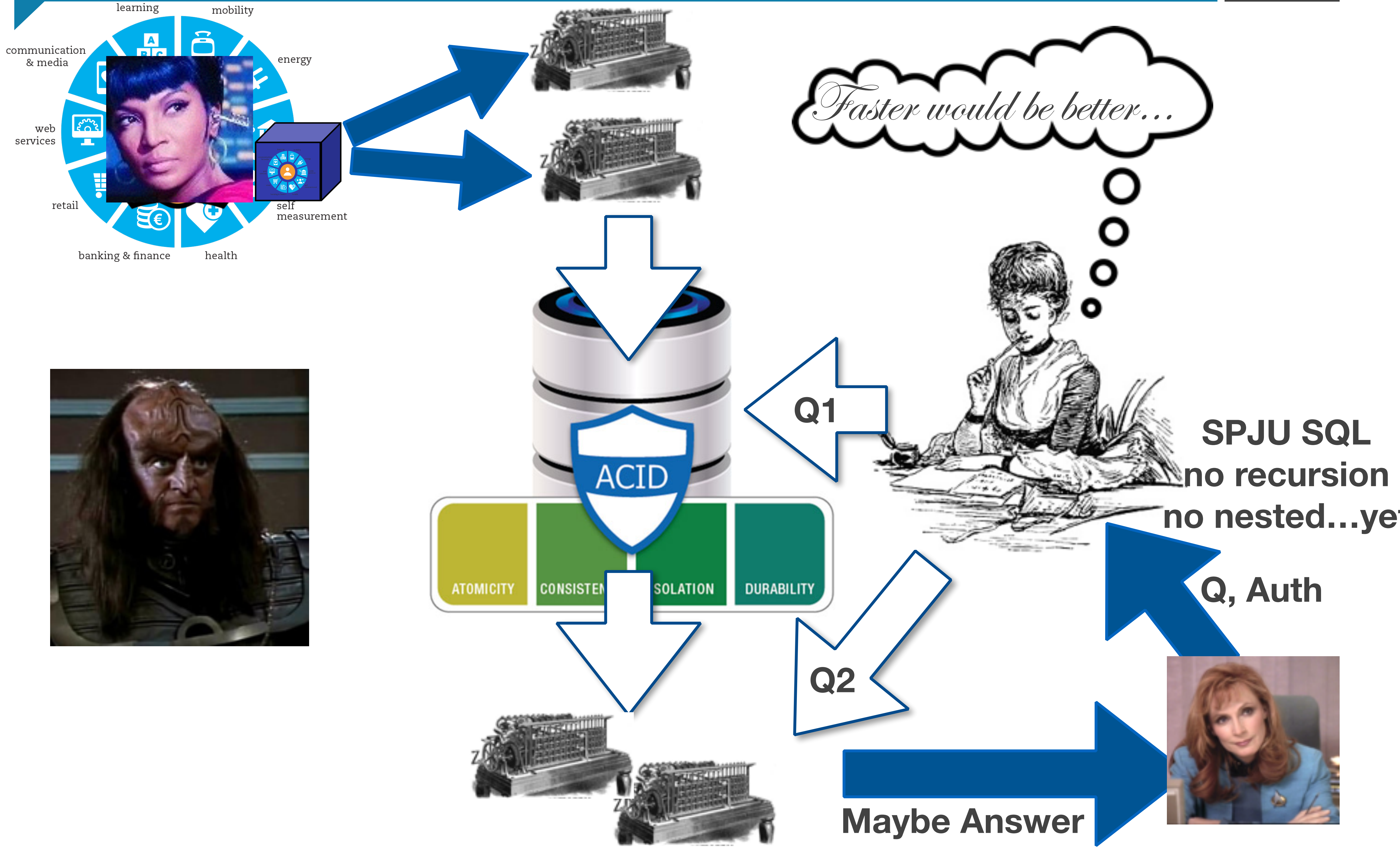


A Sample Use Case




```
SELECT
    ship_name,
    stype AS ship_type,
    count(*) AS Number_Of_Berths,
    distance/speed AS eta
FROM ship
     JOIN track ON ship.id = track.ship_id
     JOIN shiptype ON ship.shiptype_id = shiptype.id
     JOIN hasberth ON ship.id = hasberth.ship_id
     JOIN berth ON berth.id = hasberth.berth_id,
     community
WHERE (community.community_name = 'Bajor') AND (occupied = FALSE) AND (ship_time = 25)
GROUP BY ship_name, stype, ship_lat, ship_lon, community_lat, community_lon, cruisingspeed
ORDER BY eta;
```

Jana


learning
mobility
communication & media
energy
web services
retail
self measurement
banking & finance
health
Faster would be better…
Q1
ACID
ATOMICITY
CONSISTEN
SOLATION
DURABILITY
SPJU SQL
no recursion
no nested…yet
Q, Auth
Q2
Maybe Answer

Partial Query Processing in SPDZ

Decrypt to Shares

Sort

Join

Select

Project

Aggregate and apply differential privacy

Encrypt for user

Large record counts from multiple relations

if ORE unavailable in relational DB

SPDZ "plaintext RAM" computation model

- Arbitrary RAM computation (registers and heap)
- Does not hide access pattern (does it matter?)
- Memory statically allocated at compile-time
- Unbounded updates to registers and heap
- Direct/ indirect heap access, direct register access
- Unbounded re-use and update during program

- Reactive: inputs and program based on prior outputs
- Sub-linear access - not all data need be touched
- Loops need not be unrolled — reduced program size

A bit racy (in leakage),
but simpler than unrolling circuits

Simple-minded Conclusions

- RAM model often appears useful in “real-world” computations
  - Dynamic iteration count in loops
  - Sub-linear access cost for data
  - Reactive computation during run-time
- ORAM model may be too rich in security, too slow in practice
- Circuit model impractical for realistic data sizes, dynamic computation