Thématiques

1. **TRES GRANDS RESEAUX ET SYSTEMES**

2. **SYSTEMES INDUSTRIELS COMPLEXES**

3. **GRANDES MASSES DE DONNEES**

4. **SECURITE, SURETE ET RISQUES**

1. TRES GRANDS RESEAUX ET SYSTEMES

VM Scheduling for Capacity Planning in Distributed Clouds

Parties prenantes

INSTITUT Mines-Télécom

MINES Nantes

UNIVERSITÉ DE NANTES

lina

Contact

Jonathan Pastor, Mines Nantes,
Ascola research team,
Département Informatique
Mines Nantes

Partenaires

The Discovery Initiative
http://beyondtheclouds.github.io

I- Background

- Cloud computing providers (Amazon, Google, Rackspace, …) serve virtual machines (VMs) that are hosted in large data centers.
- ***Capacity Planning:*** optimize ressource usage within a cloud infrastructure.
- ***VM Scheduling:*** Overloaded VMs may be relocalized by a scheduling algorithm on underloaded servers. (one important problem of capacity planning)
- Development of distributed schedulers to scale this size (In particular multi-cluster topology).

II- Problem

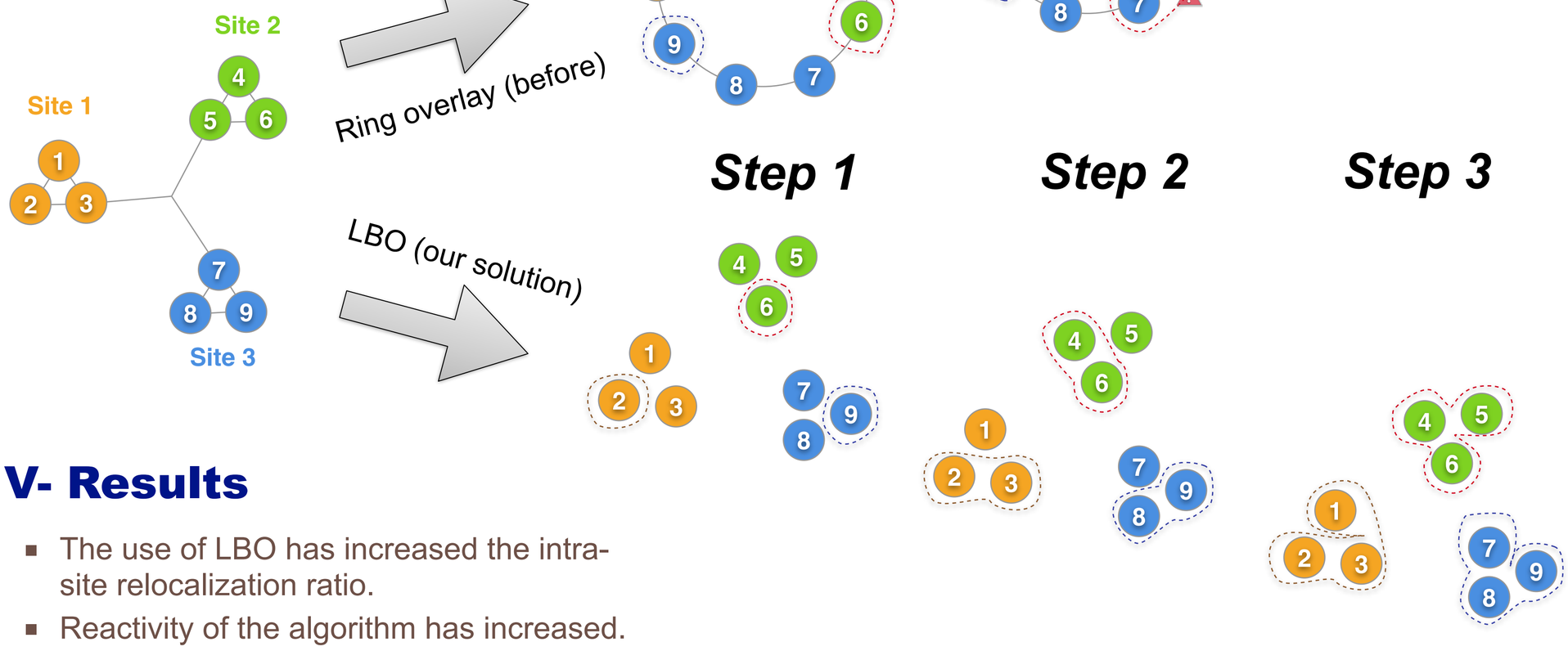
- **Current schedulers do not take into account network parameters (bandwidth, latency, …).**
- **Ineffective collaborations:**
  Inter-site collaborations ⇨ Inter-site relocalizations
- **Need of a "locality aware" VM scheduler.**

III- Objectives

- Introduce "locality properties" into DVMS (Distributed Virtual Machine Scheduler), a scalable VM scheduler, by leveraging a locality-based overlay network (LBO) instead of a ring-based overlay.
- **Maximize intra-site relocalizations and minimize inter-site relocalizations.**

IV- Example

- We take the example of a multisite configuration:
  - 3 geographical sites.
  - Each site is composed of 3 servers (nodes).
- An ISP (Iterative Scheduler Procedure) is started on nodes {2, 6, 9}.
- The example compares the use of 2 overlay networks with DVMS:
  - Ring-based overlay (Chord).
  - Locality-based overlay (Vivaldi).



V- Results

- The use of LBO has increased the intra-site relocalization ratio.
- Reactivity of the algorithm has increased.
- Inter-sites collaborations have become more efficient.

	Chord	LBO
Average	0.496	0.863
Minimum	0.378	0.798
Maximum	0.629	0.935

Table: Comparison of intra-site relocalization ratio.

VI- Towards a Fully Decentralized Cloud: the Discovery Research Initiative

- Efficiency of DVMS improved without modifying its core.
- First glimpse of the promising future of using locality properties to improve massively distributed clouds.
- This work will be included in the massively distributed cloud system developed by the ***Discovery initiative***.
- **First step toward a highly distributed cloud infrastructure that takes into account locality properties.**

Colloque Institut Mines-Télécom "Numérique : Grande échelle et complexité
Mars 2014

Vers un Nuage responsable et sécurisé

Parties prenantes

MINES Nantes

lina

UNIVERSITÉ DE NANTES

Contact

Ronan-Alexandre Cherrueau,
Mario Südholt
ASCOLA Research group
Mines Nantes

http://www.emn.fr/z-info/ascola

Partenaires

Projet Européen **A4Cloud**
(FP7, EC 317550)

The Cloud Accountability Project (A4Cloud) focuses on the Accountability For Cloud and Other Future Internet Services as the most critical prerequisite for effective governance and control of corporate and private data processed by cloud-based IT services.

www.a4cloud.eu

Le Nuage

Informatique en Nuage :
- Réservoir de services
- Très disponible (accès partout/tout le temps)
- Tolérant aux pannes
- Élastique sur les ressources allouées

Données personnelles :
- Quelques protocoles pour gérer les données perso (ex: OAuth 2.0)
- Vos données perso sont sûrement déjà dans le Nuage **!**

Problématique

Assurer les revendications des utilisateurs/fournisseurs sur l'utilisation des données pour un Nuage responable et sécurisé

Exemples :
- Alice autorise le partage de ses photos avec ses amis sur Dropbox
- Dropbox revendique la collection des métadonnées des photos pour améliorer leur recherche

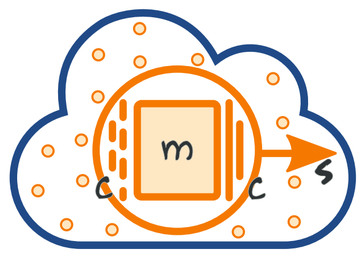
Responsabiliser le Nuage

- Requêter le Nuage pour tester si une revendication est respectée
- Empêcher la violation d'une revendication sur le Nuage

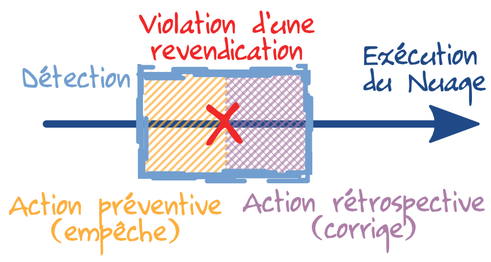
Méthode

Langage de point de coupure :
- Séquence sur l'historique d'exécution du Nuage
- Décrit la violation d'une revendication

Langage d'action :
- Modifie dynamiquement les services
- Empêche et/ou corrige la violation d'une revendication



Le langage de point de coupure prend en compte les carctéristiques des services hébergés sur le Nuage. Les séquences s'expriment aux niveaux chorégraphie (s), implémentation (m) et intercepteur (c)



En fonction du temps de la détection d'une violation, le langage d'action permet d'appliquer des actions préventives ou rétrospectives

Exemple

Revendication : OAuth Provider (OP) revendique la protection de l'identité de ses utilisateurs lors d'une authentification unique

Requête : Est-ce que le "state" est présent lors de l'authentification ?

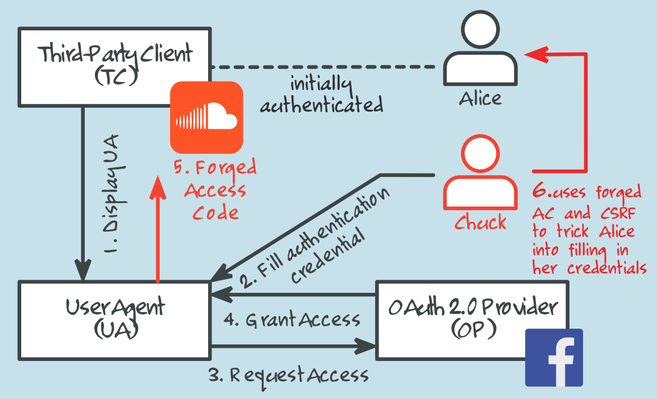
Action: Interdire l'authentification

```
pscheme AuthzExistsState? {

  // Request the Cloud:
  pscope CheckParamNotExists:
    displayUA(args, K)s ;
    [reqAcc(args, K')s &
      !exists("state", args)]
      @StateNotExists ;
    -s,c,m; K'(code, args')s ;
    K(code, args')s

  // Defines actions:
  action after StateNotExist { ... }

}
```



OAuth 2.0

- Délégue l'accès des données perso à une application tièrce
- Très largement utilisé par les acteurs du web (Facebook, Google, Reddit ...)
- Failles de sécurité dans les implementations

Conclusion & Perspectives

- Langage de détection et de correction pour appliquer des politiques sur le Nuage
- Responsabiliser et sécuriser le Nuage
- Avoir une biliothèque abstraite de solutions pour appliquer des politiques sur les données perso

Colloque Institut MInes-Télécom "Numérique : Grande échelle et complexité
Mars 2014

Synthèse de traces réelles de centres de données virtualisés

Parties prenantes

MINES Nantes

EasyVirt

Auteurs

Frédéric DUMONT
frederic.dumont@mines-nantes.fr

Jean-Marc MENAUD
menaud@mines-nantes.fr

Partenaires

FONDATION TELECOM

INSTITUT CARNOT Télécom & Société numérique

1- CONTEXTE

De par sa flexibilité, le **Cloud Computing** s'est imposé comme un nouveau modèle technique et économique au sein des entreprises. Cependant, l'effet rebond de cette flexibilité et élasticité s'est traduit par l'explosion du nombre d'environnements virtuels à gérer. Il n'est plus rare qu'un administrateur soit amené à administrer un parc de plusieurs centaines voir milliers de machines virtuelles. Sans outil adapté d'aide à la gestion du parc, cette tâche d'administration peut vite se révéler impossible à réaliser.

2- OBJECTIF

Notre objectif est de regrouper par similarité des ensembles de VM puis de déterminer celles ne pouvant être regroupées avec les autres. Cette approche classique d'analyse de données s'appuie sur une technique bien connue : le **Clustering**. Le **Clustering** consiste à regrouper un ensemble de points, caractérisés par plusieurs dimensions, en partitions (ou clusters) de points similaires. La similarité est exprimée par l'utilisation d'une mesure de distance entre les points.

L'algorithme le plus usité, K-MEANS[1], propose de diviser un ensemble de points en k partitions afin d'obtenir une similarité satisfaisante pour l'ensemble des K partitions. Cette approche itérative cherche à déterminer K centroïdes, un point de l'espace définissant le centre d'une partition. Cet algorithme possède plusieurs défauts dont la sensibilité au bruit, défaut bloquant puisque nous cherchons les VM atypiques.

[1]J. McQueen. Some methods for classification and analysis of multivariate observations. In In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, page pp. 281–297, 1967.

3- METHODE PROPOSEE

Nous avons développé un algorithme de partitionnement multicritères et multi-ressources insensible aux bruits. La distance utilisée dans notre algorithme est calculée suivant un "taux de ressemblance", paramétrable, permettant de définir les bornes minimales et maximales des intervalles des valeurs statistiques.

4- RESULTATS

La figure1 détaille le partitionnement réalisé par notre approche : un groupe composé d'un nombre important de VM et de plus petits groupes composés de 1 à 3 VMS. Sur la figure2, l'algorithme K-MEANS partitionne en un nombre important de petits groupes, assez homogènes en nombre de VM, finalement peu exploitables dans notre cas.

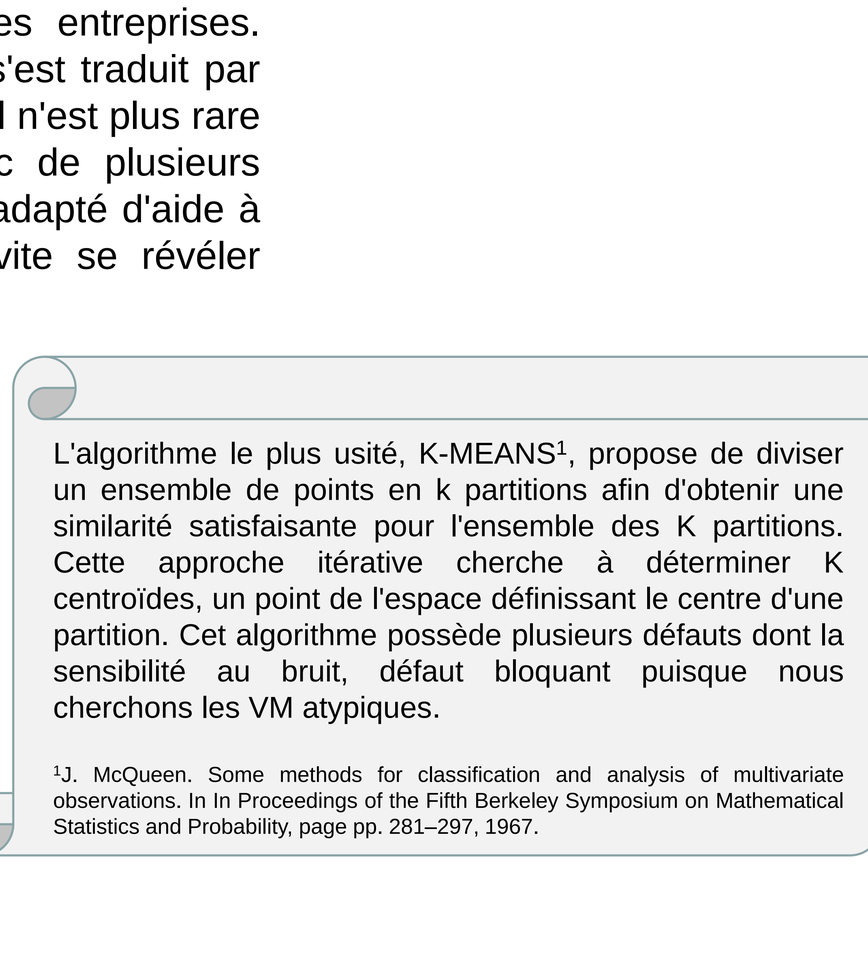


FIGURE 1

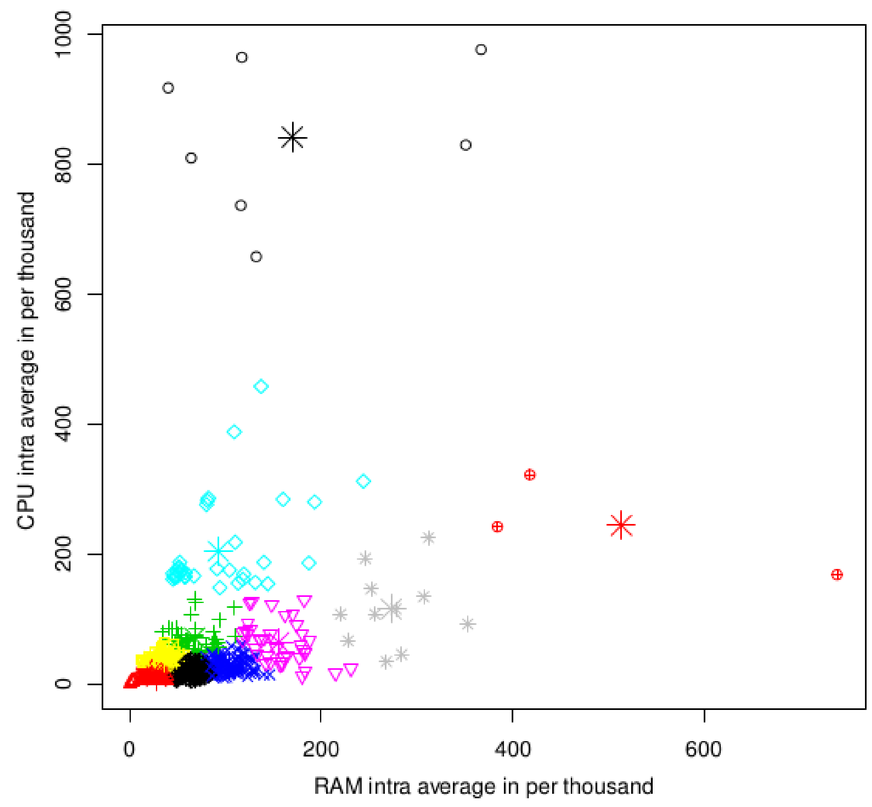


FIGURE 2

5- CONCLUSION

Le **Clustering** est une technique consistant à regrouper par partitions un ensemble de points similaires. La similarité est exprimée par une mesure de distance entre les points. Nous avons étudié les différents concepts du Clustering ainsi que les principaux algorithmes existants. De par leurs limites, aucunes méthodes existantes ne répondent à nos besoins. Nous avons alors développé notre propre algorithme de Clustering, insensible aux bruits, performants, multi-ressources et multicritères.

Colloque Numérique : Grande échelle & Complexité
Mercredi 26 et Jeudi 27 mars 2014

VM Scheduling for Capacity Planning in Distributed Clouds

Parties prenantes

INSTITUT Mines-Télécom

MINES Nantes

UNIVERSITÉ DE NANTES

lina

Contact

Jonathan Pastor, Mines Nantes,
Ascola research team,
Département Informatique
Mines Nantes

Partenaires

The Discovery Initiative
http://beyondtheclouds.github.io

I- Background

- Cloud computing providers (Amazon, Google, Rackspace, …) serve virtual machines (VMs) that are hosted in large data centers.
- ***Capacity Planning:*** optimize ressource usage within a cloud infrastructure.
- ***VM Scheduling:*** Overloaded VMs may be relocalized by a scheduling algorithm on underloaded servers. (one important problem of capacity planning)
- Development of distributed schedulers to scale this size (In particular multi-cluster topology).

II- Problem

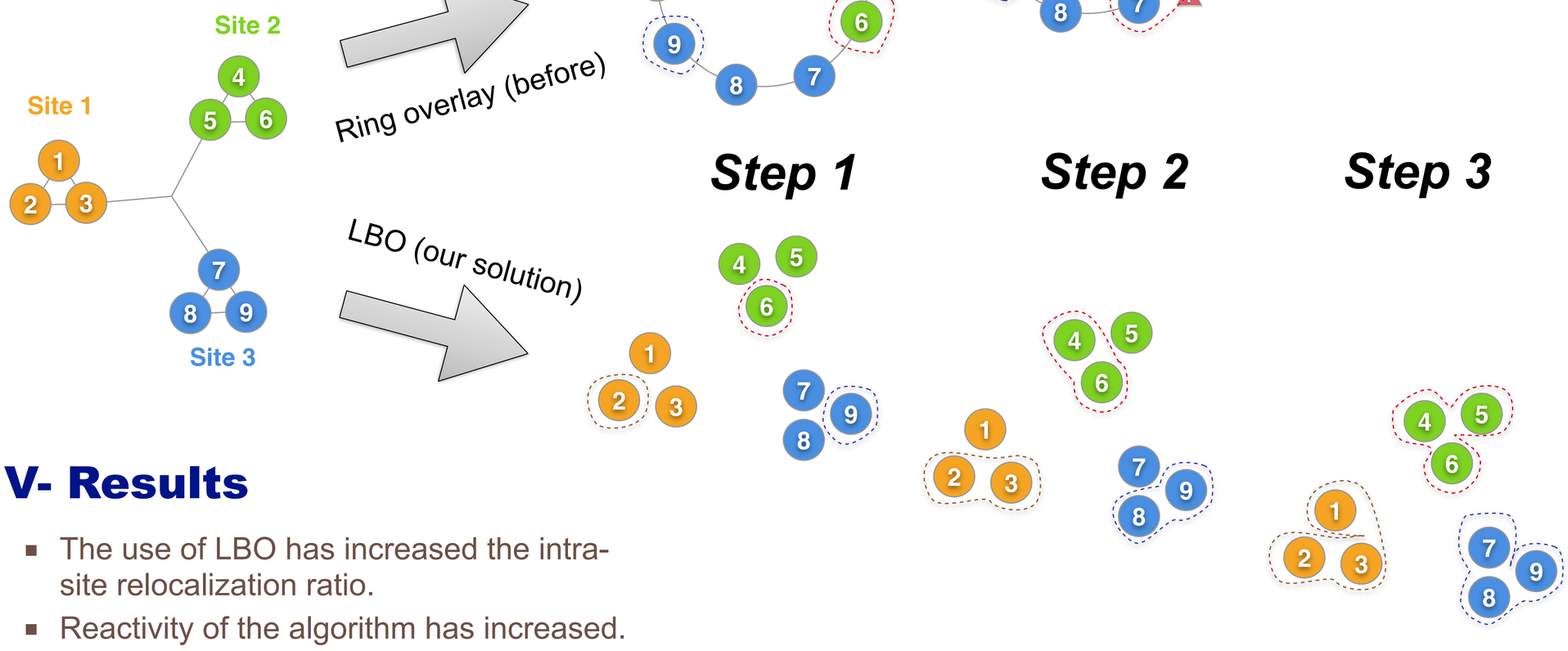
- **Current schedulers do not take into account network parameters (bandwidth, latency, …).**
- **Ineffective collaborations:**
  Inter-site collaborations ⇨ Inter-site relocalizations
- **Need of a "locality aware" VM scheduler.**

III- Objectives

- Introduce "locality properties" into DVMS (Distributed Virtual Machine Scheduler), a scalable VM scheduler, by leveraging a locality-based overlay network (LBO) instead of a ring-based overlay.
- **Maximize intra-site relocalizations and minimize inter-site relocalizations.**

IV- Example

- We take the example of a multisite configuration:
  - 3 geographical sites.
  - Each site is composed of 3 servers (nodes).
- An ISP (Iterative Scheduler Procedure) is started on nodes {2, 6, 9}.
- The example compares the use of 2 overlay networks with DVMS:
  - Ring-based overlay (Chord).
  - Locality-based overlay (Vivaldi).



V- Results

- The use of LBO has increased the intra-site relocalization ratio.
- Reactivity of the algorithm has increased.
- Inter-sites collaborations have become more efficient.

	Chord	LBO
Average	0.496	0.863
Minimum	0.378	0.798
Maximum	0.629	0.935

Table: Comparison of intra-site relocalization ratio.

VI- Towards a Fully Decentralized Cloud: the Discovery Research Initiative

- Efficiency of DVMS improved without modifying its core.
- First glimpse of the promising future of using locality properties to improve massively distributed clouds.
- This work will be included in the massively distributed cloud system developed by the ***Discovery initiative***.
- **First step toward a highly distributed cloud infrastructure that takes into account locality properties.**

Colloque Institut Mines-Télécom "Numérique : Grande échelle et complexité
Mars 2014

Gouvernance Agile de Réseaux d'équipements Machine-to-Machine (M2M)

Parties prenantes

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Auteurs

Camille Persson(*,**), Gauthier Picard (*), Fano Ramparany (**), Olivier Boissier (*)

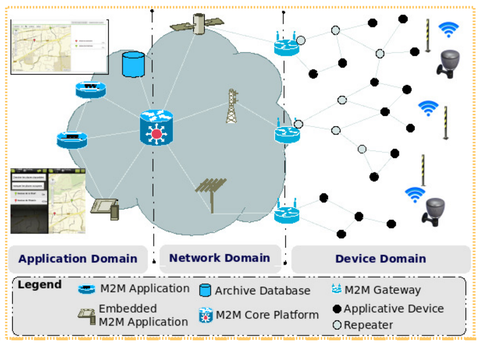
(*) Institut Henri Fayol / ISCOD ENS Mines Saint-Etienne

(**) R&D/TECH/MATIS/COSY Orange Labs Network and Carrier (France Teecom)

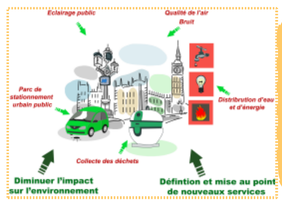
Partenaires

orange™

Contexte



Architecture d'un système M2M. Décomposition en 3 domaines : Application, Réseaux, Appareils



Senscity propose une infrastructure Machine-to-Machine (M2M) mutualisée pour la gestion et le déploiement de services basés sur des réseaux de capteurs et actionneurs interagissant directement avec le monde physique, à l'échelle urbaine (Smart City).

- *Contraintes d'une architecture M2M en mileu urbain*
- Réseaux de ***capteurs*** et ***actionneurs*** (WSAN) ultra-basse consommation et bas débit
- Durée de vie (20 ans sur 1 batterie)
- Large échelle : millions de capteurs, étendue géographique, nombreuses applications, gros volume de données ...

Il est nécessaire de fournir un modèle de gouvernance permettant le passage à l'échelle des infrastructures Machine-to-Machine

Problématique du passage à l'échelle dans le M2M

Définition : Extensibilité (Scalability)

Un système S est extensible sur un ensemble de dimensions D_S, jusqu'à un point fini d_{bound}, par rapport à un ensemble de propriétés $\mathcal{P}_S$, s'il est capable de gérer les changements d'échelles en satisfaisant les spécifications r_{P_i} du systèmes, ou s'il est capable de trouver un compromis $\mathrm{r}_{P_i,bearable}$ entre les propriétés satisfaisant globalement l'ensemble des spécifications.

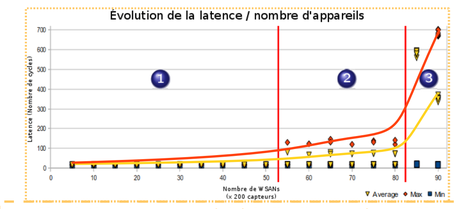
Soit :

$$\exists d_{bound} \in \mathrm{D}_S,\ \forall d \leq d_{bound},\ \forall \mathrm{P}_i \in \mathcal{P}_S \quad (a)$$
$$\exists \mathrm{r}_{P_i,bearable} \in \mathcal{P}_S \times \mathcal{D}_S :: r_{P_i,bearable} = True \quad (b)$$
$$\mathrm{r}_{P_i} = True \iff r_{P_i,bearable} = r_{P_i} \quad (c)$$

- D_S est un tuple $\langle D_0, ..., D_i, ..., D_N \rangle$
- $D_i \in \mathcal{D}_{\mathcal{M}\in\mathcal{M}}$ est défini sur un domaine continu ou discret, ordonné partiellement ou totaltement
  Soit $=_{D_i}, <_{D_i}$ deux fonction d'ordre définies sur $D_i \times D_i$
- r_{P_i} et $\mathrm{r}_{P_i,bearable}$ sont des fonctions d'évaluation des spécifications pour la propriété $P_i \in \mathcal{P}_{\mathcal{M}\in\mathcal{M}}$
  Soit $r_{P_i} : D_S \longrightarrow \{True|False\}$

On regroupe les dimensions du passage à l'échelle et les propriétés des systèmes M2M par catégories : {D1, D2, D3, D4} / {P1, P2, P3}

(D1) Taille du système
- *Nombre d'appareils*
- *Nombre d'utilisateurs*
- *Quantité de données*
- *Nombre de services*
- *Nombre de partenaires*



Simulation de passage à l'échelle d'un système M2M

1 *Fonctionnement normal* 2 *Apparition de problèmes* 3 *Système défaillant*

(D2) Hétérogénéité
- *Types d'appareils*
- *Canaux de communications*
- *Types de données*
- *Technologies de développement*

(P1) Utilisation
- *Latence*
- *Efficacité*
- *Disponibilité*
- *Privacité*
- *Simplicité*

(P2) Performance
- *Efficience*
- *Sûreté*
- *Autonomie*
- *Cohérence*
- *Robustesse*

(P3) Administration
- *Évolutivité*
- *Faisabilité*
- *Maintenabilité*
- *Observabilité*
- *Testabilité*

(D3) Topologie
- *Étendue géographique*
- *Topologie réseaux*
- *Mobilité des services/appareils*
- *Répartition des données*

(D4) Monde physique
- *Durée de vie*
- *Dynamique de l'environnement*
- *Bruits et erreurs*
- *Hostilité du milieu*
- *Comportements émergents*

Complexité de la gestion du passage à l'échelle des systèmes M2M nécessite une approche décentralisée (multi-agent)

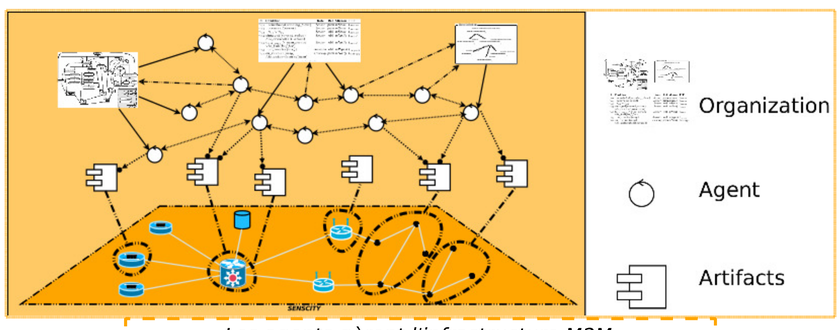
Organisation multi-agent pour la gouvernance du M2M

- ***Système muli-agent centré Organisation pour la gouvernance du M2M :***

Agents *Entités autonomes pro-actives qui gèrent le fonctionnement de l'infrastructure M2M*

Artefacts *Composants de l'architecture M2M manipulables par les agents (eg. modules de la plateforme, capteurs...)*

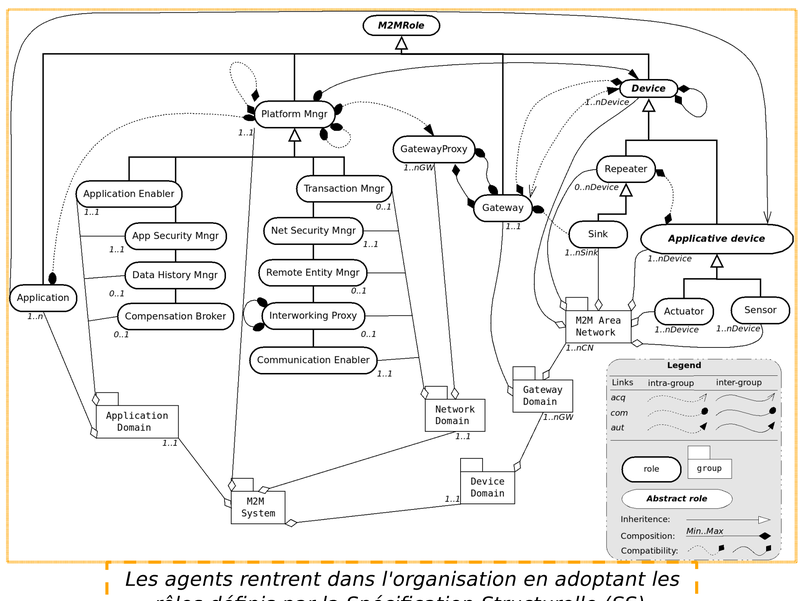
Spécifications Organisationnelles (OS) *Définition de la structure (SS), du fonctionnement (FS) et des règles (NS) de l'organisation basée sur le framework $\mathcal{MOISE}$*
Permet de garantir un fonctionnement global conforme aux spécifications r_{P_i} du système



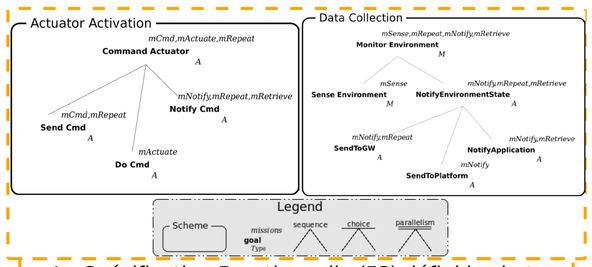
Les agents gèrent l'infrastructure M2M, dont les composants sont représentés par des artefacts.

- ***Gouvernance Agile pour le passage à l'échelle de l'infrastructure M2M :***
  - *Situations non prévues (extension du système)*
  - *Organisation non adaptée aux changement d'échelle (eg. buts non atteignables, délais trop courts...)*
  - *Besoin de redéfinir les Spécifications Organisationnelles*

Les agents peuvent détecter localement les situations nécessitant une réorganisation et proposer des solutions locales pour permettre un meilleur fonctionnement du système



Les agents rentrent dans l'organisation en adoptant les rôles définis par la Spécification Structurelle (SS).



La Spécification Fonctionnelle (FS) défini les buts du système, regroupés dans des missions, ainsi que les plans permettant d'atteindre ces objectifs.

Id	Condition	Role	Rel.	Mission	TTF
n_{01}	$scheduled(sensing\_time)$	Sensor	perm	$mSense$	t_{sense}
n_{02}	$occurred(event)$	Sensor	perm	$mSense$	t_{sense}
n_{03}	$n_{01} \vee n_{02}$	Sensor	obl	$mSense$	t_{sense}
n_{04}	$changed(sensed\_value) \wedge is\_critical(situation)$	Sensor	obl	$mNotify$	t_{send}
n_{05}	$t_{last\_msg} \geq msg\_period \vee is\_full(buffer)$	Sensor	obl	$mNotify$	t_{send}
n_{06}	$on\_receive(msg)$	Repeater	obl	$mRepeat$	t_{repeat}
n_{07}	$on\_receive(msg) \wedge is\_authenticated(msg)$	Gateway	perm	$mNotify$	t_{notify}

Les normes (NS) associent les missions aux rôles et définissent les conditions pour remplir ces missions.

