

Avoiding Genetic Racial Profiling in Criminal DNA Profile Databases

Jacob Blindenbach, Karthik Jagadeesh, Gill Bejerano, and David Wu



CODIS DNA Profiling



Check for match



FBI Criminal Database

“On April 21, 2021, [...] the 20 millionth DNA profile was contributed to the national DNA database via the CODIS software [...] This remarkable crime-solving tool has aided over 545,000 investigations”

FBI.gov, April 21th, 2021

The New York Times

How Commandos Could Quickly Confirm They Got Their Target

American officials identified the body of the Islamic State leader Abu Bakr al-Baghdadi swiftly, even after he blew himself up.

New York Times, October 27th, 2019

CODIS DNA Profiling



Check  for match



FBI Criminal
Database

After query:



Added to database

Regardless of match result



FBI Criminal
Database

What is the Problem?

N.Y.P.D. Detectives Gave a Boy, 12, a Soda. He Landed in a DNA Database.

The city has 82,473 people in its database. Many of them have no idea their genetic information is there.

New York Times, August 15th, 2019

The New York Times

'Race-Biased Dragnet': DNA From 360 Black Men Was Collected to Solve Vetrano Murder, Defense Lawyers Say

New York Times, March 31st, 2019

The Washington Post
Democracy Dies in Darkness

The Post's View · Opinion

You need a good reason to curb privacy. None exists for collecting DNA at the border.

Washington Post, January 11th, 2020

NYPD's 'Knock-and-Spit' DNA Database Makes You a Permanent Suspect

Newsweek, February 11th, 2019


Approach

- **Privacy preserving cryptographic protocol**
 - **Protects privacy of collected DNA profile**
 - **Protects privacy of DNA profiles in database**
 - **Works in the field (does not take hours)**



 only sees match result 

 does not learn what is in 

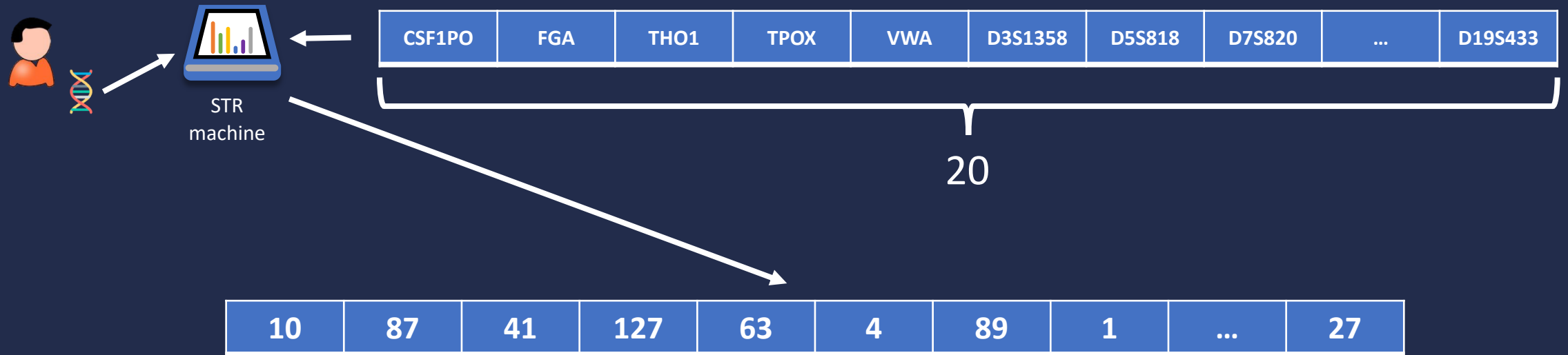
 does not learn what is 

Protocol Overview

- **Private DNA Matching → private evaluation of Finite State Machines (FSM)**
 - **Evaluating equality**
 - **Evaluating number of matching components**
- **Obliviously evaluate FSM with Oblivious Transfers (OT)**
- **Iterative protocol that uses problem's structure**
 - **Unlike Yao / FHE / Private Equality Testing**

STR Profiles

Short Tandem Repeats (STR)



STR Profiles



10	87	41	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----



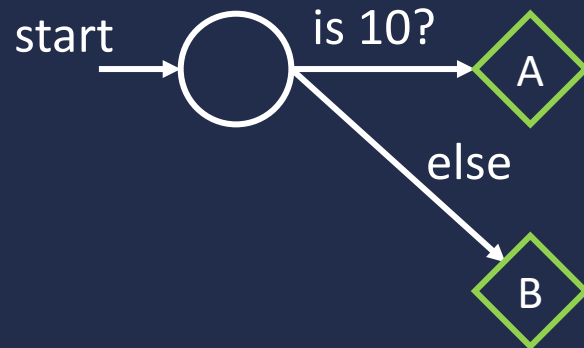
10	87	41	127	63	4	89	1	...	27
10	87	28	127	63	4	89	1	...	27

