SECURE MULTIPARTY COMPUTING FOR MEDICAL RECORD LINKAGE

JASON N. DOCTOR[1], JAIDEEP VAIDYA[2], XIAOQIAN JIANG[3], SHUANG WANG[3], LISA SCHILLING[4], TOAN ONG[4], MICHAEL E. MATHENY[5,6], LUCILA OHNO-MACHADO[3] AND DANIELLA MEEKER[7]

[1]Schaeffer Center for Health Policy & Economics, University of Southern California, Los Angeles, CA 90033

[2]Management Science & Information Systems Department, Rutgers University, Newark, NJ 07102

[3]Health System Department of Biomedical Informatics, UC San Diego, La Jolla, CA, 92093

[4]Department of Medicine, University of Colorado, Anschutz Medical Campus, CO, 80045

[5]Geriatric Research Education and Clinical Care Service, Tennessee Valley Healthcare System VA, Nashville, TN 37212

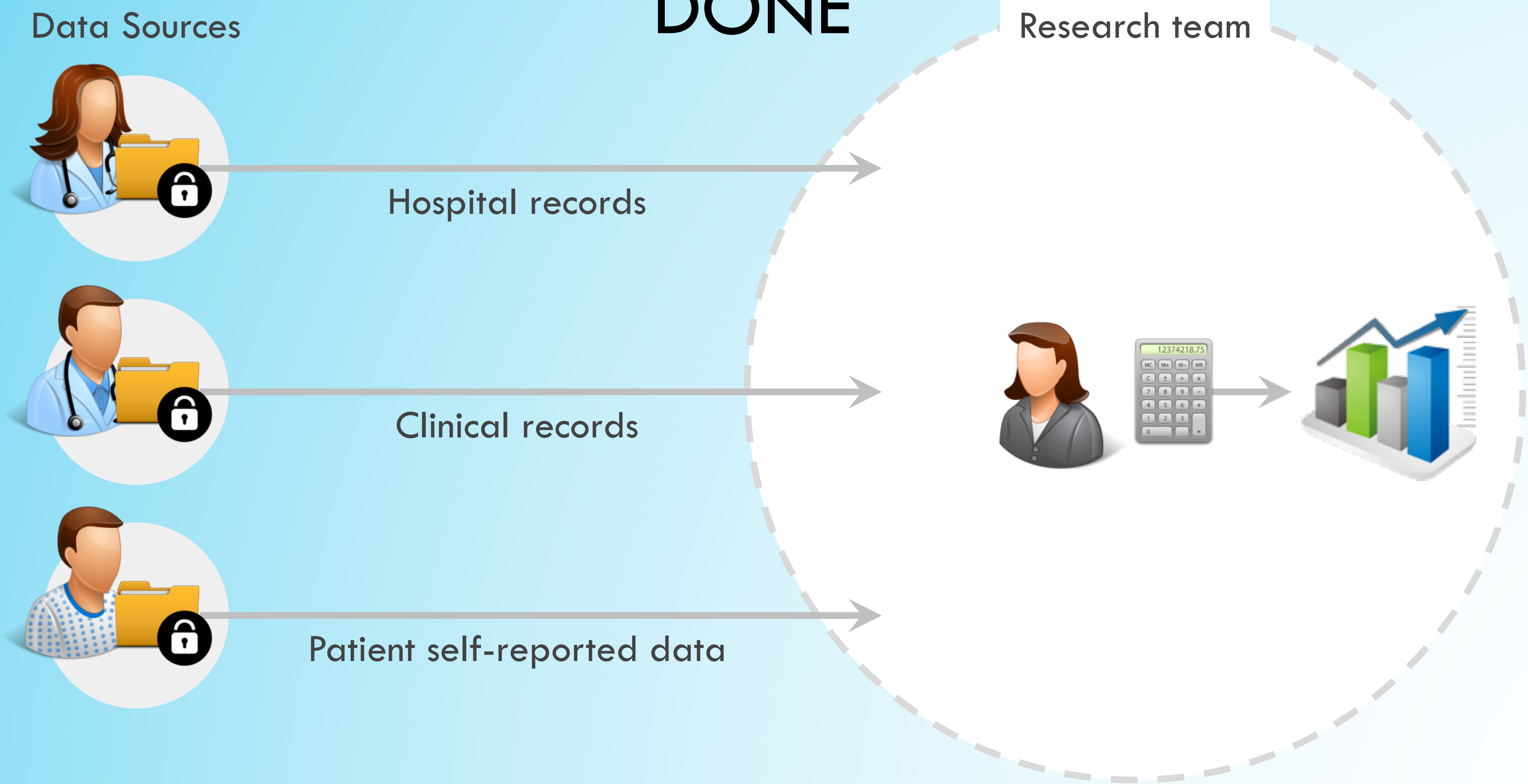
[6]Departments of Biomedical Informatics, Medicine, and Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37235