Différents niveaux de réorganisation :

Structurelle	***Fonctionnelle***	***Normative***
Cardinalités des rôles	Supprimer des buts non atteignables ou trop couteux	Renforcer/relacher les relations déontiques
Nouveaux rôles	Redéfinition des plans de buts	Redéfinition des conditions
Compatibilité des rôles	Cardinalité des missions	Redéfinition des TTFs
Renforcement des communications	Redéfinition des TTFs	

Conclusion

- **Contribution :**
  - ***Analyse du passage à l'échelle dans le M2M***
  - ***Modèle de gouvernance de systèmes M2M basé sur une organisation Multi-Agents***
- **Travaux futurs :**
  - ***Définir les comportements de réorganisation des agents : quand ? qui ? quoi ? pourquoi ?***
  - ***Garantir la cohérence des réorganisations***
  - ***Déploiement/test sur l'infrastructure SensCity***

Numérique : Grande échelle et complexité

26 et 27 mars 2014

Contact Camille.Persson@emse.fr

Site web www.emse.fr

Système de recommandation de lieux basé sur l'intention, le contexte et le réseau social

Parties prenantes

Lab-STICC

TELECOM Bretagne

Auteurs

Romain PICOT-CLEMENTE

Cécile BOTHOREL

Philippe LENCA

Lab-STICC CNRS, UMR 6285, Télécom Bretagne

Partenaires

FONDATION TELECOM

INSTITUT CARNOT Télécom & Société numérique

CréditMutuelArkéa

CREM

ENSICAEN ÉCOLE NATIONALE SUPÉRIEURE D'INGÉNIEURS DE CAEN & CENTRE DE RECHERCHE

DIGITAL AIRWAYS USER INTERFACE SOLUTIONS

SYSNAV Navigation Solutions

MONEXT

aqoba

OBJECTIF

- Recommandation de lieux à un utilisateur en condition de mobilité selon :
  - son intention de visite : visite découverte, visite efficace
  - son contexte : session de visites passées + position géographique courante
  - son réseau social : les visites de ses amis

PROPOSITION

Construction d'un modèle de règles d'association (hors-ligne)

- Extraction de règles d'association par algorithme FP-Growth à partir des sessions de visites de tous les utilisateurs, de la forme : $Règle\ R : A \rightarrow B$ (Conséquent / Antécédent)

  Exemple de règle : « un utilisateur qui visite la Tour Eiffel et l'Arc de Triomphe visite le Louvre »

Processus de recommandations pour un utilisateur (temps-réel)

- Sélection des règles d'association dont $A \subset (session\ courante\ de\ visite)$
- Classement des règles $R: A \rightarrow B$ selon une mesure de pertinence M_p

$$M_p(R) = M_g(R)\left(\alpha M_i(R) + (1-\alpha)\, M_s(R)\right)$$

Mesure géographique / Mesure d'intérêt basée sur l'intention / Mesure sociale

$M_i(R: A \rightarrow B) = confiance(R: A \rightarrow B) = \frac{P(AB)}{P(A)}$,
si intention = efficacité

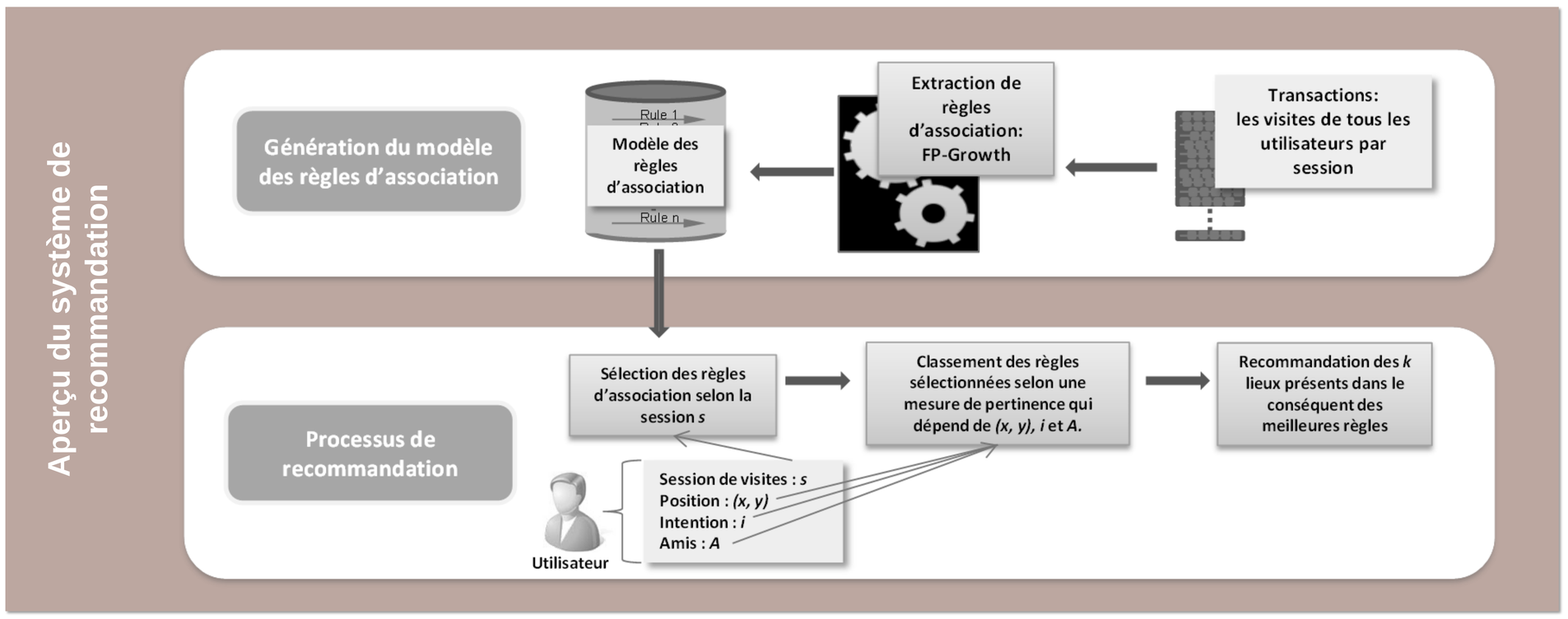
$M_i(R: A \rightarrow B) = surprise(R: A \rightarrow B) = \frac{P(AB)-P(A\bar{B})}{P(B)}$,
si intention = découverte

$M_g(R: A \rightarrow B) = \alpha.distance(B, utilisateur)^{\beta}$

$M_s(R: A \rightarrow B) = confiance_{amis}(R: A \rightarrow B)$

- Recommandation des k lieux conséquents des k meilleures règles

APERCU DU SYSTÈME DE RECOMMANDATION



APPLICATION SUR DONNEES REELLES

Recommandation de magasins, tests de montée en charge

PAY2YOU PLACES

- Données très volumineuses de paiement d'utilisateurs dans des magasins sur 1 année : cluster Hadoop de 80 machines, 1 million de transactions, 40 minutes pour la phase de génération du modèle de règles
- Algorithme FP-Growth Map-Reduce -> point d'étranglement, une étape utilise un reducer unique (construction de l'arbre)
- Expérimentation et évaluation des recommandations sur des utilisateurs réels à venir

hadoop

APACHE HBASE

Références complémentaires

PICOT-CLEMENTE Romain, BOTHOREL Cécile, LENCA Philippe**. Towards Intention, Contextual and Social based Recommender System.** ACIIDS 2014 : The 6th Asian Conference on Intelligent Information and Database Systems, 7 - 9 Avril 2014, Bangkok, Thailand, 2014.

PICOT-CLEMENTE Romain, BOTHOREL Cécile. **Un système de recommandation de lieux basé sur la mesure de Katz dans les réseaux sociaux géographiques**. MARAMI 2013 : 4ième conférence sur les modèles et l'analyse des réseaux : Approches mathématiques et informatiques, 16-18 octobre 2013, Saint-Etienne, France, 2013.

PICOT-CLEMENTE Romain, BOTHOREL Cécile. **Recommendation of shopping places based on social and geographical influences.** RSWeb 2013 : 5th ACM RecSys Workshop on Recommender Systems and the Social Web, 13 octobre 2013, Hong Kong, Hong Kong, 2013.

CHALMERS Sean, BOTHOREL Cécile, PICOT-CLEMENTE Romain**. Big Data - State of the Art.** Rapport technique sur les techniques et plateformes Big Data avec des recommandations sur le choix en fonction des besoins. Lien : http://portail.telecom-bretagne.eu/publi/public/fic_download.jsp?id=21241

Contact romain.picotclemente@telecom-bretagne.eu Site web www.telecom-bretagne.eu

Future Internet: Enabling Inter-Domain Path Diversity

Funded by ETICS, Seventh Framework Programme

Auteurs

Xavier Misseri,

Jean-Louis Rougier

Network and Computer Science Department,

Telecom ParisTech

Collaboration avec:

Damien Saucez,

INRIA, Sophia Antipolis

Ivan Gojmerac

FTW, Wien (Austria)

Helia Pouyllau,

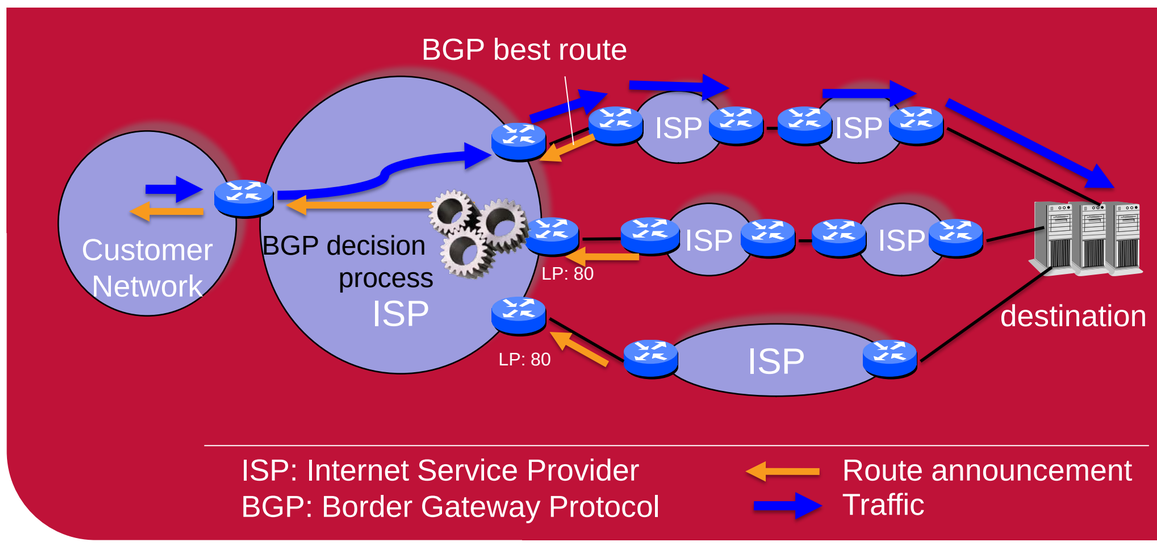
Lamine Lamali

Alcatel Lucent Bell Labs

Current Internet limitation

One single route to destination

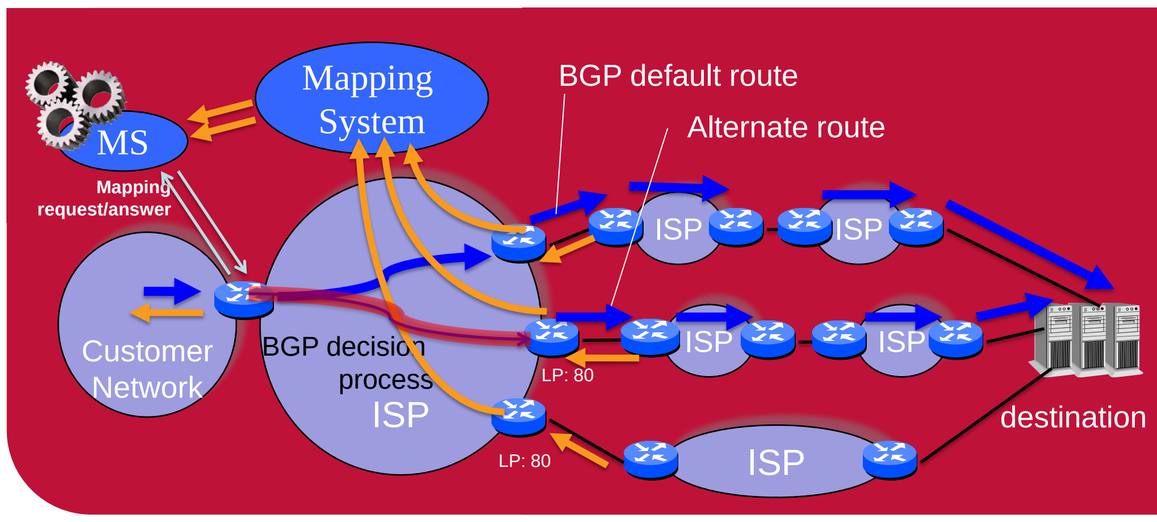
- **Reason:** BGP (current routing protocol) selects one route to any destination, based on a rigid « decision process »
  - Local Preferences,
  - Path length
- **Consequence:** « One fits all » model in contrast with variety of applications (e.g. data, streaming) and customers (eyeballs, content providers, …)
- **Fact:** Huge Potential Internet Diversity (7 routes available in average for Tier 1 providers)



Proposed Incremental Architecture

Step 1: Customer-Provider

- **Proposal:** Select the exit domain to benefit from path diversity
  - Path enforcement via encapsulation (bypass BGP default route)
  - Path diversity management via a Mapping System
  - IETF LISP architecture can be used to implement our scheme [1]



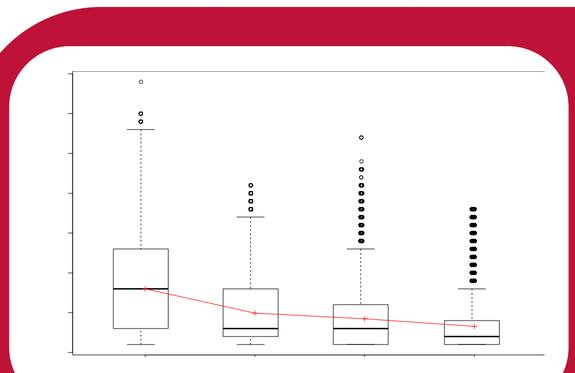
Step 2: Provider-Provider

- **Proposal:** Interconnection of Mapping Systems for propagation of diverse paths
- **Issue:** Global Internet Routing Stability insured with well-known Rules (Gao & Rexford). Risk: This rules are no longer valid when considering path diversity.
- **Contribution:** Simple and Generic Stability rules for path diversity routing [2]. More flexible rules (more flexibility allowed in the choice of routes).
- Approach is scalable [3] and incremental: One ISP starts to benefit from path diversity without any cooperation with other ISPs (Step 1). Diversity then further increase with Mapping interconnections (Step 2)

Results (Evaluation on Internet topology)

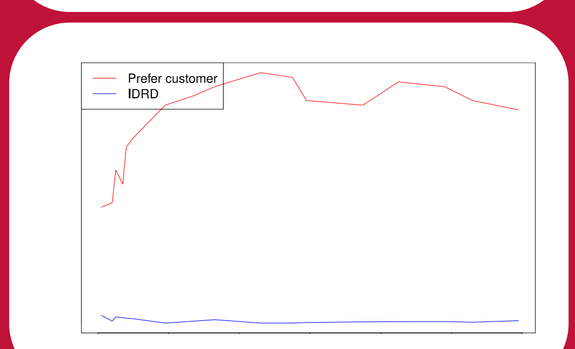
Route Diversity: a Key Feature for Traffic Engineering

- Robustness. Possibility to use disjoint routes (fast path switching, without waiting for global routing re-convergence)
- Flexibility. Allows ISP to announce the “best” routes to its customers, based on specific customer needs and/or flow requirements [4].



Step1: Distribution of Route Diversity for a Tier 1 Network,
based on different route filtering: All, cheapest route only (Local Pref), “shortest path” only (AS Path Length), both LP+AS

(CAIDA & Route Views databases)



Step2: Probability that no alternative disjoint route is available
In the the current Internet (red), with proposed scheme and relaxed stability conditions (blue).

(CAIDA database)

Some References:

[1] X. Misseri, J.-L. Rougier, and D. Saucez, “Internet routing diversity for stub networks with a Map-and-Encap scheme,” in IEEE International Conference on Communications (ICC), 2012

[2] X. Misseri, I. Gojmerac, and J. L. Rougier, “IDRD: Enabling Inter-Domain Route Diversity,” in IEEE International Conference on Communications (ICC), 2013

[3] X. Misseri, I. Gojmerac, and J. Rougier, “Internet-wide multipath: a scalability analysis of path identification schemes,” in Network Of the Future (NOF), 2012

[4] H.Pouyllau, M.L.Lamali, X.Misseri, J.L.Rougier, “Method and system for advertising inter-domain routes”. European Patent No. 13176712.1-1853. 2013.

Contact rougier@telecom-paristech.fr

REFLEX
Multiband Digital Predistortion

Parties prenantes

TELECOM ParisTech

ESIEE PARIS

Auteurs

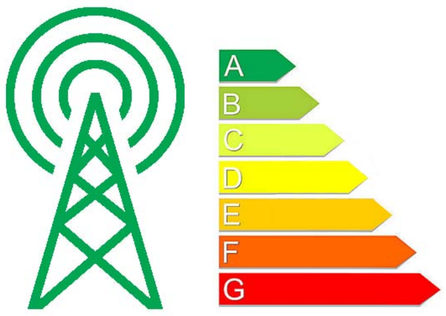
Dang-Kièn Germain Pham
Patricia Desgreys
Mazen Abi Hussein
Olivier Venard
Patrick Loumeau

Partenaires

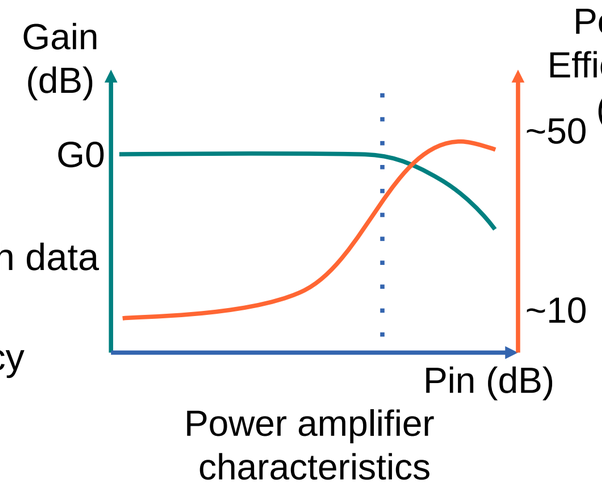
FONDATION TELECOM

INSTITUT CARNOT Télécom & Société numérique

MOTIVATIONS

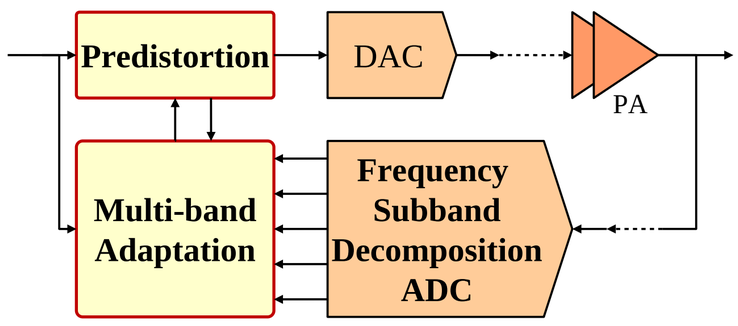


- Future telecommunication networks: More and more wireless devices, High data rate transmissions, Reduce energy footprint
- Main contributor: Base station power amplifier, Trade-off Linearity/Efficiency



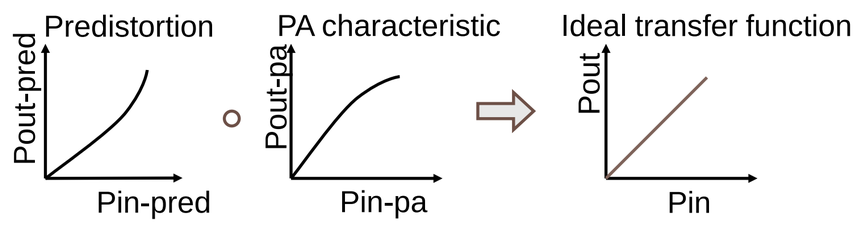
Power amplifier characteristics

Multirate Digital Predistortion



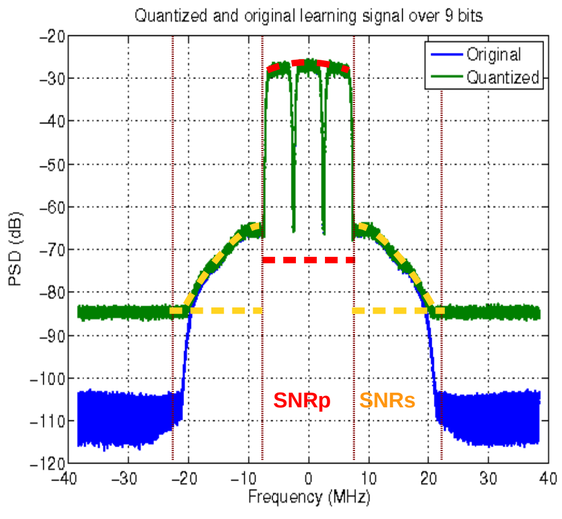
- Develop an algorithm for digital predistortion adapted to a multi-band ADC

- DPD principle



- ADC : Optimized parallel bandpass ΣΔ ADC
  - MSNBC architecture [patented in 2012]

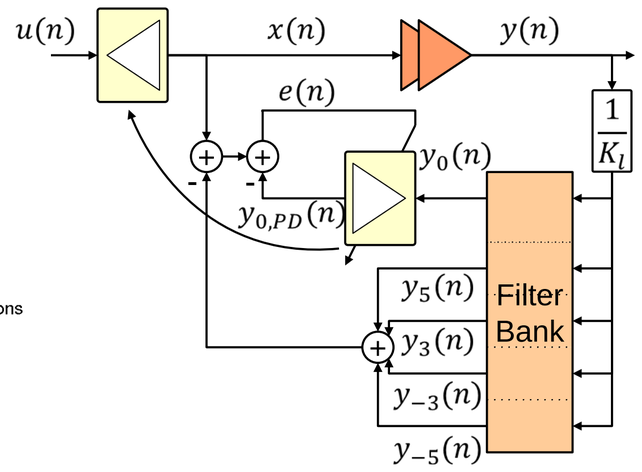
SUBBAND QUANTIZATION EFFECT ON A CLASSIC DPD SYSTEM



Simulation results summary

- Case 1 : Uniform quantization
  - Same « quantum » for each subband
  - Optimum resolution : 10 bits
- Case 2 : Fix resolution of the high power subband
  - Set the SNR of subband 'P'= 64dB
  - Correction perf. very sensitive to the quantization of adjacent bands
- Case 3 : Fix resolution of the adjacent subbands
  - Set the SNR of subband 'S'= 22dB
  - The resolution of the high pow. subband can be reduced to 8 bits

SUBBAND DIGITAL PREDISTORTION ALGORITHM



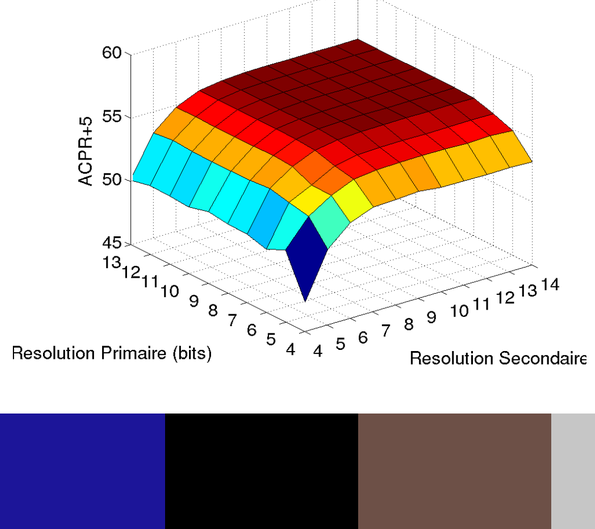
- PA model : Memory-polynomial
- DPD model : Memory-polynomial

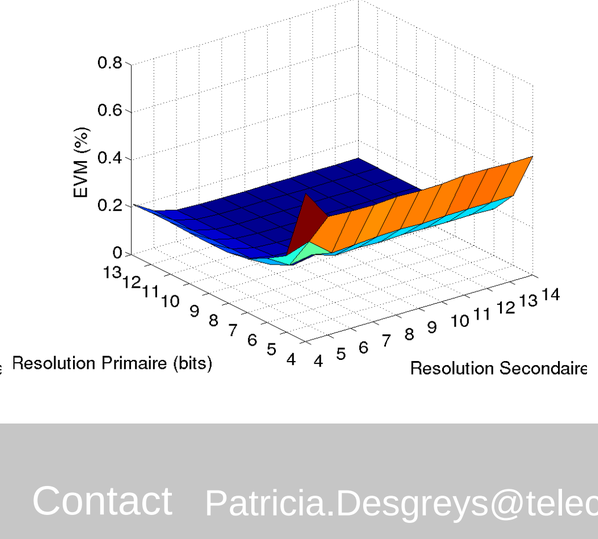
- Performance metrics :
  - ACPR:

    More sensitive to quantization of adjacent subbands for low resolutions
  - EVM:

    Depends mostly on quantization of the high power subband

Performance metrics simulated results





Future work

- Multirate implementation
- Feasibility study on the implementation on digital processor (DSP / FPGA)
- Resource gain estimation

Contact Patricia.Desgreys@telecom-paristech.fr

SMART GRID

PhDs: BLIMAN Federico, HORTA Jose, KADDAH Rim, OBADIA Mathis
Advisors: KOFMAN Daniel and ROUGIER Jean-Louis

Why a smarter grid?

- Reduce CO2 emission (The 20-20-20 targets)
- Energy self sufficiency
- Enhance reliability
- Reduce capex and opex costs
- Advanced service models

Means

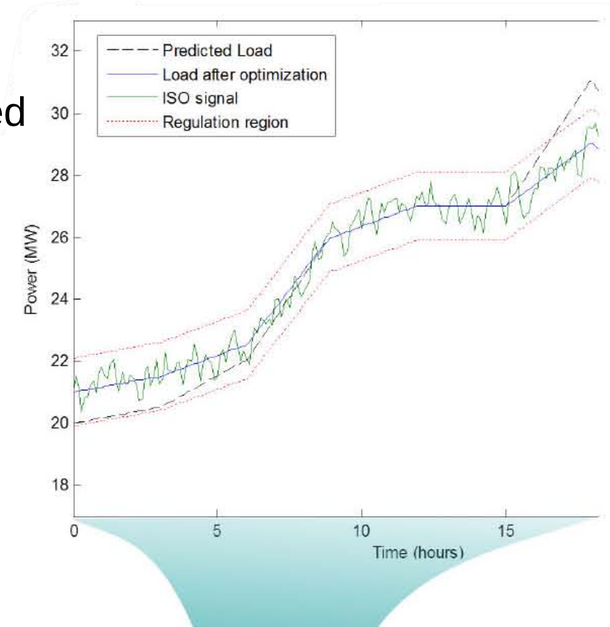
- Change the load curve shape (reduce peak, lower consumption)
- Distributed energy resources
- Renewable energy sources (wind, PV,…)
- Enhance efficiency

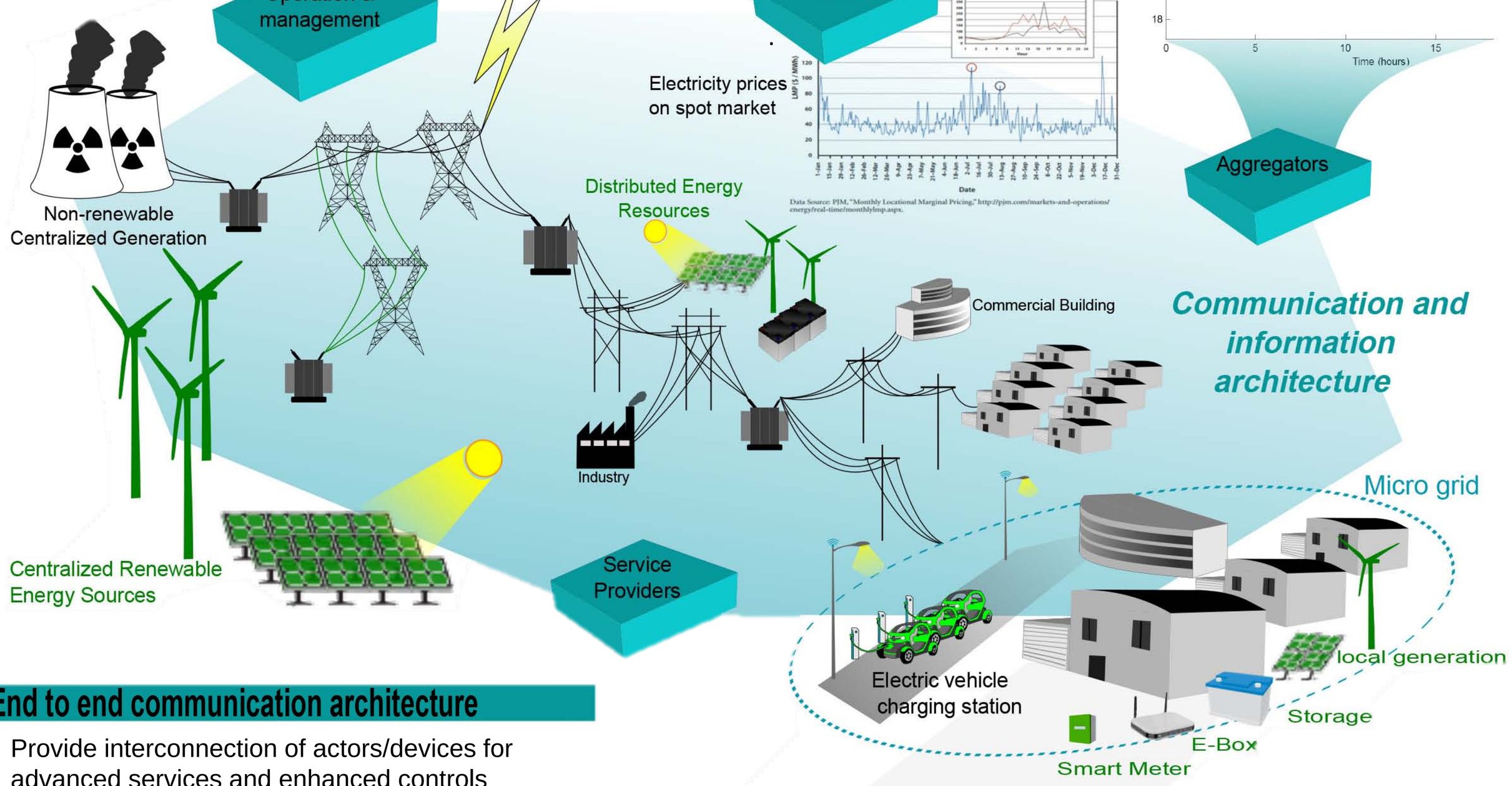
Crisis management (power shortage)

- Traditional approach: Rolling blackout
- Our approach: **Differentiated services**
  - Continuous supply for **critical** loads
  - Take into account **utility** for users depending on their characteristics, environmental conditions and appliances' operation
  - **Fairness**

Aggregators

- Provide advanced DR mechanism to leverage consumers' storage capabilities and load and generation flexibility
- Enable prosumers' participation in the electricity market, including ancillary market
- Dynamically optimize Aggregator's decisions based on: load forecasts, client policies, market prices, flexibility capabilities and ISO signals



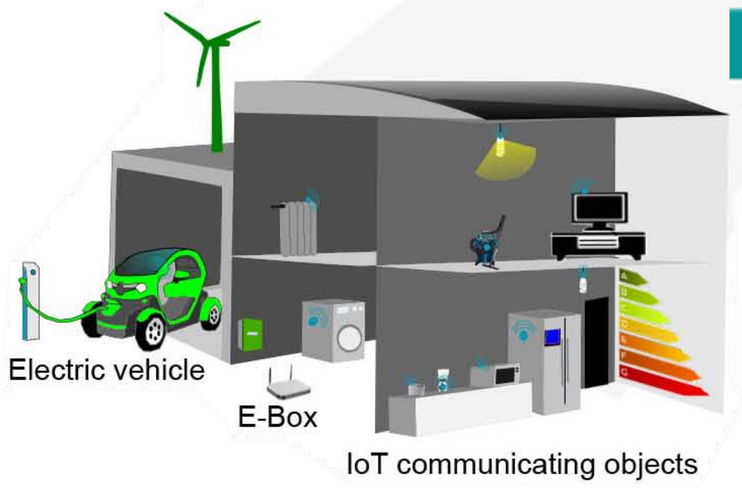


End to end communication architecture

- Provide interconnection of actors/devices for advanced services and enhanced controls
- Optimal distribution of overall system intelligence
- Requirements: Interoperability, Flexibility, Reliability, Security, CAPEX & OPEX.
- Based on ESOs work for M/490 mandate

Internet of Things

- Architecture for customer energy management system targeting autonomic policies' implementation:
  - auto-discovery, self-configuration and self-healing
- Solutions for advanced grid monitoring and control
- Smart grid, vehicles, cities and homes convergence



Microgrid Management

- Manage cooperatively electricity production and consumption locally on a neighborhood or campus level
  - Leverage local storage and renewable energy sourcing capabilities
  - Enhance efficiency (e.g., less transport losses)
- Ensure overall system visibility, stability and predictability

Lisp-Lab

Building Today the Internet of Tomorrow

Institut Mines-Télécom

AGENCE NATIONALE DE LA RECHERCHE ANR

Auteurs

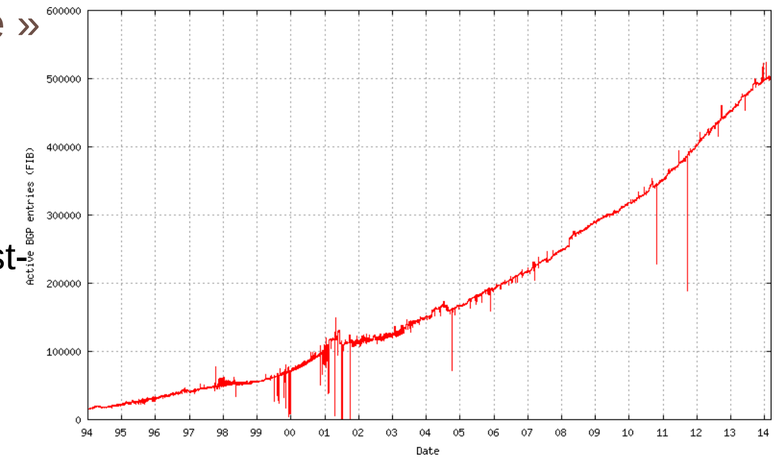
Luigi Iannone

Jean-Louis Rougier

Internet Scalability

When « large scale » is synonym of « complex and expensive »

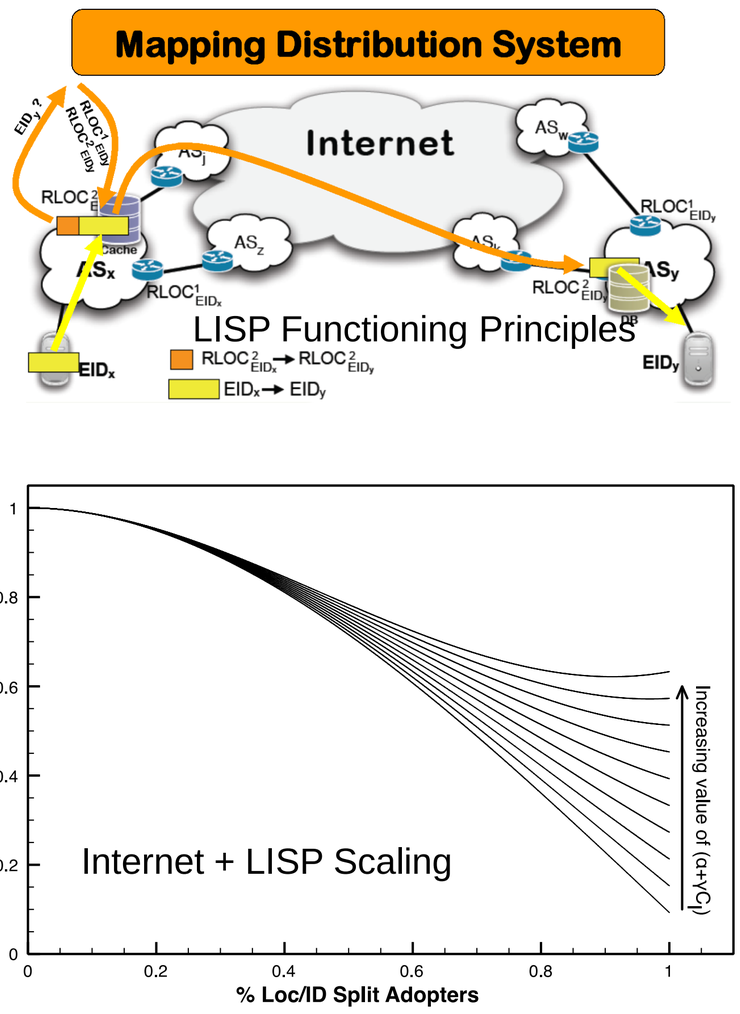
- BGP Routing Information Base (RIB) in the Default Free Zone is growing at fast rate, causing scalability problems.
- The “opex” costs for maintaining, updating, provisioning, and managing this large amount of entries, makes the Internet less cost-effective.
- Why this BGP Inflation?
  - **Single numbering space:** for both host transport sessions identification and network routing.
  - **Traffic Engineering:** BGP announces only the best-path, hence traffic engineering is performed by de-aggregating prefixes.



“Addressing can follow topology or topology can follow addressing. Choose one.”

Y.Reckter

Partenaires



LISP Functioning Principles

LISP: Locator/ID Separation Protocol

Toward a thinner Internet Core

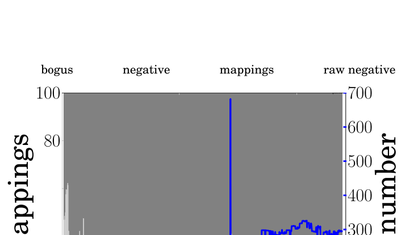
- LISP Principles: **Map-and-Encap**
  - Different addressing spaces to identify end-hosts and locate routing’s infrastructure end-points (stub domain’s border routers).
    - **End-system Identifiers (EIDs)**: End-systems are identified by their IP address, which lays in a separated space in respect of the inter-domain routing infrastructure.
    - **Routing LOCators (RLOC)**: The IP address of border router(s) locate, in the routing infrastructure, the attachment point of the domain to which a certain EID pertains.
- Map between the two spaces and tunnel (encap) packets in the core Internet.
  - **Mapping EIDs to RLOCs:** To set up end-to-end communication a mapping function is needed to associate the EIDs (the who) with the RLOCs (the where).

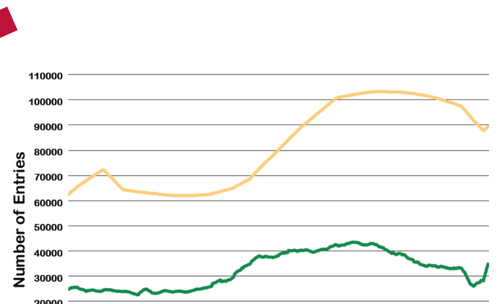
Internet + LISP Scaling

The LISP-Lab Approach

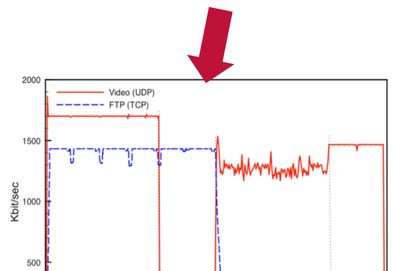
Leading the way to Future Internet Services

- The LISP-Lab Platform aims at providing an environment for high quality research and the design, development, and assessment of new services and use-cases.
- Technical tasks planned in the LISP-Lab project range from cloud networking, to access technology, through inter-domain connectivity, traffic engineering, and network management, has a large scope to boost innovation beyond the LISP technology itself.





LISP-Lab Planed Platform Infrastructure



Partenaires: LIP6, Alphalink, border 6, NONSTOP Systems, Orange Labs, Renater, UCOPIA COMMUNICATIONS, rezopole

Contact Luigi.Iannone@telecom-paristech.fr

Site web www.lisp-lab.org

Studying Information Dissemination in a D2D network using Call Details Records

Labs

TELECOM SudParis

s@movar CNRS Institut TELECOM

Authors

Rachit Agarwal
(rach.agarwal@gmail.com)

Vincent Gauthier
(vincent.gauthier@telecom-sudparis.eu)

Monique Becker
(monique.becker@telecom-sudparis.eu)

Thouraya Toukabrigunes
(thouraya.toukabrigunes@orange.com)

Hossam Afifi
(hossam.afifi@telecom-sudparis.eu)

Partner

orange™

Abstract

In a network of devices in close proximity such as Device to Device (D2D) communication, we study the dissemination of public safety information at country scale level. In order to provide a realistic model for the information dissemination, we extract spatial distribution of the population of Ivory Coast from census data and determine migration pattern from the call detail records obtained during the Data for Development (D4D) challenge [1]. We latter apply epidemic model towards the information dissemination process. We then propose enhancements to the dissemination model by adding latent states and beamforming to the epidemic model. In this paper, we study the transient states towards the evolution of the population having the information for different cases. Through the results we show that enhancements in the dissemination process can be achieved in large and realistic scenarios.

Context: Dissemination of emergency information in metapopulation and dynamic network using epidemic model.

