



Efficient and Secure Outsourcing of Genomic Data Storage and Processing

Cédric Lefebvre, Joao Sousa, Zhicong Huang, Jean Louis Raisaro,
Carlos Aguilar, Marc-Olivier Killijian, Jean-Pierre Hubaux

Université de Toulouse, EPFL, LAAS

April 25th 2017



Genomes for noobs

DNA sequencing and mutation

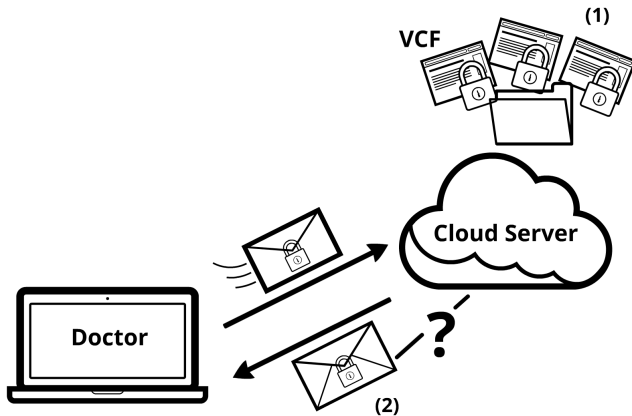
- DNA sequencing: succession of letters that indicate the order of nucleotides: A, C, G, T
- Mutation: Variation in alleles of genes in a gene pool

Operations

There are four operations: substitution, multiple substitution, insertion, deletion

- Substitution: A**A**AGATCA - > A**G**AGATCA
- Multiple substitution: A**A**AGATCA - > A**G****T**GATCA
- Insertion: AAAGATCA - > AA**A****G**TGATCA
- Deletion: A**A**AGATCA - > AGATCA

Why ?



Idash track 3

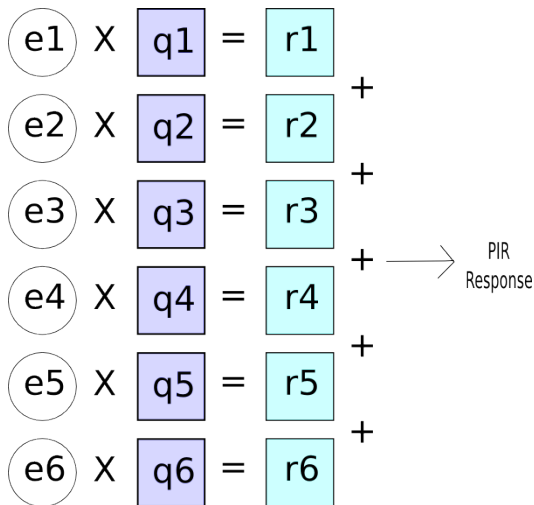


Requirements

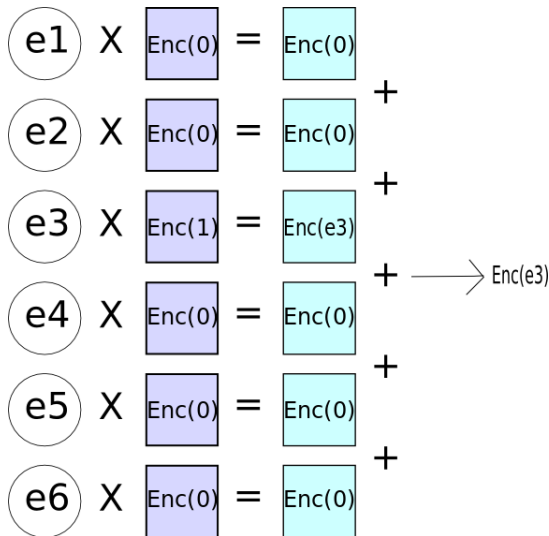
- Client-server model
- 10 Mb/s network link
- 1 round query/reply
- Hide data, query and access patterns from the cloud
- Employ homomorphic encryption
- >80 bits security
- Search for SNP, DEL, INS, SUB
- Reveal less than 20 variants during each search



