CENTER FOR CYBER SECURITY

Privacy-Preserving Disease Susceptibility Test with Shamir's Secret Sharing

Guyu Fan and Manoranjan Mohanty

جامعة نيويورك أبوظبي
NYU | ABU DHABI

THE PROBLEM

- Personalized medicine based on genomic data has many advantages, but it also brings serious privacy concerns since genomic data contains highly sensitive information.
- Genome-based Disease Susceptibility Tests (DSTs) use the patient's genetic makeup to determine the patient's susceptibility to certain diseases.

OUR CONTRIBUTION

- We propose an efficient and secure DST scheme that preserves the privacy of patient genomic data and pharmaceutical trade secrets (mainly SNP weights, see below) using Shamir's Secret Sharing.

DISEASE SUSCEPTIBILITY TESTS

- The computation of disease susceptibilities depends on certain Single-Nucleotide Polymorphisms (SNPs) in the patient's genome:
  - ✓ The state of each SNP, determined by the number of major/minor alleles at that SNP position.
  - ✓ Weights associated with each SNP position.
- The weighted-averaging method for computing disease susceptibility uses the following equation:

$$S_P^X = \frac{1}{\sum_{i\in L_X} C_i^X} \times \sum_{i\in L_X} C_i^X \left[\frac{p_0^i(X)}{(0-1)(0-2)}(\text{SNP}_i-1)(\text{SNP}_i-2)+\frac{p_1^i(X)}{(1-0)(1-2)}(\text{SNP}_i-0)(\text{SNP}_i-2) + \frac{p_2^i(X)}{(2-0)(2-1)}(\text{SNP}_i-0)(\text{SNP}_i-1)\right]$$

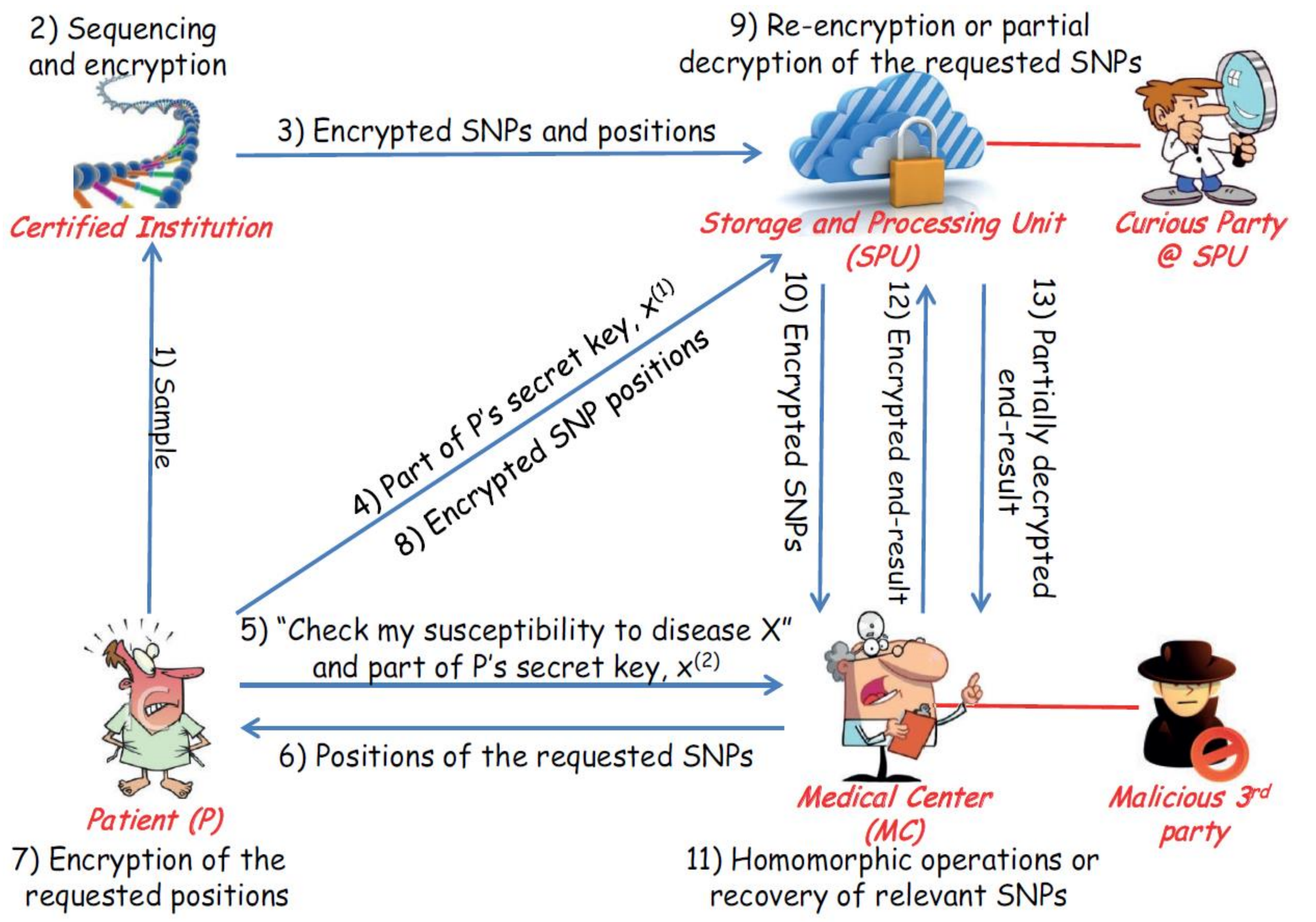
SHAMIR'S SECRET SHARING

- Shamir's secret sharing hides the secret by creating shares with a polynomial in the final field GF(p):