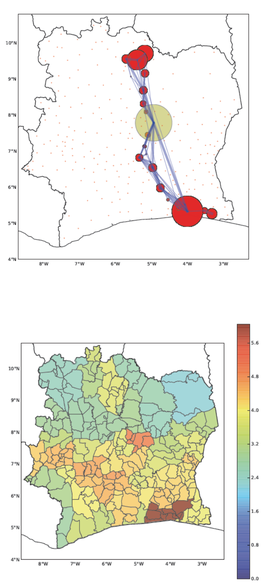
Data Analysis

- **Extract User's movement at the country level from Call Details Records provided by Orange [1].**
- **Generate transition probability matrix (ν) from all movement patterns.**
- **Determine population density from Census data.**

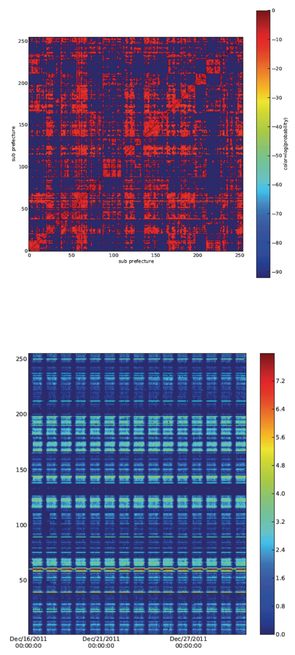
Model

- **Split the country into metapopulation [2,6] (subprefecture).**
- **Generate mobility between each meatpopulation base on our analysis of the CDR dataset of Orange.**
- **Add latent states to the initial SIR model in order to modelize a variable density of user in each metapopulation.**
- **Generate the epidemic process in order to simulate the spreading of information accross the country.**

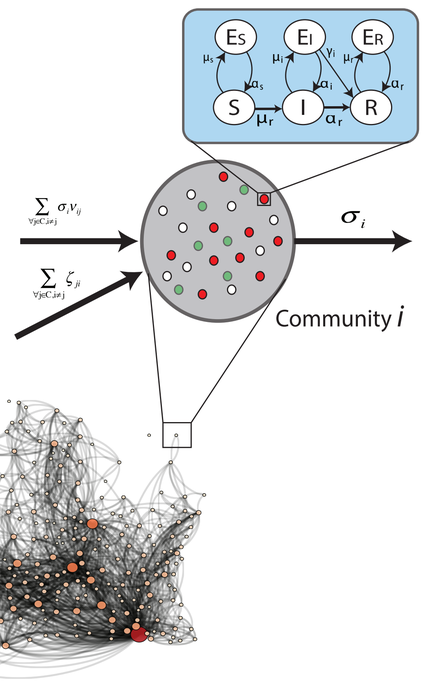
Raw data



Analysis

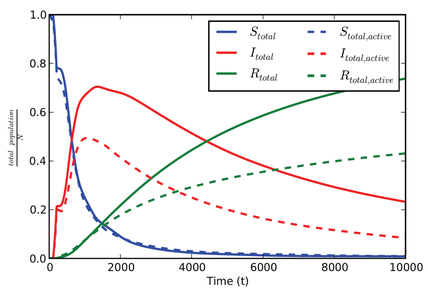


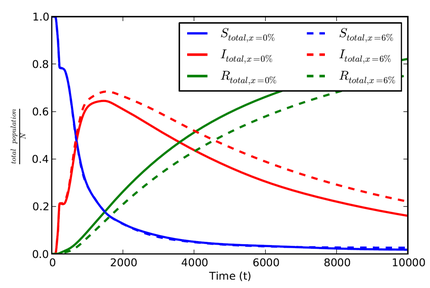
Diffusions model
(base on SIR spreading process)



Results

- **Variable people density affects the information spreading in mobile environment.**
- **Information spreading through local interaction could lead diffusion at country scale in a timely maner (Cf. Video [3]).**
- **We solve numerically a large system of differential equations to compute the spatio-temporal evolution of the diffusion.**
- **We validate the result by simulations using the Gillespie algorithm (Tau-Leap).**





Conclusion

We first display as supplementary material a movie [3] that shows the diffusion process in Ivory Coast. We can see that the diffusion that initially takes place in the East side of the country, is spreading quickly into the major cities of Ivory Coast through: Abidjan (the economic captital), Bouak (the second largest city), Youkousono, (Political Capital) Soudre. Later on, the information is spreading more slowly into less populated areas, mostly from Est to West. The West side of the country is well know to be mostly an agricultural region (Cooa, coffee, rice). We can also notice that the diffusion of the information takes a very long time to spread over the northern part of the country. As Suggested by [4] whom have been working on the same datasets, the fact that the northern part of the country is less diffusive might be the consequence of socio-economic disparity in place inside the country. Highlighting on the fact that this part of the country is still relatively "disconnected from the main economic and political center of Côte d'Ivoire".

References

[1] V. Blondel, M. Esch, C. Chan, F. Clerot, P. Deville, E. Huens, F. Morlot, Z. Smoreda, and C. Ziemlicki, **Data for Development: The D4D Challenge on Mobile Phone Data,** http://arxiv.org/pdf/1210.0137v1.pdf, 2012.

[2] D. Watts, R. Muhamad, D. Medina, and P. Dodds, "**Multiscale, Resur- gent Epidemics in a Hierarchical Metapopulation Model,**" Proceedings of the National Academy of Sciences of the United States of America, vol. 102, pp. 11157–11162, August 2005.

[3] R. Agarwal, V. Gauthier, and M. Becker, **Diffusion process using mobility data available for D4D challenge**, http://dx.doi.org/10.6084/m9.figshare.69817, 2013.

[4] C. Andris, L. Bettencourt, **Development, Information and Social Connectivity in Côte d'Ivoire**, *Netmob*, 2013.

[5] R. Agarwal, A. Banerjee, V. Gauthier, M. Becker, C. K. Yeo, and B. S. Lee, **Achieving Small-World Properties using Bio-Inspired Techniques in Wireless Networks**, The Computer Journal, vol. 55, pp. 909–931, March 2012.

[6] V. Colizza and A. Vespignani, **Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations**, Journal of theoretical biology, vol. 251, no. 3, pp. 450–467, 2008.

Le colloque **Numérique : Grande échelle et complexité**, 26 et 27 mars 2014, Institut Mines-Télécom

Contact: vincent.gauthier@telecom-sudparis.eu

http://complex.luxbulb.org/content/d4d-challenge

Understanding the Evolution of Multimedia Content in the Internet through BitTorrent glasses

Reza Farahbakhsh*, Angel Cuevas*, Ruben Cuevas, Roberto Gonzalez**, Noel Crespi***
* Institut Mines-Telecom, Telecom SudParis, France, CNRS UMR5157, RS2M department, Service Architecture Group.
** Universidad Carlos III de Madrid, Spain

Measurement Methodology & Dataset

- Large scale measurement over **The Pirate Bay (TPB)** portal
- The tool subscribes to TPB's RSS service to get a notification for any new content
- The RSS feed provides the .torrent file
- Retrieves IP of the tracker from the .torrent and connects to it immediately
  - Identify the first IP (first seeder) as publisher
  - captures the IP address of a majority of consumers.
- We use MaxMind to determine the location of Publishers and Consumers.
- In Summary: (i) publisher's username and IP address (ii) list of majority of consumers

	pb09	pb10	pb11	pb12
Crawling Period	11/28/09–12/18/09	04/09/10–05/05/10	10/21/11–12/13/11	01/28/12–02/12/12
Duration (days)	21	27	54	16
Torrents	15.8K	38.2K	72.0K	21.0K
Downloads	—	95.6M	79.0M	11.1M

Analysis & Results

Content Evolution Analysis

Content Availability Evolution

Category	pb09 (%)	pb10 (%)	pb11 (%)	pb12 (%)
AUDIO	15.958	15.208	12.535	13.884
Music	10.118	10.796	7.984	8.414
Audio Books	0.376	0.728	0.579	0.608
Sound Clips	0.162	0.076	0.095	0.120
FLAC	1.757	1.218	1.894	1.910
Other	3.546	2.390	1.984	2.833
VIDEO	39.234	41.266	52.260	46.272
Movies	23.004	20.084	20.623	19.924
Movies DVDR	–	1.625	1.448	2.029
Music Videos	1.646	2.340	1.151	1.608
Movie Clips	–	0.433	0.237	0.493
TV shows	11.913	14.216	21.996	15.435
Handhled	0.207	0.258	0.353	0.110
Highres – Movies	1.348	0.644	1.842	1.728
Highres – TV shows	–	0.603	3.690	4.039
3D	–	–	0.072	0.014
Other	1.115	1.062	0.849	0.890
APPLICATIONS	16.788	9.922	3.986	5.006
Windows	13.514	9.283	3.371	3.647
Mac	0.726	0.258	0.238	0.345
UNIX	0.071	0.089	0.136	0.235
Handheld	0.292	0.133	0.031	0.014
IOS(Ipad/Iphone)	–	–	0.051	0.302
Android	–	–	0.097	0.349
Other OS	2.184	0.159	0.061	0.115
GAMES	4.997	3.253	3.084	4.236
PC	3.636	2.599	2.642	3.039
Mac	0.039	0.037	0.043	0.072
PSx	0.181	0.063	0.088	0.254
XBOX360	0.201	0.099	0.070	0.148
Wii	0.389	0.198	0.141	0.168
Handheld	0.551	0.258	0.102	0.053
IOS(Ipad/Iphone)	–	–	0.026	0.211
Android	–	–	0.232	0.177
Other	0.402	0.279	0.092	0.115
PORN	8.264	21.553	21.140	23.007
Movies	5.950	10.767	9.097	10.386
Movies DVDR	–	0.532	0.014	0.057
Pictures	1.232	1.688	0.971	1.206
Games	0.091	0.026	0.015	0.077
Highres – Movies	0.201	0.511	1.878	2.422
Movie Clips	–	7.308	8.670	8.313
Other	0.791	0.720	0.494	0.546
OTHER	14.759	8.798	6.994	7.595
E-books	5.185	4.352	3.865	5.068
Comics	0.421	1.059	1.316	1.278
Pictures	2.930	2.173	1.227	1.163
Covers	0.058	0.016	0.021	0.005
Physibles	–	–	–	0.005
Other	6.164	1.198	0.565	0.077

Proportion of each content type (portion of available content)

- Movies/TV shows (in VIDEO) are the most available contents. (>34% of the total content)
- if we add PORN-Movies subcategory, 40%-50% for Movies and TV Shows.
- Increment of the High Resolution content (from 1% to 10%)

Content Popularity Evolution

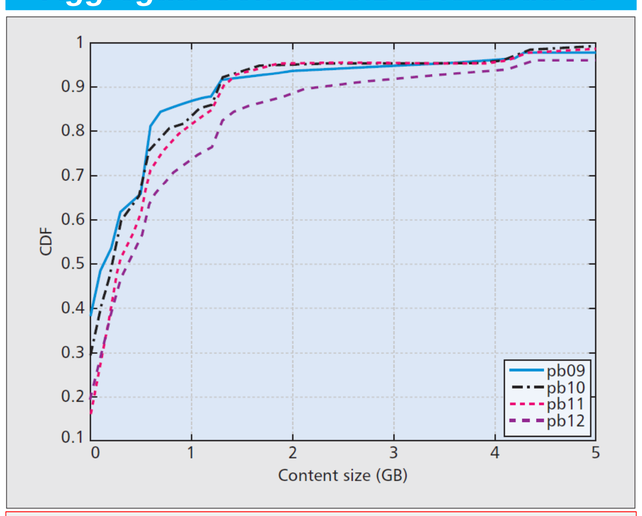
Categories	pb10 (%)	pb11 (%)	pb12 (%)
AUDIO	4.671	5.574	4.972
Music	3.814	3.977	1.036
Audio Books	0.119	0.213	0.093
Sound Clips	0.011	0.065	0.053
FLAC	0.208	0.297	0.292
Other	0.518	1.021	3.498
VIDEO	71.299	64.080	58.925
Movies	41.394	29.874	22.667
Movies DVDR	0.937	1.027	0.943
Music Videos	0.443	0.245	0.284
Movie Clips	0.066	0.037	0.097
TV shows	26.448	27.010	28.349
Handhled	0.127	0.040	0.014
Highres — Movies	0.766	3.533	3.702
Highres — TV shows	0.723	2.205	2.826
3D	–	0.025	0.000
Other	0.396	0.086	0.043
APPLICATIONS	2.117	0.996	0.810
Windows	2.041	0.934	0.725
Mac	0.050	0.041	0.027
UNIX	0.002	0.002	0.000
Handheld	0.018	0.001	0.000
IOS(Ipad/Iphone)	–	0.003	0.002
Android	–	0.012	0.054
Other OS	0.006	0.001	0.001
GAMES	1.274	2.182	1.013
PC	0.790	1.747	0.756
Mac	0.003	0.003	0.000
PSx	0.018	0.023	0.006
XBOX360	0.027	0.119	0.165
Wii	0.144	0.102	0.019
Handheld	0.216	0.022	0.001
IOS(Ipad/Iphone)	–	0.005	0.006
Android	–	0.154	0.056
Other	0.075	0.007	0.004
PORN	17.256	24.300	31.012
Movies	11.259	13.209	17.685
Movies DVDR	0.034	0.014	0.025
Pictures	0.740	0.255	0.598
Games	0.007	0.004	0.009
Highres — Movies	0.385	1.727	3.089
Movie Clips	4.559	8.827	8.388
Other	0.272	0.264	1.218
OTHER	3.383	2.868	3.268
E-books	1.337	2.099	2.604
Comics	0.326	0.225	0.115
Pictures	1.307	0.266	0.258
Covers	0.003	0.000	0.000
Physibles	–	–	0.000
Other	0.410	0.278	0.291

Distribution of content popularity (proportion of download sessions)

- VIDEO is the most popular category by attracting more than 60%
- PORN appears as the 2nd most popular category among BitTorrent users.
- 90% of the total downloads comes from VIDEO and PORN
- High-resolution PORN and VIDEO popularity increase (from 1.87% to 9.62%)

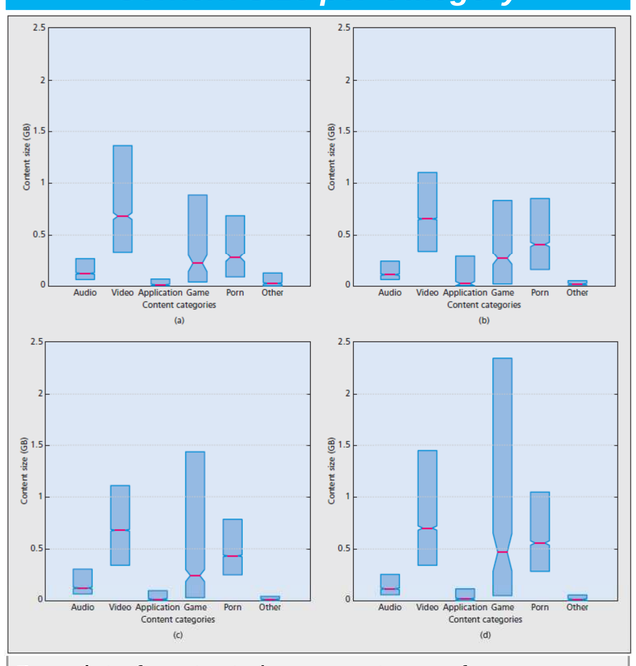
Content Size Analysis

Aggregate Content Size Distribution



- BitTorrent content has doubled its size in a period of 2 years
- in median from 223 MB to 458 MB

Content Size per Category



Box plot of content size per category for (a) pb09, (b) pb10, (c) pb11 and (d) pb12 (25^{th}, 50^{th} 75^{th} percentile)

Main Finding & Conclusion

- This work is a thorough analysis on the evolution of multimedia content available in the most popular BitTorrent portal over a two years period between Nov. 2009 and Feb. 2012.
- The major part of the Internet traffic, sustained in four main findings:
  - Multimedia content has doubled its size in a period of only 2 years.
  - The major part (80%) of the consumed multimedia content corresponds to TV Shows and Movies (including porn) that belong to those categories with the largest size.
  - High-resolution content, which has very large size, is increasing its presence by 5 times in two years and it already represents 8% of the available content and 10% of the downloads in our most recent snapshot dated at the beginning of 2012.
  - Audio represents 12%-15% of the available content but only attracts only 5% of the downloads.

Reference:

- R. Farahbakhsh, A. Cuevas, R. Cuevas, R. Gonzalez, N. Crespi, "Understanding the evolution of multimedia content in the Internet through BitTorrent glasses". IEEE Networks Magazine, November 2013.

Plateforme CREDO «FTTH: De l'infrastructure aux usages »

Auteurs

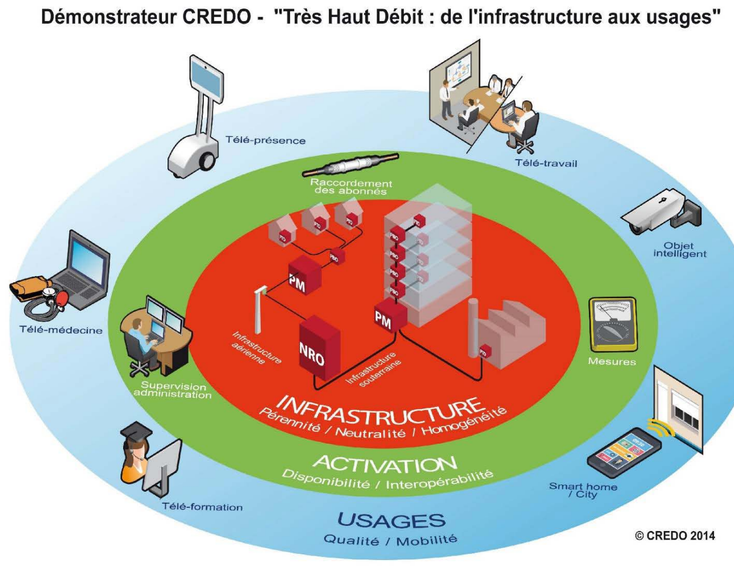
Eric Gangloff

Laurent Bernard

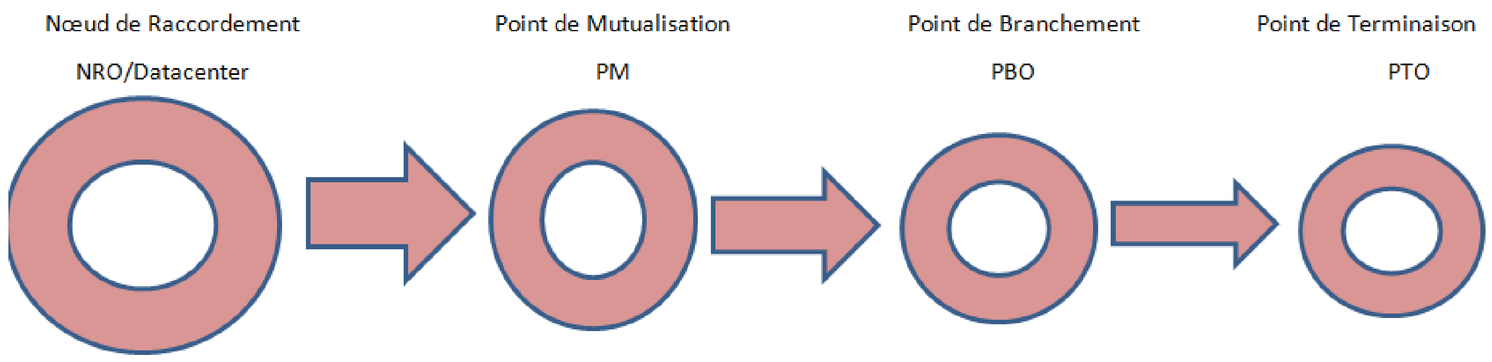
Partenaires

La Fibre: une révolution en marche

Les Infrastructures de réseaux d'accès
Les nouveaux usages



Infrastructures de réseaux d'accès

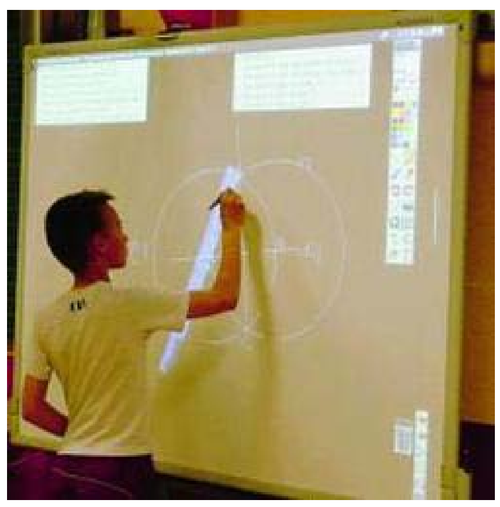


La plateforme CREDO modélise un vrai réseau optique fonctionnel en établissant des chemins physiques et logiques du point de présence des opérateurs dans les nœuds de raccordements optiques (NRO) à la prise habitation (PTO). Il permet ainsi d'appréhender et de comprendre toutes les fonctions du réseau

Les nouveaux usages

La plateforme met en valeur l'apport du FTTH pour de nouveaux usages tous consommateurs de bande passante:

- Les services à la personne
- La télé médecine
- Le télé travail et la télé formation (skype haute définition) ;
- Le divertissement et les média sociaux.



Les services associés

La plateforme est ouverte à l'accueil:

- de présentations: collectivités territoriales, ….
- de projets étudiants
- d' expérimentations de nouveaux usages avec des partenaires académiques et industriels

Contact eric.gangloff@ telecom-sudparis.eu Site web

Projet VELCRI

Véhicule électrique **Communication Véhicule Infrastructure (V2G)**

Télécom SudParis

Hossam Afifi
hossam.afifi@it-sudparis.eu

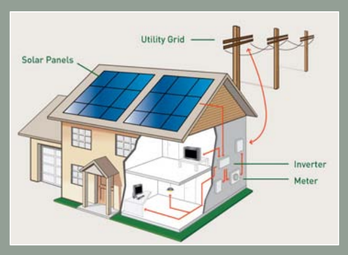
Michel Marot
Michel.marot@it-sudparis.eu

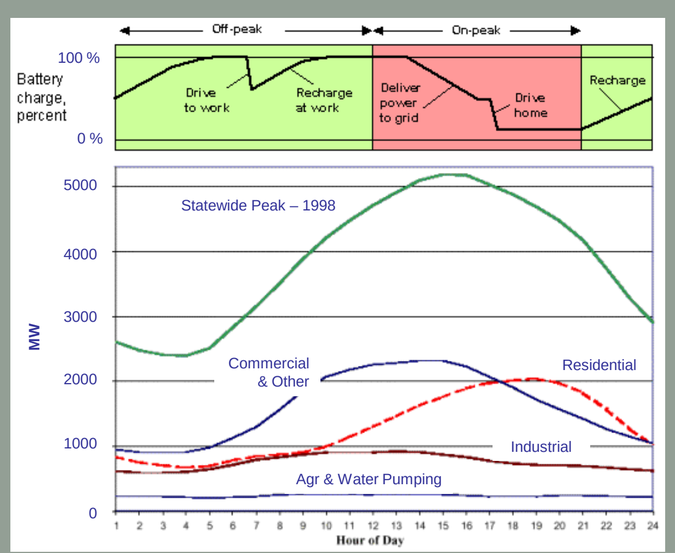
Télécom ParisTech

Houda Labiod

Un nouvel écosystème

Fournis de l'énergie &
Échange de l'information

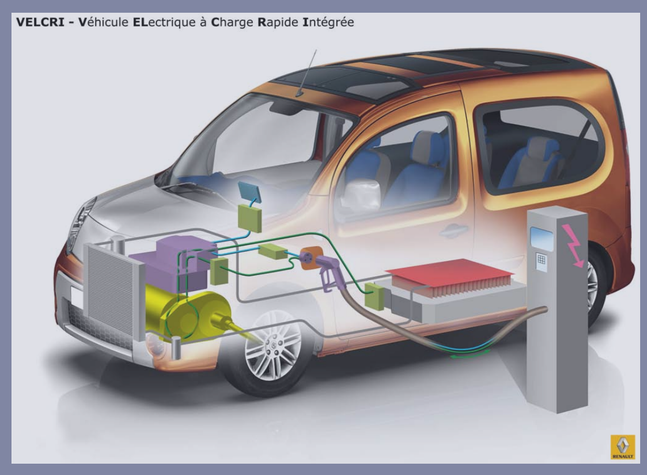




Fournisseur d'énergie (distribué) &
Echange de l'informations

Besoin de Communications

- Mesurer, échanger
  - La consommation des véhicules
  - L'état de charge de la batterie
- Optimiser la recharge des véhicules en fonction :
  - De la localisation
  - De la demande en énergétiques
  - De la disponibilité de la production d'énergie
  - Du type de batterie
- Fournir la tarification adapté à tout type de recharge/échange de batteries



Communications de machine à machine

Un environnement multi-utilisateur, multi opérateurs, et distribué

- Développement d'une interface sécurisée de communication entre le véhicule et l'infrastructure de distribution d'électricité (sans fil, filaire)
  - Communication Multimodale
  - Inspiration du système de roaming du réseau cellulaire pour l'identification, l'authentification et la facturation (sur facture domicile)
  - Pervasivité des moyens de communications dans le véhicule
  - Authentification des véhicules, et le paiement sécurisé des recharges
- Communication du niveaux de la batterie aux différents éléments du réseau
  - Stations de recharge
  - Infrastructure de production d'énergie
- La batterie devient un élément de stockage d'énergie faisant partie du réseau de distribution d'énergie

Janvier 2012

eCOUSIN

enhanced COntent distribUtion with Social INformation

eCOUSIN designs a novel social-aware network architecture that exploits social-content interdependencies with built-in content dissemination functionalities to improve its efficiency.

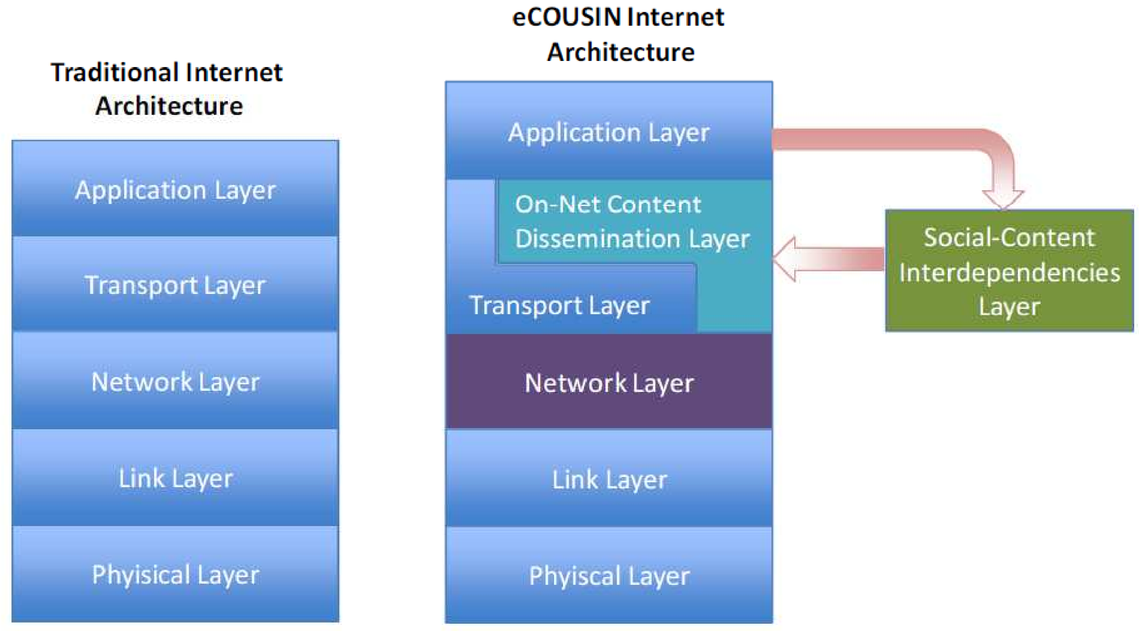
Context : Social-Content Revolution

- Online Social Networks (OSNs) have drastically changed the way contents are consumed on the Internet:

Users consume contents based on the information shared through OSNs.

The popularity of a content is highly impacted and often dictated by its "social" success.

- Operators need to evolve and optimize their network to avoid being overwhelmed by the ever growing traffic volumes resulting from this paradigm change



eCOUSIN Objectives

Design a novel social-aware network architecture that exploits the social-content interdependencies with built-in content dissemination functionalities to improve its efficiency

- Implement high performance distributed tools for collecting necessary data to study and model the social-content interdependencies
- Improve the scalability of network infrastructures when handling contents by exploiting social information
- Design an on-net operational framework that tightly integrates network functionalities and content-related service functionalities
- Design of algorithms that exploit social information for placing and delivering contents in an optimized manner with a special focus on mobile environments

Key Challenges

- Model social-content interdependencies based on gathering information of users' real-time interactions, and on the interdependencies between user interaction in OSNs and the resulting behaviour over content distribution services
- Extend content replication, placement, search and retrieval techniques with additional information extracted from OSNs
- Investigate proper naming schemes for OSN traffic delivered onto Information Centric Network (ICN); how OSNs can adapt them to the ICN paradigm, and how ICN routing can benefit from the OSNs' social links to improve its routing and forwarding strategy
- Develop and evolved a management system for content placement and delivery to mobile users by exploiting statistical patterns derived from mobility-, connectivity- and social information.

Expected Impacts

- Offer to European citizens a vastly improved content delivery experience
- By placing the right content closer to the user, media streams can be delivered at higher transfer rates and with lower delay, without increasing the burden on the network infrastructure

SEVENTH FRAMEWORK PROGRAMME

eCousin
http://www.ict-ecousin.eu

PROJECT DATA

-Start Date: 11/2012
-Duration: 30M
-EU Funding: 2,998M€

CONSORTIUM

ORANGE, France;
TELECOM ITALIA, Italy;
TELECOM SUD-PARIS, France;
IMDEA NETWORKS, Spain;
ALCATEL LUCENT, Belgium/Germany;
TECHNISCHE UNIVERSITAT Darmstad, Germany;
UNIVERSITY OF CAMBRIDGE, United Kingdom;
UNIVERSIDAD CARLOS III DE MADRID, Spain

Contact:
Yannick Le Louedec, ORANGE, France
Email: yannick.lelouedec@orange.com
Noel Crespi, IMT-TSP, France
noel.crespi@mines-telecom.fr

Services Réseaux Programmables SDN for cloud networking: optimal provisioning and instantiation

Authors

Marouen Mechtri and Djamal Zeghlache

In collaboration with Hadji Makhlouf (now with IRT SystemX)

Optimal selection & provisioning

Objectives

- Optimal selection of virtual services (compute, storage and communications) according to users' and tenants' requests and requirements (QoS and SLA)
- Address the entire workflow from requests to instantiation and adaptation and rely on SDN services to achieve instantiation.

Contributions

- Exact model and heuristic algorithm to scale to thousands of nodes and links
- Convergence of clouds and networks
- Rely on sharing, virtualisation and SDN principles

Mathematical Models

Exact Virtual Network Mapping

- Joint node and link selection
- Mathematical Programming Formulation

$$\min Z = \sum_{i \in V_P} \sum_{k \in V_T \setminus R} d(i,k) x_{ik} + \quad \text{Placement}$$

$$\sum_{(ij) \in E_P} \sum_{\substack{k_1 \in V_T \setminus R \\ k_1 \neq k_n}} \sum_{k_n \in V_T \setminus R} d_1(ij, P_{k_1,k_n}) y_{ij,k_1,k_n} + \quad \text{Inter domain Path}$$

$$\sum_{(ij) \in E_P} \sum_{k_1 \in V_T \setminus R} d_2(ij,k_1) y_{ij,k_1,k_1} \quad \text{Intra domain Path}$$

$$d_1(ij, P_{k_1,k_n}) = \begin{cases} 1, & \text{if } CPU(i,k_1) \text{ and } CPU(j,k_n) \,\& \\ & STO(i,k_1) \text{ and } STO(j,k_n) \& \\ & MEM(i,k_1) \text{ and } MEM(j,k_n) \& \\ & lat_{ij} \geq lat_{k_1,k_n}; \\ 0, & \text{otherwise.} \end{cases} \quad \text{and} \quad d_2(ij,k_1) = \begin{cases} 1, & \text{if } cpu_i + cpu_j \leq CPU_{k_1}; \\ 0, & \text{otherwise.} \end{cases}$$

V_T is the set of vertices and E_P the set of edges of the physical or reference or target graph, d(i, k), x_{ij} and y_{ij,k_n,k_m} are Boolean variables indicating if a virtual resource is mapped on a physical one (nodes & links)

Additional Constraints

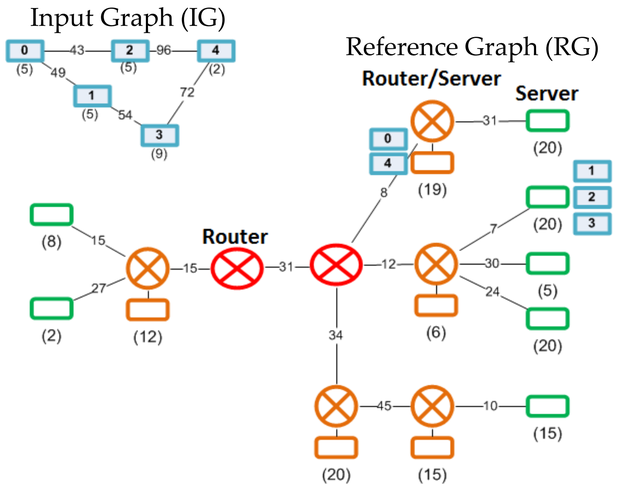
- Node mapping $\sum_{k \in V_T \setminus R} x_{ik} = 1, \forall i \in V_P$
- Limited storage $\sum_{i \in V_P} sto_i x_{ik} \leq STO_k, \forall k \in V_T \setminus R$
- Node & Link mapping

$$\sum_{k_1 \in V_T \setminus R} \sum_{k_n \in V_T \setminus R} y_{ij,k_1,k_n} = 1, \forall (ij) \in E_P$$

$$\sum_{k_n \in V_T \setminus R} y_{ij,k_1,k_n} = x_{ik_1}, \forall (ij) \in E_P, \forall k_1 \in V_T \setminus R$$

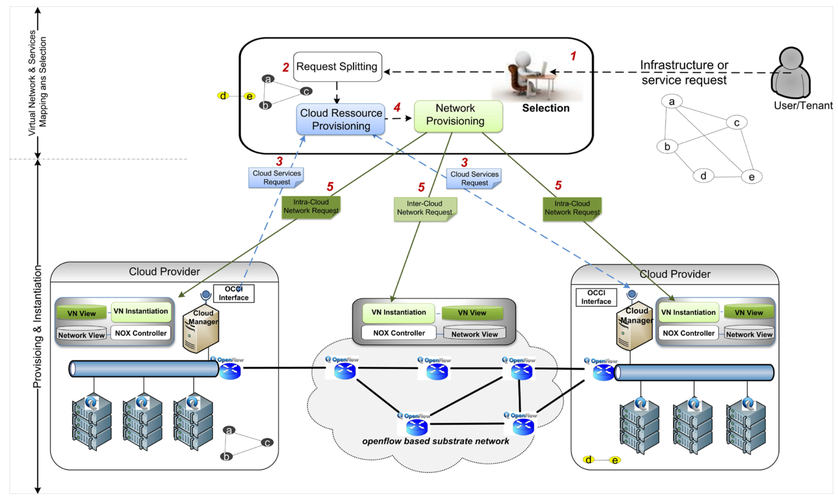
- Latency $lat_{k_1,k_n} y_{ij,k_1,k_n} \leq lat_{i,j}, \forall (ij) \in E_p, \forall k_1, k_n \in V_T \setminus R, k_1 \neq k_n$
- Localisation $x_{ik} + x_{jk} \leq 1, \forall i,j \in Sep, \forall k \in V_T \setminus R$

$$or \sum_{k \in V_T \setminus R} z_{ij}^k = 1, \forall i,j \in J \qquad x_{ik} + x_{jk} = 2 z_{ij}^k, \; \forall i,j \in J, \forall k \in V_T \setminus R$$

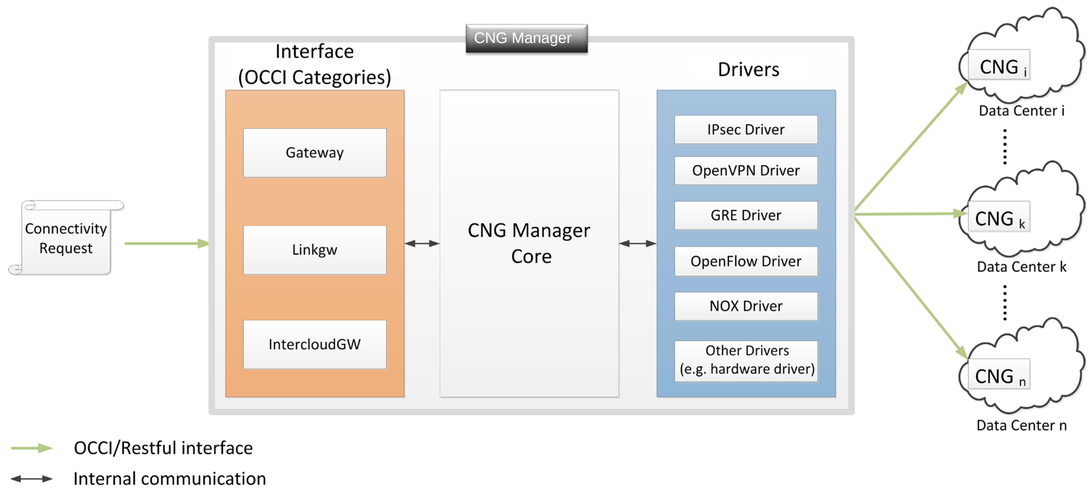


Example of resources localization constraints :
- virtual resources 0 and 4 in same node
- virtual resources 1, 2 & 3 in another node