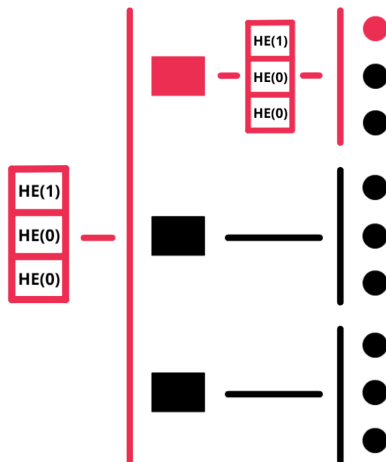
Response generation



Request generation



Recursion



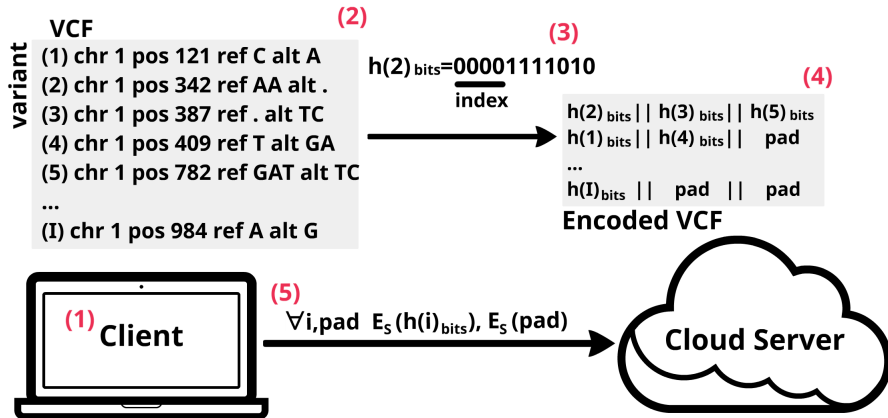
VCF file



```
##fileformat=VCF4.2
##INFO=<ID=SVTYPE,Number=1,Type=String,
Description="Type of structure variant">
##INFO=<ID=END,Number=1,Type=Integer,
Description="End position of the variant described in this record">
#CHROM POS ID REF ALT QUAL FILTER INFO
```

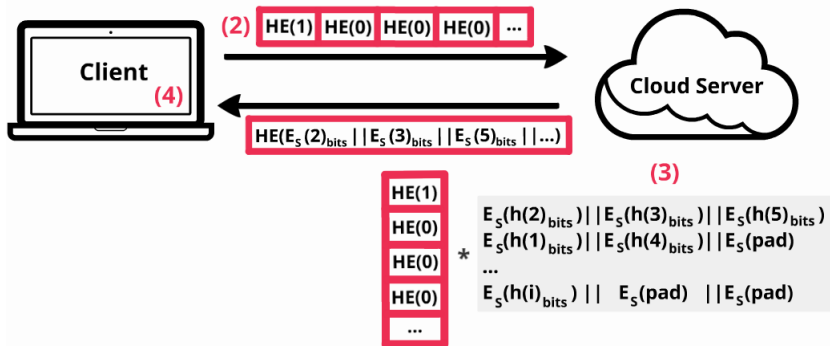
```
1 160929435 rs7520618 G A . SVTYPE=SNP;END=160929436
1 160932043 rs113387749 A . SVTYPE=INS;END=160932043
1 160932206 rs5778188 C . SVTYPE=DEL;END=160932207
1 160932771 rs2256505 A G . SVTYPE=SNP;END=160932772
1 160934077 rs2481074 T A . SVTYPE=SNP;END=160934078
1 160934818 rs1023115 A G . SVTYPE=SNP;END=160934819
1 160935328 . AAA TGC . SVTYPE=SUB;END=160935331
1 160935334 rs75452934 AA TC . SVTYPE=SUB;END=160935336
```

Initialization Phase

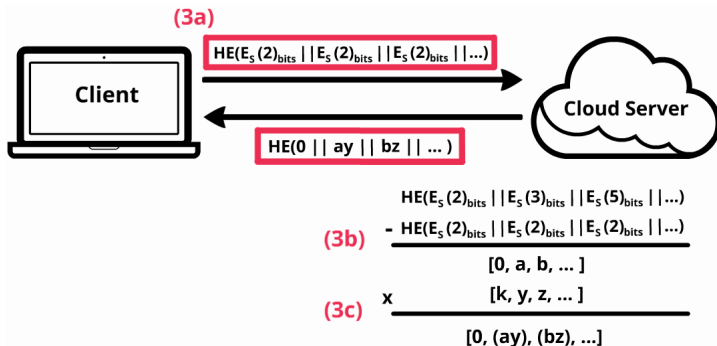


Querying Phase

(1)
 (2) chr 1 pos 342 ref AA alt . $\longrightarrow$ $h(2)_{\text{bits}}$ $\longrightarrow$ pos = 0



Subtraction Step



Parameters



Table : List of parameters.

Parameters	Description
<code>data_hash_size</code>	Length in bits of a variant's hash to be stored.
bits for mapping = x	Number of bits extracted from the hash that maps to a specific index.
<code>row_size</code>	Number of elements per row of <i>data_hash_size</i> bits.
encryption mode	Cryptographic parameters for the FV scheme.
aggregation	
dimensionality	Level of recursion.

Choice



data_hash_size	48
bits for mapping	13
num_entries	8192
row_size	716
encryption mode	80:1024:62:14
aggregation	3
dimensionality	2 (53x52)

Results



Data preparation (s)	19.6
Size of VCF file (Mbytes)	35
Importation (s)	0.71
PIR query generation (s)	0.011
Sending query (s)	1.49
PIR reply generation (s)	0.46
Sending reply (s)	1.26
Reply extraction (s)	0.49
Round-trip-time (s)	2.7

Conclusion



Conclusion

- Far better than the state of the art
- One variant / query
- <https://github.com/JoaoAndreSa/XPIR-iDash>