Full
Partial

Creating the FSM



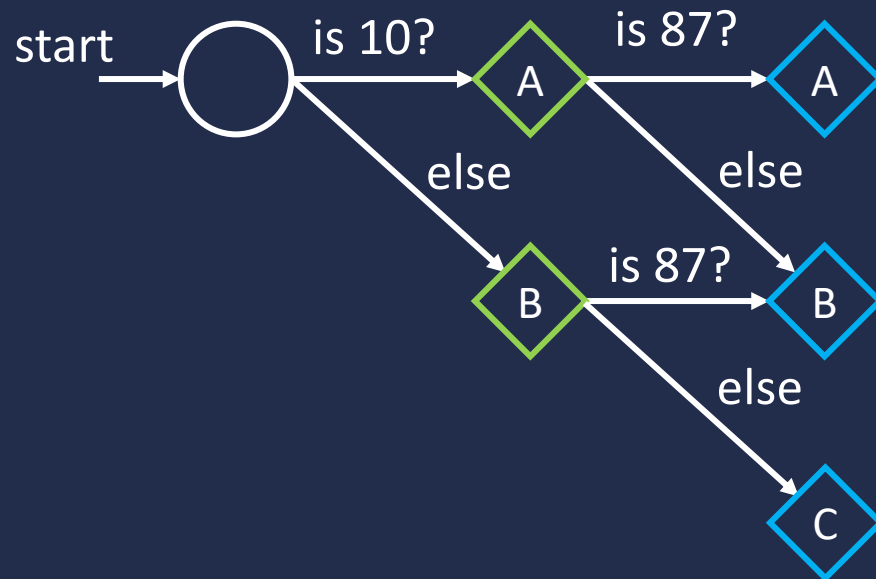
10	87	28	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----



Creating the FSM



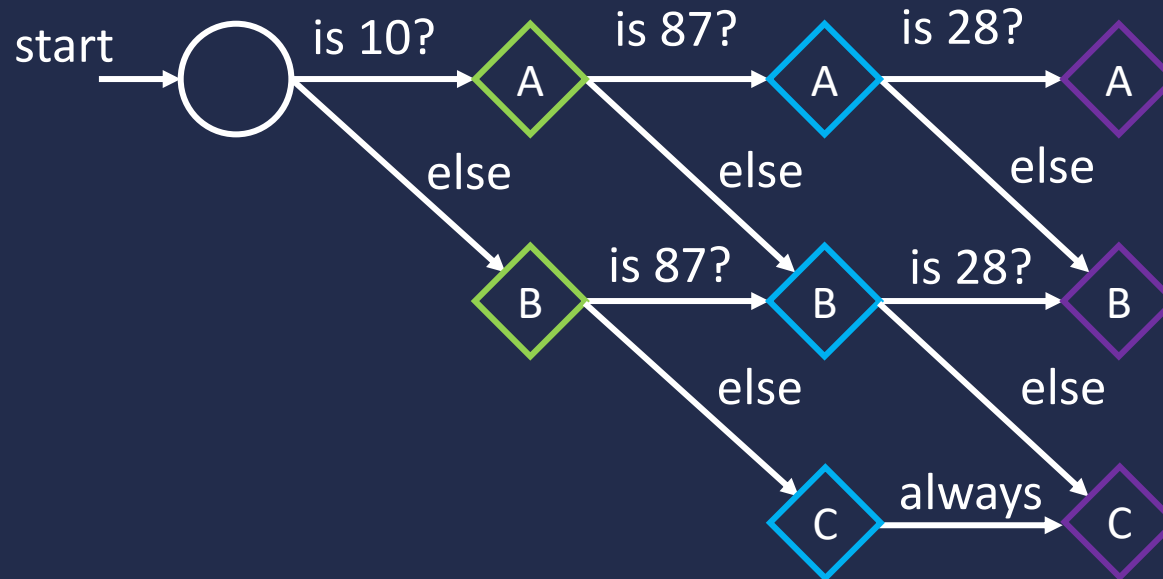
10	87	28	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----



Creating the FSM



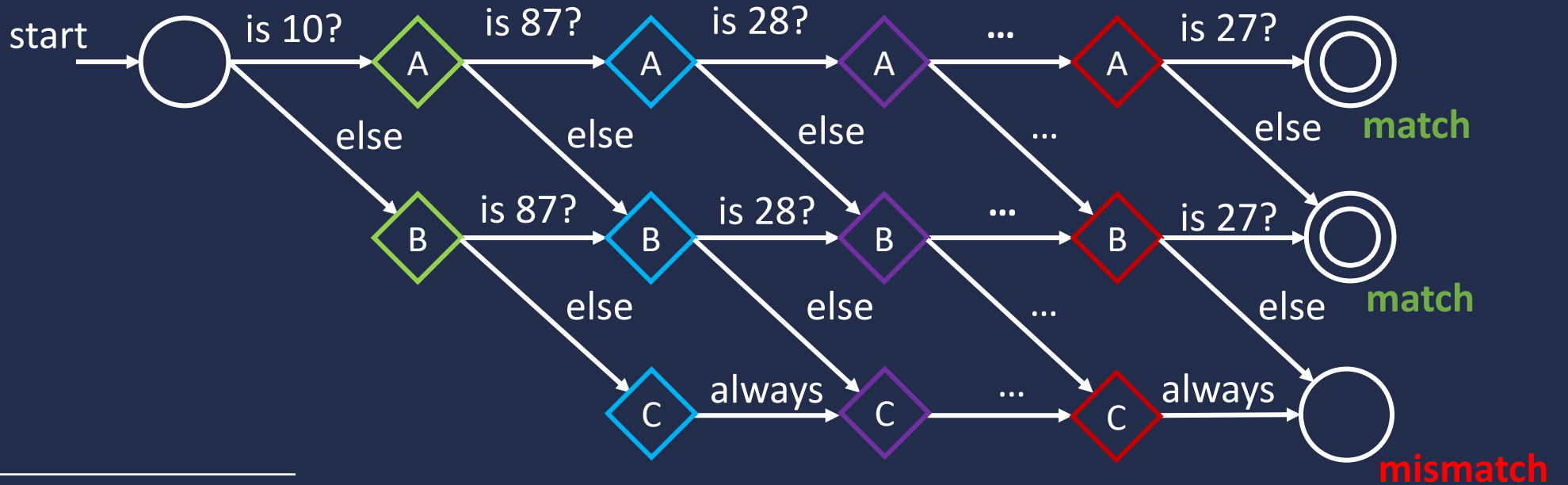
10	87	28	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----