Provisioning and instantiation process

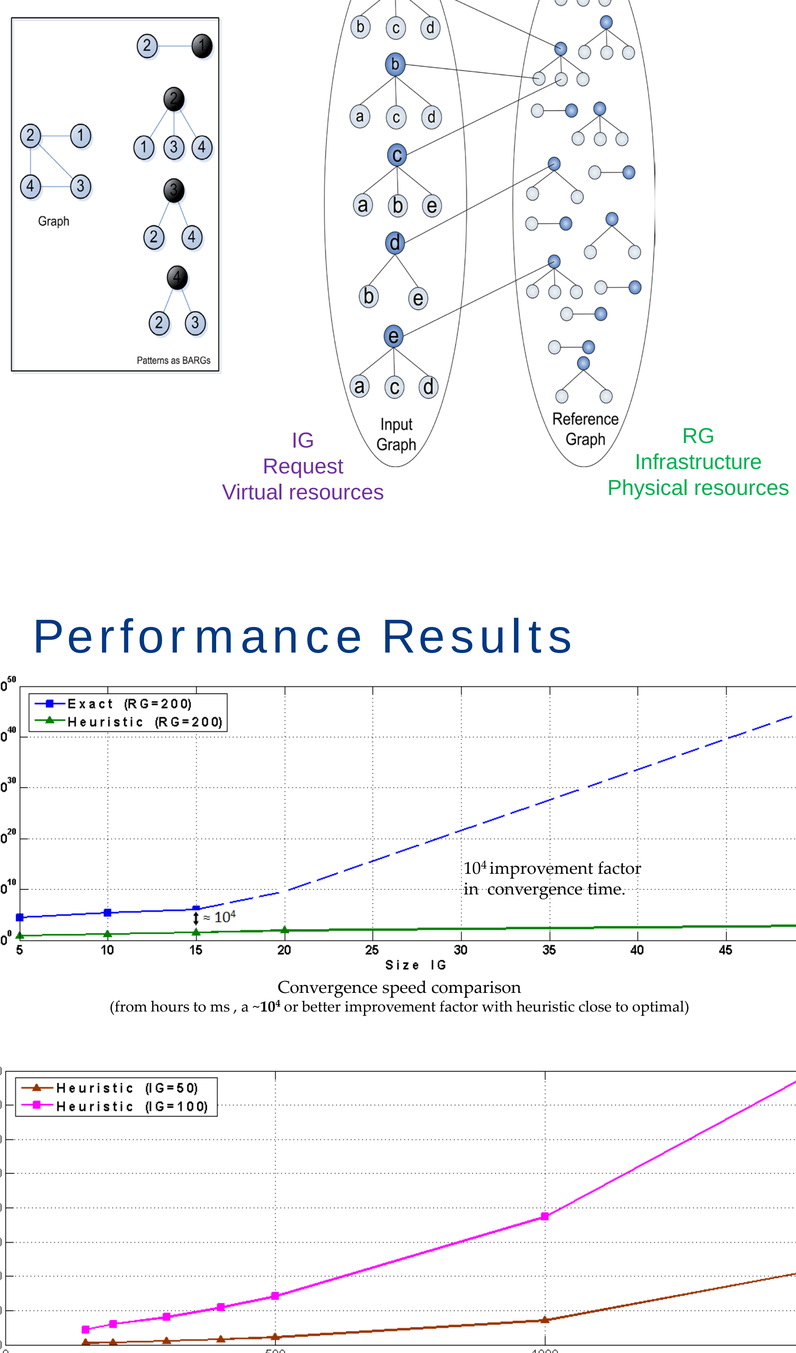


SDN compliant Instantiation/Networking System



Heuristic approach

- Based on graph patterns and bipartite request and reference graph mappings



Performance Results

Convergence speed comparison
(from hours to ms , a ~10^4 or better improvement factor with heuristic close to optimal)

Scalability of Heuristic Algorithm as a function of input and reference graph sizes

Distributed Architecture of Multiscale Context Management for the IoT

Objective: Design of a distributed context management framework for IoT context-aware applications

Parties prenantes

TELECOM SudParis

Auteurs

Chantal Taconet (responsable pour TSP)

Amel Bouzeghoub

Sophie Chabridon

Denis Conan

Léon Lim

Samer Machara Marquez

Sam Rottenberg

Partenaires

IRIT CNRS - INPT - UPS - UT1 - UTM

artal technologies

Financement

Projet INCOME

INfrastructure de gestion de COntexte Multi-Échelle pour l'Internet des Objets

ANR-11-INFR-009, 2012-2015

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR

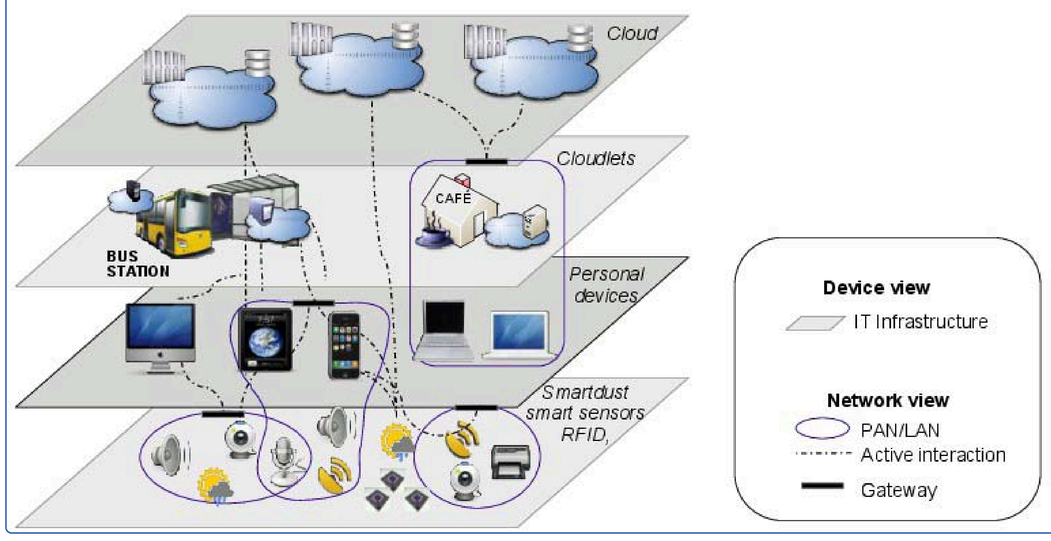
Context and problematic

IoT context-aware applications

- Smart cities, intelligent transport, leisure and entertainment, etc.

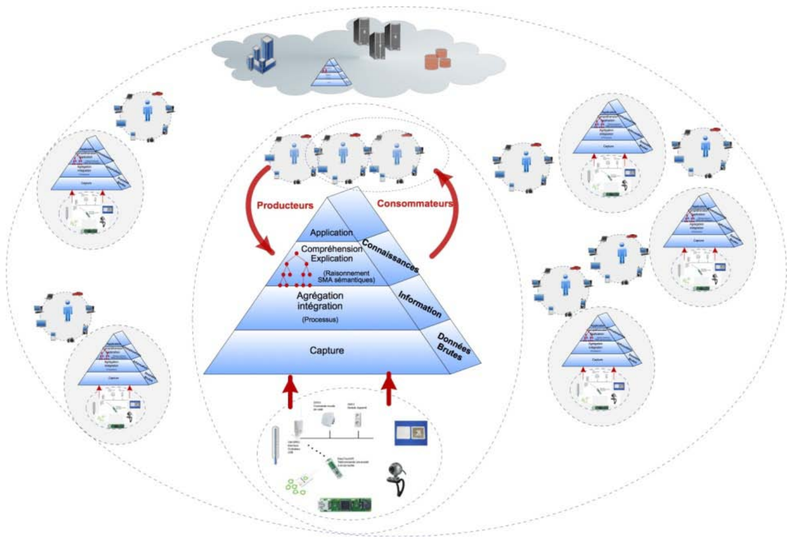
Infrastructure for the IoT

- Complex systems distributed over several levels of ICT
  - Smart objects, personal computers, proximity servers, cloud servers.



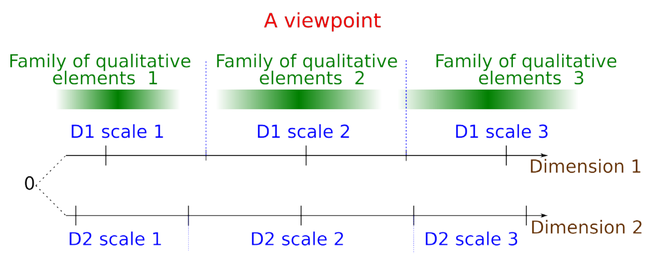
Context management for IoT applications

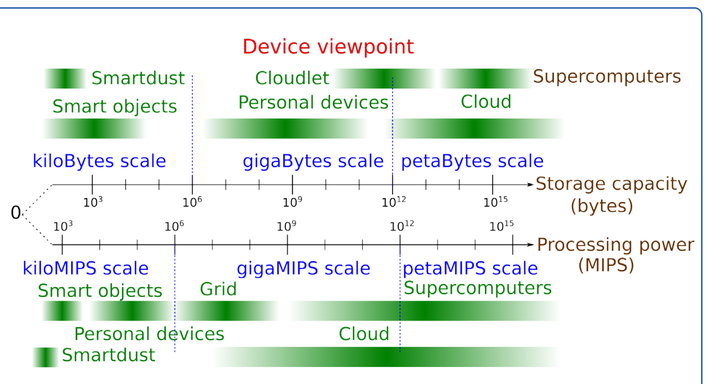
- Context management: data delivery, processing, and presentation
- Context management for the IoT
  - Context data perceived from ambient space but also from other spaces
  - Distribution of context data processing
  - Quality of context and privacy protection concerns

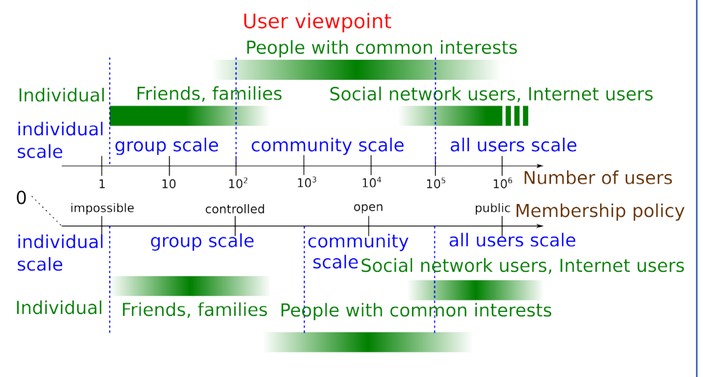


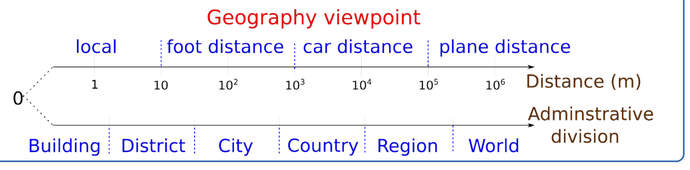
Approach: multiscale distributed systems

- Multiscalability ≠ scalability
  - Dealing with heterogeneity
- **Multiscale system characterization**
  - Model driven process to define
    - Viewpoints / Dimensions / Scales
- Constraints for context data delivery in terms of a multiscale characterization
  - Examples
    - Geography / Distance / Foot distance
      - Limit the car park information delivery to those at foot distance
    - Geography / Administrative division / City
    - \+ User / Membership policy / Group scale
      - Limit the delivery of GPS position to friends in the same city









Multiscale context management infrastructure

Architecture

- Distributed event-based system with push and pull modes
- Construction of spaces according to viewpoints, dimensions, and scales
- Intra-space brokering service using content-based filtering

  ⇒ Powerful and expressive filtering of context data
- Inter-space brokering service using topic-based filtering

  ⇒ Scalable filtering of context data

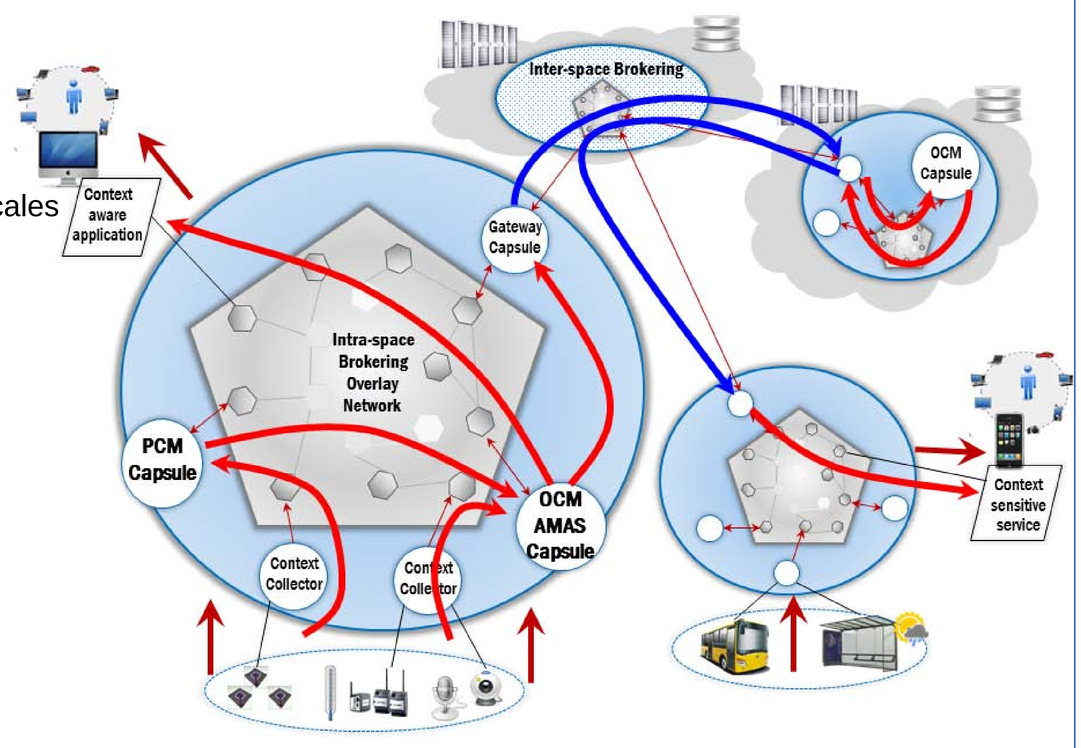
Functionalities

- Context data delivery, processing, and presentation

Extra-functionalities

- Quality of context and privacy protection

  ⇒ Rule-based filtering for controlling the distribution of context data



Monitoring and Reconfiguration of Cloud Resources

Authors

Mohamed Mohamed
Djamel Belaïd
Samir Tata

1. Context

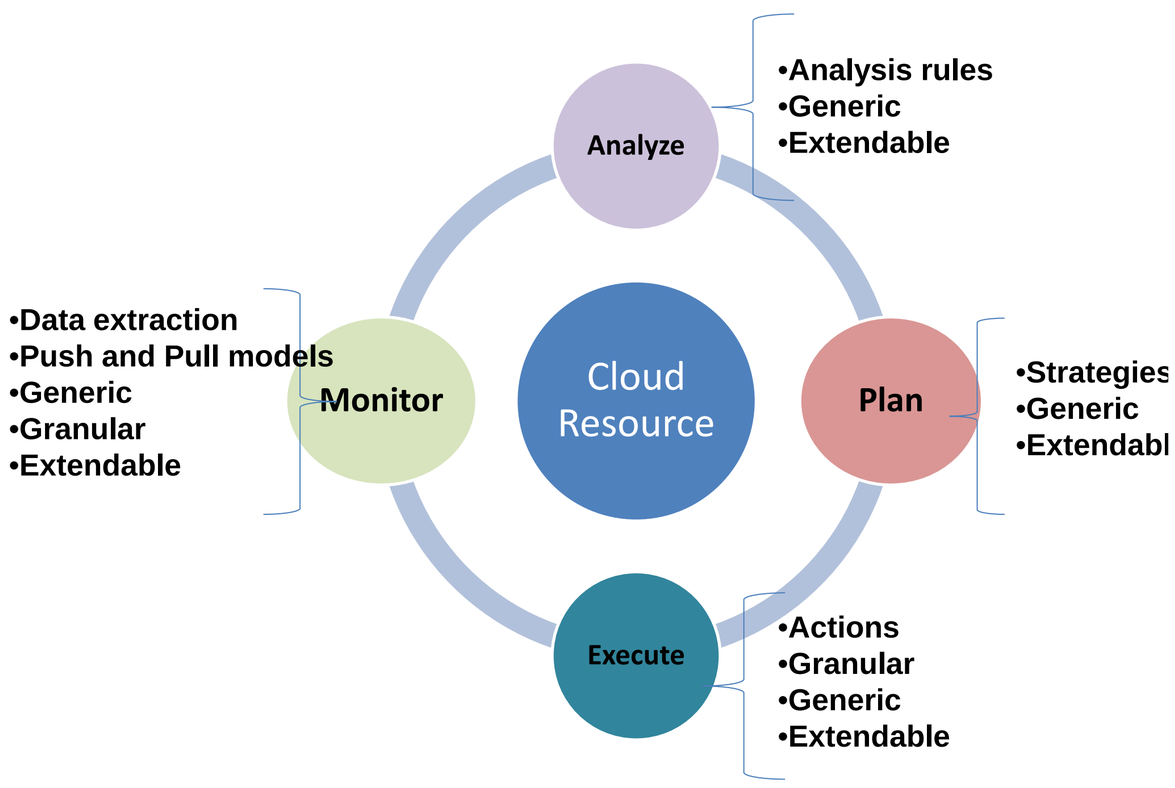
Cloud Computing environments:

- Massively scalable
- Dynamically configured
- Delivered on demand
- Heterogeneous resources

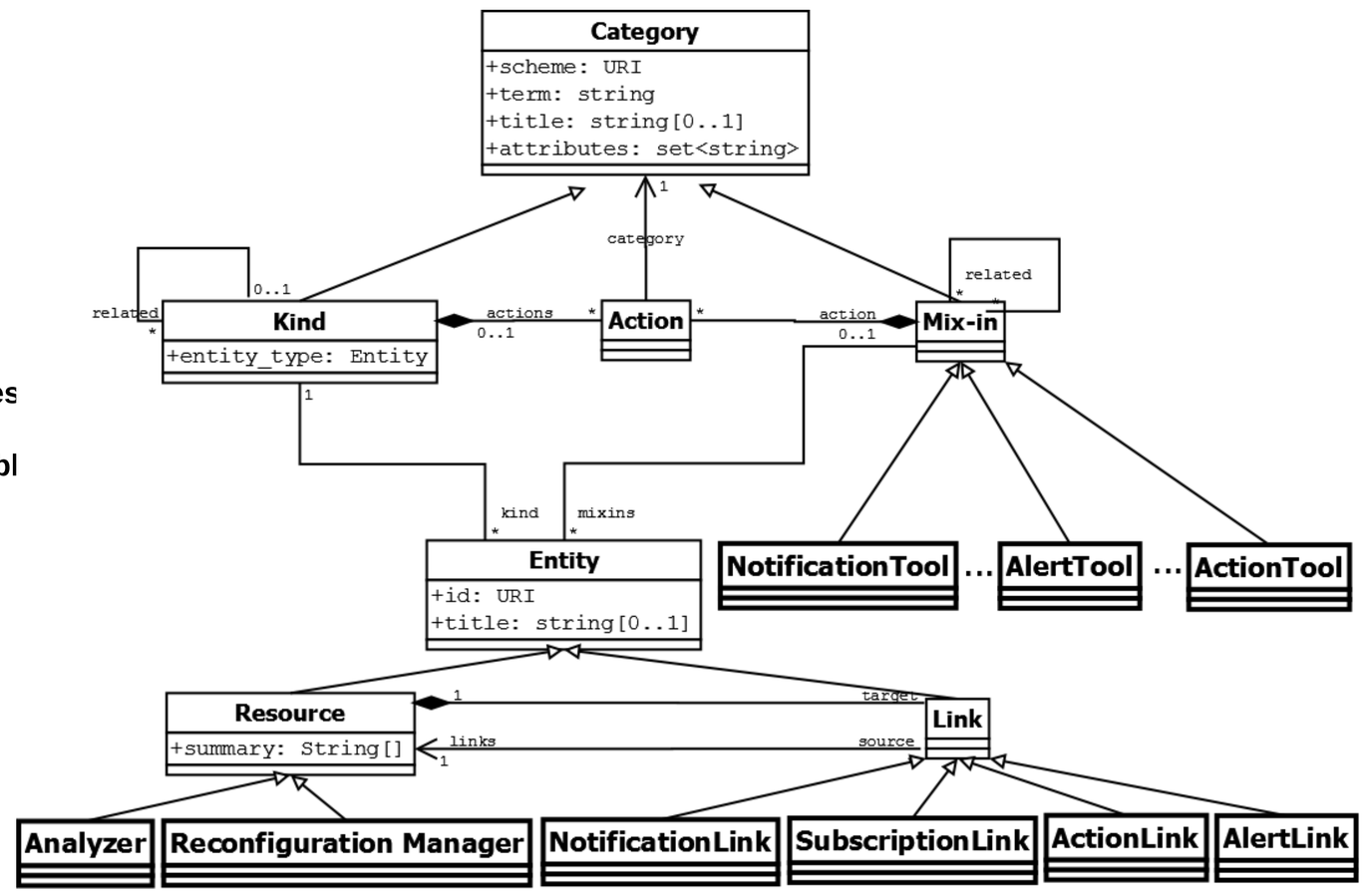
2. Objectives

- Define a model for a standard description of Monitoring and Reconfiguration requirements
- Generic Monitoring and Reconfiguration solution independent of the Cloud Service layer
- Extensible and granular solution

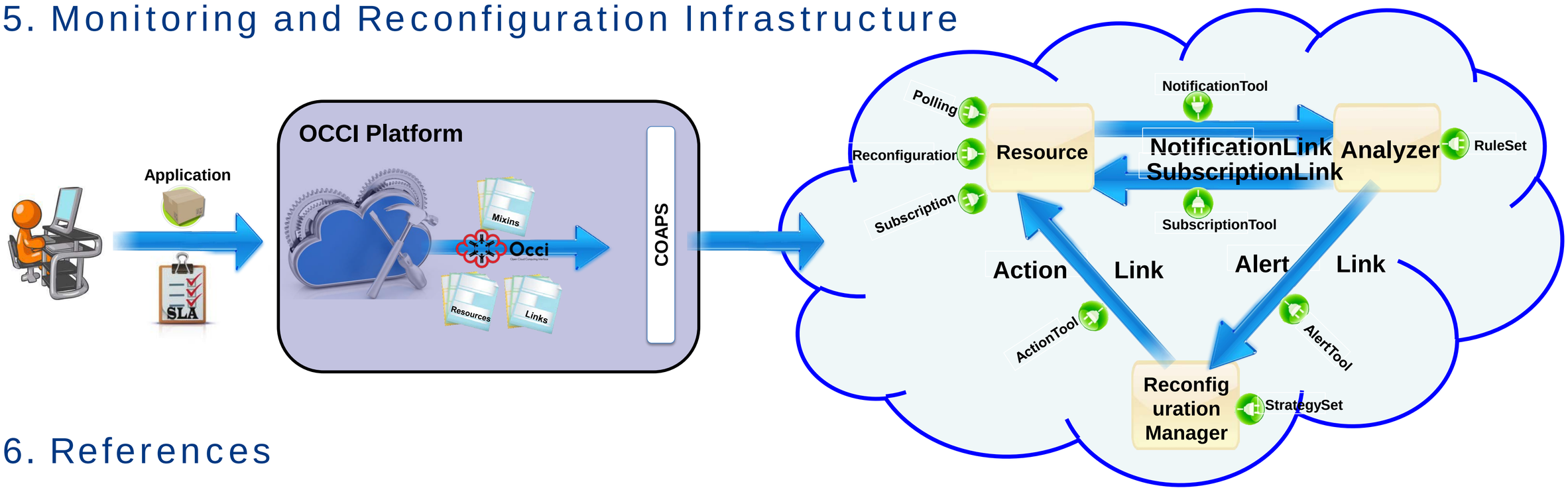
3. Monitoring and Reconfiguration requirements



4. OCCI defined Entities and Mixins



5. Monitoring and Reconfiguration Infrastructure



6. References

- M. Mohamed, D. Belaïd, S. Tata, "Monitoring and Reconfiguration for OCCI Resources", in *IEEE International Conference on Cloud Computing Technology and Science, CloudCom'2013* Bristol, UK, 2-5 December 2013.
- M. Mohamed, D. Belaïd and S. Tata, "Adding Monitoring and Reconfiguration Facilities for Service-based Applications in the Cloud", in IEEE International Conference on Advanced Information Networking and Applications, AINA'2013, Barcelona, Spain, March 25-28, 2013.
- R. Nyren, A. Edmonds, A. Papaspyrou, and T. Metsch, “Open Cloud Computing Interface - Core,” Tech. Rep., 2011.
- COAPS: a Generic Cloud Application Provisioning and Management API, http://www-inf.telecom-sudparis.eu/SIMBAD/tools/COAPS/
- OCCI4Java Platform and Application, http://www-inf.telecom-sudparis.eu/SIMBAD/tools/OCCI/

Contact: { Mohamed.Mohamed, Djamel.Belaid, Samir.Tata } @telecom-sudparis.eu

2. SYSTEMES INDUSTRIELS COMPLEXES

Interoperability requirements for collaborative process analysis

INSTITUT Mines-Télécom

Centre de Recherche LGI2P

Auteurs

Nicolas Daclin
Sihem Mallek
Vincent Chapurlat

ISOE
Interoperable Systems and Organization Engineering

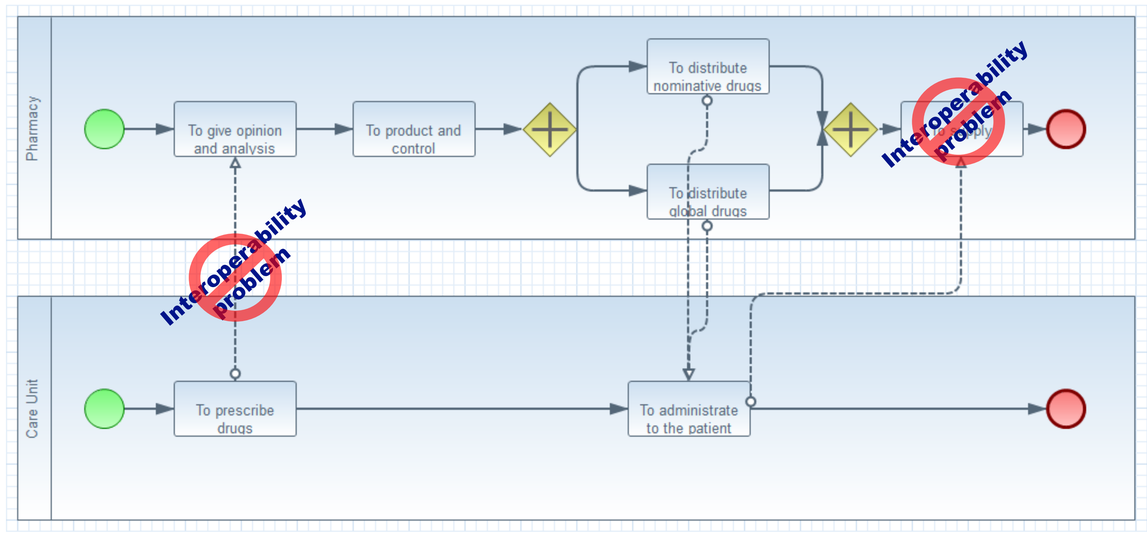
Partenaires

Université Bordeaux 1

Overall context

- To **verify interoperability** requirements within **collaborative processes** models using **formal verfication techniques** to detect and anticipate interoperability problems.

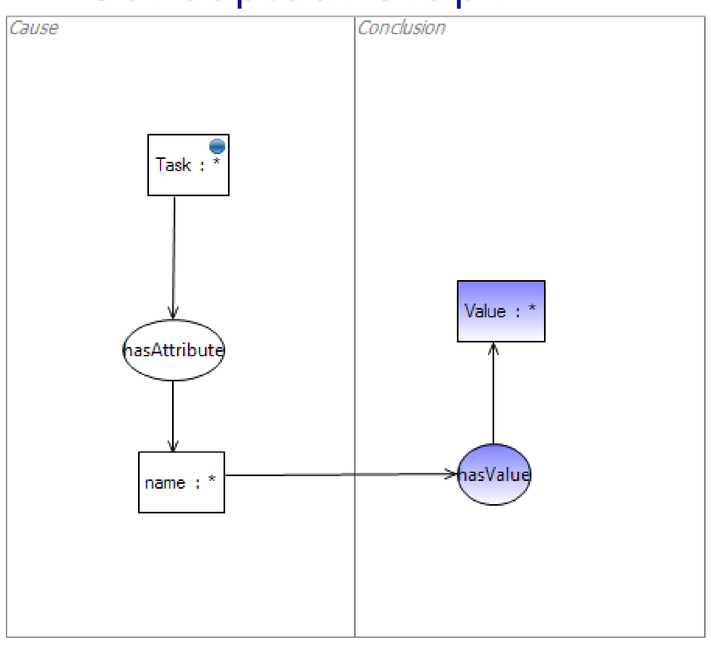
Collaborative process model (BPMN 2.0)



Problematic

- How to guide and facilitate end-users to **select** their own interoperability requirements?
- How to allow end-users to **write** their own interoperability requirements?
- How to ensure that interoperability requirements are **well written**?

Conceptual Graph

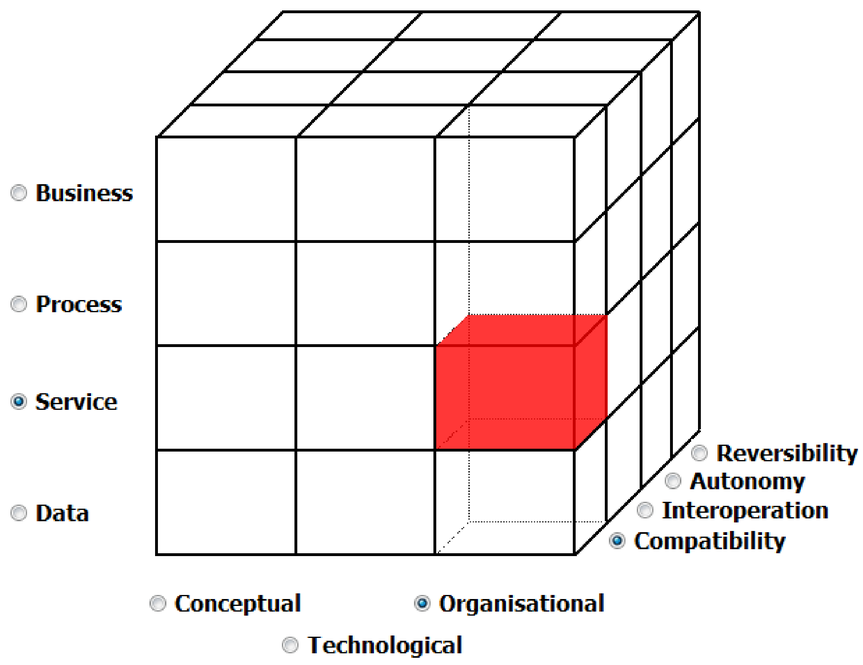


TCTL

E<> Task.Working and T>5 and T<10

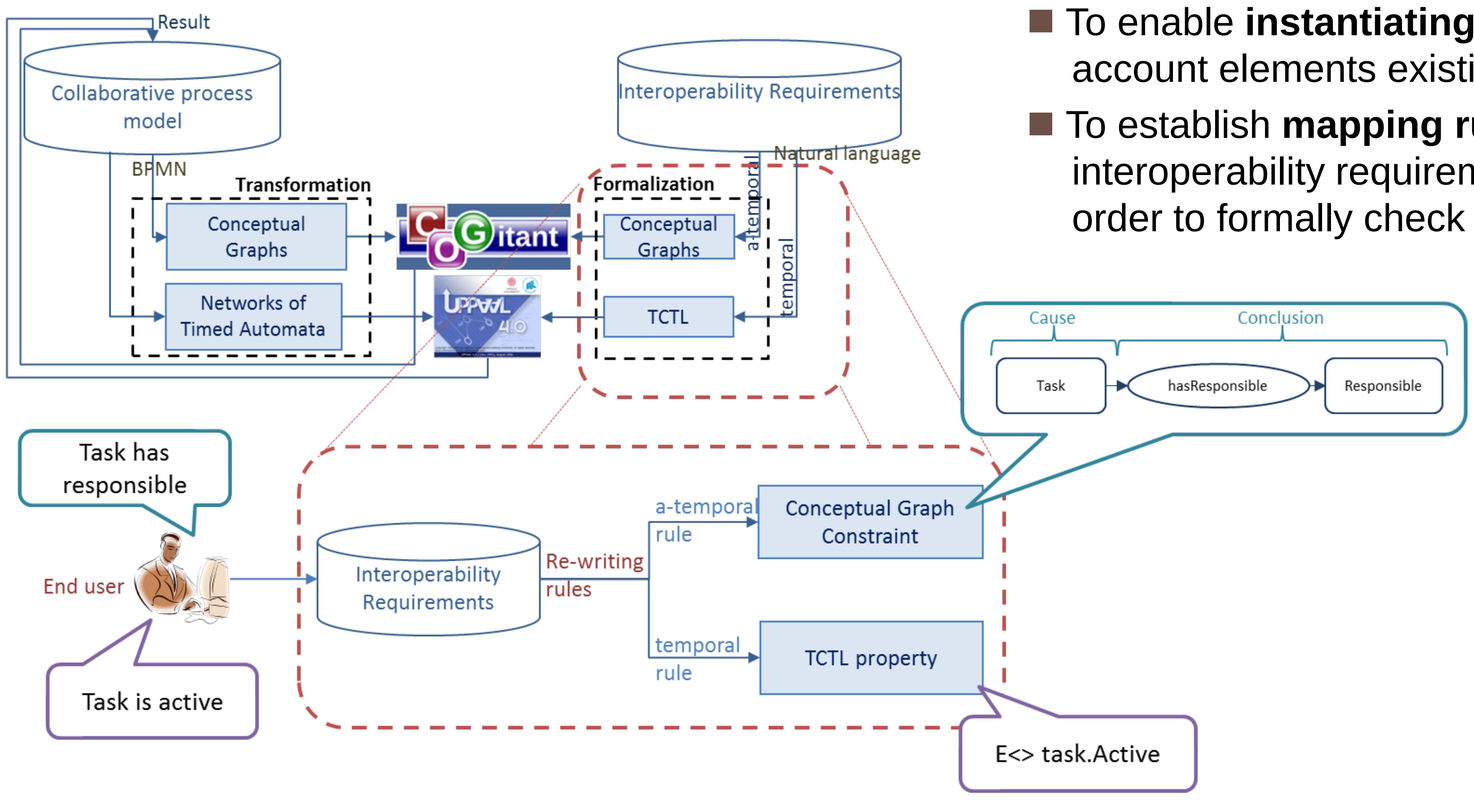
A<> Resource.Active and Resource.T<10

Interoperability requirements framework



Interoperability requirements writing

- To allow the selection of predefined interoperability requirements consistently positioned in a **human readable repository**.
- To enable **instantiating selected requirements** taking into account elements existing in the collaborative process model.
- To establish **mapping rules** to re-write correctly these interoperability requirements into TCTL and Conceptual Graphs in order to formally check them.



Future works

- To enable end-users to write directly their own interoperability requirements with a dedicated Domain Specific Language.
- To propose interoperability solutions relative to the identification of not checked requirements.

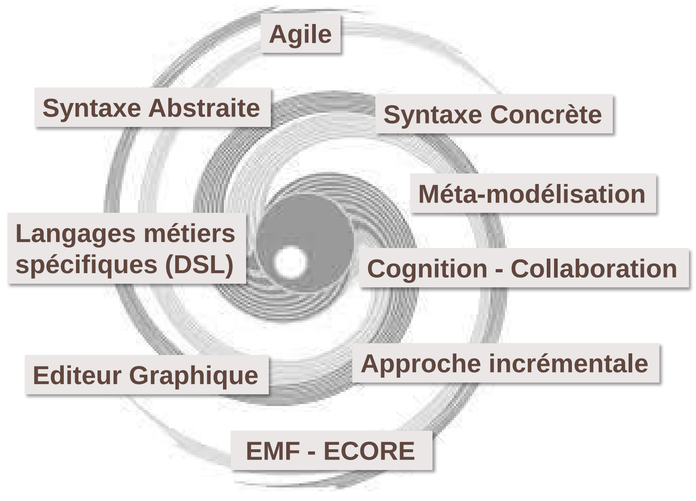
Contact nicolas.daclin@mines-ales.fr
Vincent.chapurlat@mines-ales.fr

Site web http://lgi2p.mines-ales.fr

Colloque IMT, Numérique: grande échelle et complexité

Mars 2014

Ingénierie des langages de modélisation: syntaxes concrètes graphiques avec DIAGRAPH



Problématique

Comment définir aisément des notations graphiques en IDM (Ingénierie Dirigée par les Modèles) ?

Modéliser des systèmes sociotechniques complexes nécessite des langages :

- UML, DSML, ontologies, simulation multi-physique, etc..