[7]Keck School of Medicine, University of Southern California, Los Angeles, CA 90089

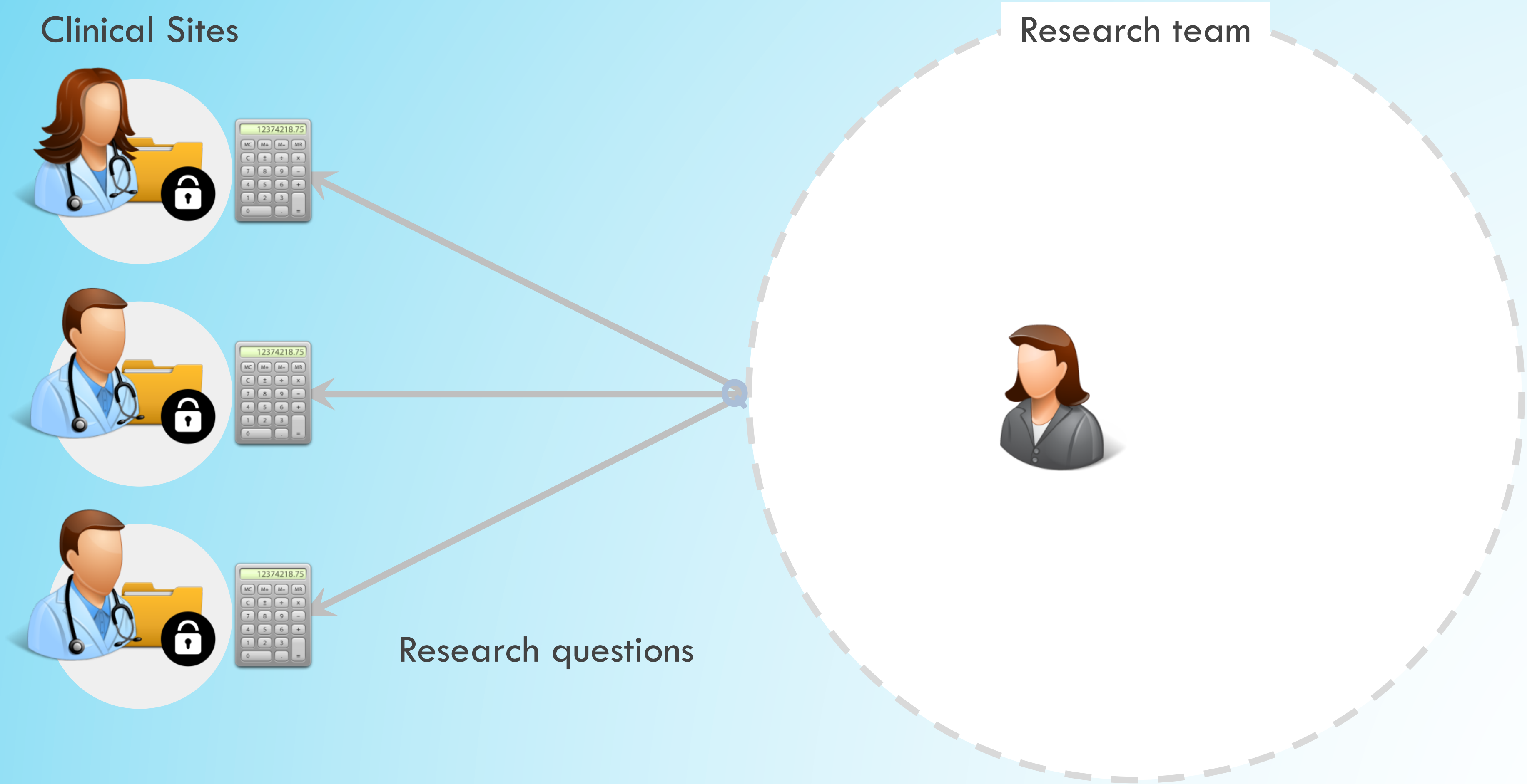
OVERVIEW

- Ways in which data are stored
- Secure multiparty computation: What is it?
- Set Intersection Method for Secure Computation
- Experiment
- Discussion
- Future Directions and New Applications