Creating the FSM



10	87	28	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

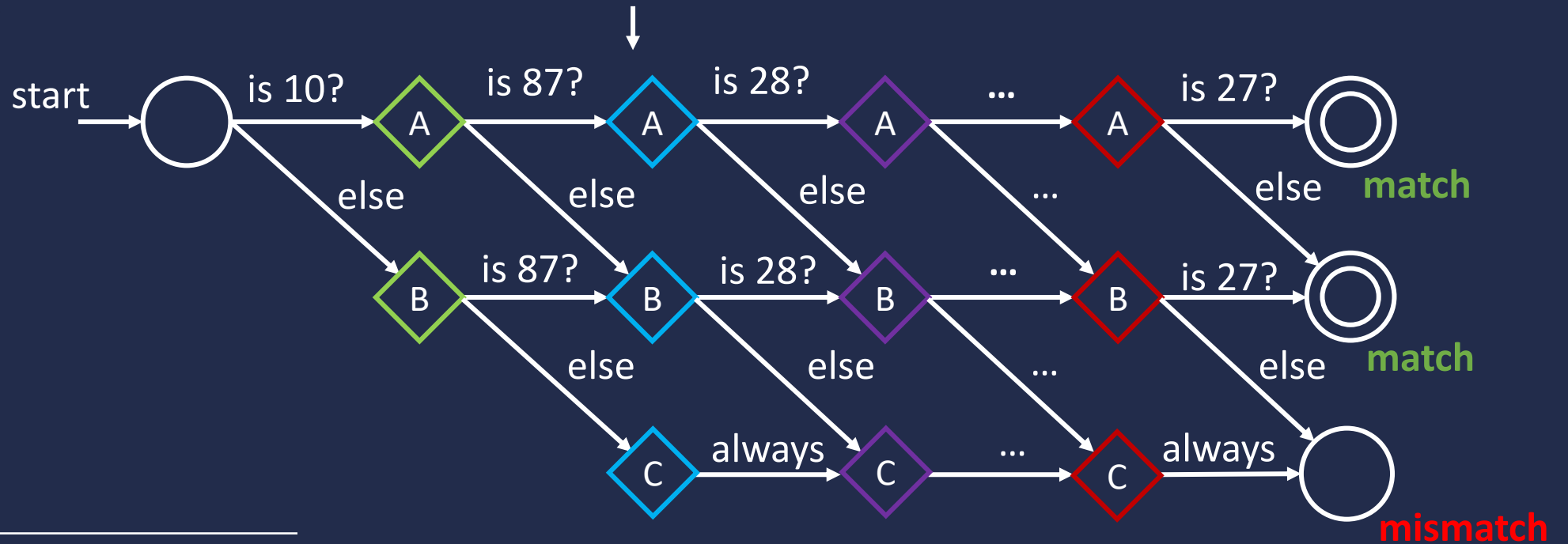


STR Matching with FSM



10	87	28	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

10	87	41	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

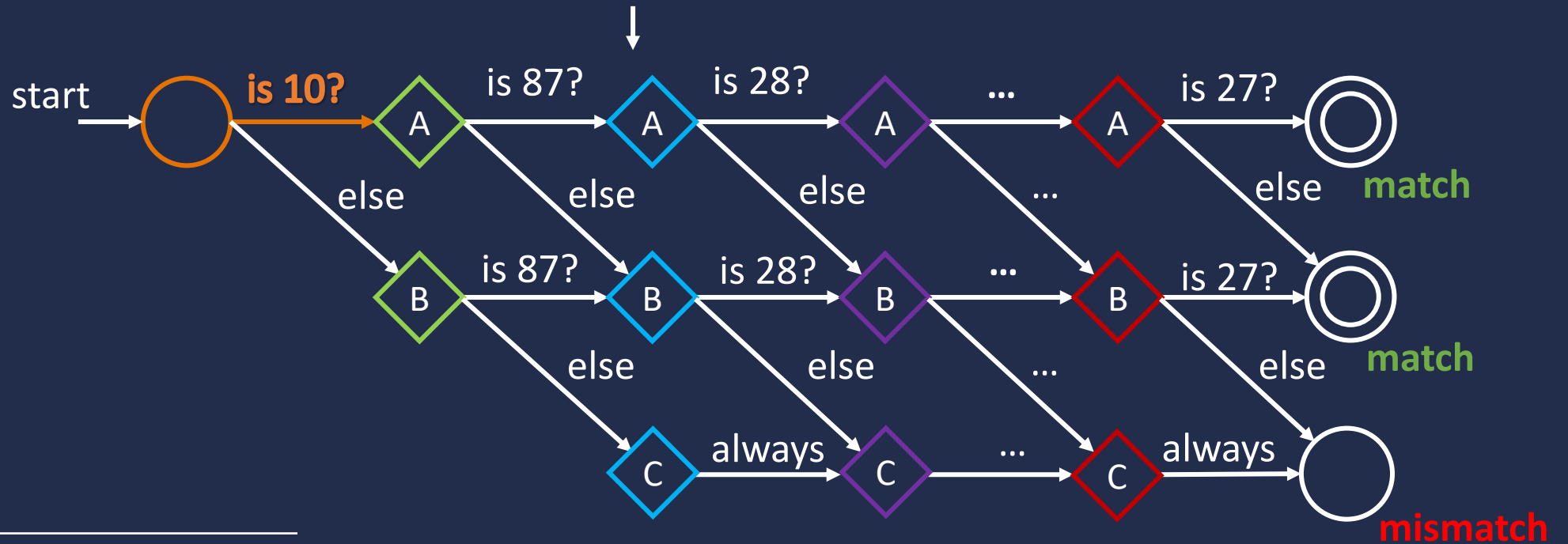


STR Matching with FSM



10	87	28	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

10	87	41	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