Pour exprimer des points de vue :

- statique, dynamique, fonctionnel, architectural, exigences, etc…

Nous nous intéressons aux DSML (Domain Specific Modeling Language) graphiques en proposant dans ce contexte une méthodologie outillée : DIAGRAPH, consistant en :

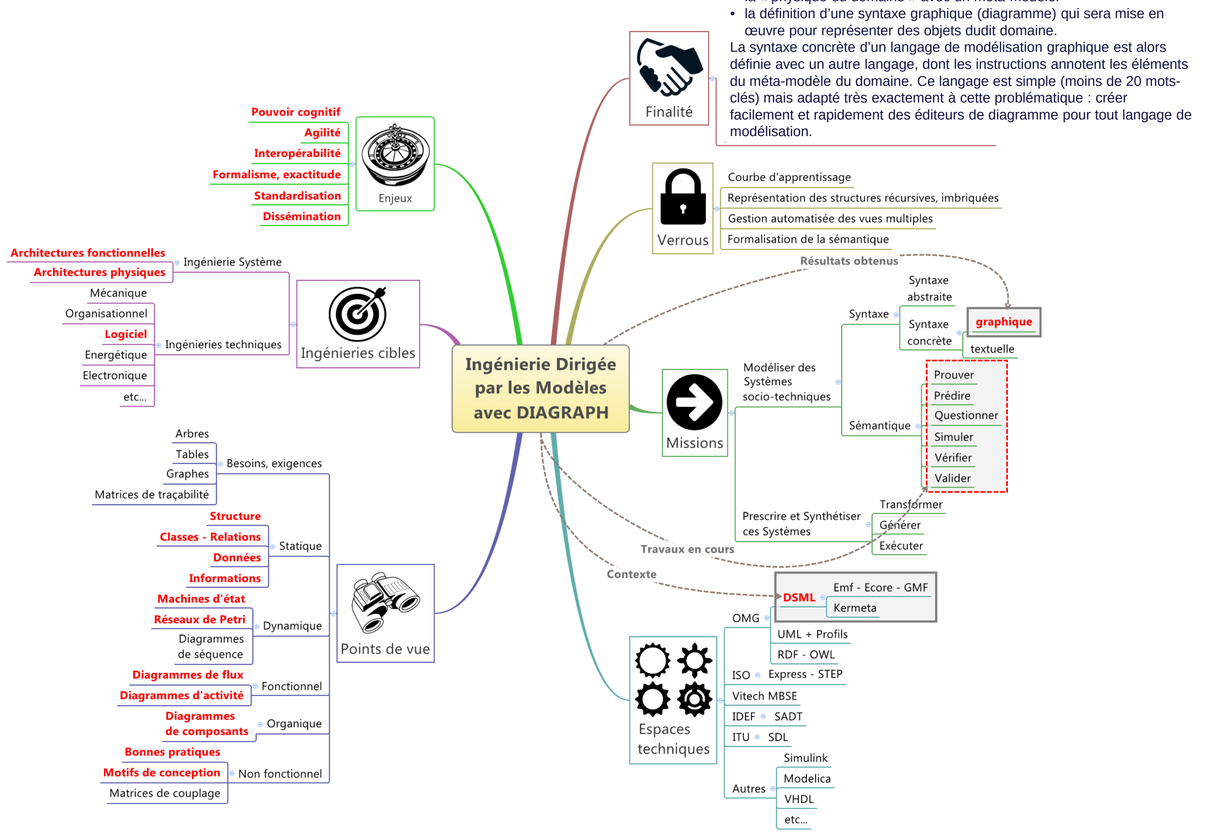
- Un processus de définition du langage, faisant actuellement défaut. (Les pratiques sont centrées sur la conception de méta-modèles, mais la création d'éditeurs graphiques est un point difficile).
- Un outillage compatible avec les environnements déjà existants dans l'écosystème Eclipse/EMF.

Contexte

Diagraph est un environnement de méta-modélisation permettant de créer des langages de modélisation graphiques adaptés à des domaines spécifiques (DSML, Domain Specific Modeling Languages).
La particularité de cet environnement est de traiter en ***un artefact unique***

- la « physique du domaine » avec un méta-modèle.
- la définition d'une syntaxe graphique (diagramme) qui sera mise en œuvre pour représenter des objets dudit domaine.

La syntaxe concrète d'un langage de modélisation graphique est alors définie avec un autre langage, dont les instructions annotent les éléments du méta-modèle du domaine. Ce langage est simple (moins de 20 mots-clés) mais adapté très exactement à cette problématique : créer facilement et rapidement des éditeurs de diagramme pour tout langage de modélisation.



Contributions

Id	Aspect	Description
1	open-source	Publié avec une licence libre et gratuite.
2	intégration des standards	Intégré dans l'écosystème Emf/Ecore.
3	interaction utilisateur	Editeur de diagramme de classes étendu permettant la saisie des rôles graphiques pour les entités du méta-modèle.
4	indépendant de la plateforme	Les diagrammes cibles sont abstraits et indépendants d'une plateforme spécifique.
5	abstraction	La structure d'un diagramme est définie par un méta-modèle publié.
6	artefact grammatical	La syntaxe concrète graphique et la syntaxe abstraite sont définies dans un artefact unique (une grammaire)
7	composition	Les langages peuvent être composés grâce à un mécanisme de fusion.
8	navigation entre les vues	Un mécanisme de vues multiples, incluant la navigation entre les vues, fait partie intégrante de la syntaxe concrète.
9	support des hiérarchies	Un support natif des hiérarchies est fourni.
10	grammaire positionnelle	La partie positionnelle de la grammaire est indépendante de la partie stylistique.
11	sémantique statique basée sur des règles	La sémantique statique est supportée par un langage de type OCL.
12	héritage graphique	L'héritage graphique est implémenté (propagation automatique des compositions graphiques).
13	inférence de constructions	Des constructions basées sur des motifs sont inférées

Un langage pour décrire des diagrammes

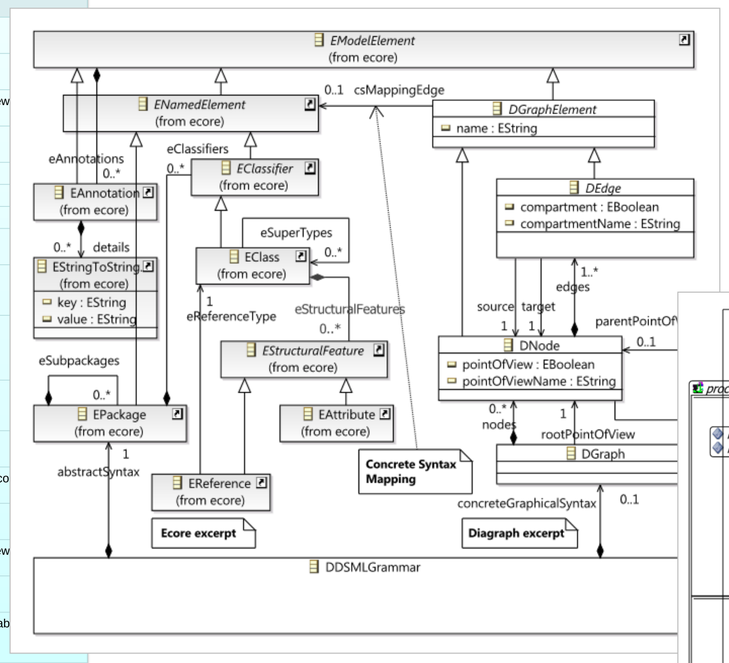
Les mots-clés

(sa syntaxe concrète)

Mot-clé	Sémantique	Correspondance méta-modèle
node	Une figure graphique (polygone, ellipse?) qui joue le rôle de la classe associée.	DNode
link	Une ligne vers la classe cible (notée sur la classe associante).	DEdge
pov	Un diagramme (et non une figure) joue le rôle de la classe.	DNode.pointOfView
lnk	Identique à link, mais notée sur la classe source.	DEdge
lsrc	Définit la source d'un link.	DEdge.source
ltrg	Définit la cible d'un link.	DEdge.target
cont	Définit le conteneur d'un link.	DNode.edges
cref	Un compartiment graphique contenant le nœud cible.	DEdge.source
kref	Identique à cref, mais plusieurs compartiments sont créés, à raison d'un compartiment par type d'association	DEdge.source
ref	Une ligne vers la classe cible. Référence associante (par conséquent, pas d'étiquette possible)	DEdge
afx	Définit une ligne affixée à travers un compartiment adjacent externe au nœud source (paradigme du port).	DEdge.source
shape	Définit une figure pour le nœud (node) ou une forme de flèche (pour link ou ref).	DShape
icon	Définit une icône associée au nœud (node) ou à la ligne (link ou ref).	DGraphElement.ico
ordered	Définit la mise en page (empilée ou libre) dans un compartiment.	DNode.layout
view	Identifie le diagramme propriétaire du nœud (dans le cas multi-vues).	DNode.pointOfView
nav	Un hyperlien de navigation vers un autre diagramme (une autre vue).	DEdge.target
label	Une étiquette graphique décorant le nœud (node) ou la ligne (link). Joue le rôle de l'attribut de la classe indiqué par l'argument.	DGraphElement.lab
style	Référence vers un style déclaré dans le modèle stylistique.	DBaseStyle (méta-modèle Diastyle)

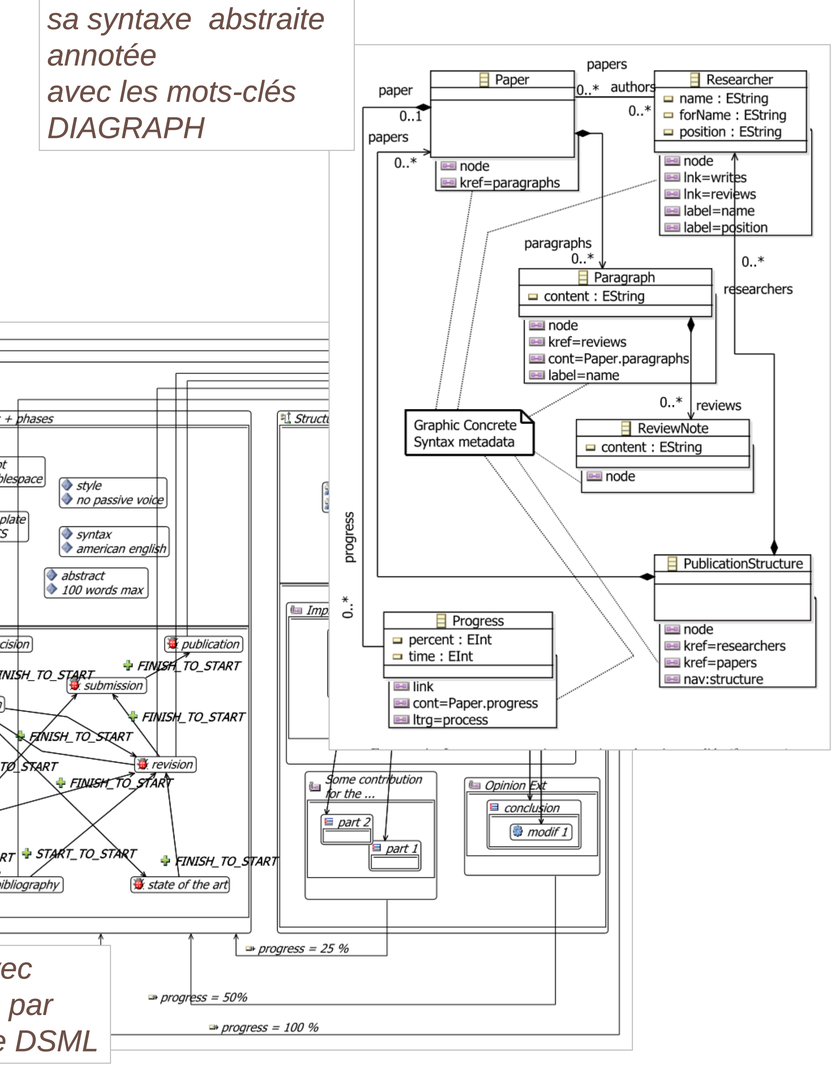
Le méta-modèle

(sa syntaxe abstraite)



Un DSML obtenu avec DIAGRAPH

sa syntaxe abstraite annotée avec les mots-clés DIAGRAPH



Un modèle édité avec le modeleur généré par DIAGRAPH pour ce DSML

Mise à disposition de l'environnement DIAGRAPH : Logiciel libre (Licence Eclipse EPL)
https://code.google.com/p/diagraph/

Auteurs

François Pfister

Vincent Chapurlat

LGI2P, Ecole des Mines d'Alès, site de Nîmes, Parc Scientifique G. Besse, 30000 Nîmes

Partenaires

Marianne Huchard

Clémentine Nebut

LIRMM, CNRS – Université Montpellier 2, 161 rue Ada, 34095 Montpellier Cedex 5, France

Colloque IMT : « Numérique : Grande échelle et complexité »

Mars 2014

Guide pour la modélisation et la vérification de la cohérence modes de fonctionnement / scénarios opérationnels en Ingénierie Système : **GEMOS**

Parties prenantes

MINES Alès

LGI²P

Laboratoire de Génie Informatique et d'Ingénierie de Production

Auteurs

Chapurlat Vincent

Daclin Nicolas

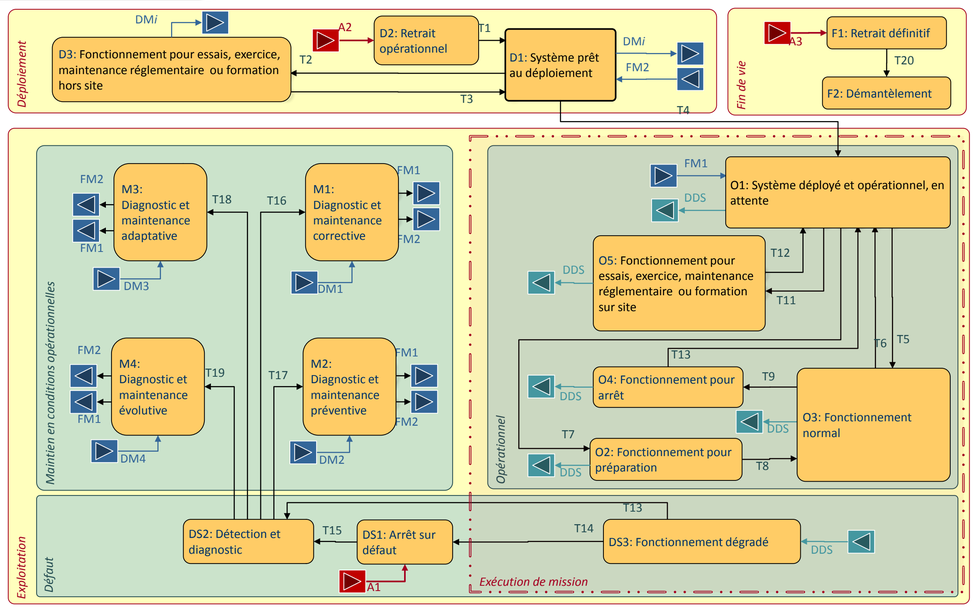
Contexte

Problématique

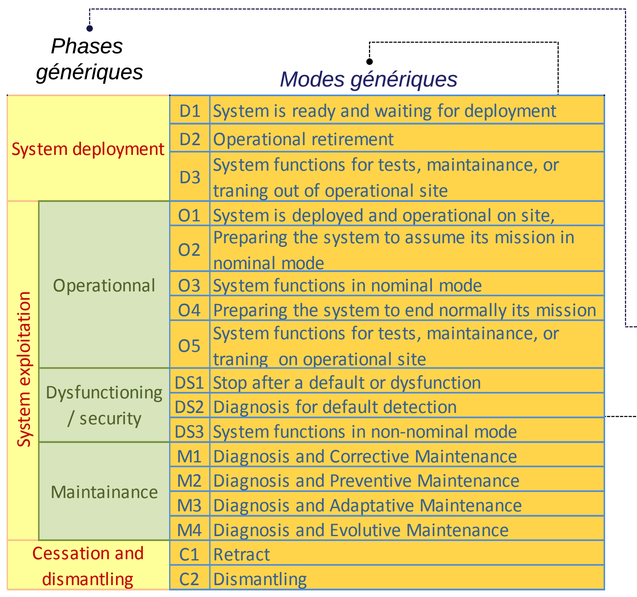
- Guider la modélisation des modes de fonctionnement d'un système complexe en cours de conception
- Aider à l'émergence d'une ou plusieurs architectures fonctionnelles correspondantes
- S'assurer de la cohérence de l'ensemble

Objectifs : un guide outillé, le GEMOS…

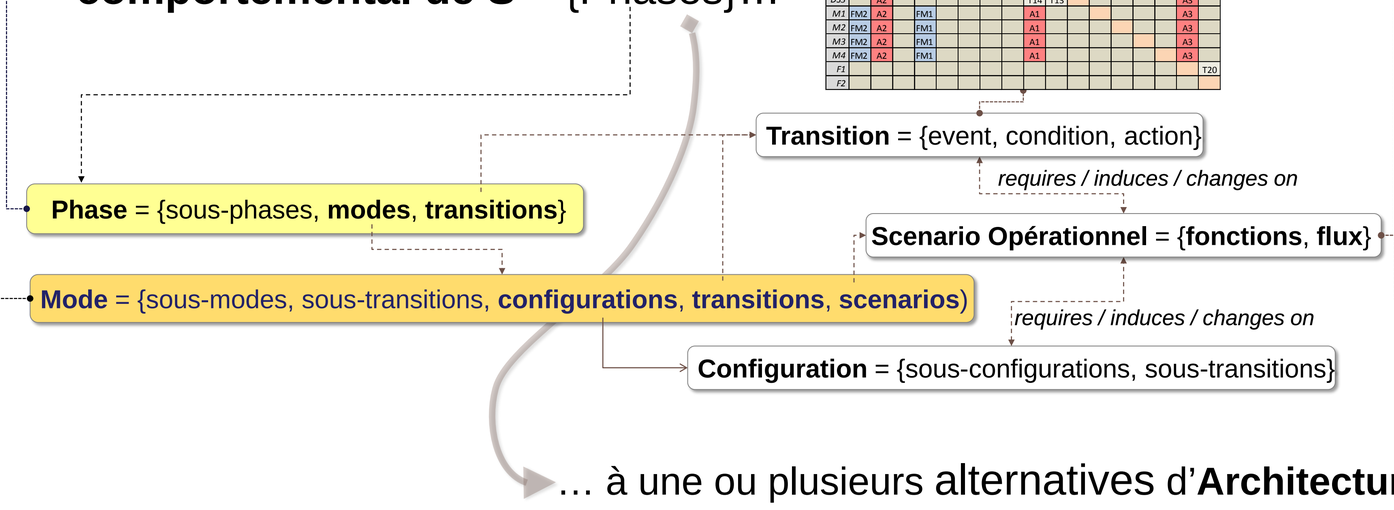
- Définir un langage de modélisation du comportement d'un système liant naturellement les modes de fonctionnement, les scénarios opérationnels et l'architecture fonctionnelle résultante
- Gagner du temps de modélisation par réutilisation d'un modèle de comportement de référence applicables à tous les systèmes
- Gagner en confiance par une vérification assistée de la qualité du modèle du système en tenant compte de divers types d'exigences
- Disposer d'un outillage support

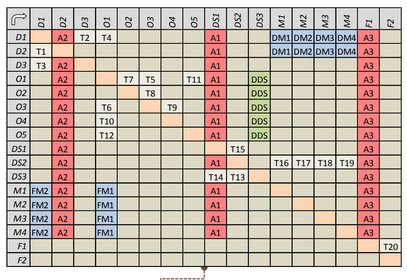


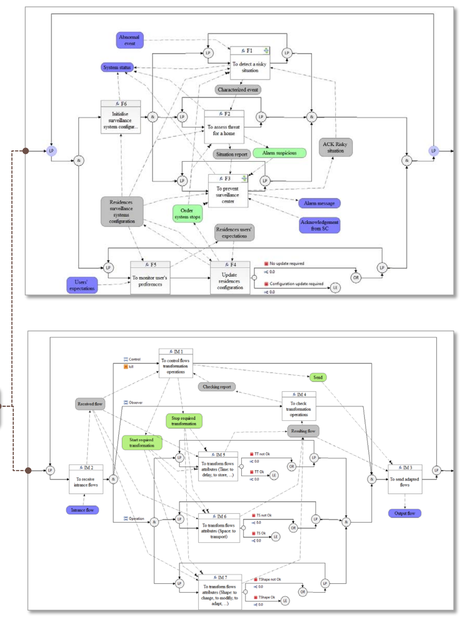
Modélisation



- **Principe** : passer du **modèle comportemental de S** = {Phases}…

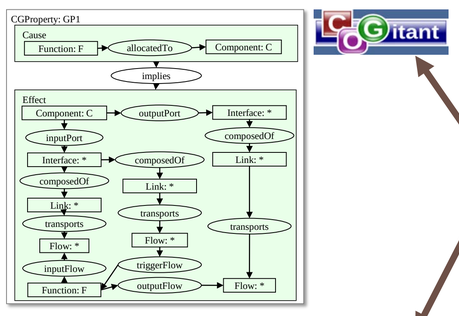


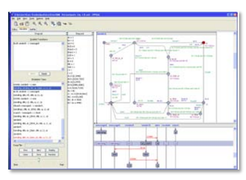


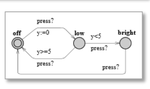


… à une ou plusieurs alternatives d'**Architectures fonctionnelles de S**

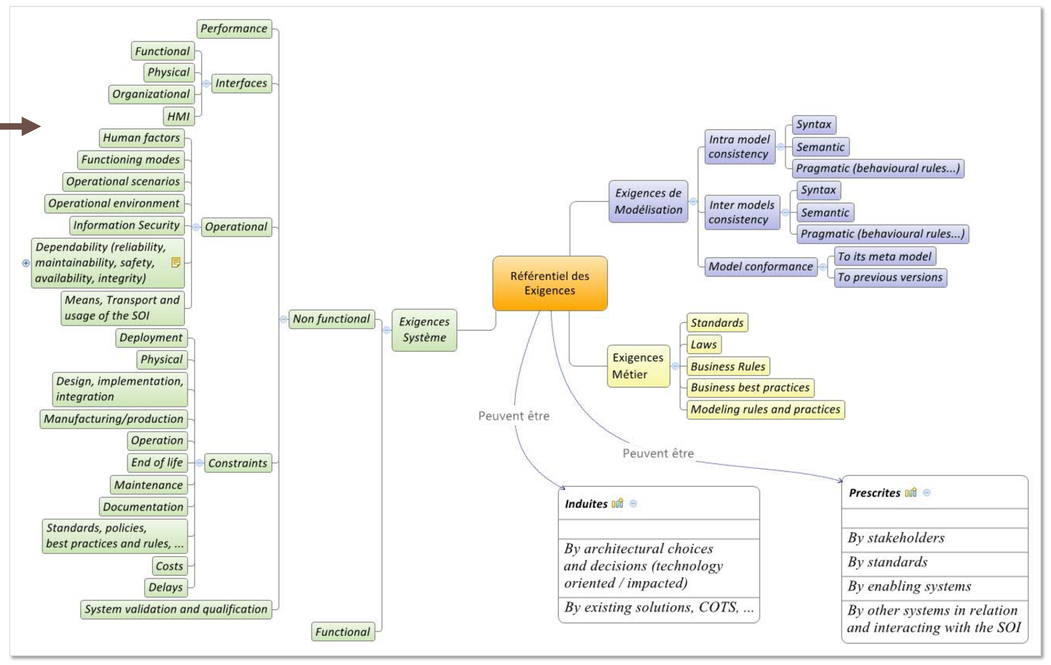
Vérification







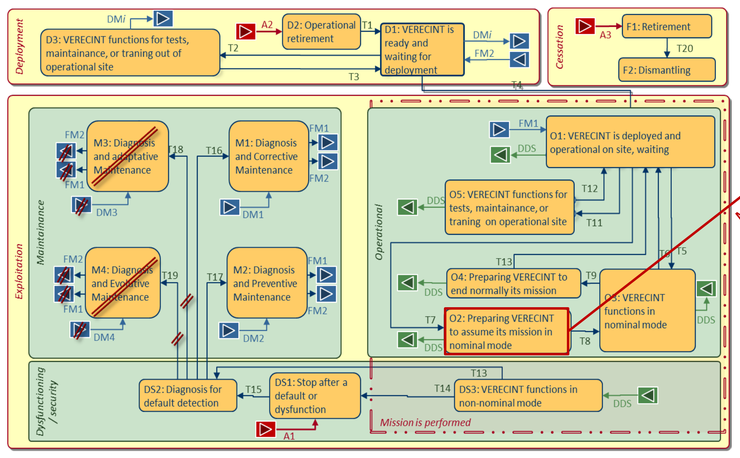
- **Principe** : traduire les **exigences de modélisation**, **métier** et **système** sous forme de **propriétés** et les prouver
- **Prouver les propriétés** : transformation vers le model checker UPPAAL (TCTL et Automates à états temporisés) et/ou vers COGITANT (Graphes Conceptuels)
- **Simuler le comportement** : sémantique opérationnelle du GEMOS



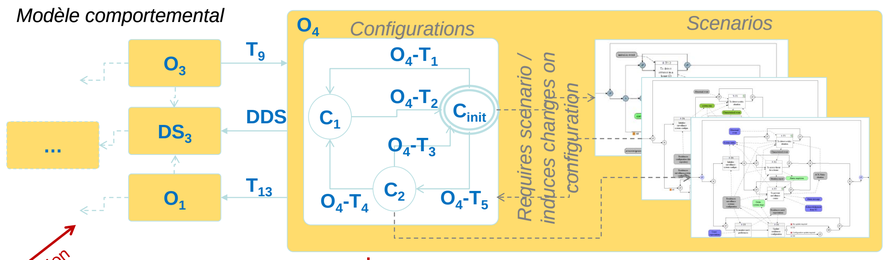
Processus opératoire actuel

http://pompiersdechateauneuf.com/lexique.php

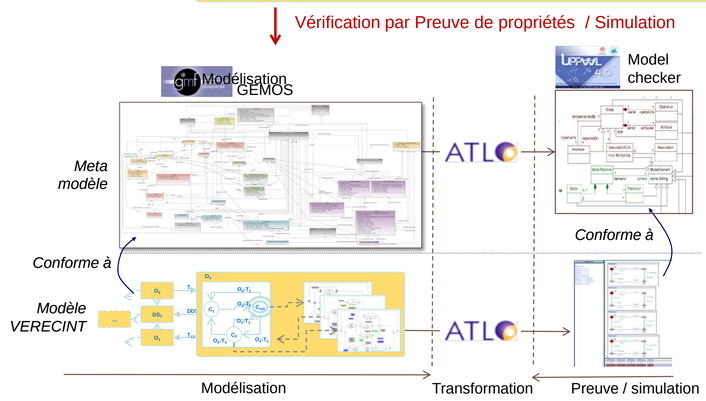
VERECINT :
véhicule de recherche et d'intervention NRBC



Modèle comportemental



Vérification par Preuve de propriétés / Simulation



Partenaires

FONDATION TELECOM

INSTITUT CARNOT Télécom & Société numérique

AFIS Association Française d'Ingénierie Système

Contact : Vincent.Chapurlat@mines-ales.fr

Site web : www.lgi2p.ema.fr

Contribution à l'évaluation d'architectures en Ingénierie des Systèmes Mécatroniques

Parties prenantes

MINES Alès

LGI²P

Laboratoire de Génie Informatique et d'Ingénierie de Production

Auteurs

Couturier Pierre
Chapurlat Vincent
Lô Mambaye

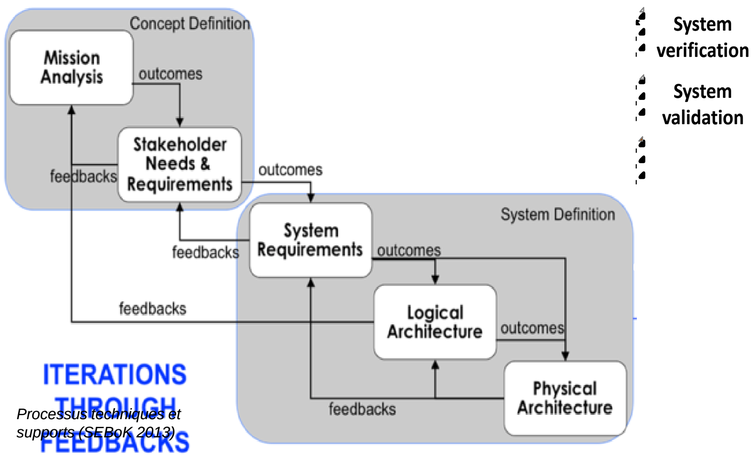
PROBLEMATIQUE ET VERROUS

Ingénierie Système (IS) et Mécatronique

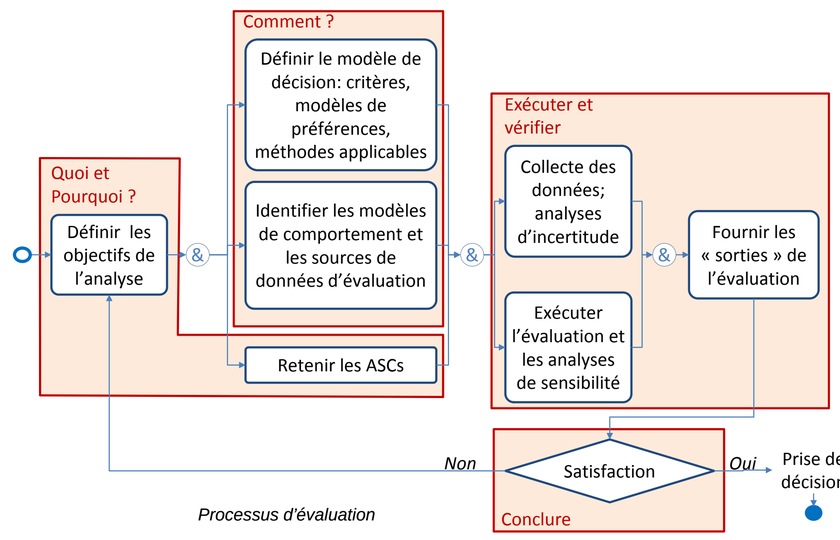
- Ingénierie interdisciplinaire / couplage fort entre disciplines
- Evaluation d'architectures organiques multi technologies : limites des modèles, des méthodes applicables et des outils.
- Nombreux paramètres / nombreux critères dont certains antagonistes
- Caractère itératif de la conception

Verrous

- Manque de vision partagée et unifiée de l'évaluation d'architectures candidates en IS : **System Analysis**
- Prise en compte et traçabilité des exigences et des connaissances métier de la Mécatronique au plus tôt
- Incertitude et imprécision des modèles d'analyse dans les premières phases de conception



Processus techniques et supports (SEBoK 2013)

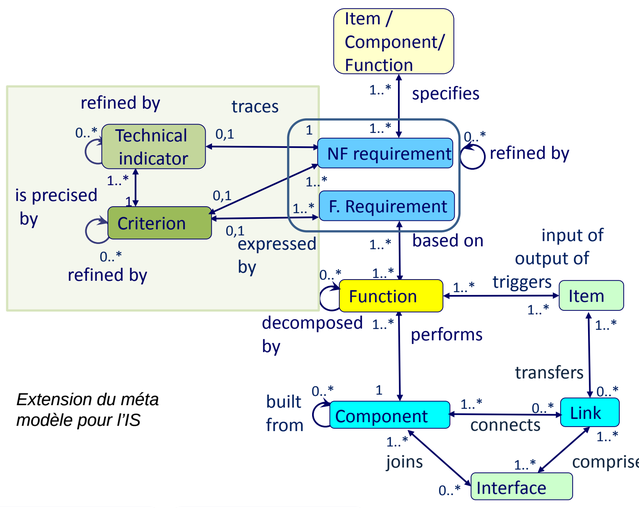


Processus d'évaluation

CONTRIBUTIONS *(Lô et al. 2013)*

Conceptuelle

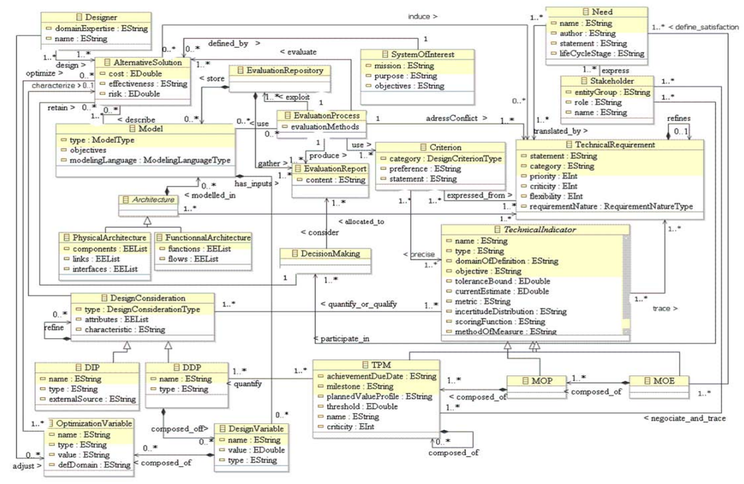
- Modèle conceptuel des données pour l'évaluation d'architectures en IS



Extension du méta modèle pour l'IS

Méthodologique

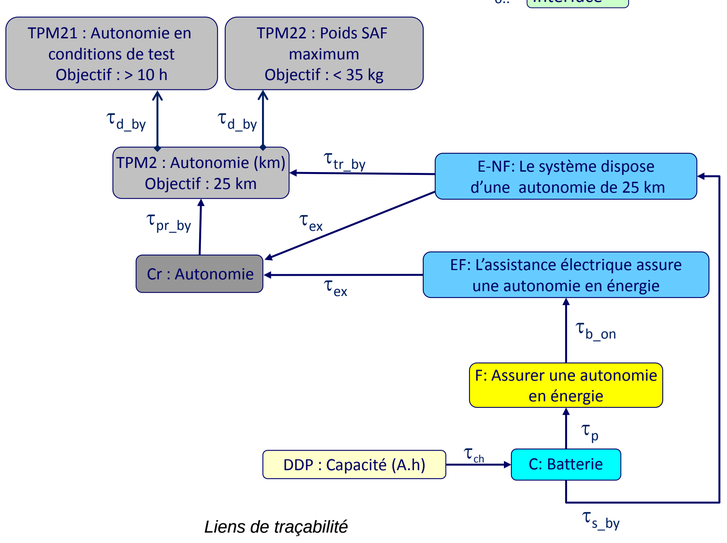
- Traçabilité des choix de conception sur les modèles issus d'un projet de conception en IS
- Méthode qualitative d'analyse multicritère



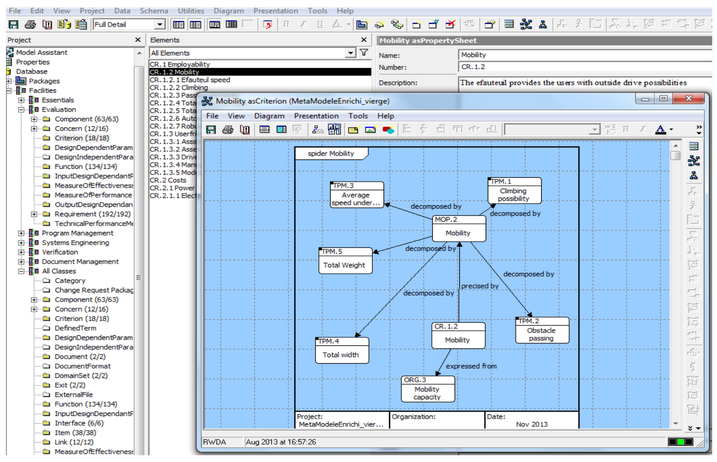
Modèle conceptuel des données pour l'évaluation

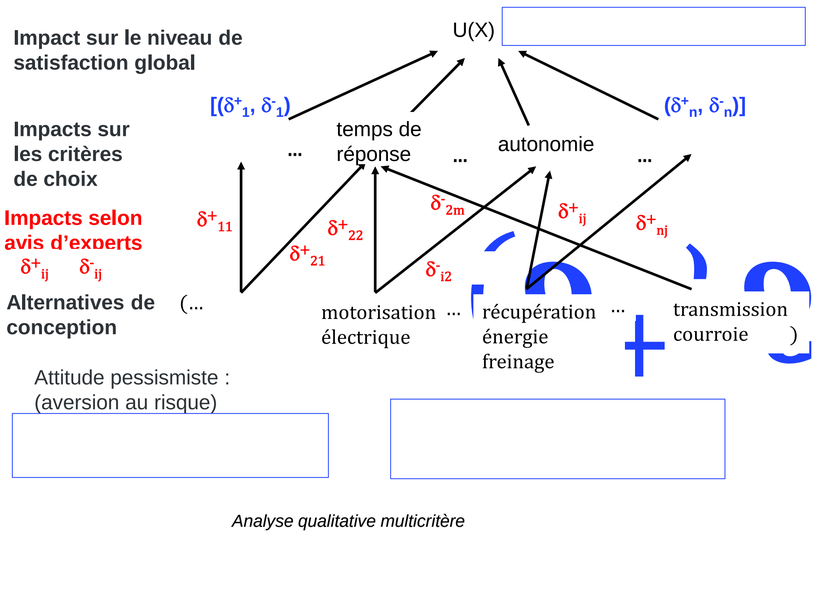
Technique

- Enrichissement de l'atelier d'IS CORE® pour l'évaluation d'alternatives de solutions



Liens de traçabilité





Analyse qualitative multicritère

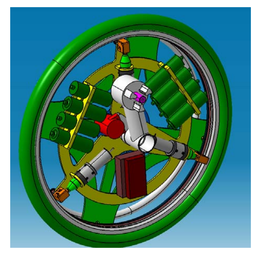
Partenaires

FONDATION TELECOM
INSTITUT CARNOT Télécom & Société numérique
AFIS Association Française d'Ingénierie Système
ED IAS ECOLE DOCTORALE
AREVA l'avenir pour énergie
Vitech Insight through integration

APPLICATION

Conception d'une assistance électrique pour fauteuil roulant

- Permettre l'accessibilité ou le maintien dans l'emploi de personnes à mobilité réduite
- Déclenchement de l'assistance électrique lors de la poussée sur la main courante
- Conduite habituelle d'un fauteuil roulant conservée
- La méthode d'évaluation permet d'identifier les solutions prometteuses.



26-27 mars 2014 Numérique : Grande échelle et complexité

Contact Pierre.Couturier@mines-ales.fr Site web www.lgi2p.ema.fr

La robotique bio-inspirée : quand la robotique s'inspire du vivant.

Parties prenantes

Auteurs

Frédéric BOYER
Professeur
Ecole des Mines de Nantes

Vincent LEBASTARD
Maitre assistant
Ecole des Mines de Nantes

Mathieu POREZ
Maitre assistant
Ecole des Mines de Nantes

Partenaires

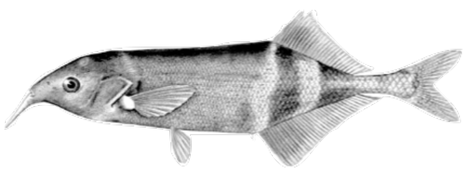
La bio-inspiration : un nouveau paradigme pour la robotique ...

■ D'un point de vue conceptuel, la bio-inspiration est un modèle de pensée dans lequel la conception de nouvelles technologies est basée sur l'étude de la nature ou du vivant.

→ En particulier, pour la robotique, ce paradigme consiste à s'inspirer des animaux pour lever les verrous de l'autonomie : c-à-d l'aptitude à percevoir, interpréter, décider et agir sur son environnement de manière adaptée sans interventions d'une volonté humaine extérieure. Dans ce contexte, l'autonomie est conçue comme le produit de :

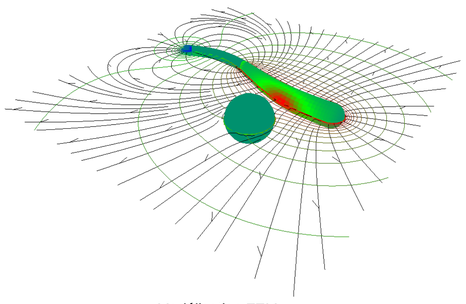
- l'intelligence incarnée dans la morphologie du corps ;
- l'intelligence collective.

