Lecture 11: MPC in Practice

MIT - 6.893

Fall 2020

Henry Corrigan-Gibbs

Plan

* Breakout rooms (discussion on last guest lecture)
* Recap
* Examples of MPC in practice

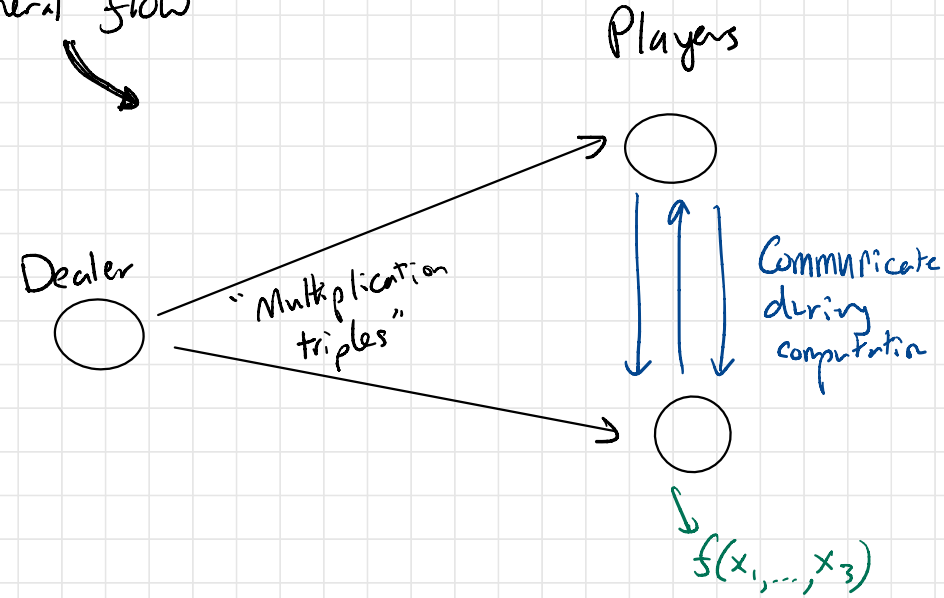
Logistics

* HW4 out now. Due 10/19 at 5pm via Gradescope

Recap: 3PC Protocol

- Express computation $f(x_1, x_2, x_3)$ as an arithmetic ckt

- General flow



- Key idea:
  - * Players start out holding additive shares of all inputs
  - * Players walk through ckt one gate at a time
    - INVARIANT: At t-th time step, Players hold additive shares of values on wires $\{1, \ldots, t\}$ in ckt.
  - * Finally, players broadcast shares of output wire val'
    - ↳ $f(x_1, \ldots, x_3)$

- To show security: Adv that controls one player sees only random vals until end... easy to simulate.

Today: Cover a few example applications of MPC.

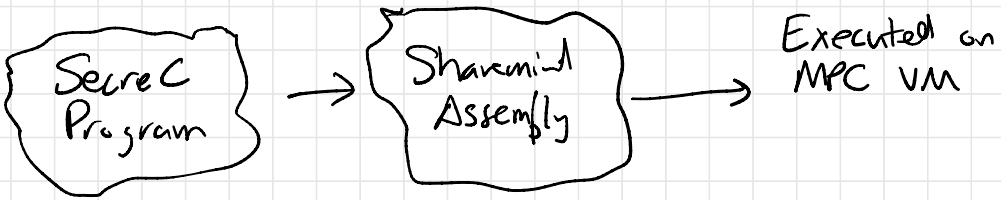
Warning: This is not an exhaustive list.

Things you will notice...

* many of these are proofs of concept, rather than deployed systems
* there are not too many examples (you'll see the same handful over and over)
* it's not easy to get even these pilot systems working in practice.

Sharemind (Estonia)

- Basic protocol:
  - 3PC w/ info-theoretic security
  - Semi-honest secure, honest majority
  - Quite similar to 3PC protocol we saw
- Programmer writes MPC routine in "SecreC" and compiles it to Sharemind assembly



SecreC is like C but with "public" and "private" types
... here "private" = cryptographically private

↳ Limitations on what you can do w/ private values (no branches, etc.)