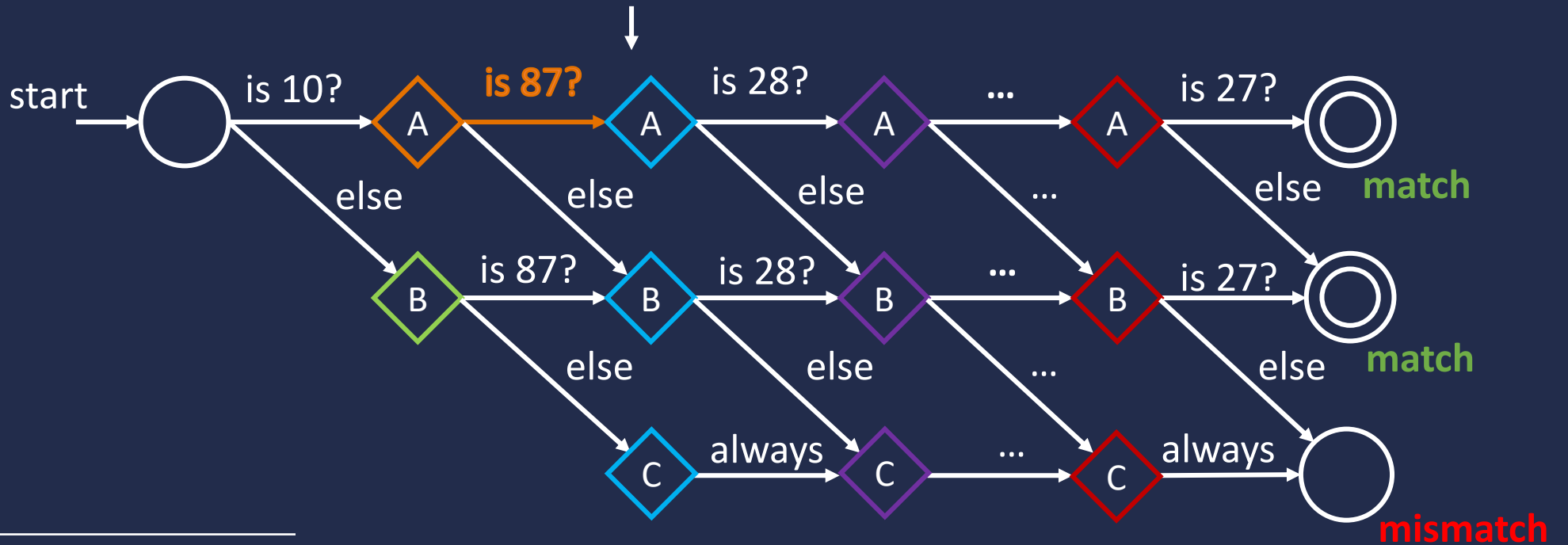


STR Matching with FSM



10	87	28	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

10	87	41	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

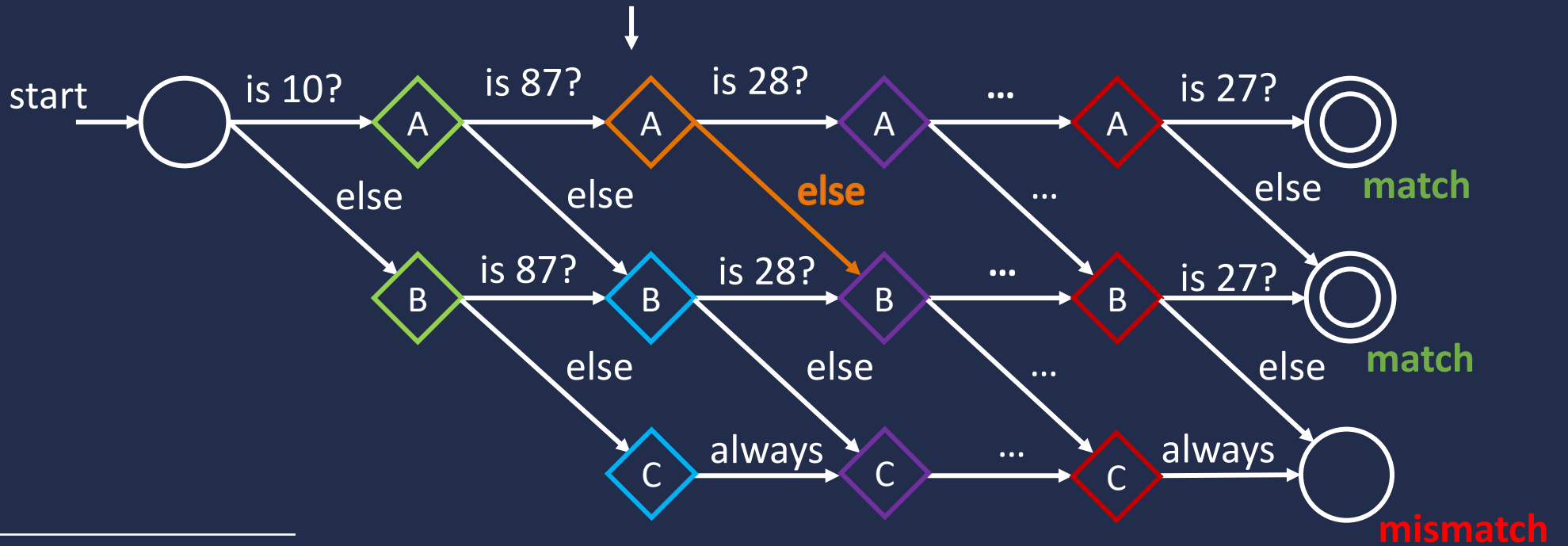


STR Matching with FSM



10	87	28	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

10	87	41	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

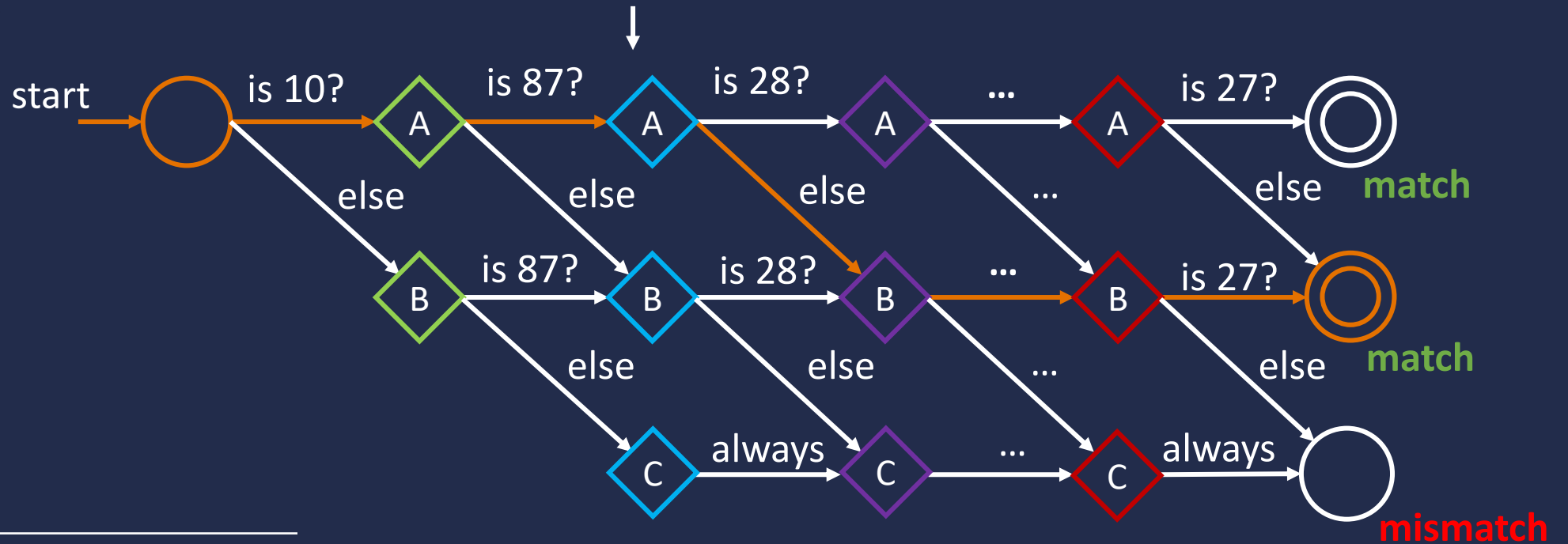