Observer



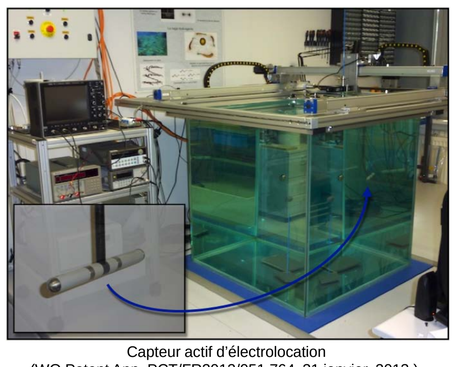
Gnathonemus petersii

↓ Comprendre

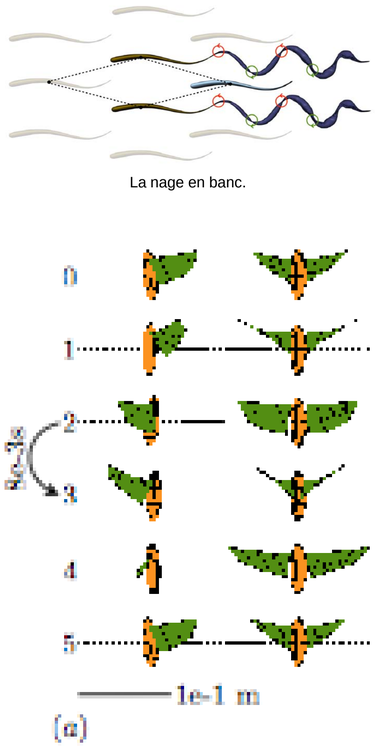


Modélisation FEM.

↓ Reproduire



Capteur actif d'électrolocation
(WO Patent App. PCT/FR2012/051,764, 31 janvier, 2013.).



La nage en banc.

Le vol battant des insectes.

Nos thèmes de recherche :

- La locomotion (depuis 2003) :
  - élaboration d'une théorie générale de la locomotion bio-inspirée en robotique ;
  - conception d'outils de modélisation et de simulation dédiés à la commande ;

  → applications à la nage des poissons, la reptation des serpents, le vol battant des insectes, etc …
- La perception (depuis 2007) :
  - la perception inspirée des poissons électriques ;
  - modélisation du sens électrique ;
  - conception de capteurs innovants pour la robotique ;
  - commande pour la navigation de robots sous-marins ;

  → brevet WO Patent App. PCT/FR2012/051,764, 31 janvier, 2013.

Les projets (passés et présents) :

■ CNRS ROBEA Robot Anguille (2003-2006) - Etude et réalisation d'un robot anguille.

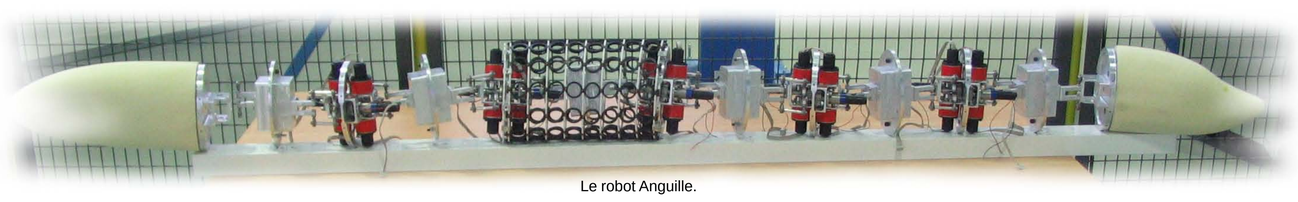
■ ANR PSIRob RAAMO (2007-2011) - Etude et réalisation d'un robot anguille autonome doté du sens électrique.

■ FP7 FET ANGELS (2009-2012) - Etude des interactions entre morphologie, perception et locomotion : application à la robotique sous-marine.

■ ANR Blanc EVA (2008-2013) - Etude et réalisation d'un robot volant autonome inspiré de l'insecte.

■ Projet Région et Carnot (2012-2013) - Equipements de laboratoire.

■ Projet Région CEA-Tech (2014-2017) - Télé-manipulation par retour électro-haptique dans l'eau et l'air.



Le robot Anguille.

Un module du robot ANGELS.

Le laboratoire de robotique bio-inspirée de l'Ecole des Mines de Nantes.

26 & 27 Mars 2014 Numérique : Grande échelle et complexité.

Contact frédéric.boyer@mines-nantes.fr Site web www.irccyn.ec-nantes.fr

Filtering atMostNValue with difference constraints: application to the Shift Minimisation Personnel Task Scheduling Problem

Jean-Guillaume FAGES jean-guillaume.fages@mines-nantes.fr Tanguy LAPEGUE tanguy.lapegue@mines-nantes.fr

1) Problem & Applications

Objective: minimise resource consumption
C1: Overlapping tasks need different resources
C2: Tasks require qualified resources

Applications:
Assignment of classes to rooms
Assignment of fixed jobs to machines
Assignment of fixed tasks to workers

A simple example of the SMPTSP

$W_{t_3} = \{w_1, w_3\}$ $W_{t_5} = \{w_1, w_2, w_5\}$
$W_{t_2} = \{w_1, w_2, w_3\}$ $W_{t_4} = \{w_3, w_4, w_5\}$
$W_{t_1} = \{w_2, w_3, w_4\}$

5 workers: $w_{1\to5}$
5 tasks: $t_{1\to5}$
qualifications: $W_{t_{1\to5}}$
Time

2) Straightforward CP Model

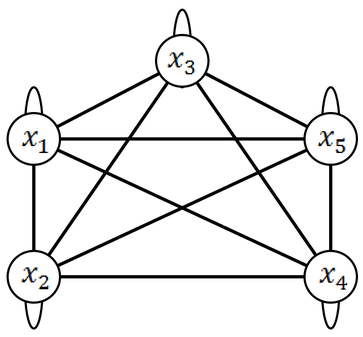
$\text{minimise}(z)$
s.t. : $AllDifferent(x_i | i \in K) \quad \forall K$ — No Communication
$AtMostNValue(X, z)$
$Dom(z) = [LB_{\neq}; |W|]$
$Dom(x_i) = W_{t_i} \quad \forall t_i \in T$

K: a maximal set of overlapping tasks
T: set of tasks W: set of workers
$LB_{\neq}$: size of the largest set K

3) AtMostNValue filtering

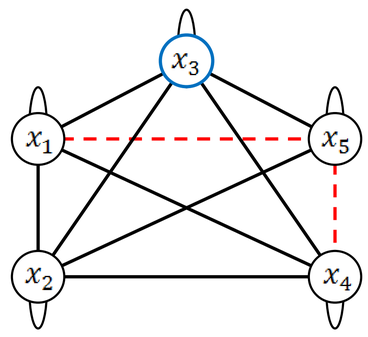
Given the intersection graph $G_I = (V, E_I)$ of the set of variables X, along with an independent set A in G_I

R_1: $\underline{z} \leftarrow \max(\underline{z}, |A|)$
R_2: $|A| = \overline{z} \Rightarrow \forall i \in V,\ Dom(x_i) \leftarrow Dom(x_i) \cap \bigcup_{a \in A} Dom(x_a)$



$z = 1$
$x_1 = \{w_2, w_3, w_4\}$
$x_2 = \{w_1, w_2, w_3\}$
$x_3 = \{w_1, w_3\}$
$x_4 = \{w_3, w_4, w_5\}$
$x_5 = \{w_1, w_2, w_5\}$

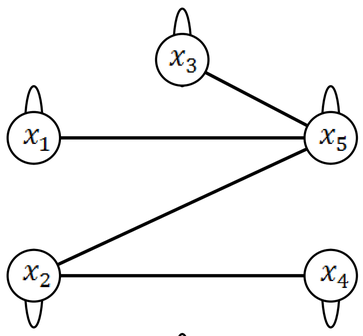
R_1 R_2



$z = 1$
$x_1 = \{w_2, w_3, w_4\}$
$x_2 = \{w_1, w_2, w_3\}$
$x_3 = \{w_1, w_3\}$
$x_4 = \{w_3, w_4, w_5\}$
$x_5 = \{w_1, w_2, w_5\}$

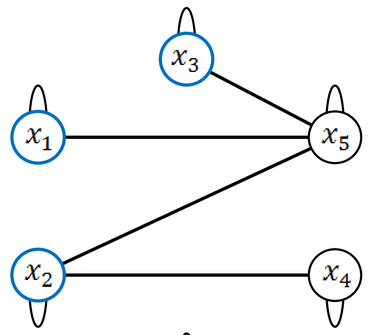
4) Improving filtering (Graph & Rules)

The constrained intersection graph G_{CI} of the set of variables X and the set of differences constraints C is deduced from G_I by removing edges (i, j) whenever $neq(i, j) \in C$



$z = [\![1; 3]\!]$
$x_1 = \{w_2, w_3, w_4\}$
$x_2 = \{w_1, w_2, w_3\}$
$x_3 = \{w_1, w_3\}$
$x_4 = \{w_3, w_4, w_5\}$
$x_5 = \{w_1, w_2, w_5\}$

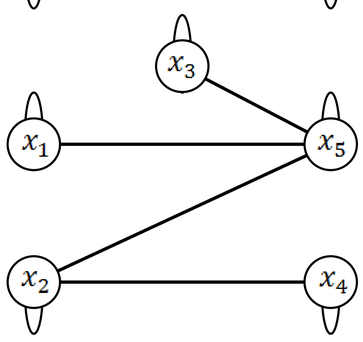
R_1 R_2



$z = 3$
$x_1 = \{w_2, w_3, w_4\}$
$x_2 = \{w_1, w_2, w_3\}$
$x_3 = \{w_1, w_3\}$
$x_4 = \{w_3, w_4, w_5\}$
$x_5 = \{w_1, w_2, w_5\}$

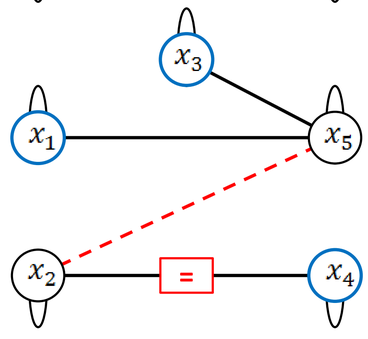
Given an independent set A of $G_{CI} = (V, E_{CI})$

R_3: $|A| = \overline{z} \Rightarrow \forall i \in V \backslash A,$
$\begin{cases} A_i = \{a\} \Rightarrow Dom(x_a) \leftarrow Dom(x_a) \cap Dom(x_i) \\ Dom(x_i) \leftarrow Dom(x_i) \cap \bigcup_{a \in A} Dom(x_a) \end{cases}$



$z = 3$
$x_1 = \{w_2, w_3, w_4\}$
$x_2 = \{w_1, w_2, w_3\}$
$x_3 = \{w_1, w_3\}$
$x_4 = \{w_3, w_4, w_5\}$
$x_5 = \{w_1, w_2, w_5\}$

R_1 R_3



$z = 3$
$x_1 = \{w_2, w_3, w_4\}$
$x_2 = \{w_1, w_2, w_3\}$
$x_3 = \{w_1, w_3\}$
$x_4 = \{w_3, w_4, w_5\}$
$x_5 = \{w_1, w_2, w_5\}$

5) Diversifying filtering

Rules rely on independent sets:
- Finding **large** independent sets is important
- Finding **different** independent sets is important

What is done in the literature?
→ minDegree algorithm (MD)
1) Fast
2) Effective
3) Deterministic → **No diversification**

How to get diversification?

1) Breaking ties randomly in MD? **(no impact)**
2) Computes k pseudo-random independent sets? **(not effective)**
3) Computes k random independent sets: R^k **(improves filtering)**
- Complements MD
- Provides control over the tradeoff time/filtering

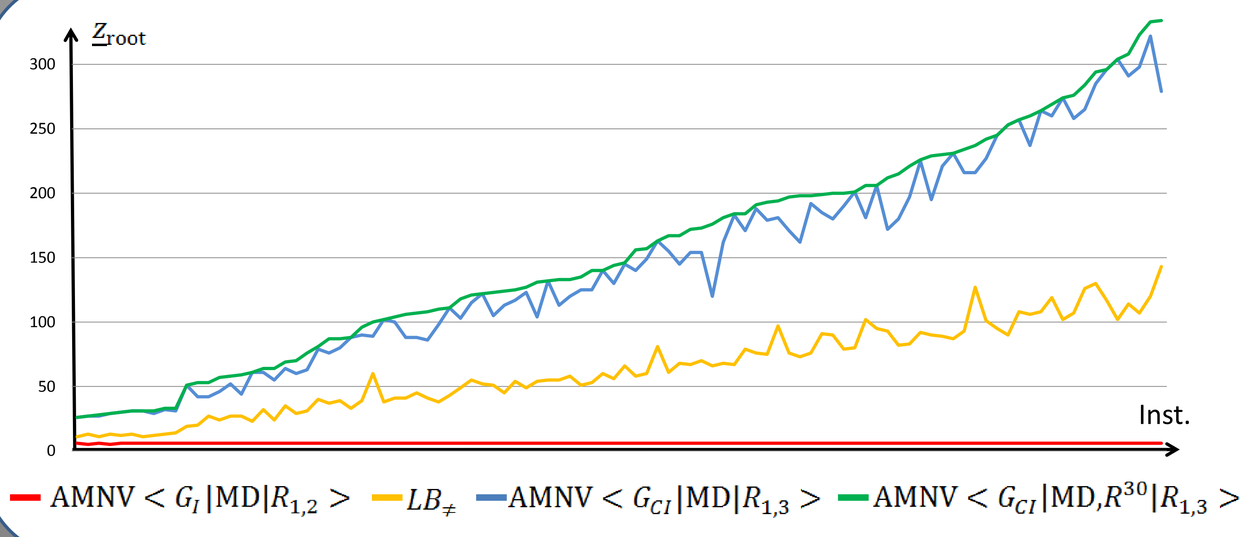
↳ Improvements opportunity

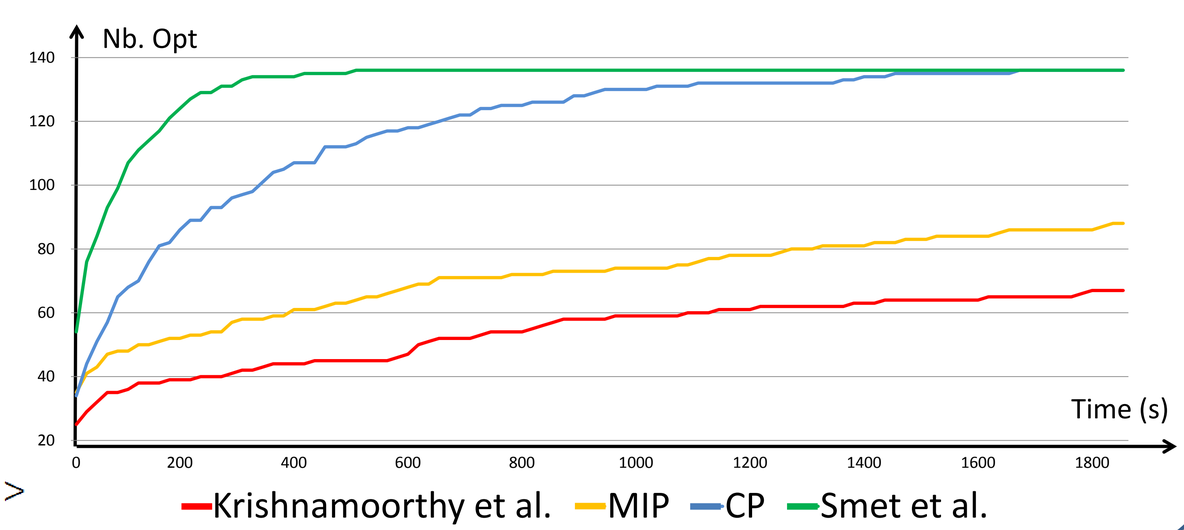
	Data 137				
k	10	30	50	70	90
Nb. Opt	101	**109**	106	99	95

	Data 100				
k	100	200	400	800	1600
Nb. Opt	19	22	28	**35**	31

Results after 5 min with $\text{AMNV} < G_{CI} | \text{MD}, R^k | R_{1,3} >$

6) Results & Literature





Programme Gaspard Monge pour l'Optimisation et la Recherche Opérationnelle

Learning Constraints for Reducing Combinatorics

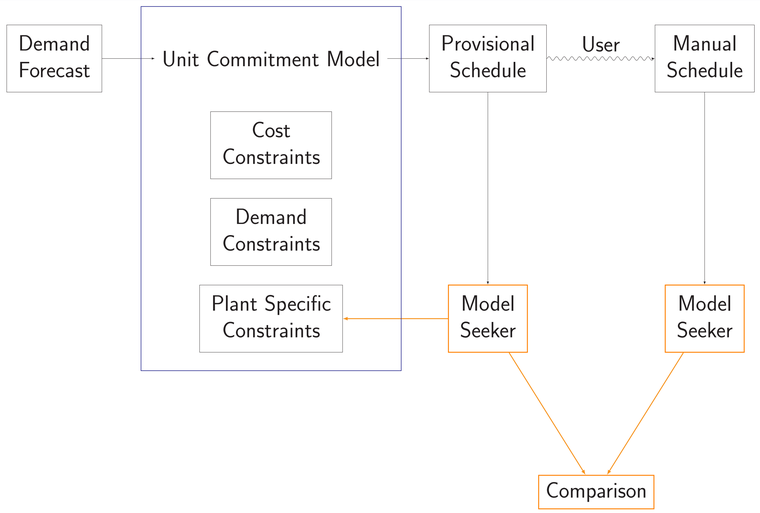
Nicolas Beldiceanu, Helmut Simonis, Georgiana Ifrim, Arnaud Lenoir, Jean-Yves Lucas

TASC team (CNRS/INRIA), Mines de Nantes, France
Cork Constraint Computation Centre, University College Cork, Ireland
EDF Research, France

Why, What, How?

- Learn plant specific constraints from production planning data
- Discover known or perhaps new, hidden constraints
- Use output of Unit Commitment Problem (UCP)
- Consider different plant types (nuclear, thermal, hydro)
- Learn from both the provisional schedule and manually modified solutions; compare
- Using, adapting and extending existing ConstraintSeeker and ModelSeeker tools
- New, specialized UCP-ModelSeeker tool combining Constraint Programming and Machine Learning
- Adding new global constraints to Global Constraint Catalogue
- Run on large datasets (1.5 million samples)

Learning in the EDF Unit Commitment Problem



Main Components of UCP-ModelSeeker

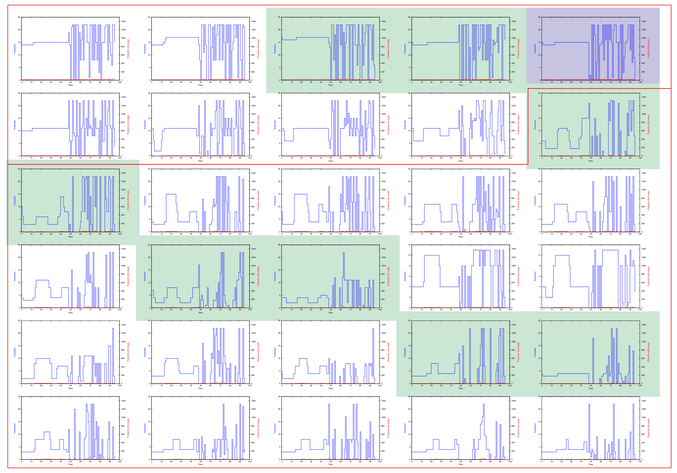
CLUSTERING TIME SERIES
(to learn stronger constraints)

LEARNING DOMAINS
- For each period
- For each functional dependency parameter

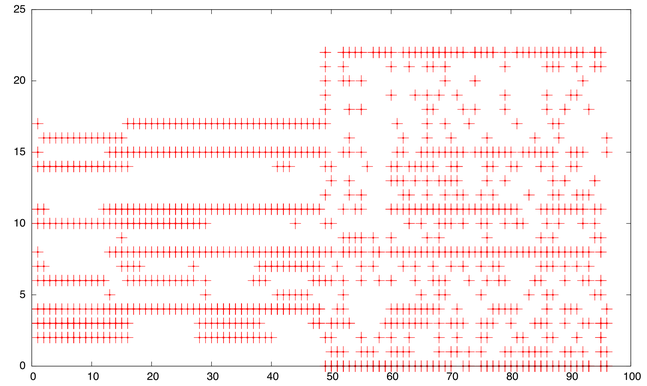
LEARNING CONSTRAINTS
- Functional dependency constraints
- Constraints without functional dependency
- Binary constraints (on consecutive periods)
- Variable partitions generator (find breakpoints)

Execute Model
(generate similar time series)

Clustering of Power Output of Example Plant, April 2010. (Clusters in red, Weekends/Holidays in green/violet)



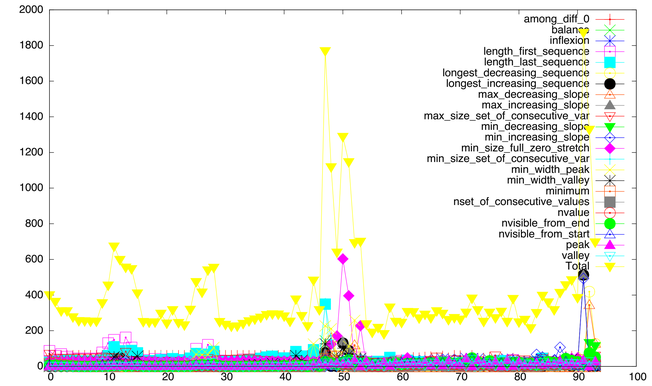
Variable Domains for Example Plant



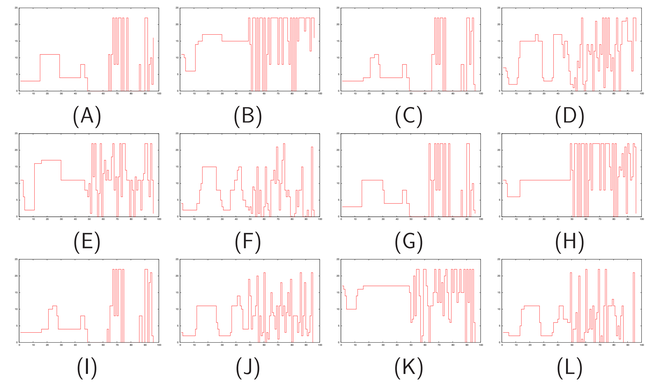
New Constraints to Characterize Structured TIme Series

among_diff_0	: *number of values different from* **0**,
max_nvalue	: *number of occurrences of the most used value*,
min_nvalue	: *number of occurrences of the least used value*,
balance	: *difference in count of the most and least used values*,
change	: *number of consecutive values that are different.*
peak	: *number of peaks*,
highest_peak	: *altitude of the highest peak*,
min_width_peak	: *smallest width of any peak*,
nvisible_from_start	: *number of peaks visible from the start*,
nvisible_from_end	: *number of peaks visible from the end*,
inflexion	: *number of peaks and valleys*,
min_dist_between_inflexion	: *minimum distance between consecutive inflexions*,
longest_increasing_sequence	: *range of the longest increasing subsequence*,
max_increasing_slope	: *maximum slope on the strictly increasing subsequences*,
min_increasing_slope	: *minimum slope on the strictly increasing subsequences*,
big_peak	: *number of big peaks of* .

Identify Timepoint(s) When Profile Was Manually Updated: Find Sub-Sequences with Different Behaviour



UCP-ModelSeeker Generated Profiles for Example Plant



Comparing Constraints, Search Strategies and Solution Quality Relative to Input Data

Variant	Search	Cluster	Split	MAE	MSE	Time (sec)
A	Frequent	All	no	452.30	**87.53**	1.10
B	Frequent	1	no	449.67	104.62	0.53
C	Frequent	2	no	298.43	70.90	0.87
D	Random	All	no	649.97	114.20	1.43
E	Random	1	no	492.90	106.68	0.54
F	Random	2	no	422.33	82.48	0.89
G	Frequent	All	yes	**445.10**	87.82	2.30
H	Frequent	1	yes	**431.33**	**101.70**	1.03
I	Frequent	2	yes	**294.00**	**70.70**	1.74
J	Random	All	yes	547.37	97.23	2.32
K	Random	1	yes	510.22	111.71	1.03
L	Random	2	yes	397.86	78.38	1.73

For More Information

http://4c.ucc.ie/~hsimonis/edfcp2013.pdf
http://4c.ucc.ie/~hsimonis/modelseeker.pdf

ADVANCED IMAGE PROCESSING IN VISUAL INSPECTION

INSTITUT Mines-Télécom

MINES ParisTech

ARMINES

Auteurs

Petr Dokladal

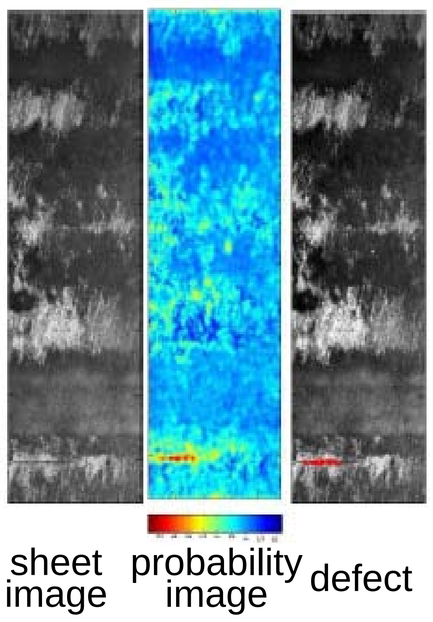
Center for Mathematical Morphology,

35, rue Saint-Honoré, 77 300 Fontainebleau,
http://cmm.mines-paristech.fr

Real-Time Control of Metal Sheet Lamination Process

SIEMENS VAI

Defect Detection in Heavily Textured Surfaces



- Every pixel is described using a feature vector (30 linear and morphological filters and 26 curvelets).
- A statistical learning is used to discriminate the defects.
- The processing is optimized (cascaded) to verify the needs of real-time processing.

Cord A., Bach F., Jeulin D. *Texture classification by statistical learning from morphological image processing. Application to metallic surfaces, Journal of Microscopy, 239, pp. 159-166, 2010*

contact: Dominique.Jeulin@mines-paristech.fr

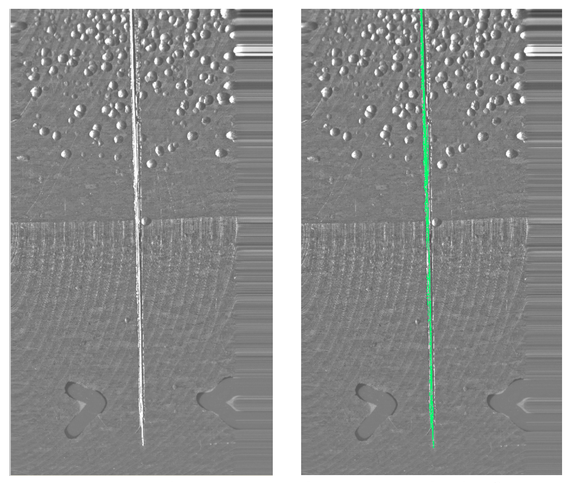
Automated Visual Inspection Of Industrial Parts

Adaptive, Cost-Optimal Defect detection

- We propose an original method to replace the dye penetrant inspection using toxic chemicals.
- The technique is fully adaptive and can detect fatal defects and ignore benign anomalies.
- Optimal algorithms have been developed to limit the processing time.

1. ***Morard V, Dokládal P, Decencière E****, Parsimonious path openings and closings. IEEE TIP, 2014*
2. ***Morard V, Dokládal P, Decencière E****, One-dimensional openings, granulometries and component trees in O(1) per pixel. JSTSP, 2012*

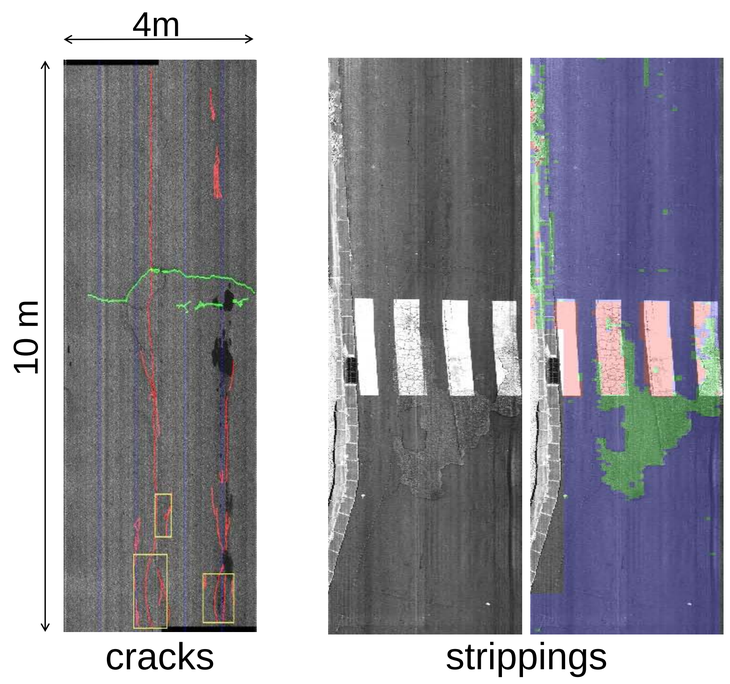
contact: Petr.Dokladal@mines-paristech.fr


crack detected defect

Paved-Road Aging Evaluation

COLAS

Detection and analysis of cracks and strippings.



cracks strippings

- Open and sealed cracks are separately detected and categorized according to : width, length, grouping and position and the cumulative length if reported for each category.
- Asphalt strippings are detected by means of texture analysis and classification.

E. Coquelle, J.-L. Gautier, P. Dokládal*. Automatic Assessment of a Road Surface Condition, Surf 2012.*

contact: Petr.Dokladal@mines-paristech.fr

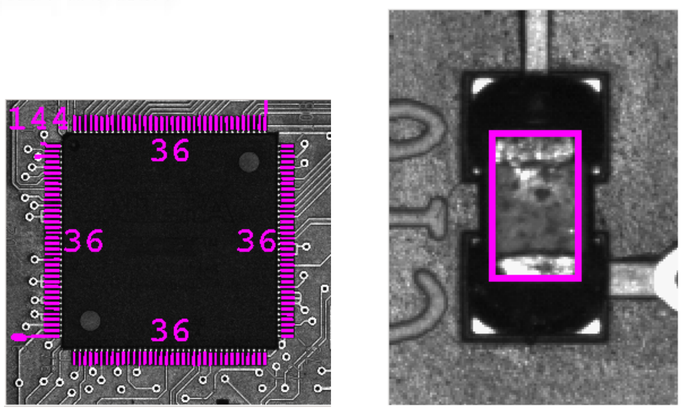
Automated Visual Inspection of Electronic Cards

TECHNOLOGY

Detection of anomalies in IC mounted PCB cards

- Automatic detection of various IC housings.
- Detection of incorrectly placed or missing IC.
- Automatic detection of IC leads.

contact: Serge.Beucher@mines-paristech.fr



detected IC leads detected IC housing

OpenETCS Project

Institut Mines-Télécom

Authors

Phd student

João Santos

Post-Doc

Huu-Nghia Nguyen

Professor

Ana Cavalli

Context

ITEA 2

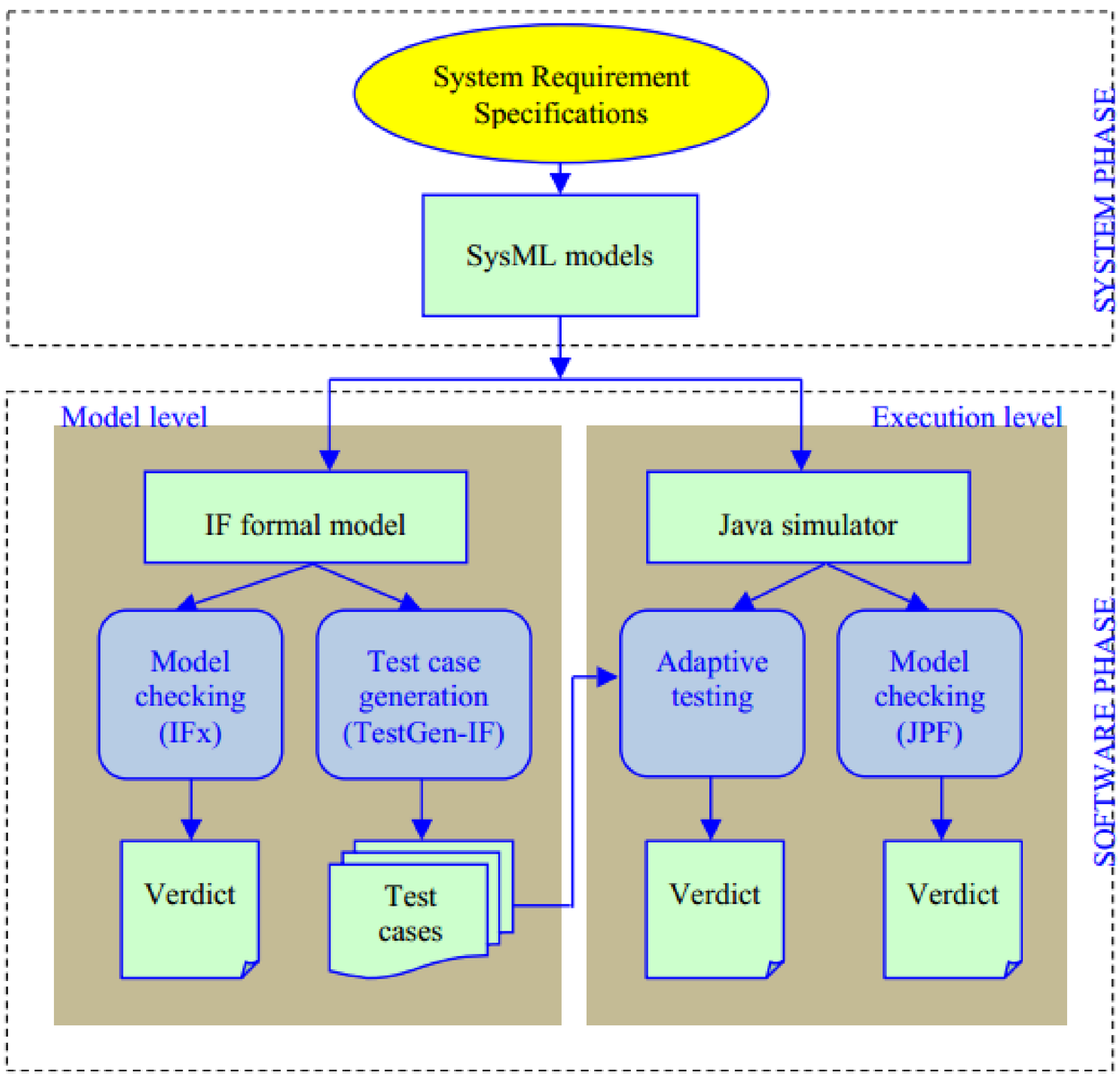
INFORMATION TECHNOLOGY FOR EUROPEAN ADVANCEMENT

Objectives

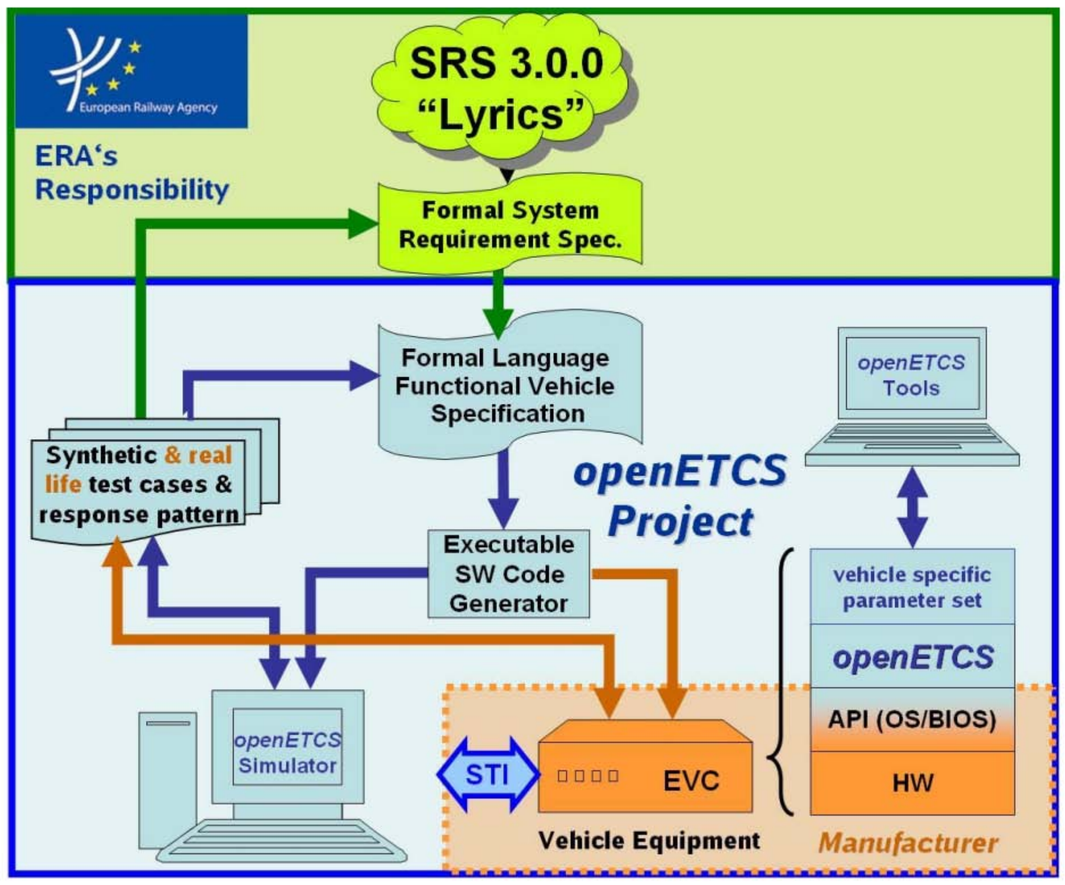
The main objective of the OpenETCS project is to develop an integrated modeling, development, validation and testing framework for leveraging the cost-efficient and reliable implementation of the European Train Control System (ETCS).

