PRIVGEN
Privacy-preserving sharing and processing of genetic data

LaTIM Inserm UMR 1101
LS2N CNRS UMR 6004
Inserm UMR 1078

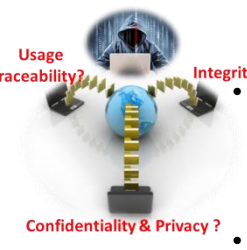
in collaboration with

Labex Genmed

Partners

Context

- **Cloud Computing and data outsourcing -** A successful paradigm to flexibility store, share and process large amount of data while minimizing costs
- **Security needs of outsourced applications and data are worsened**
  - Owners loss the control on their data and applications (***confidentiality***, ***integrity***, ***availability***?)
  - Service provider may in turn transmit data to third-party service providers (***traceability***, ***intellectual/scientific ownership protection***?)
  - Storage by the service providers of data issued from different sources (***privacy***?)



- **Sharing of outsourced genetic data and applications – more than an experimental framework**
  - Needs for international sharing of genetic data for better human genome decryption to improve diagnosis …
  - Data highly personal, covering a large security spectrum needs (*privacy*, *data reliability – integrity + authenticity -*, *scientific ownership* …)
  - Distributed applications
  - Different initiatives (e.g. beacons) with identified security weaknesses …

Objectives

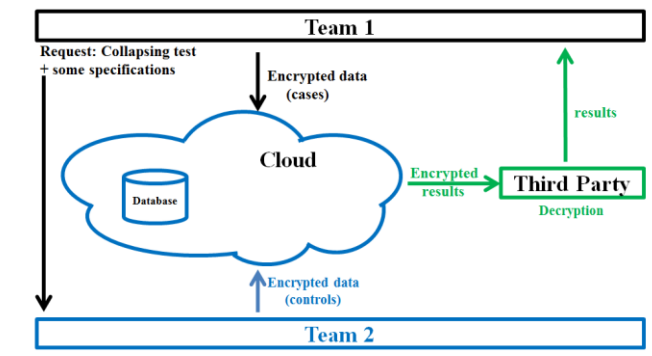
- **Respond to actual security solutions limitations**
  - Cloud applications impose satisfying many security properties at once ➔ Needs to make interacting different security mechanisms
  - Cloud applications are distributed computations executed on behalf of multiple stakeholders
- **Two research axis**
  - Composition of security and privacy mechanisms applied to compositions of complex computations
  - New multipurpose security mechanisms able to satisfy several security objectives at once.

Challenge 1 - Mechanisms for a continuous digital content protection

- **Objective:** Merging different security mechanisms into one configurable digital content protection tool for multipurpose security purposes.
- **Contributions :** Provide continuous data protection with joint security mechanisms configurable by a composition language.

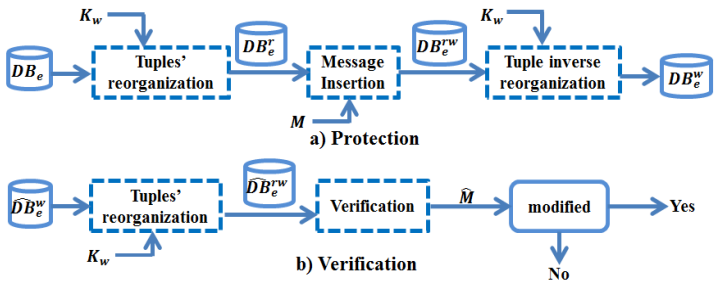
Processing of encrypted genetic data

- **Objective:** Allow two or more research teams to perform genetic association studies while preserving data confidentiality and privacy.
- **Contributions :** Homomorphic encryption based genetic association study using secure χ^2 test.



Controlling the integrity of encrypted genetic data

- **Objective:** Allow the cloud to control the integrity of homomorphically encrypted outsourced data.
- **Contributions :** A dynamic joint homomorphic encrytion-watermarking scheme able to detect and identify altered data under user data update constraints.



Challenge 2 - Composition of security and privacy-protection mechanisms

- **Objective:** Provide a development approach for privacy-preserving distributed genetic applications
- **Contributions :** A composition theory for security and privacy properties - Programming support.

Sharing architecture

- **Contributions :** A multi-cloud based architecture with a trusted party for data processing. Geneticists' data storage is delegated to the Clouds which are independent and non communicating for privacy reasons.