STR Matching with FSM



10	87	28	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

10	87	41	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----



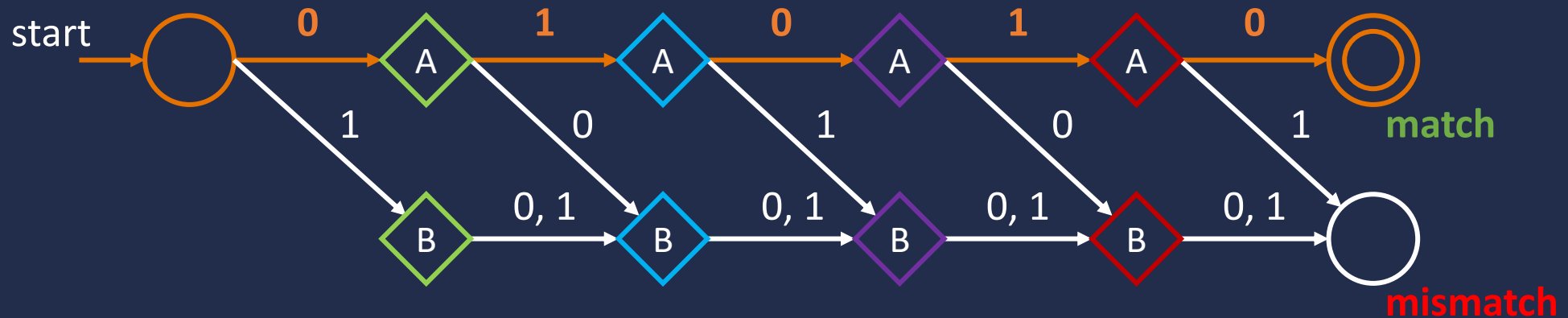
Equality FSM



10	87	28	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

10	87	41	127	63	4	89	1	...	27
----	----	----	-----	----	---	----	---	-----	----

Send query through FSM to get match result



Oblivious Transfer (1 out of k)