Our approach

Our work focuses on the validation and verification of an ETCS formal model. We resort to Model Checking, Simulation and Testing to achieve this goal.



OpenETCS Architecture



What is new?

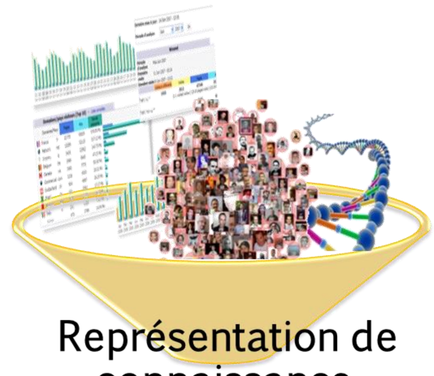
❑ OpenETCS is based on Open Standards at all levels, including hardware and software, interfaces definition, design tools , verification and validation as well as embedded control software.

❑ The avionics sector has already developed its own source tools chain and created and eco-system. The similarity between the requirements of the railway and aviation safety equipment, make such tool chains a good basis for the project.

Partners

3. GRANDES MASSES DE DONNEES

De la connaissance à la décision
Exploiter de gros volumes de données par l'analyse sémantique



Parties prenantes

MINES Alès

LGI2P – équipe KID

Equipe

Sylvie Ranwez
Jacky Montmain
Michel Crampes
Gérard Dray
Stefan Janaqi
Michel Plantié
François Trousset
Nicolas Fiorini
Sébastien Harispe

Partenaires

Montpellier SupAgro

Universitat Rovira i Virgili

Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

Vincent Ranwez[1]

Montserrat Batet[2]

David Sánchez[2]

[1] Equipe DAVEM, SupAgro Montpellier,

[2] Université Rovira i Virgili de Tarragone, Espagne

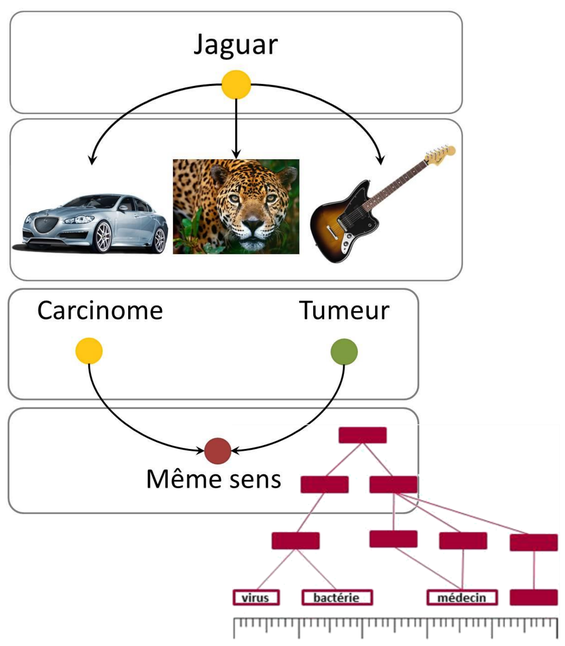
Sources d'information diverses, hétérogènes

Données non structurées

- Corpus de textes (publications, pages Web, réseaux sociaux)
- Ressources multimédia (images, sons, vidéos…)

Données structurées

- Données liées, massives et à caractère sémantique
- Représentation de la connaissance (ontologies), graphes RDF(S), OWL,…



Recherche d'information conceptuelle

Améliorer la pertinence des résultats

- Désambiguïser, généraliser, spécialiser les requêtes en utilisant une ontologie de domaine
- Assurer de meilleurs taux de précision et de rappel par des mesures sémantiques appropriées

Personnalisation, visualisation et interactivité

- Paramétrisation et personnalisation du système et de l'interface : pondération, reformulation, lentilles
- Visualisation globale des résultats sur une carte sémantique : affichage en 2D des résultats en fonction de leur degré de pertinence
- Justification des résultats

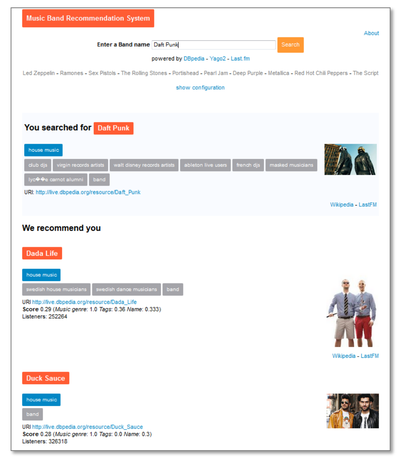
Développements logiciels

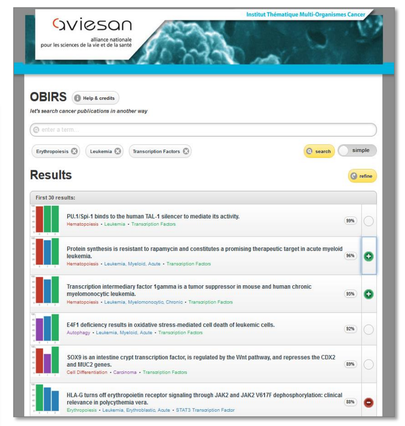
SML – *Semantic Measures Library*
Librairie logicielle dédiée au calcul de similarité sémantique
- Open-source, langage Java, exécutable en ligne de commande
- Traitement de gros volumes de données

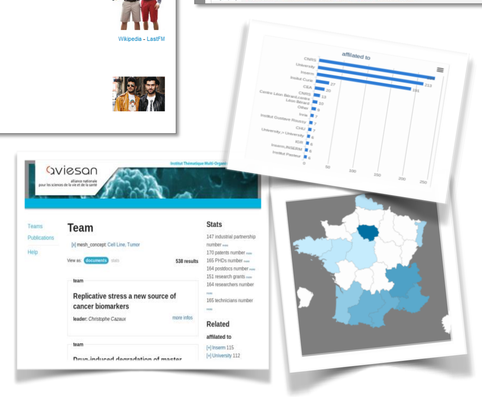
OBIRS – *Ontology Based Information Retrieval System*
Recherche d'information conceptuelle centrée sur l'utilisateur
- Application à la recherche de gènes indexés par la Gene Ontology
- Recherche de publications scientifiques biomédicales

Système de recommandation
Exploitation des mesures de similarité dans le contexte des données liées pour un système de recommandation

Kalitmo
Visualisation de données structurées pour la gestion de collectifs







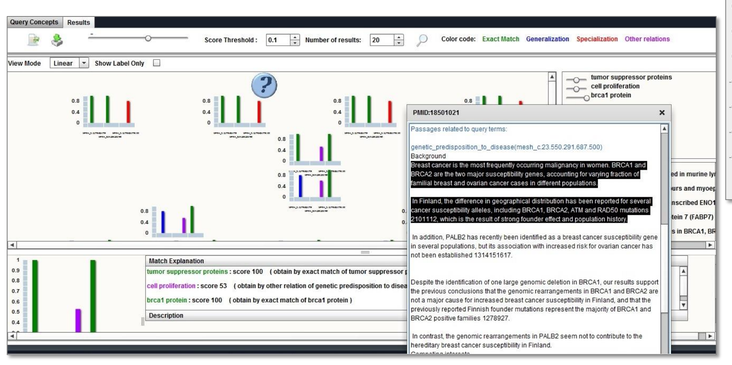
Extraction d'opinion

- Fouille de textes et apprentissage supervisé sur de larges corpus (Web) : TAL, segmentation, analyses statistiques
- Identification d'un lexique, détection de critères relatifs à un domaine
- Polarisation contextuelle, extraction conceptuelle et évaluation de critères

Développements logiciels

Synopsis
Outil de détection d'opinion, prise en compte d'une évolution temporelle

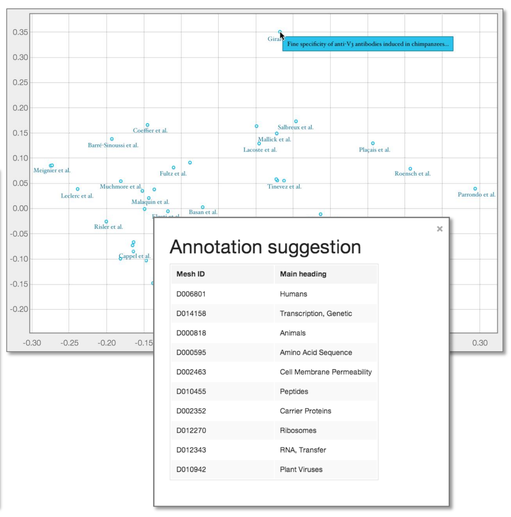
CoLexIR
Environnement hybride de RI associant deux approches conceptuelle et lexicale
Application à la fouille de corpus scientifiques avec mise en évidence des passages pertinents pour l'utilisateur



Indexation conceptuelle

Indexation par propagation

- Interface interactive d'assistance à l'indexation
- Annotation conceptuelle semi-automatique



Complexité" 26 et 27 mars 2014 colloque de l'Institut Mines-Télécom **"Numérique : Grande échelle &**

Contact sylvie.ranwez@mines-ales.fr Site web http://kidknowledge.wp.mines-telecom.fr

Enabling Large-scale Model Driven Engineering

Parties prenantes

Auteurs

Benelallam Amine[1]

Tisi Massimo[1]

Sunyé Gerson[2]

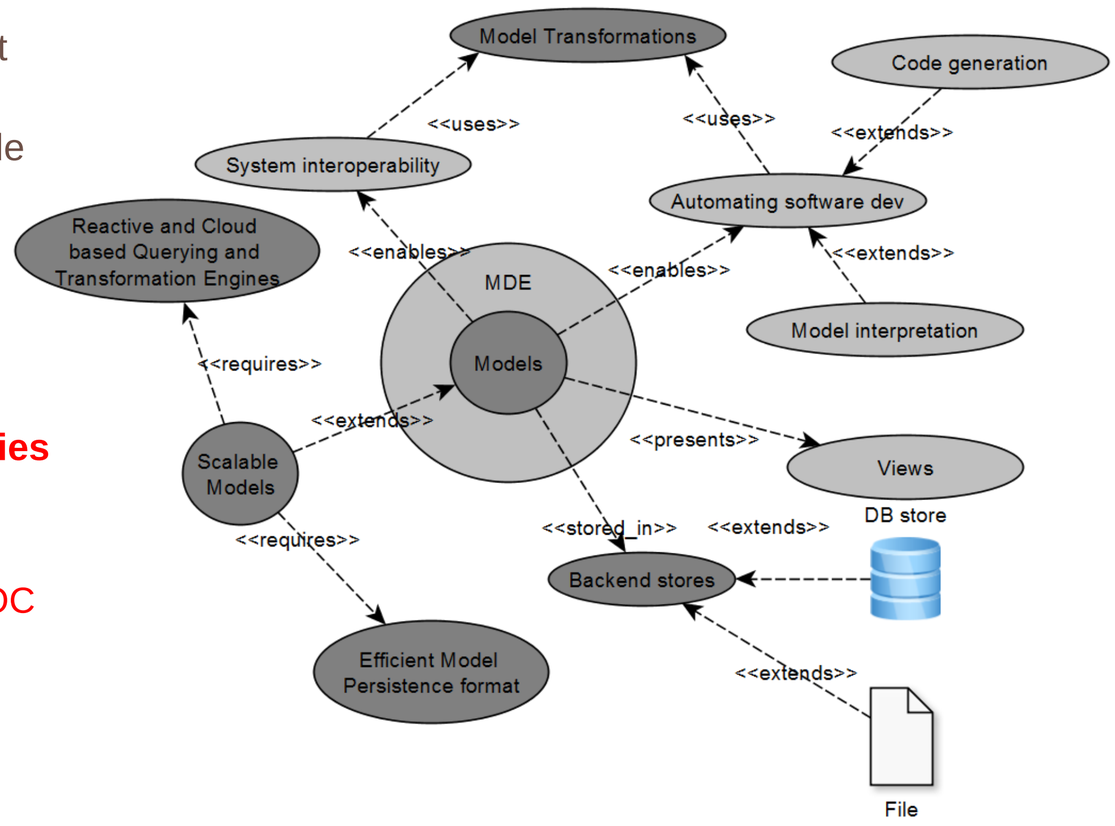
Gomez Llana Abel[1]

[1] AtlanMod, Mines Nantes, INRIA, LINA

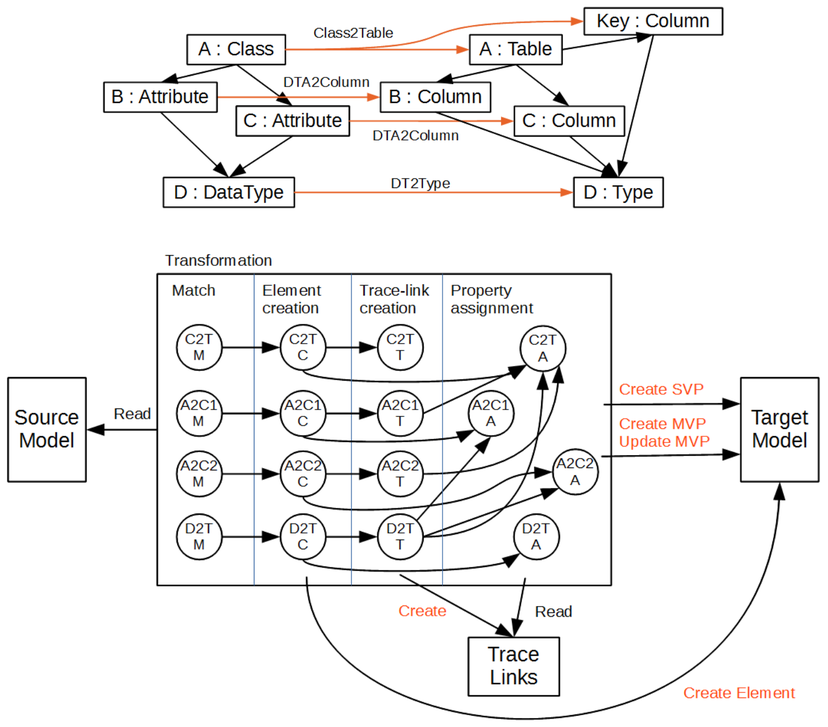
[2] AtlanMod, Université de Nantes, LINA

I- Model Driven Engineering

- Applying modeling to assist software development during the entire life-cycle
- Automating software development throughout code generation, validation, visualization etc.
- Large MDE community around Eclipse Modeling project
- Large set of tools in the Eclipse Modeling Project from industry and academics
- **The current generation of modeling technologies is stressed to its limits**
- **Examples from industry :**
  - Reverse-engineering systems with millions of LOC
  - Synchronizing views on building information models of several Gbs
- **Need of scalable MDE solutions for very large and complex systems**



II- Enabling scalability in Model Transformations (MT)

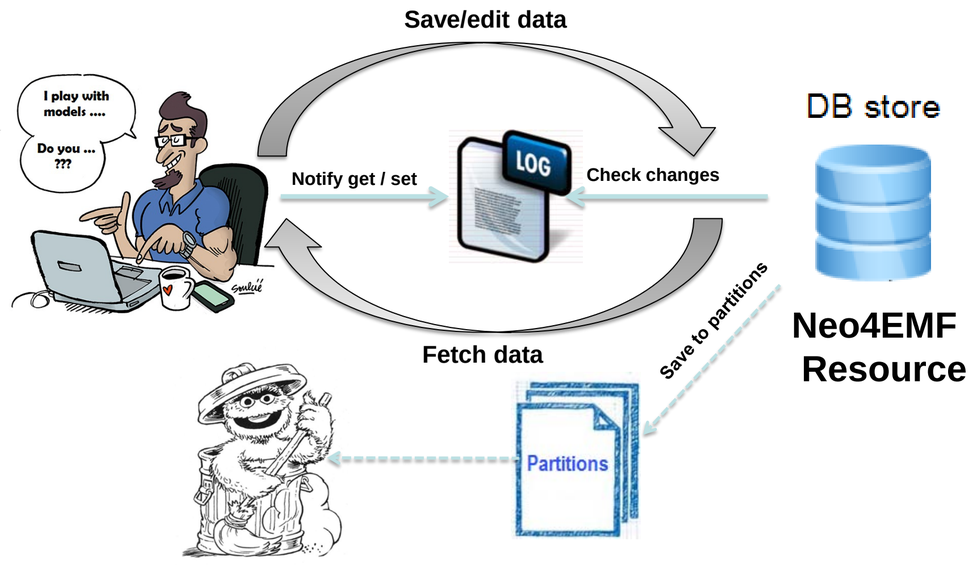


- The ATL transformation language:
  - is a model transformation language and toolkit developed in the AtlanMod research team
  - provides a parallel engine for faster transformations
  - enables change propagation and model synchronization using an incremental execution
  - reduces memory footprint and computation using lazy transformations
  - enables infinite transformations using lazy transformations

III- Enabling scalability in Model Persistence

Neo4EMF

- A model persistence framework (MPF) is a middleware that assists the storage of models
- The Neo4EMF model persistence framework:
  - provides a No-SQL backend using a graph database
  - enables loading large models using an on-demand loading mechanism
  - enables a lightweight first time loading by fetching objects not their data
  - involves a change (access) log to unload (save) models elements



Contact amine.benelallam@inria.fr

Site web http://www.emn.fr/z-info/atlanmod/index.php/User:Benelallam

Predicting personalized response to drug and environmental chemicals from genomic data with machine learning

[1,2] Erwan Scornet, [1] Elsa Bernard, [1] Yunlong Jiao, [1] Veronique Stoven, [1] Thomas Walter, [1] Jean-Philippe Vert

[1]Center For Computational Biology, Mines ParisTech ; INSERM U900 ; Institut Curie

[2]Paris VI

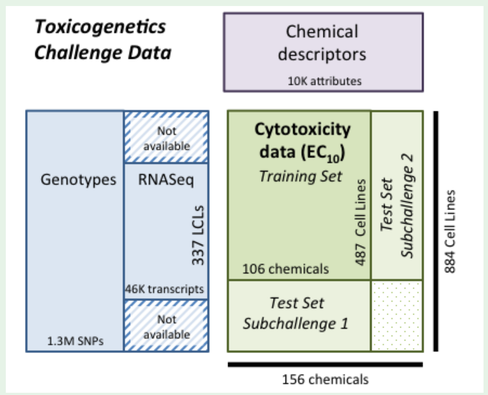
March 2014

Abstract : The response to drugs and environmental chemicals changes with people. Variation in genotype can explain different reaction to drug : while some treatment may be effective for some people, it can be useless and even harmful for others. The same mechanisms are at stake regarding the reaction to environmental agents, such as allergens. Recent advances in high-throughput sequencing open the way to personalized treatment based on genotype data. In this work, we predict the toxicity level of 106 chemicals for each patient using both information on patients and on chemicals.

Challenge of Toxicogenetic

- Provide personalized information about chemical toxicity for each patient
- Use genetic infomation of each patient to predict the chemical toxicity
- Use chemical information (substructures, compounds,...) to learn toxicity across chemicals

Data description



Data were available thanks to the DREAM challenge 8.0 which gathers :

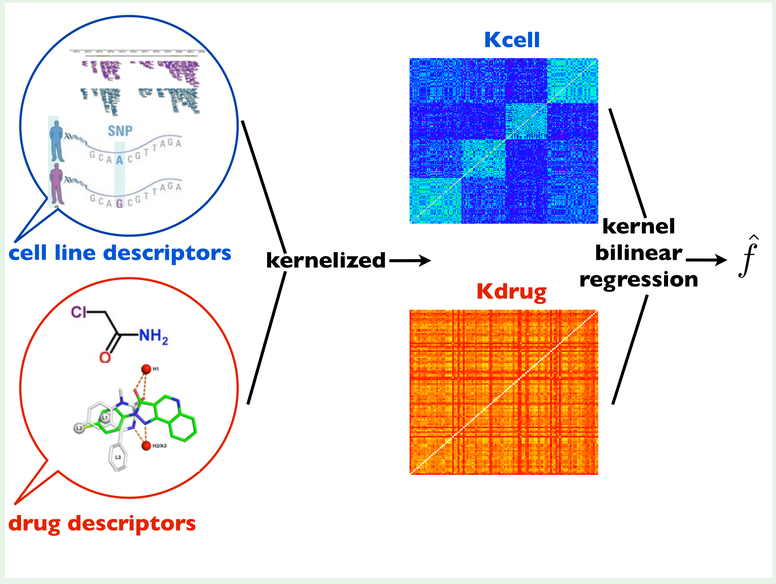
- Genotype data from the 1000 Genomes project (aiming at finding most genetics variants with frequency of 1%)
- RNA sequencing data from the Geuvadis Project (sequencing of some cell lines of the 1000 Genomes Project)
- Three covariates (sex, population and batch)

Toxicity $\longrightarrow$ The drug concentration that reduces the ATP synthesis of 10%.

Methods

To predict the toxicity values, we aim at

- Creating a measure of similarity between chemicals
- Creating a measure of similarity between genotypes



Let $\mathbf{x_i}$ be a vector of descriptors of cell line $\mathbf{i}$ and $\mathbf{y_j}$ be a vector of descriptors of chemical $\mathbf{j}$

- We model the toxicity $\mathbf{t_{ij}}$ of the chemical $\mathbf{j}$ on the cell line $\mathbf{i}$ by

$$\mathbf{t_{ij}} = \underbrace{\mathbf{f(x_i, y_j)}}_{\text{"bilinear"}} + \underbrace{\mathbf{b_j}}_{\text{offset}} + \underbrace{\epsilon_{\mathbf{ij}}}_{\text{noise}}$$

- We estimate $\mathbf{f}$ and $\mathbf{b}$ by penalized least-square regression

$$\min_{\mathbf{f}\in\mathcal{H},\mathbf{b}\in\mathbb{R}^p} \sum_{i=1}^{n}\sum_{j=1}^{p}\{\mathbf{t_{ij}} - \mathbf{f(x_i, y_j)} - \mathbf{b_j}\}^2 + \lambda\|\mathbf{f}\|^2_{\mathcal{H}}.$$

To specify $\|\mathbf{f}\|^2_{\mathcal{H}}$ we just have to choose two kernels :

$\longrightarrow$ $\mathbf{K_{cell}(x_l, x_{l'})}$ which measures the similarity between cell lines $\mathbf{x_l}$ and $\mathbf{x_{l'}}$.

$\longrightarrow$ $\mathbf{K_{drug}(y_m, y_{m'})}$ which measures the similarity between chemicals $\mathbf{y_m}$ and $\mathbf{y_{m'}}$.

- Then, the solution of the previous optimization problem is given by

$$\hat{\mathbf{f}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n}\sum_{j=1}^{p} \hat{\alpha}_{\mathbf{ij}}\mathbf{K_{cell}(x, x_i)K_{drug}(y, y_j)}$$

where $\hat{\alpha}_{\mathbf{ij}}$ depend only on

$\longrightarrow$ $\mathbf{K_{cell}} = \left(\mathbf{K_{cell}(x_l, x_{l'})}\right)_{\mathbf{l,l'=1...,n}}$

$\longrightarrow$ $\mathbf{K_{drug}} = \left(\mathbf{K_{drug}(y_m, y_{m'})}\right)_{\mathbf{m,m'=1,...,p}}$

$\longrightarrow$ The toxicity matrix.

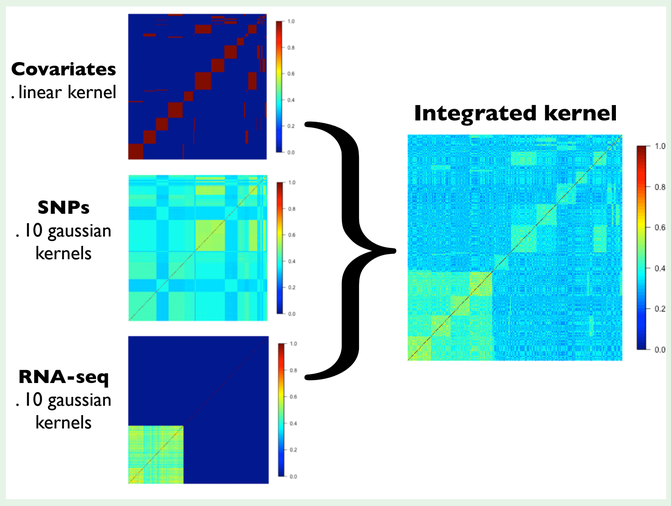
We also know that $\hat{\mathbf{b}}$ only depends on these three matrices.

References

1. B. Chem and al. *Journal of Chemical Information and Modeling* 49 :2044–2055, 2009.
2. Y. Yamanishi and al. *Journal of Chemical Information and Modeling* 51 :1183-1194, 2011.
3. P. Mahé and al. *Journal of Chemical Information and Modeling*, 45 :939–951, 2005.
4. T. Evgeniou and al. *Journal of Machine Learning Research* 6 :615-637, 2005.

What type of kernel do we choose ?

Kernel on cell lines :



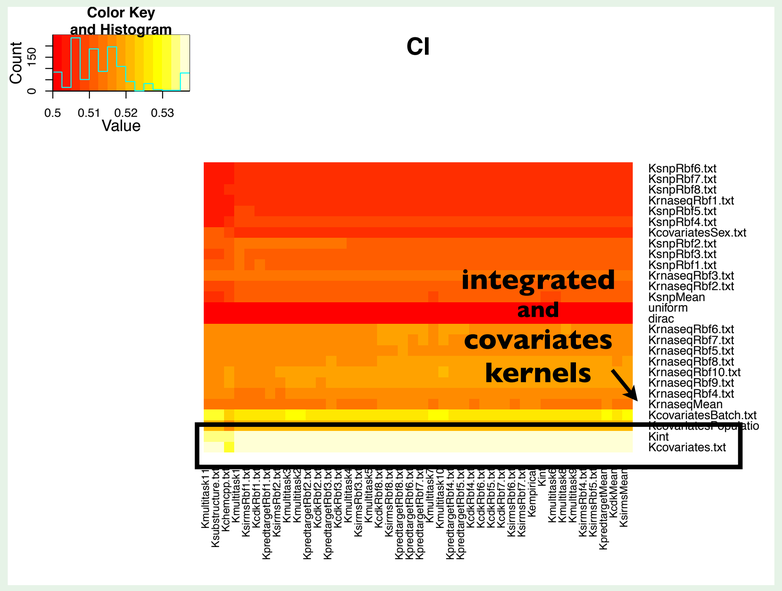
$\longrightarrow$ **29** kernels whose one that incorporate all informations (integrated kernel).

Kernel on chemicals :

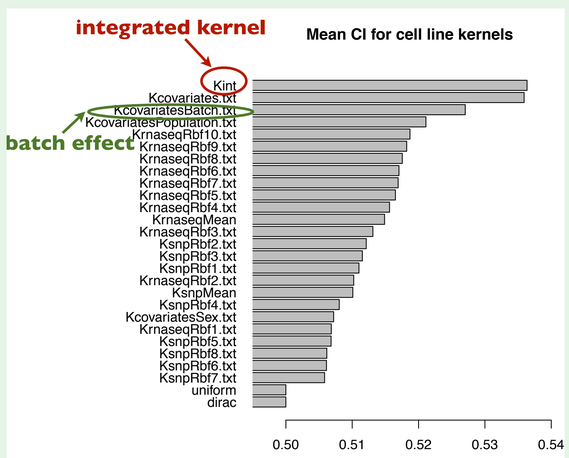
- Dirac kernel (build one model per chemical)
- Uniform kernel (build one single model)
- Kernel based on chemical features (using link to human protein, 2D substructures,...)
- Empirical kernel (evaluate the link between chemicals)
- An integrated kernel

Results

$\longrightarrow$ Evaluate the performance of each model by cross validation

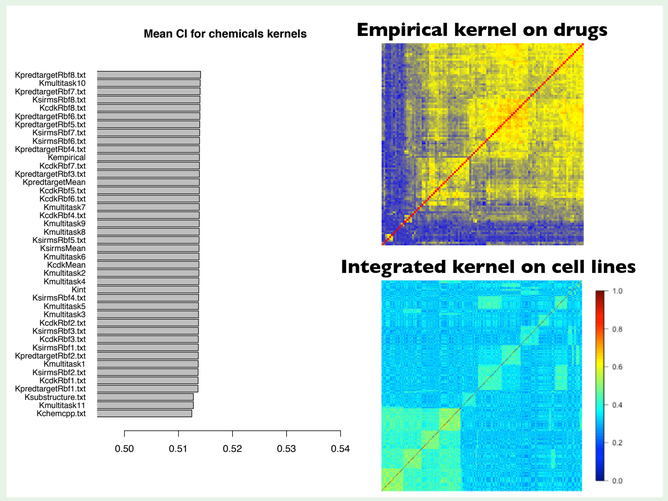


$\longrightarrow$ Integrated kernel and covariate kernel have the lowest concordance index.



- The integrated kernel has the highest prediction accuracy.
- The covariate kernel has a high prediction accuracy whereas it does not use any genomic information.
- This is due to a strong batch effect.

$\longrightarrow$ We choose the integrated kernel for cell lines.

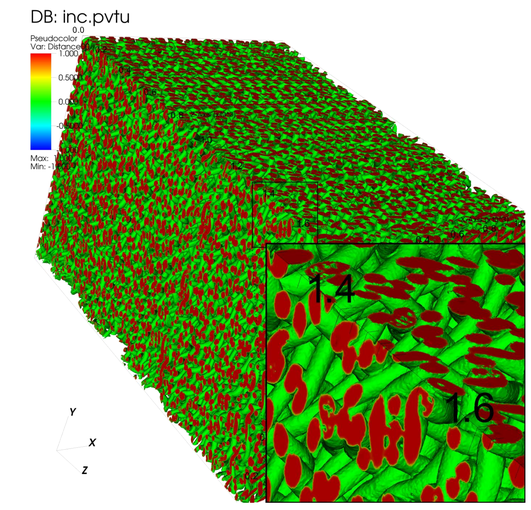


- All chemical kernels have the same accuracy : Dirac kernel performs as good as kernels using information on chemicals.
- Maybe the chemical descriptors do not fit this particular toxicogenetic problem.

$\longrightarrow$ We choose the empirical kernel for chemicals : since no kernel seems to be the best, we evaluate the link between chemicals instead of using some predetermined kernel.

$\longrightarrow$ This model ranked second out of 100 models submitted, in the DREAM challenge 8.0.

Simulation directe à partir de la datamasse des images 3D



1200x1200x1791
780 millions de nœuds
4000 cœurs de calcul

STRATEGIE GENERALE

Objectifs

- Construire des maillages de façon pertinente et efficace à partir d'imagerie 3D
- Compression de l'image, sans perte d'information, par manipulation au format maillage
- Utilisation de ces maillages dans différents domaines d'application

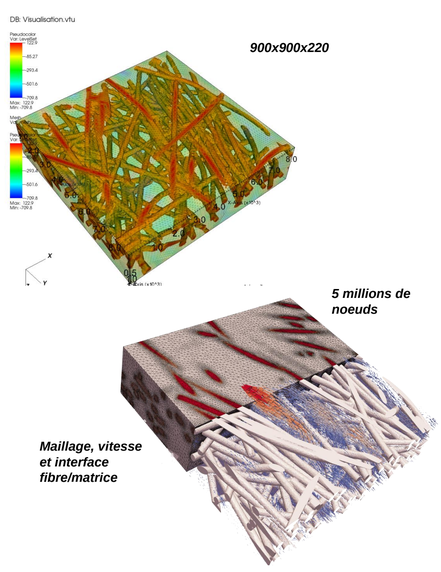
Résultats atteints / prototypes / démonstrateurs

- « Mailleur d'images »: génération avec un mailleur topologique et adaptation de maillage par minimisation de l'erreur d'interpolation de l'image sur le maillage, associée à une technique de réinitialisation de la valeur du voxel/pixel
- Au bout d'une année: génération de microstructures virtuelles, de géométries élancées en sous-sol, d'environnements urbains; calculs éléments finis et éléments frontière sur certaines applications
- Plateformes logicielles: CimLib, Morph-M, Neper, Zebulon, …

.

MATERIAUX COMPOSITES

Perméabilité d'un milieu fibreux



- Arrangement 3D irrégulier obtenu par imagerie 3D [Orgéas et al, 3S-R], caractéristiques renfort: R=0.1mm, L=10mm, ϕ=0.83
- Images acquises par microtomographie-X (900x900x220 voxels), taille du voxel = 10 μm3
- Génération du maillage éléments finis (~5 millions de nœuds): interpolation directe de l'image et adaptation, sur 96 cœurs
- Calcul d'écoulement sur le système fibre- matrice (sur 96 cœurs) et homogénéisation pour obtention de la perméabilité

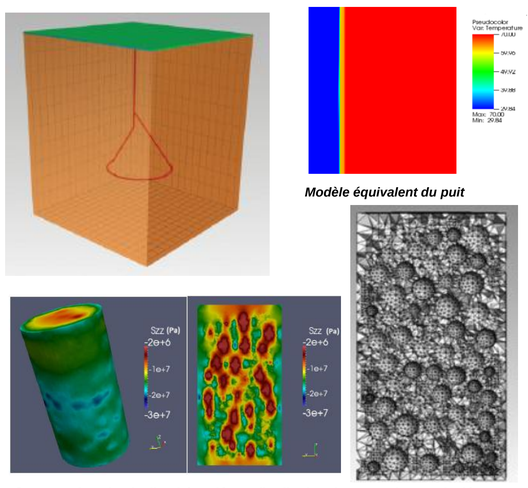
MICROSTRUCTURES METALLIQUES

Déformation d'un polycristal

- Image obtenue par tomographie de contraste à diffraction (DCT)
- Adaptation d'un algorithme de génération de cellules de Voronoï aux grains de l'image et maillage associé avec Neper
- Calculs de déformation sur l'échantillon avec Zebulon

Image obtenue par DCT

Maillage et cellules de Voronoï obtenues avec Neper



Compression simple d'un béton léger: distribution des contraintes calculées dans la méso-structure

GEOSCIENCES

Modélisation d'un échangeur géothermique

- Ecoulement d'eau dans un tube avec gainage (et couplage avec la température), validation CimLib/Fluent en 2D pour un modèle simple

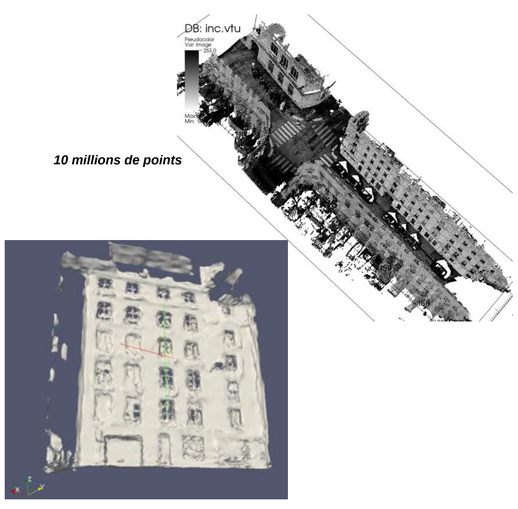
Modélisation de la de la méso-structure des bétons légers

- Compression d'un béton avec un calcul à l'échelle de la méso-structure pour obtenir le module de Young et proposer un mécanisme pour la rupture des bétons légers

MODELISATION DES ENVIRONMENTS URBAINS

Génération de maillages 3D à partir de nuages de points

- Rue de Paris, capture obtenue avec un Velodyne, 10 millions de points
- Compression de l'information par utilisation d'un maillage 3D surfacique, mais aussi 3D volumique (immersion dans une géométrie volumique)
- Applications: calculs sur ces environnements



Parties prenantes

MINES ParisTech
AGERE ECOLE DES MINES DE PARIS CENTRE DE ROBOTIQUE COOPERATIVE PERVASIVE
Cemef CENTRE DE MISE EN FORME DES MATERIAUX
CMM
GEOSCIENCES
CENTRE DES MATERIAUX PIERRE-MARIE FOURT
MINES Albi-Carmaux
Institut Clément Ader
MINES Alès
POLE MATERIAUX ET STRUCTURES DU GENIE CIVIL
MINES Douai
Technologie des Polymères et Composites Ingénierie Mécanique
SMS SCIENCES DES MATERIAUX ET DES STRUCTURES
Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Auteurs

Luisa Silva
Rima Ghazal
Daniela Craciun
Jia-Xin Zhao
Min Quan Thai
Sébastien Nadler
et al.

Partenaires

INSTITUT Mines-Télécom
INSTITUT CARNOT M.I.N.E.S
GENCI GRAND EQUIPEMENT NATIONAL POUR LE CALCUL INTENSIF
PRACE

Numérique : Grande échelle & Complexité

Mars 2014

Contact luisa.silva@mines-paristech.fr
Site web www.cemef.mines-paristech.fr

MoGDIW, an integrated workflow for cell motility genes discovery in high-throughput time-lapse screening data

[1,2,3]Alice Schoenauer Sebag, [2]Céline Raulet-Tomkiewicz, [2]Robert Barouki, [1]Jean-Philippe Vert, [1]Thomas Walter

[1]Center For Computational Biology, Mines ParisTech ; INSERM U900 ; Institut Curie

[2]INSERM U747 ; Paris V

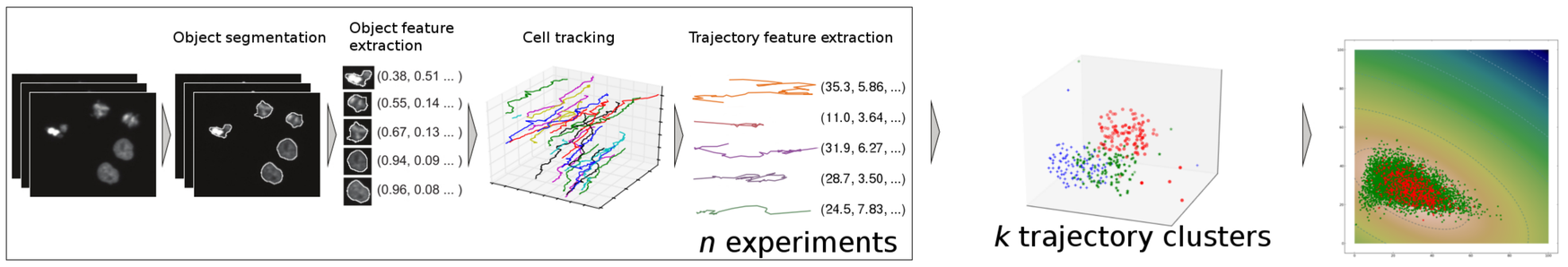
[3]Agro ParisTech

December 2013

Abstract : Cellular migration is a fundamental biological process. Progress in the fields of gene silencing and high-throughput (HT) microscopy has only recently made its study possible on a large scale. However, all existing HT migration screens measure motility at the level of cell population. Here, we present MoGDIW, a generic integrated workflow which addresses cell motility genes discovery in HT time-lapse screening data at single cell level. It is composed of cell tracking, cell trajectory mapping to an original feature space, migration pattern identification, and discriminant characterization of each experiment in terms of migratory behaviours. In comparison with an existing migration screen, MoGDIW application to a genome-wide time-lapse screen shows little overlap. However, its results are enriched in migration and adhesion-related (MAR) genes, and could be visually confirmed.