THE OLD WAY: BRING DATA TO A CENTRAL DATABASE WHERE COMPUTATIONS ARE DONE



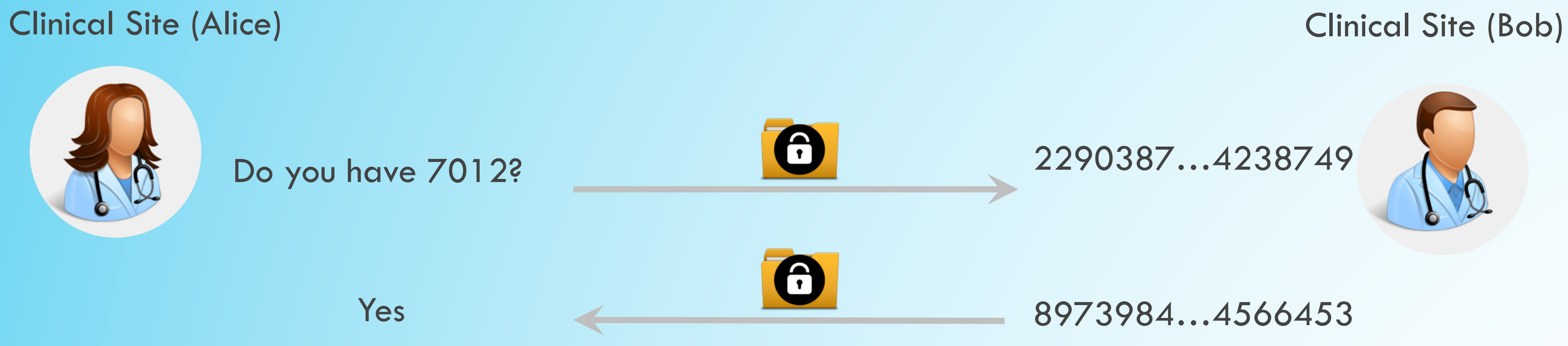
THE NEW WAY: BRING THE COMPUTATION TO THE DATA



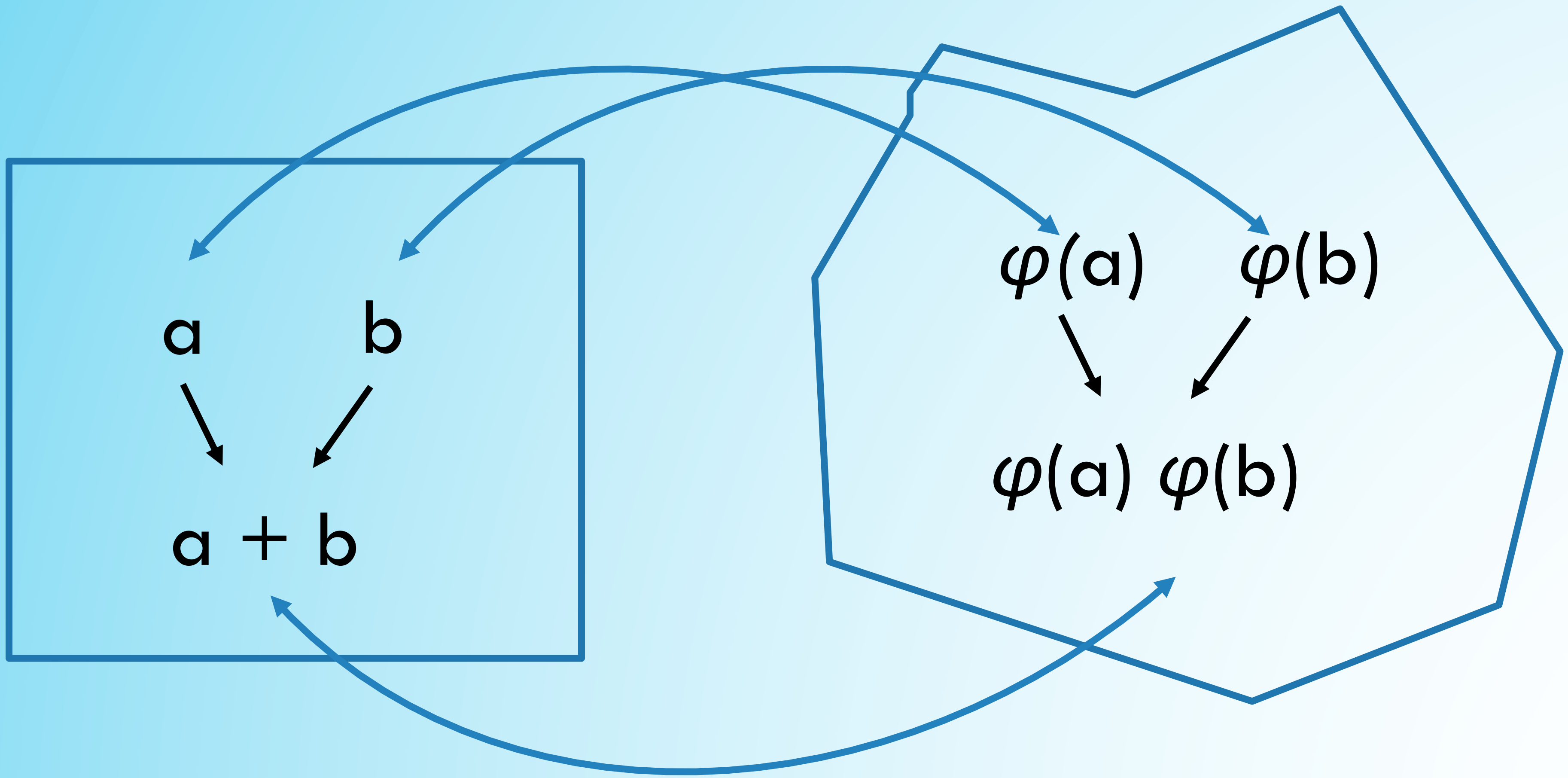
”The secret to strong security: less reliance on secrets.”