↳ But, nicer to work with than ckts, since can interleave arbitrary computation with MPC

Application: Data Analysis

Bogdanov et al. (2016)

- Real use described in PETS 2016 paper
- In years 2008-2012 43% of students in ICT course in Estonia dropped out
  - ↳ B/c they got hired away from school?
  - ↳ or B/c courses were too hard?

(ICT = Inf. Comm Tech)

- Could get education data by students from Ministry of Ed
- Could get employment data from Min of Finance

↳ Legal barriers prevent analysis of joint data set

Ed

PersonID	ICT?	# months in school

Finance

PersonID	# months worked	ICT job?

$f($ ↓ , ↙ $) =$ { % of students working in ICT jobs who are/aren't studying ICT in school.

Application: Data Analysis

- Lots of legal barriers
  - Is an MPC "processing personal data"?
- Computation b/w 3 parties
  - Est Inf System Authority
  - RMIT (Finance)
  - Cybernetica (co behind Sharemind)
- Computation: ≈ 600k ed records
  ≈ 10M tax records

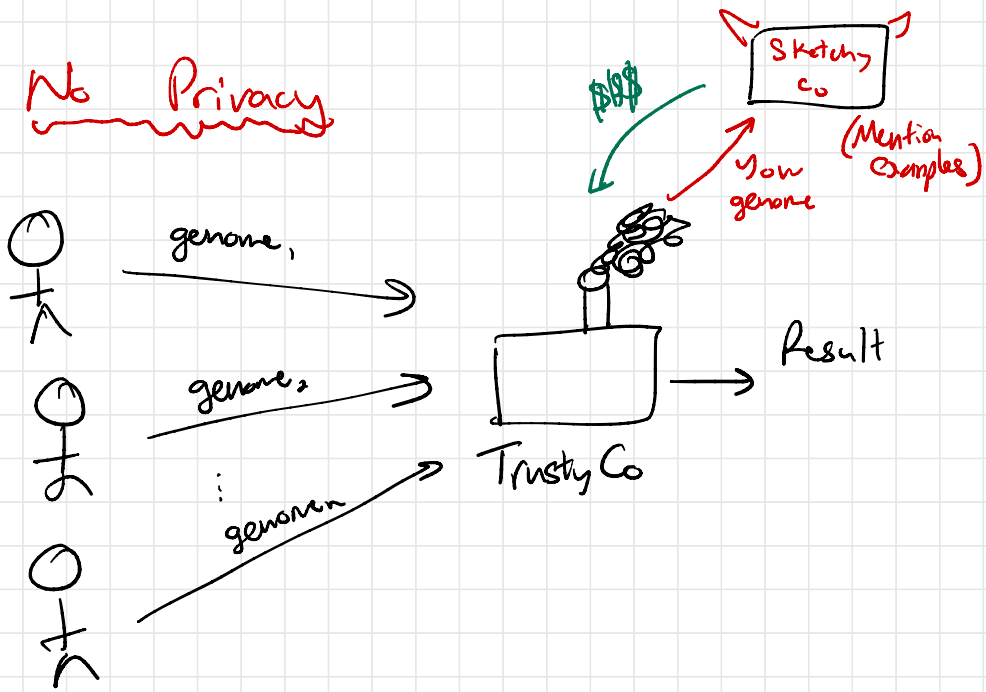
  → 16 days of computation over WAN
- How do you know whether you got the right answer???
  - cross-validation is important, but sometimes not so easy.

Genomics (proof of concept)

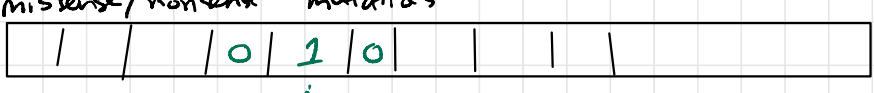
Jagadesh et al. (2017)

Problem: There are rare genetic mutations that cause rare diseases.

If you have a group of 15 people with rare disease X, how can you identify if there's a genetic mutation that they all share?

