Interactive Communication InterCom

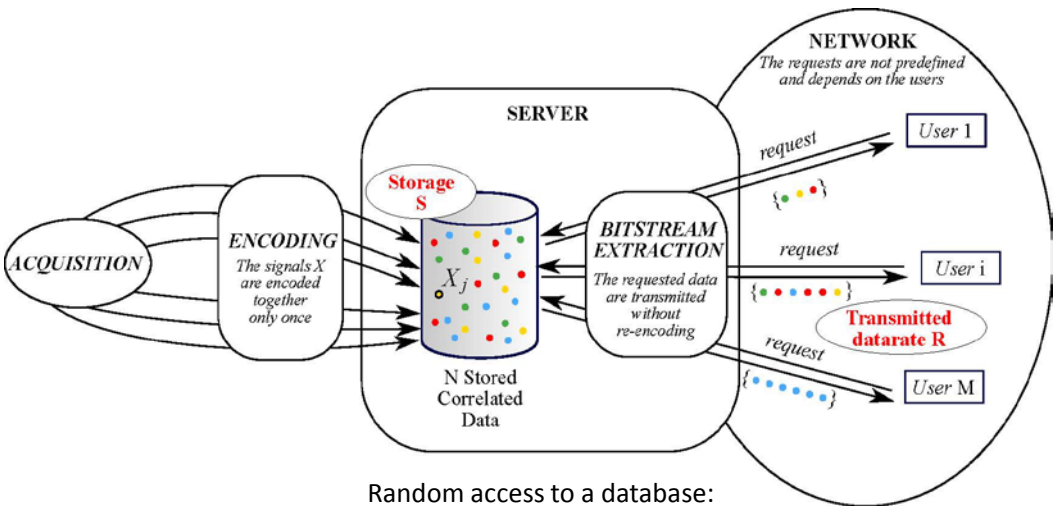
TEAMS
- Inria, Sirocco team
- LabSTICC, Télécom Bretagne
- Inria, i4S team
- External partner: L2S, CentraleSupelec.

Massive random access to subsets of compressed correlated data

The **interCom** project aims to develop novel **compression** techniques allowing **massive random access** to **large** databases.

- **large** database: to be stored on a single server, the data have to be ***compressed*** efficiently, i.e. the redundancy/correlation between the data have to be exploited.
- **random access:** The dataset is then stored on a server and made available to users that may want to access only a subset of the data. Such a request for a subset of the data is indeed **random**, since the choice of the subset is user-dependent.
- **massive** requests: upon request, the server can only perform low complexity operations (for instance no decompression/compression).

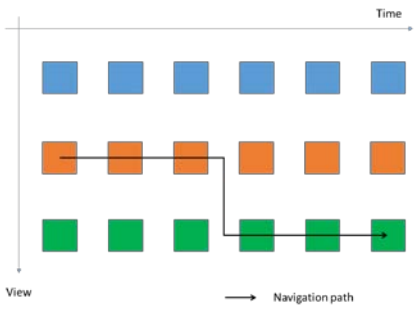
Algorithms for two emerging applications of this problem will be developed:
Free-viewpoint Television (FTV) and massive requests to a database collecting data from a **large-scale sensor network** (such as Smart Cities).



Random access to a database:
the user can choose any subset of the compressed correlated data.

Applications: 1.Free-viewpoint Television (FTV)

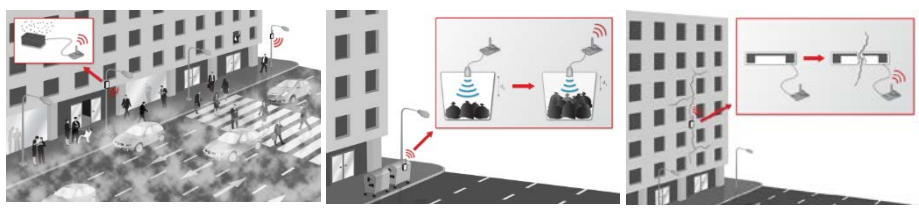
FTV is a system for watching videos in which the user can choose its viewpoint freely and change it at anytime.