-Whitfield Diffie

SECURE MULTI-PARTY COMPUTATION: PUBLIC FUNCTION WITH PRIVATE INPUTS

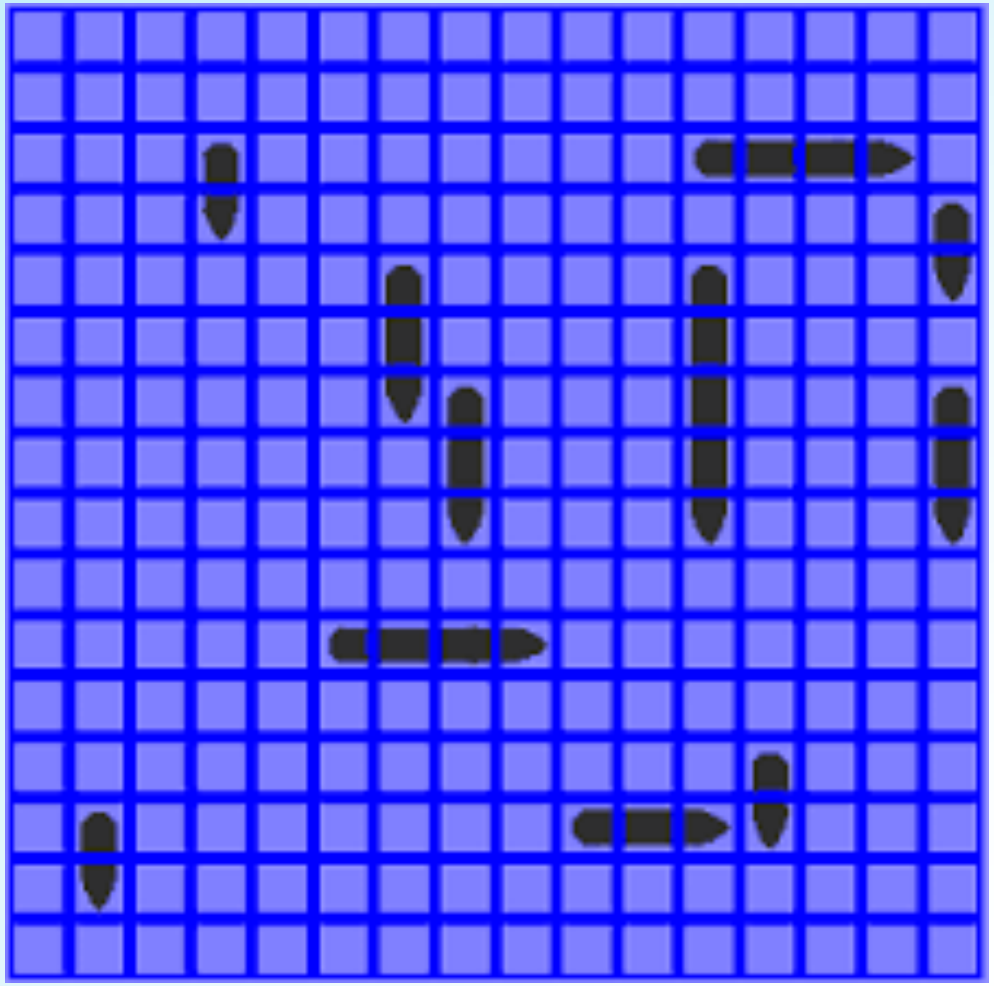


HOMOMORPHIC ENCRYPTION



SET INTERSECTION (LIN & TZENG, 2005) LIKE "BATTLESHIP"

- ALICE'S NUMBER IS REPRESENTED BY A PARTICULAR CONFIGURATION OF SHIPS
- BOB'S NUMBER IS REPRESENTED BY ANOTHER CONFIGURATION OF SHIPS
- ALICE'S SHIPS ARE EACH ENCRYPTED ZEROS
- OPEN WATER IS REPRESENTED BY RANDOM NUMBERS
- BOB MULTIPLIES THE ENCRYPTED NUMBERS ALICE GIVES HIM THAT CORRESPOND TO HIS SHIP LOCATION AND RETURN THESE TO ALICE
- SHE DECRYPTS, IF SHE GETS A "0" THEY HAVE A MATCHING SHIP AND ALICE'S NUMBER IS BIGGER THAN BOB'S



SET INTERSECTION EQUIVALENCE

- Matching has more applications
  - Buying and selling prices in market transactions
  - Authentication in Communication
  - Medical Record Linkage
- Developing an application would have many uses

OUR MAIN RESULT

- If Bob returns messages on Z (target identifier) and Z – 1, then we can match identifiers

Case 1 (null set): Y < Z ; Y < Z -1	Case 2 (nonempty on both): Y > Z ; Y > Z -1
Case 3 (null set): Y < Z ; Y = Z -1	**Case 4** **(nonempty on 1):** **Y = Z ; Y > Z -1**

EXPERIMENT

DATA SET

- Data represent 20,000 cases, 10,000 at each of two clinical sites (6,000 of which match)
- Metro Community Provider Network in Denver, Colorado between January 1, 2007 and January 31, 2013.

COMPUTING ENVIRONMENT

- The algorithm was carried out on a Apple, inc. MacIntosh computer with a 2.4 Ghz Intel Core i5 processor and 8 GB (1699 MHz DDR3) of memory (My Laptop).