Receiver

Start: $c \in \{0, 1, \dots, k-1\}$

Finish: m_c



Sender

m_0, m_1, \dots, m_{k-1}

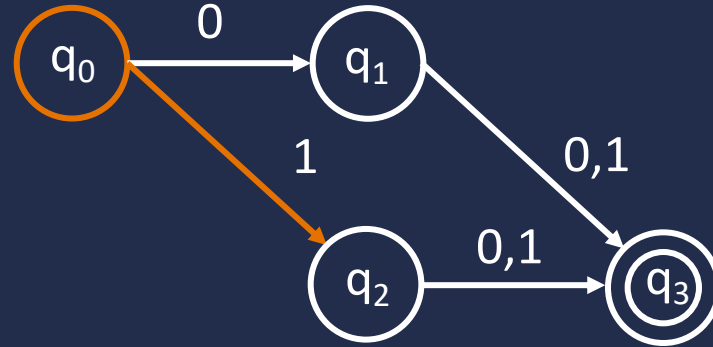
Nothing



FSM Transition with Oblivious Transfer



Client



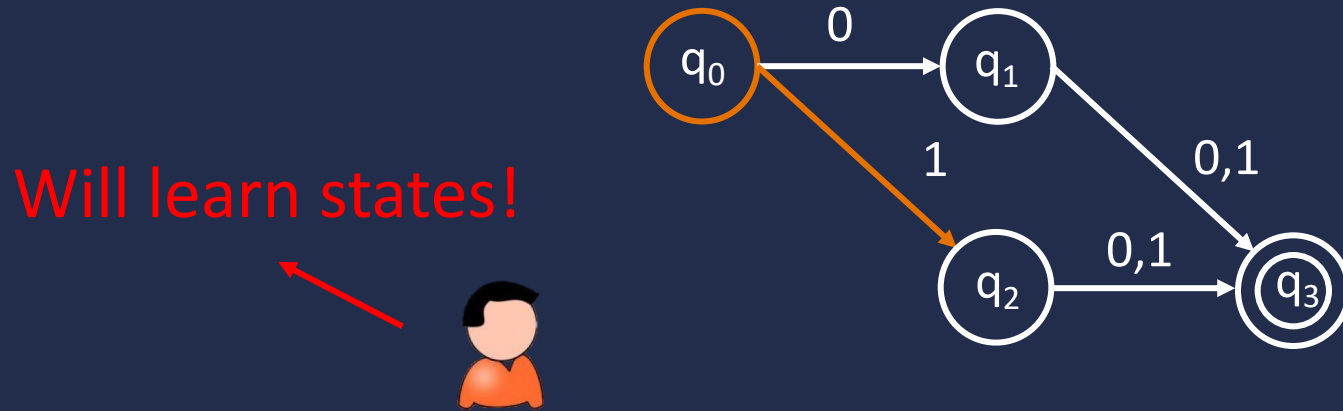
Server

$c = (\text{current state, transition})$

$c = (q_0, 1) = q_2$

	0 transition	1 transition
q_0	$M_{(q_0,0)} = q_1$	$M_{(q_0,1)} = q_2$
q_1	$M_{(q_1,0)} = q_3$	$M_{(q_1,1)} = q_3$
q_2	$M_{(q_2,0)} = q_3$	$M_{(q_2,1)} = q_3$
q_3	$M_{(q_3,0)} = \text{end}$	$M_{(q_3,1)} = \text{end}$

FSM Transition with Oblivious Transfer



Client



Server

$c = (\text{current state, transition})$

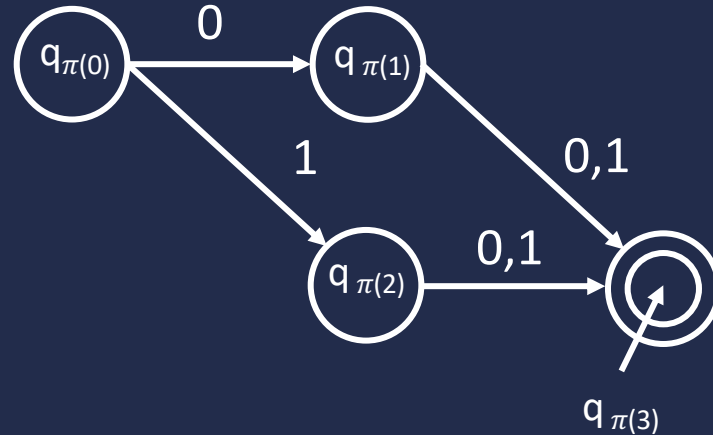
$c = (q_0, 1) = q_2$

	0 transition	1 transition
q_0	$M_{(q_0,0)} = q_1$	$M_{(q_0,1)} = q_2$
q_1	$M_{(q_1,0)} = q_3$	$M_{(q_1,1)} = q_3$
q_2	$M_{(q_2,0)} = q_3$	$M_{(q_2,1)} = q_3$
q_3	$M_{(q_3,0)} = \text{end}$	$M_{(q_3,1)} = \text{end}$

FSM Transition with Oblivious Transfer



Client



Permute States!