Technical characteristics of FTV:
- **large database**: multiple views (100 in the MPEG working group in charge of FTV)
- **random access:** request one view (subset of the data) according to viewer choice
- **massive access:** huge number of viewers such as in the 2022 FIFA World Cup

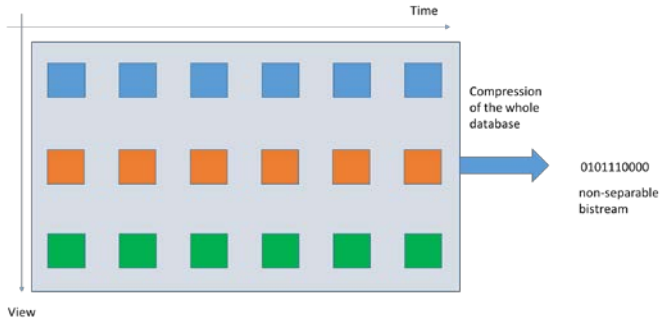
2. Large-scale sensor network (Smart City)

Smart city is an emerging concept, where a city relies on information and communication technologies to enhance quality and performance of urban services. For instance atmospheric pollution control, structural health monitoring of large structures or fill level detection of garbage.

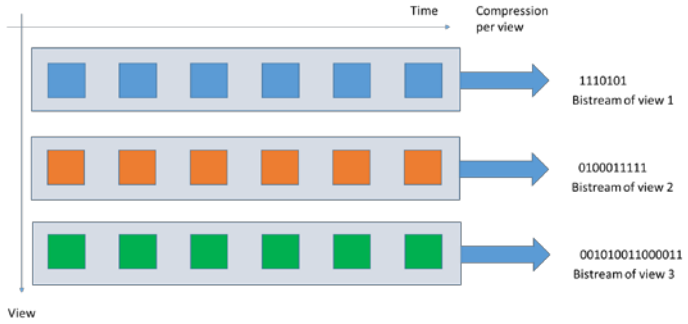


Technical characteristics of smart city monitoring:
- **large** database: acquisition of vast amounts of data relevant to city systems
- **random access:** database made available to everyone (private users)
- **massive access:** available also to companies (systematic requests i.e. huge number)

Limitations of traditional compression schemes

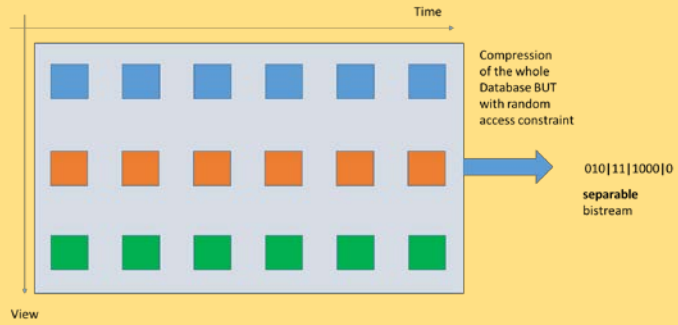


1. Send the whole database
   ⇒ inefficient transmission datarate
2. Decode the whole database
   and re-encode the request only
   ⇒ too complex for server



3. Split the database into chunks
   and compress each chunk separately
   ⇒ inefficient storage and datarate

Preliminary theoretical results



Novel coding scheme:
- Encoding into a separable bitstream
- Bit extractor at the server

Theorem. (The rate-storage region). *Let* (X,Y) *be an i.i.d. source determined by* $\{P(X,Y_j),p_j\}_j$

Consider universal lossless source coding under random user access as in the figure above. For a request j, the region of achievable rate-storage pairs (R,S) *is*

Better than expected ⟵ $R \geq H(X|Y_j)$

As expected ⟵ $S \geq \max_{k:p_k>0} H(X|Y_k)$

where $H(X|Y_j)$ *stands for the conditional entropy of the source* X *given* Y_j .

Project Overview

Task 1. Information Theory for FTV: the information theoretical bounds and coding scheme for FTV.

Task 2. Towards an end-to-end FTV compression scheme

Task 3. Information theory for Massive random access: the information theoretical bounds of massive random access and application to meteorological sensor networks