Composition theory

- **Algebraic laws**: extend the theory for security mechanisms combination (watermarking, encryption, fragmentation) with classical queries for correct security query formulation.

$$id \equiv detectw_a \circ wat_a$$

$$decrypt_{(s,a)} \circ crypt_{(s,a)} \circ detectw_a \circ wat_a \equiv detectw_a \circ decrypt_{(s,a)} \circ crypt_{(s,a)} \circ wat_a$$

$$\pi_a \circ detectw_a \equiv detectw_a \circ \pi_a$$

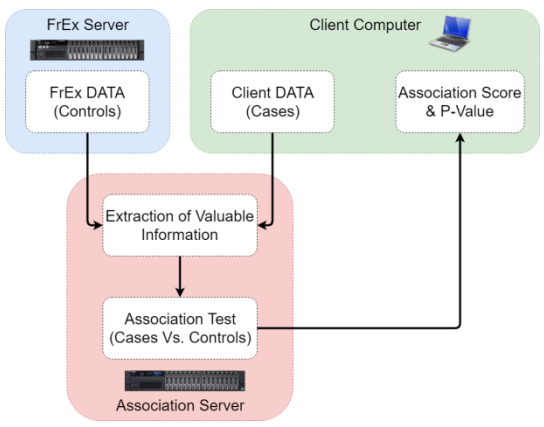
$$detectw_a \circ \sigma_p = \sigma_p \circ detectw_a$$

```
scenario : GeneticQuery [SubjectId,ZIP,Gender,DoB,
                         Variant,TypeVar,MyTattoo]
scenario =  do

  G1  'SendRequest' (TP,[Q1])
  G1  'SendRequest' (TP,[Q2,Q2'])
  G1  'SendRequest' (TP,[Q3,Q3'])

  TP  'SendRequest' (LeftCloud,[Q1])
  TP  'SendRequest' (RightCloud,[Q2,Q2'])
  TP  'SendRequest' (RightCloud,[Q3,Q3'])

  let q1 = LeftCloud  'executeRequest' [Q1];
  let q2 = RightCloud 'executeRequest' [Q2,Q2'];
  let q3 = RightCloud 'executeRequest' [Q3,Q3'];

  demData1       <- LeftCloud  'SendData' (TP,q1)
  demData2       <- RightCloud 'SendData' (TP,q2)
  vcfFiles       <- RightCloud 'SendData' (TP,q3)

  let r1 = decrypt VariantWE (AESD "key2") vcfFiles;
  let r2 = decrypt TypeVarE (AESD "key1") r1;
  let vcfFiles = detectw VariantW (RGIG "wkey1") r2;
  let Data = defrag (defrag demData1 demData2) vcfFiles

  TP 'ReturnResults' (G1, TP 'Compute' Data)
```

- **Implementation**: an abstract implementation in Idris shows the exchange workflow and security operations to perform a GWAS-like analysis in the suggested architecture.

Challenge 3 - Distributed processing of genetic data

- **Objective** - a platform for: i) sharing relevant genomic information while maintaining privacy; ii) supporting the distributed execution of applications over shared genetic data.



People

- **G. Coatrieux**, Pr., LaTIM Inserm UMR 1101, IMT Atlantique
- **M. Südholt**, Pr., Ascola, LS2N CNRS UMR 6004, IMT Atlantique
- **E. Genin**, Dr., Inserm UMR 1078
- **J-F. Deleuze**, Dr., CNG
- **D. Niyitegeka**, Ph.D Student, LaTIM Inserm UMR 1101, IMT Atlantique
- **F. Boujdad**, Ph.D Student, Ascola, LS2N CNRS UMR 6004, IMT Atlantique
- **R. Bellafqira**, IR, LaTIM Inserm UMR 1101, IMT Atlantique
- **T. Ludwig**, IR, Inserm UMR 1078

Publications

- FZ Boujdad, M Südholt. ***Constructive Privacy for Shared Genetic Data***. CLOSER 2018-8th International Conference on Cloud Computing and Services Science, 2018.
- J. Franco-Contreras, G. Coatrieux. ***Protection of Relational Databases by Means of Watermarking: Recent Advances and Challenges***. Advances in Security in Computing and Communications, Intechopen, pp. 101-123, 2017.