- Software was written in the statistical computing language R, Version 3.3.0, released May 3rd, 2016 (Ihaka, Ross, & Robert, 1996) and utilized the package HomomorpheR for additive homomorphic encryption.

SORTING

We utilized a sorting algorithm to reduce the number of necessary comparisons from a x b to a + b.

RESULTS

- It took 63 minutes for Alice to encrypt her data
- Alice’s single vector of encrypted IDs required 462.5 MB of space (about 46.3 GB per million lives)
- Bob’s dynamic multiplication of messages, Alice’s decryption of Bob’s Messages took 6 hours, 32 minutes and 27 seconds.

RESULTS (HH:MM:SS)

Time	Alice encryptions	Bob's Message Generation	Decryption by Alice
User	*00:00:04*	*00:09:57*	*00:00:05*
System	*00:00:02*	*00:00:13*	*00:00:02*
Elapsed	*00:00:05*	*00:11:06*	*00:00:05*

DISCUSSION

- Secure Multiparty Record Linkage is Feasible
- Alice's encryptions can be done offline which adds efficiency
- Methods for generating global identifiers are well-known (Kho et al. 2015)
- Two-party example can be extended easily to multiple parties

FUTURE DIRECTIONS AND NEW APPLICATIONS

- Apply this method across a distributed data network
- Clinical medical record sharing peer-to-peer (through certification matching)
- Authentication in secure communication between clinicians (e.g., Secure texting)
- Genetic marker matching for the protection of personal genetic data

ACKNOWLEDGMENTS

- This research funded under the Patient Scalable National Network for Effectiveness Research (pSCANNER) supported by the Patient Centered Outcomes Research Institute (PCORI), Contract CDRN-1306-04819 (PI: Dr. Ohno-Machado).