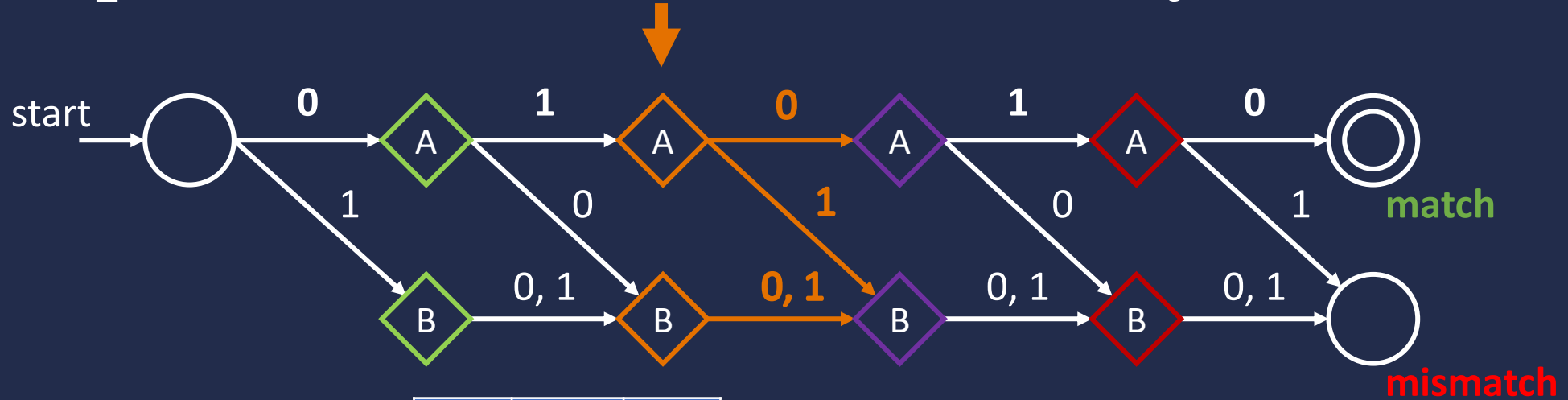


Server

$c = (\text{current state, transition})$

	0 transition	1 transition
$\pi(q_0)$	$M_{(\pi(q_0),0)} = \pi(q_1)$	$M_{(\pi(q_0),1)} = \pi(q_2)$
$\pi(q_1)$	$M_{(\pi(q_1),0)} = \pi(q_3)$	$M_{(\pi(q_1),1)} = \pi(q_3)$
$\pi(q_2)$	$M_{(\pi(q_2),0)} = \pi(q_3)$	$M_{(\pi(q_2),1)} = \pi(q_3)$
$\pi(q_3)$	$M_{(\pi(q_3),0)} = \text{end}$	$M_{(\pi(q_3),1)} = \text{end}$