$$F(x) = (S + \alpha_x) \bmod p$$

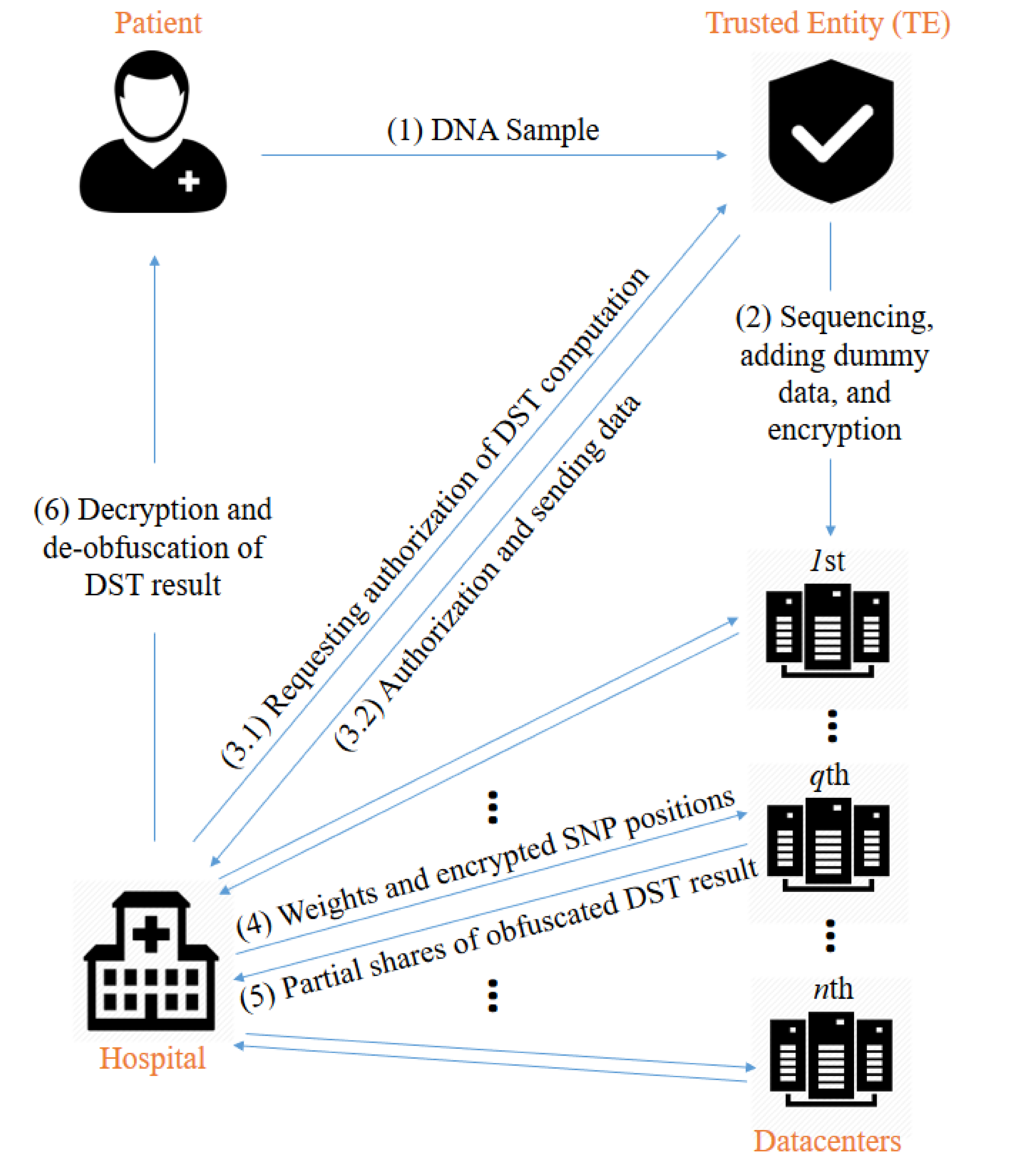
- Shamir's secret sharing is homomorphic to additions.
- With enough shares, Shamir's secret sharing can also be homomorphic to a fixed number of multiplications.

SCHEME PROPOSED BY AYDAY ET AL.



- SNP positions are encrypted via symmetrical encryption.
- SNP states are encrypted via Paillier cryptosystem.

OUR SCHEME



- SNP positions are encrypted via symmetrical encryption.
- SNP states are encrypted via Shamir's secret sharing.
- Dummy SNPs and dummy weights are introduced to obfuscate the number of SNPs and the SNP weights.
- DST is split into multiple parts to enhance obfuscation.

ANALYSIS AND EXPERIMENT

- Our scheme minimizes patient involvement.
- The hospital is able to verify the integrity of test results thanks to data redundancy.
- With access to plaintext SNP weights (while **not** being able to distinguish real weights from dummy weights), datacenters are able to prevent a malicious hospital from inferring patient SNP states through malicious SNP weight attacks (Barman et al.)
- Compared to Ayday et al.'s scheme, storage requirement is reduced **40**-fold and traffic per DST reduced **4**-fold.
- Computation is also much more efficient. Experiments show that our scheme runs **10,000** times faster.