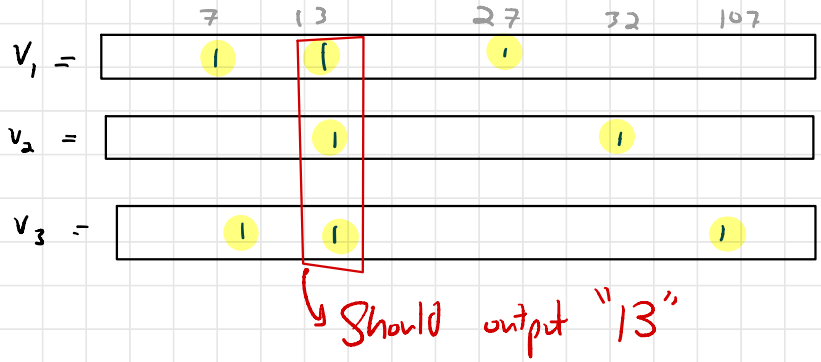


Each genome is encoded as dim = 28m vector of missense/nonsense mutations



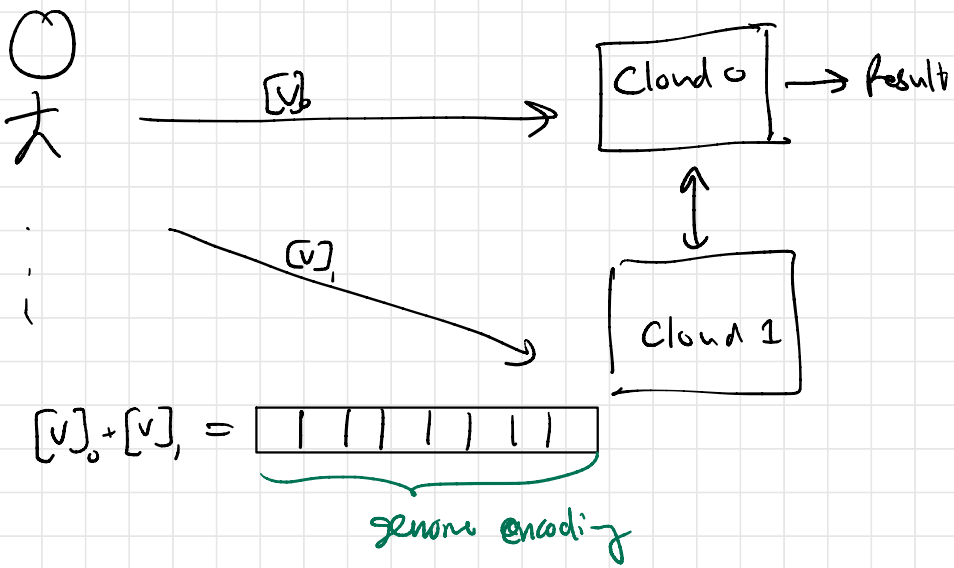
Genomics

Example: 3 participants



MPC Approach

Based on Yao's Garbled Circuits (2PC)
↳ semi-honest



Computation

$$f\left(\begin{matrix}[v_1]_0,\dots,[v_n]_0\\ [v_1]_1,\dots,[v_n]_1\end{matrix}\right) = \text{argmax} \sum_{i=1}^{n} [v_i]_0 + [v_i]_1$$

Genomics

Executed over WAN with ≥10 genome encodings

< 60 seconds, < 100 MB traffic

REMINDER...

Please use your cryptography skills to do good in the world.

Google (apparently deployed but hard to get specifics)

Ad conversion measurement...

"Ad Supplier" (Google)

Knows which ads you saw & clicked

Advertiser (e.g. Macy's)

Knows how much you spent on shoes.

Q: How many users saw an ad, and spent $, and how much did they spend total?

Notice that you might see an ad online but make a purchase in store, or on diff device.... using referrer headers may not be enough.

	Google	Macy's
INPUT	$Viewers \subseteq Users$	$\{(buyer_i \in Users, spend_i)\}$
OUTPUT	$\lvert Viewers \cap Buyers \rvert$	$\sum_{buyer_j \in Viewers} spend_j$

Google

Construct special-purpose semi-honest secure 2PC for this functionality.

↳ Cheaper + faster to avoid malicious security (Legal enforcement)

Daily: ≈ 1,000 protocol executions per advertiser
≈ 100k records per run

Each execution: ≈ 10 MB of data
≈ 12 mins of compute