Optimize Communication with Layered FSM



	0	1
A	B	A
B	B	B
...
A	⊙	○

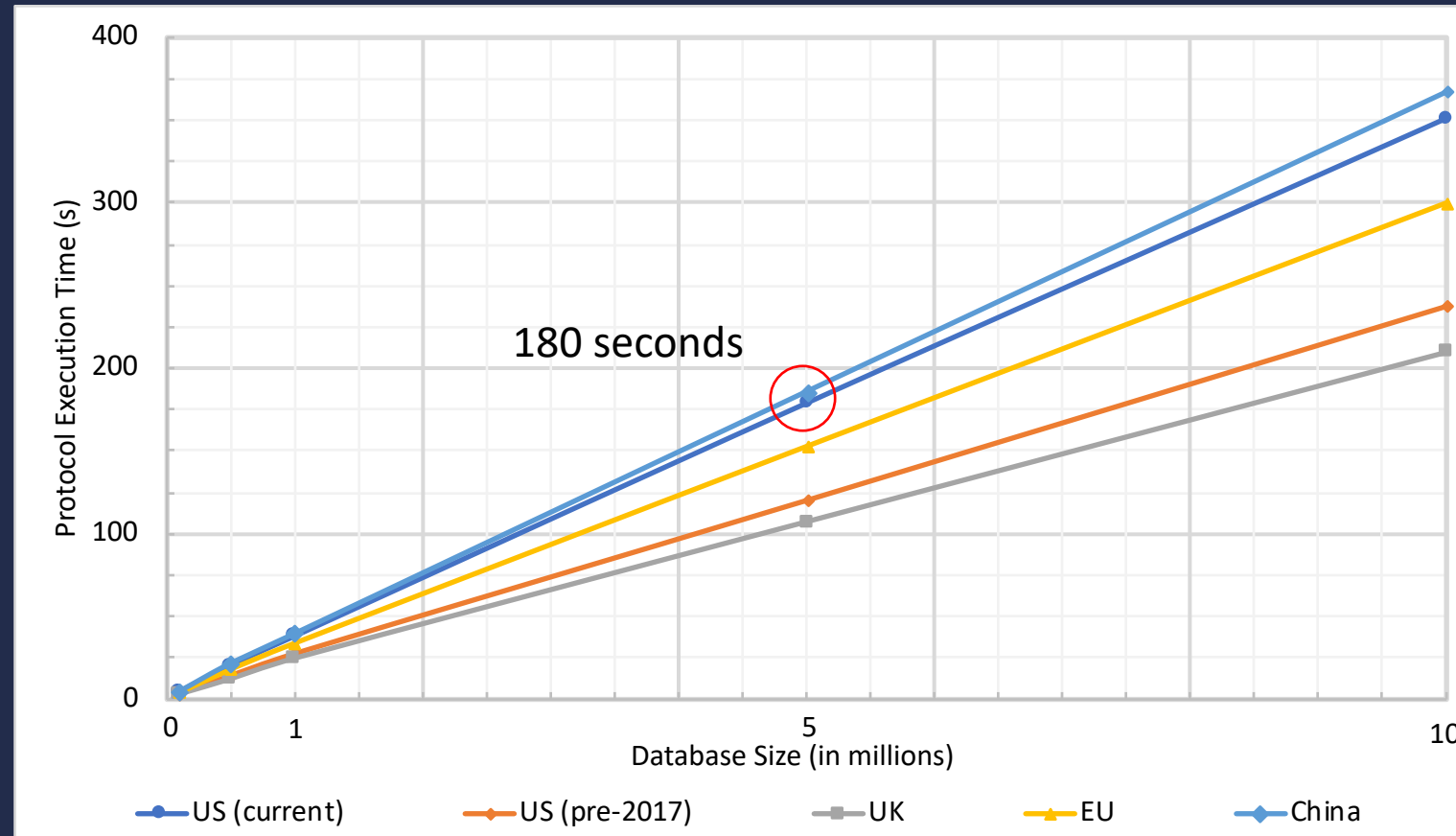


	0	1
A	A	B
B	B	B

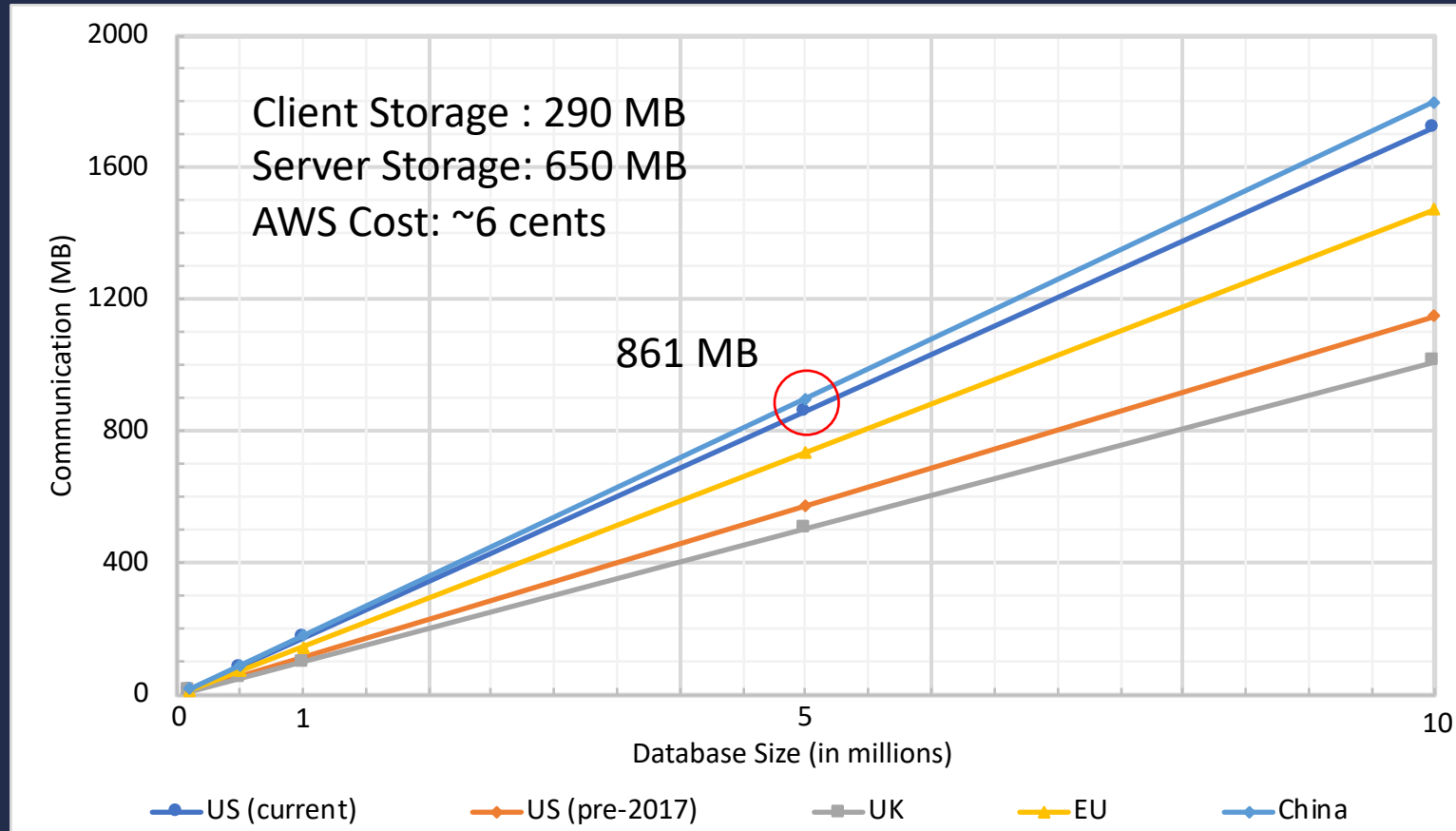
Implementation

- **Precomputed OT correlations**
 - **OTs require only symmetric operations**
- **Ran protocol in parallel for all simulated database entries**
- **Used different STR profile types for countries that use CODIS like systems**
- **Benchmarked on AWS**

Time vs. Database Size



Communication vs. Database Size

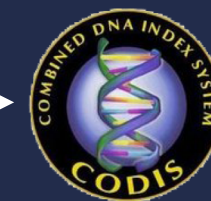


Summary

- Performs privacy preserving matching
 - Privacy for suspect and database
- Operates efficiently
 - **3 minutes** for five million database entries
- Feasible on mobile device
 - Moderate end-to-end communication cost (**860 MB**)
 - Low client side storage requirement. (**290 MB**)



Match



only sees match result



does not learn what is in



does not learn what is



Conclusion

- **Generic search for various genomic data**
 - **Low stringency matching**
 - **23 and Me genomic data**



The New York Times
***Genealogy Sites Have Helped
Identify Suspects. Now They've
Helped Convict One.***

A new forensic technique sailed through its first test in court, leading to a guilty verdict. But beyond the courtroom, a battle over privacy is intensifying.

New York Times, July 1st, 2019

Questions

- **Email:** jab7dq@virginia.edu
- **Code:** <https://github.com/jBlinden/private-codis>
- **Paper (Nature Computational Science):** <https://rdcu.be/cjq70>