Motility **G**enes **D**iscovery **I**ntegrated **W**orkflow

Aim : quantitavely assess and compare single cell migration under different chemical perturbations in time-lapse microscopy data



Cell tracking

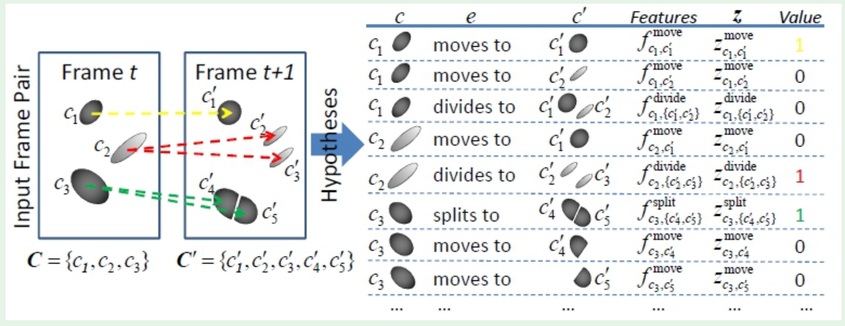
Challenges :

- High population density and high phenotypic variability
- Frequent absence of overlaps in consecutive frames (time interval : 30')
- Appearing, disappearing, merging and splitting objects
- Minimum use of prior knowledge
- *Usable for non-experts*

$\Longrightarrow$ Strategy : use of Machine Learning

Structured learning for cell tracking

- Idea (from [3]) :



- Prior knowledge : possible events **e** which can occur to an object between two frames

$$\{move,\ appear,\ disappear,\ split\ in\ 2\ or\ 3,\ merge\ at\ 2\ or\ 3\}$$

- Data : consecutive frames with already segmented objects
- Target : learn the model on annotated data and use it.
  - Model :

$$\hat{\mathbf{z}}(\mathbf{t}) = \arg\max_{\mathbf{z}}\ \mathbf{L}(\mathbf{x}(\mathbf{t}), \mathbf{z}; \mathbf{w}) = \sum_{\substack{\mathbf{e} \\ \mathbf{Obj_{i,t}} \\ \mathbf{Obj_{j,t+1}}}} < \mathbf{w^e}, \mathbf{f^e_{i,j}} > \mathbf{z^e_{i,j}}$$

$$\mathbf{st}\ \ \forall \mathbf{i} \in \{\mathbf{0}, \ldots, \mathbf{N(t)}\} \sum_{\substack{\mathbf{e} \\ \mathbf{Obj_{j,t+1}}}} \mathbf{z^e_{i,j}} = \mathbf{1}$$

$$\mathbf{and}\ \ \forall \mathbf{j} \in \{\mathbf{0}, \ldots, \mathbf{N(t+1)}\} \sum_{\substack{\mathbf{e} \\ \mathbf{Obj_{i,t}}}} \mathbf{z^e_{i,j}} = \mathbf{1}$$

  - Learning **w** : Support Vector Machine (SVM, algorithm : bundle method [6])

Results

- Training set : $\sim$ **32 000** links with 0.5% appear ; 0.5% disappear ; 1% merge ; 2% split

Algorithm (software)	Mean recall	Mean precision	Ref
Constrained nearest neighbour (Cell Cognition)	72.7%	62.8%	[1]
Linear assignement problem (Cell Profiler)	78.3%	73.0%	[2]
Structured learning	91.1%	91.5%	[3]

Providing a graphical user-interface for annotating videos

- Integration in Cell Cognition [1], an open source software platform for the analysis of live cell imaging data, with the IMBA, Vienna
- Extensions with regard to cell track annotation :
  - Generation of cell tracks rough estimation using a Nearest Neighbor tracker
  - GUI extension to support manual correction of the Nearest Neighbor trajectories

Experiment characterization in terms of migratory patterns

Trajectory mapping to an original feature space :

TABLE: Feature types and examples of corresponding features

Goal	Examples of feature
Track characterization	Diffusion coefficient, persistence, track evenness
Static quantification	Convex hull area, largest move, total path length
Dynamic quantification	Mean acceleration, mean instantaneous speed

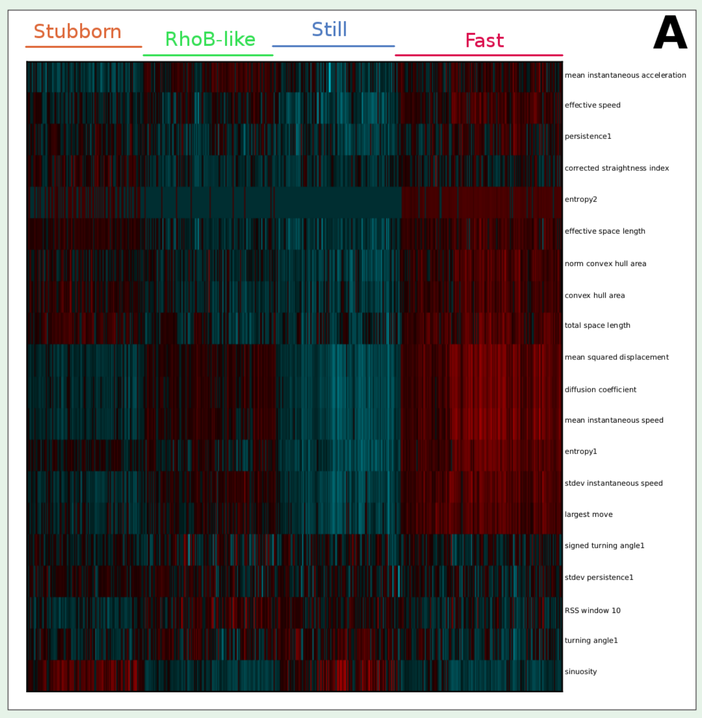
Experiment characterization :

1. Clustering of all trajectories across experiments
2. Characterization of an experiment by the cluster histogram of its trajectories
3. Pearson's χ^2 test for testing significant deviation from control histogram

Analysis of Mitocheck screen

- Data : MitoCheck data set [4], $\sim$ **200,000** videos of HeLa cells, produced by high-throughput live cell imaging, following selective down-regulation of all protein coding genes, one by one, by RNA interference (RNAi)
- Gene subset : 1,081 genes previously selected by [5] consisting of
  - phosphatases
  - kinases
  - migration and adhesion-related genes (MAR genes, a priori selected by the Geiger Laboratory, Weizmann Institute, Israël)
- Method : MoGDIW with K-means, k=4
- Results in comparison with [5] :
  - Small overlap of selected gene lists
  - No enrichement in known migration genes in either case
  - Enrichment of MoGDIW high-confidence list in MAR genes (61% vs 34% for [5])

FIGURE: A. Identification of MoGDIW cluster characteristics using single linkage hierarchical clustering ; B. Venn diagram comparing MoGDIW's and [5]'s high-confidence genes ; C. Trajectory examples



Perspectives

- Application to the whole Mitocheck dataset and biological validation of hit genes
- Application to newly generated Environmental Toxicology data to assess the consequences of chemical exposure on single cell migration

References

1. *Held et al.*, Nature Methods, 7(9) :747-54, 2010.
2. *Jaqaman, K. et al*, Nature Methods, 5(2008) :695-702.
3. *Lou et Hamprecht*, NIPS, 2011.
4. *Neumann et al*, Nature, 464 :721-727, 2010.
5. *Simpson et al*, Nat. Cell Biol., 10 :1027-1038, 2008.
6. *Tsochantaridis et al*, JMLR, 6 :1453-1484, 2005.

De l'amont vers l'aval

Modélisation et Simulation Numérique
Calcul Intensif et ses Applications

Adaptation de maillage anisotrope, calcul parallèle et capture d'interface

Applications au matériaux et à la mise en forme, mais pas seulement...

PARTENAIRES

Entre autres...

PUBLICATIONS

Quelques unes...

Digonnet, *App Math Model* (2000)
Bruchon, *Int Journ Num Meth Eng* (2009)
Hachem, *Journ of Comp Physics* (2010)
Coupez, *Journ of Comp Physics* (2011)
Ville, *Int Journ Num Meth Fluids* (2011)
Silva, *Int Journ Mat Form* (2012)
Bernacki, *Scripta Mat* (2011)
Tillier, *Int Journ Or Maxi Surg* (2012)
Carozzani, *Met Mat Trans* (2013)

REMERCIEMENTS

GENCI, PRACE, ANR

CONTACT

luisa.silva@mines-paristech.fr

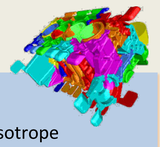
www.cemef.mines-paristech.fr

CEMEF MINES ParisTech
CS 10207
06904 Sophia Antipolis - France

CimLib

C++ et calcul scientifique

Caractéristiques phare



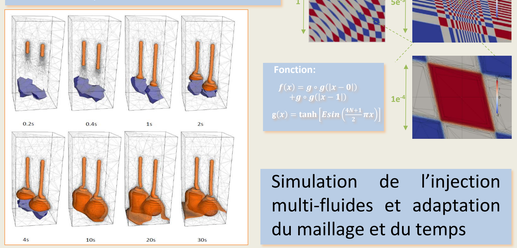
- une base éléments finis
- haut degré de parallélisme
- adaptation de maillage et du temps anisotrope
- méthodes de stabilisation performantes
- calcul de surfaces libres ou interfaces par des méthodes robustes
- validation automatique et mise à disposition sur un serveur de partage de projet
- utilisée pour le développement de logiciels industriels: Rem3D, Thost, Ximex, Forge, Thercast, Transweld

Adaptation anisotrope

Maillage et temps

Automatique, en parallèle et en dynamique

2h, 23 millions de nœuds, 10 itérations, 256 cœurs

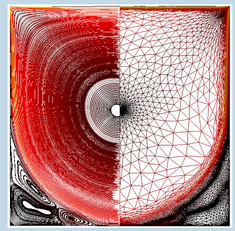


Simulation de l'injection multi-fluides et adaptation du maillage et du temps

Méthodes numériques avancées

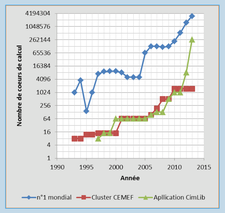
▶ Génération et adaptation de maillages anisotropes

- Adaptation anisotrope en dynamique et en parallèle (3D) basée sur une carte de métrique et sur des estimateurs d'erreurs anisotropes 3D sur le gradient des fonctions ou sur la distribution de la longueur des arêtes, avec contrôle du nombre d'éléments



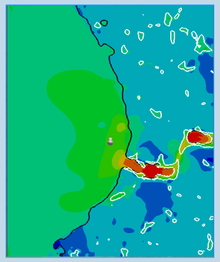
▶ Calculs massivement parallèles

- Haute performance des simulations grâce à l'exploitation efficace de la parallélisation de la gestion du maillage et de la résolution des systèmes linéaires
- Repartitionnement dynamique, solveurs itératifs parallèles et multigrilles
- Benchmarking et applications dans des supercalculateurs du Tiers1 (Genci) et Tiers0 (Prace)



▶ Ecoulements multiphasiques, interactions fluide-structure et interfaces mobiles

- Approche monolithique et méthode des volumes immergés
- Méthodes éléments finis stabilisées pour les écoulements du très petit (très visqueux) au très haut (peu visqueux) nombre de Reynolds, avec couplages thermiques et cinétiques
- Interactions thermomécaniques entre les phases, changement de phase
- Evolution des interfaces par des approches type level-set

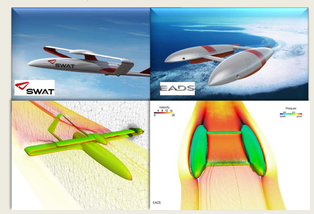


Aérodynamique et aérothermie

Ecoulements turbulents, transferts thermiques

Interactions avec les structures, fixes ou mobiles

Ecoulement d'air
Haut Re (1^{e8})

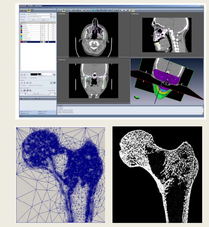


Drone léger à aile bouclée et dirigeable à 7000 m

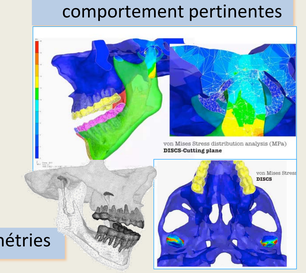
Simulation et le vivant

Biomécanique

Tissus mous ou structures osseuses



Calcul mécanique et lois de comportement pertinentes

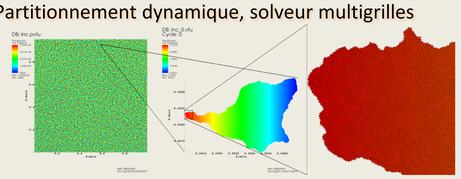


Reconstruction des géométries

Massivement parallèle

Supercalculateurs et supercalculs

Partitionnement dynamique, solveur multigrilles



Génération, adaptation de maillage

dim	# éléments (milliards)	# coeurs
2D	66,8	100 000
3D	81,4	100 000

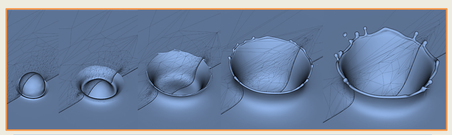
Résolution d'un très grand système linéaire

dim	# inconnues (milliards)	temps calcul	# coeurs
2D	100	319 s	262 144
3D	55	447 s	65 536

Ecoulement multiphasique

Interfaces implicites évolutives

Approche monolithique et méthode level-set



Evolutions complexes: chute d'une goutte de lait

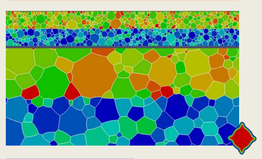
Injection assistée-eau: transport multi level-set

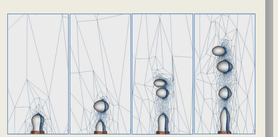
Physique numérique

Changements de phase

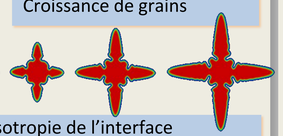
Germination et croissance

Ebullition
Croissance et ascension de bulles de vapeur





Recristallisation
Croissance de grains

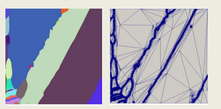


Solidification
Croissance dendritique et anisotropie de l'interface

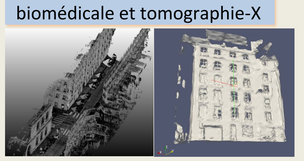
Simulation et le réel

Imagerie et Données Massives

3D volumique ou surfacique



Microstructures numériques, biomédicale et tomographie-X



Environnements urbains et nuages de points

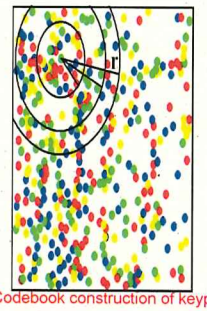
Visual textures as realizations of multivariate log-Gaussian Cox processes

Huu-Giao Nguyen, Ronan Fablet, Jean-Marc Boucher
Institut Telecom / Telecom Bretagne / LabSTICC (CS 83818 - 29238 Brest Cedex 3 – France)
{huu.nguyen, ronan.fablet, jm.boucher}@telecom-bretagne.eu

Abstract: In this paper, invariant texture characterization and recognition are addressed from the characterization of the spatial distribution of image. Visual keypoint sets in visual textures are here regarded as realizations of spatial point processes. We show that empirical second-order statistics considered in [4] relate to a non-parametric form of a log-Gaussian Cox model and investigate the relevance of parametric Cox models for texture recognition issues. Reported results validate the proposed descriptor compared to state-of-the-art approaches[1] with three datasets: UIUC, KTH-Tips, Brodatz.

Descriptive statistics of multivariate point processes

Local feature extraction



Codebook construction of keypoints[2]

A *spatial point process* **S** is defined as a locally finite random subset of a given bounded region **B** in $\mathbf{R}^2$. A realization of such a process is a spatial point pattern {s = $\{s_1,...,s_n\}$ of n points in **B**. Given realizations of apoint process, the moments of random variable are relevant descriptive statistics. The p^{th}-order moment is defined by:

$$\mu^{(p)}(B_1\times..\times B_n)=E\{N(B_1)..N(B_n)\}$$

For a marked spatial point process $\{(s_i,m_i)\}$ in *given bounded regions* B_i, where m_i is a mark associated to point s_i [3], the second-order moment is given by:

$$\alpha^2(B_1\times B_2)=E\sum_{s_1\in S}\sum_{(s_2\neq s_1)\in S}I_{B_1}(s_1)I_{B_2}(s_2)$$

Ripley's K function considers circular analyzing regions and resorts to the mean numbers of points of type **j** in a region of radius **r** centered at the points of type **i** :

$$K_{ij}(r)=(\lambda_i\lambda_j)^{-1}\alpha_{ij}^2(r)$$

$$=(\lambda_i\lambda_j)^{-1}E\sum_h\sum_{l\neq h}\delta_i(m_h)\delta_j(m_l)\mathbb{I}(\|s_h-s_l\|\leq r)$$

Log-Gaussian Cox processes

Cox processes $\{X_i\}$ with random intensity functions $\{Z_i\}$ are point processes such that $X_i|Z_i$ is a Poisson process, where $Z_i=\exp(Y_i)$. $\{Y_i\}$ is a multivariate Gaussian field on **S** characterized with mean function μ=EY(s) and covariance function $c_{ij}(r)=\mathrm{Cov}(Y_i(s_1),Y_j(s_2),r)$ [4].

• Intensity function: $\lambda_i=\exp(\mu_i+\sigma_i^2/2)$

• Pair correlation function: $g_{ij}(r)=\exp(\beta_{ij}(r))$

• Ripley's K function: $K_{ij}(R)=2\pi\int_0^R r\,g_{ij}(r)dr$

• The estimation of the pair correlation function is given by:

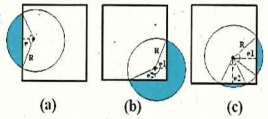
$$g_{ij}(r)=(2\pi r\lambda_i\lambda_j)^{-1}\sum_h\sum_{l\neq h}\delta_i(m_h)\delta_j(m_l)\xi(\|s_h-s_l\|,r)b_{s_h}$$

• Given a parameterization **L(β,r)**, namely Exponential, Hyperbolic or Cardinal sine, model parameters are estimated from the minimization of the following criterion:

$$\int_0^R\{\sigma_{ij}L(\beta_{ij},r)-c_{ij}(r)\}^2dr$$

The proposed descriptor is formed by: $(\lambda_i, \sigma_{ij}, \beta_{ij})$

Edge-effect corrections:



Feature dimension reduction: A codebook of keypoint pairs $u=M(s_i,s_j)$ from two categorized keypoint s_i,s_j is considered, such that:

$$g_u(r)=(2\pi r\lambda_u)^{-1}\sum_h\sum_{l\neq h}\delta_u(M(s_h,s_l))\xi(\|s_h-s_l\|,r)b_{s_h}$$

Scaling effects:The actual radius of image is estimated by a reference radius r_{ref} and scale factor φ_{ref} (the rate of average point densities per surface unit).

$$r_i=\frac{\varphi_{ref}}{\varphi_i}r_{ref}$$

Application to texture recognition

Parameter setting:

• Categories of visual keypoints k = {60,120,150}, pairs of keypoints k*=60.

• r_{ref}= min(w,h)/2 ln(x) where **w**(idth),**h**(ight) the size of image, x=(1:0.1:exp(1)].

• Covariance function **L(β,r)** : Gaussian function.

• Classifier: random forest.

Result:

	N_t	Gabor filter	Cooc matrix	BoK[4]	Keypoint cooc.[15]	Multi-Fractal [27]	Keypoint combination[28]	Descriptive stat.[21]	log-Gaussian Cox
UIUC	1	31.22±3.14	45.33±3.03	67.25±2.75	37.62±2.63	61.14±2.90	72.53±2.45	**76.66±1.65**	75.21±1.75
	10	57.37±1.93	70.67±1.72	8[illegible]2±1.45	54.14±1.72	83.68±1.65	93.7±1.15	94.23±0.78	**95.42±0.71**
	20	67.78±1.23	80.12±1.30	9[illegible]9±1.15	91.67±1.38	93.65±1.31	93.37±0.95	97.34±0.25	**97.84±0.32**
Brodatz	1	78.52±1.72	75.42±1.73	83[illegible]4±1.50	54.33±1.63	85.95±0.91	83.35±1.35	87.67±0.81	**88.81±0.92**
	3	85[illegible]4±1.41	83.22±1.04	92.73±0.91	93.17±0.87	92.41±0.73	94.34±0.45	95.67±0.33	**96.14±0.41**
KTH-Tips	5	62.32±2.41	62.83±2.42	64.42±2.81	71.67±2.63	72.63±2.45	78.7±2.35	**81.34±1.53**	80.17±2.15
	20	82.65±1.39	80.45±1.67	8[illegible]2±1.45	82.47±1.48	87.18±1.53	90.23±1.51	92.15±1.26	**92.42±1.11**
	40	89.87±0.88	88.95±0.77	90[illegible]9±0.65	91.15±1.05	91.33±0.97	94.35±0.87	95.05±0.41	**95.74±0.45**

Classification rates and standard deviations of proposed descriptor over 50 random selections compared to state-of-the-art approaches.

N_t	DoG+Sift	FH+Surf	FH+Brief	(Hes-Lap)+Daisy	(Har-Lap)+(Sift-Spin)
1	75.21±1.75	75.05±1.54	**75.43±1.71**	75.13±1.69	74.85±1.87
5	91.95±1.13	93.73±1.11	91.42±1.23	**92.13±1.19**	91.15±1.4[illegible]
10	95.42±0.71	95.15±0.91	95.22±0.85	**95.47±1.08**	95.23±[illegible].72
15	**96.87±0.65**	95.14±0.63	96.43±0.51	96.75±0.53	96.37±[illegible].6[illegible]
20	**97.84±0.32**	95.75±0.41	97.25±0.34	97.67±0.35	97.14±[illegible].37

Comparison performance of proposed model with the different detector-descriptor types on UIUC dataset.

References:

[1].J.Zhang et al. "Local features and kernels for classification of texture and object categories: a comprehensive study". IJCV, 73(2), p.213–238, 2007.
[2].G.Csurka et al. "Visual categorization with bags of keypoints". ECCV, p.1–22, 2004.
[3] H-G.Nguyen et al. "Spatial statistics of visual keypoints for texture recognition." ECCV, pp. 764–777,2010.
[4].J.Møller et al.. "Log-Gaussian cox processes," SJS, 25(3), p. 451–482, 1998.

RANDOM WALK MODELS FOR GEOMETRY-DRIVEN IMAGE SUPER-RESOLUTION

APPLICATION TO REMOTELY SENSED GEOPHYSICAL FIELDS AT OCEAN SURFACE

Ronan Fablet
Professor, Institut Mines-Télécom/Télécom Bretagne
Brest, France

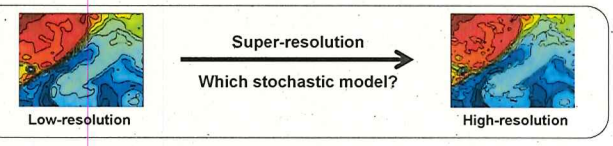
Co-authors

Brahim Boussidi
Institut Mines-Télécom/Télécom Bretagne
LabSTICC/TOMS
Brest, France

Emannuelle Autret, Bertrand Chapron
Ifremer/LOS (Space oceanography)
CERSAT
Brest, France

ABSTRACT

This paper addresses stochastic geometry-driven image models and their application to super-resolution issues for textured geophysical fields. Whereas most stochastic image models rely on some priors on the distribution of grey-level configurations (e.g., patch-based models, Markov priors, multiplicative cascades,...), we here focus on geometric priors. Regarding image level-lines as realizations of 2D random walks, we introduce a stochastic geometry-driven model for 2D textures and consider an application to image super-resolution. The targeted application is the stochastic interpolation of missing data in muti-sensor sea surface observation.



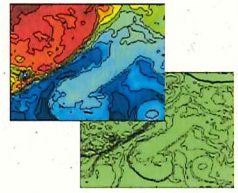


Image level-lines as realizations of 2D random walks

Why level-lines?

Image level-lines provide a contrast-invariant image representation [5] and fully characterize the geometry of an image.

Correlated random walk model [1]:

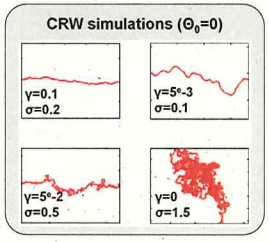
$$d\theta(s) = -\gamma\,(\theta(s) - \theta_0) + \sigma dW(s) \quad (1)$$

Directional drift — Brownian process

Parameters γ and σ control the regularity of the random walk, in terms of regularity along the walk and of oscillation around the directional drift. The CRW model is associated with a Fokker-Plank representation, from which one can derive the stationary statistics of the random direction θ and turning angle δθ [1].

$$p(\theta) \propto exp\left(-\frac{\gamma}{\sigma^2}(\theta - \theta_0)^2\right)$$

$$p(\delta\theta) \propto exp\left(-\alpha\frac{\gamma}{\sigma}\delta\theta^2\right) \quad \text{with} \quad \alpha = \Delta s\,(1 - exp(-\gamma\Delta s))$$



Stochastic geometry-driven texture model

Stochastic geometry field model as a generalization of (1) for orientation fields:

$$\nabla_\theta\theta = -\gamma\,(\theta - \theta_0) + \sigma\nabla_\theta W \quad (2)$$

Reference orientation — 2D Brownian sheet

Where ∇_θ is the gradient in direction θ

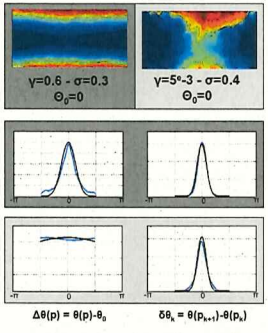


Image I such that its level-lines are everywhere tangent to random vector field u_θ:

$$\langle \nabla I(p), u_\theta(p) \rangle = 0, \forall p \quad (3)$$

Implementation:

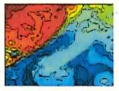
1)Simulate an orientation field from a numerical integration of the stochastic equation,

2)Solve for (3) as a variational minimization (cf. [3]).

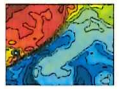
Application to texture-based super-resolution of geophysical fields

Problem statement

I_{HR}: High-resolution image
NxM grid



I_{LR}: Low-resolution image
N/KxM/K grid (here, K=2⁴)



Stochastic super-resolution model

Given a low-resolution image I_{LR}, sample a high-resolution image such that:

$$\begin{cases} \nabla_\theta\theta(p) = -\gamma(p)\,(\theta(p) - \theta_{LR}(p))\,dp + \sigma(p)\nabla_\theta W(p) \\ \tilde{I} = \arg\min_I \int \| \langle \nabla I(p), u_\theta(p) \rangle \| dp \quad \text{subject to } I_{LR} = \mathcal{P}\left[\tilde{I}\right] \end{cases}$$

with W a Brownian sheet and θ_{LR} the orientation field of the low-resolution image (i.e., the angle of the local tangent to the level-lines). $\mathcal{P}$ is an orthogonal projection operator (here a wavelet operator) such that:

$$I_{LR} = \mathcal{P}\,(I_{HR}) \quad \text{and} \quad \mathcal{P}\,[I_{HR} - \mathcal{P}(I_{HR})]$$

The low-resolution gradient drives the high-resolution geometrical variability

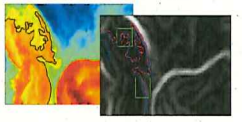
Parameter fields γ and σ are set according to the following observation :

- Large gradients in I_{LR} result in more regular level-lines in I_{HR};
- Conversely, weak gradients in I_{LR} involve regular level-lines in I_{HR}.

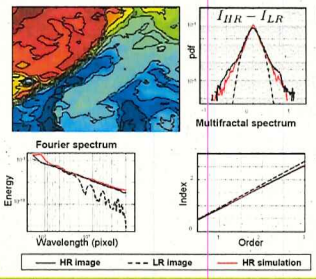
$$\gamma(p) = \gamma_0 \|\nabla I_{LR}(p)\|^\nu \quad \sigma(p) = \sigma_0 \|\nabla I_{LR}(p)\|^{-\beta}$$

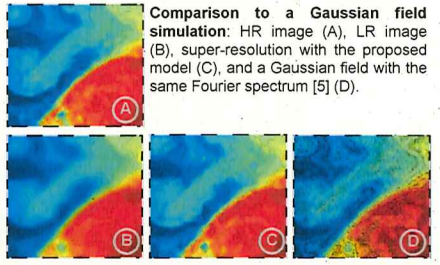
(Here, we set empirically γ_0=2.14, ν=0.54, σ_0=0.14, β= 0.13)

Level-lines vs. $\|\nabla I_{LR}\|$



Simulation for sea surface temperature observations





Comparison to a Gaussian field simulation: HR image (A), LR image (B), super-resolution with the proposed model (C), and a Gaussian field with the same Fourier spectrum [5] (D).

References

[1] Codling, E.A., Plank, M.J., Benhamou, S. Random walk models in biology. J. R. Soc. Interface, 5:813-834, 2008.

[2] Fablet, R., Pujolle, S., Chessel, A., Benzinou, A., Cao F. 2D Image-based reconstruction of shape deformation of biological structures using a level-set representation. CVIU, 111(3):295-306, 2008.

[3] Freeman, W.T., Liu, C. Markov Random Fields for Super-resolution and Texture Synthesis. In A. Blake, P. Kohli, and C. Rother, eds., Advances in Markov Random Fields for Vision and Image Processing, Chapter 10. MIT Press, 2011.

[4] Galerne, B., Gousseau, Y., Morel., J., 2011. Random phase tex- tures : Theory and synthesis. IEEE Transactions on Image Processing. 20(1), 257–267.

[5] Monasse, P., Guichard, F. Fast Computation of a Contrast Invariant Image Representation. IEEE TIP, 9:860- 872, 2000.

Contact ronan.fabet@telecom-bretagne.eu Webpage: perso.telecom-bretagne.eu/ronanfablet

New Decoding Strategy For Underdetermined MIMO Transmission Using Sparse Decomposition

Yasser Fadlallah, Abdeldjalil Aïssa-El-Bey, Karine Amis, Dominique Pastor and Ramesh Pyndiah
Telecom Bretagne, Brest, France

Parties prenantes

Auteurs

Yasser Fadlallah

Abdeldjalil Aïssa-El-Bey

Karine Amis

Dominique Pastor

Ramesh Pyndiah

Partenaires

Introduction

- Maximum Likelihood joint detection enables to detect at once the symbols transmitted in the same time intervals.
- ML detector selects the closest point to the received signal in the receive constellation.
- ML is optimal for medium to high SNR values.
- Computation cost increases exponentially with the signal dimension.
- Alternative solution such as sphere decoder keeps an exponential increase of the computation cost.
- **Goal:** find out an efficient detection problem of the transmitted symbols with moderate computation cost.

System Model

- MIMO flat fading channel.
- Perfect knowledge of the channel matrix at the receiver.
- Transmit symbols belong to a finite alphabet constellation.
- Received signal is defined as

Sparse Decomposition

- Let $Q=\{q_1,...,q_M\}$ the finite alphabet transmit constellation. Let $\mathbf{q}=[q_1,...,q_M]$ the vector space in which the finite alphabet vector can be cast, and $\mathbf{B}_q$ the decomposition matrix defined as

$$\mathbf{B}_q=\begin{pmatrix}\mathbf{q} & \mathbf{0}_M & \cdots & \mathbf{0}_M\\ \mathbf{0}_M & \mathbf{q} & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}_M\\ \mathbf{0}_M & \cdots & \mathbf{0}_M & \mathbf{q}\end{pmatrix}$$

- The transmitted symbol vector $\mathbf{x}$ can be rewritten after symbol decomposition on $\mathbf{q}$ as : $\mathbf{x}=\mathbf{B}_q\,\mathbf{s}$
- The received signal is reformulated as: $\mathbf{y}=\mathbf{H}\mathbf{B}_q\,\mathbf{s}+\mathbf{z}.$

New MIMO Detector based on ℓ_1-norm minimization

- The new problem is the decoding of the binary source vector $\mathbf{S}$. To this end, we propose to solve it using the following minimization problem

$$\arg\min_{s\in R^{NM\times 1}}\|\mathbf{s}\|_0,\quad \text{subject to}\ \mathbf{s}\in\left\{\left\|\mathbf{y}-\mathbf{H}\mathbf{B}_q\mathbf{s}\right\|<\varepsilon,\ \text{and}\ \mathbf{B}_1\mathbf{s}=1_N\right\}$$

where ε is a constant defined later, and $\mathbf{B}_1=\begin{pmatrix}1_M & \mathbf{0}_M & \cdots & \mathbf{0}_M\\ \mathbf{0}_M & 1_M & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}_M\\ \mathbf{0}_M & \cdots & \mathbf{0}_M & 1_M\end{pmatrix}$

- In the literature of sparse reconstruction, the ℓ_0-norm can be relaxed by the ℓ_1 -norm minimization, and the problem is resolved using

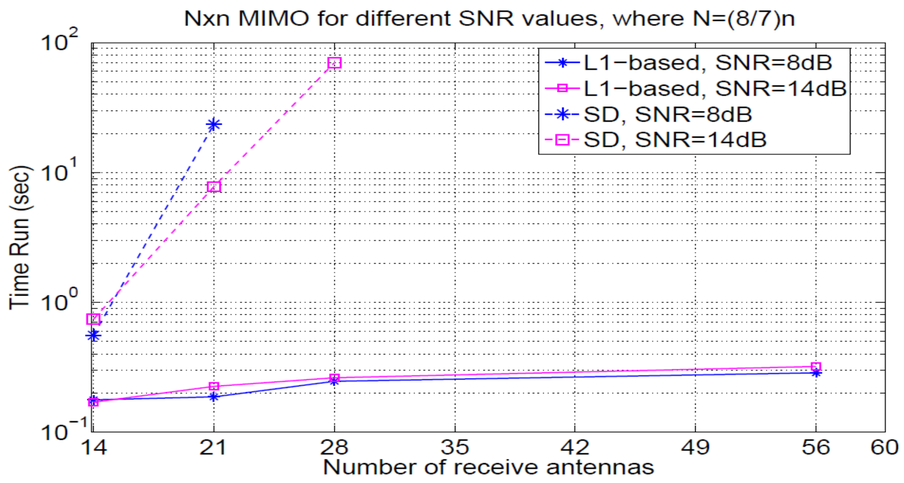
$$\arg\min_{\mathbf{s}\in R^{NM\times 1}}\|\mathbf{s}\|_1,\quad \text{subject to}\ \mathbf{s}\in\left\{\left\|\mathbf{y}-\mathbf{H}\mathbf{B}_q\mathbf{s}\right\|<\varepsilon,\ \text{and}\ \mathbf{B}_1\mathbf{s}=1_N\right\}$$

Applications

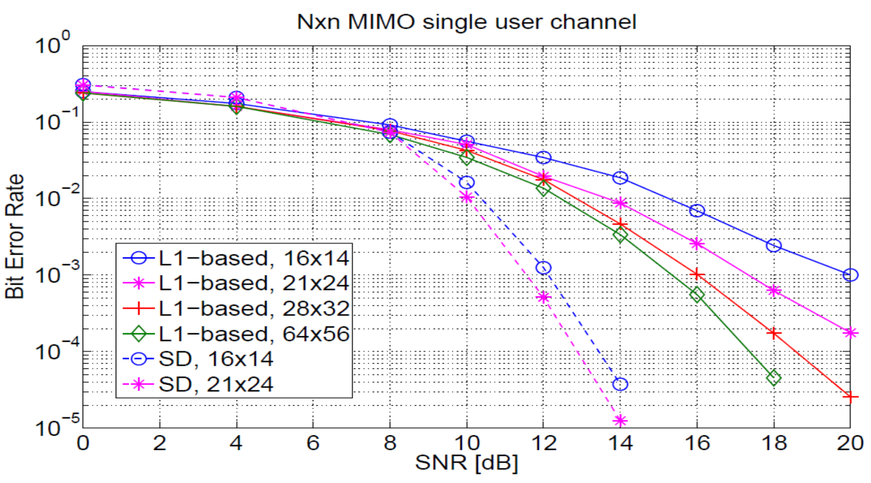
- Large MIMO systems
  - In a noiseless channel, the equivalence between the ℓ_0-norm and the ℓ_1-norm hold for large dimensions of ***s***.
- MIMO frequency selective channel
  - The received signal can be written as

$$\begin{pmatrix}\mathbf{y}(1)\\ \vdots\\ \mathbf{y}(T_f+L)\end{pmatrix}=\begin{pmatrix}\mathbf{H}_0^T & \cdots & \mathbf{H}_L^T & \mathbf{0}^T_{n\times N}\cdots\mathbf{0}^T_{n\times N}\\ \mathbf{0}^T_{n\times N} & \ddots & \ddots & \ddots\ddots\vdots\\ \vdots & \ddots & \ddots & \ddots\ddots\vdots\\ \mathbf{0}^T_{n\times N} & \cdots & \mathbf{0}^T_{n\times N} & \mathbf{H}_0^T\cdots\mathbf{H}_L^T\end{pmatrix}\begin{pmatrix}\mathbf{x}(1)\\ \vdots\\ \mathbf{x}(T_f)\end{pmatrix}+\begin{pmatrix}\mathbf{z}(1)\\ \vdots\\ \mathbf{z}(T_f+L)\end{pmatrix}$$

Simulations Results



- We assume 4-QAM modulation.
- The computational complexity keeps almost invariant with the system dimensions and the SNR level, whereas the SD time-run increases exponentially with these two factors



Conclusions

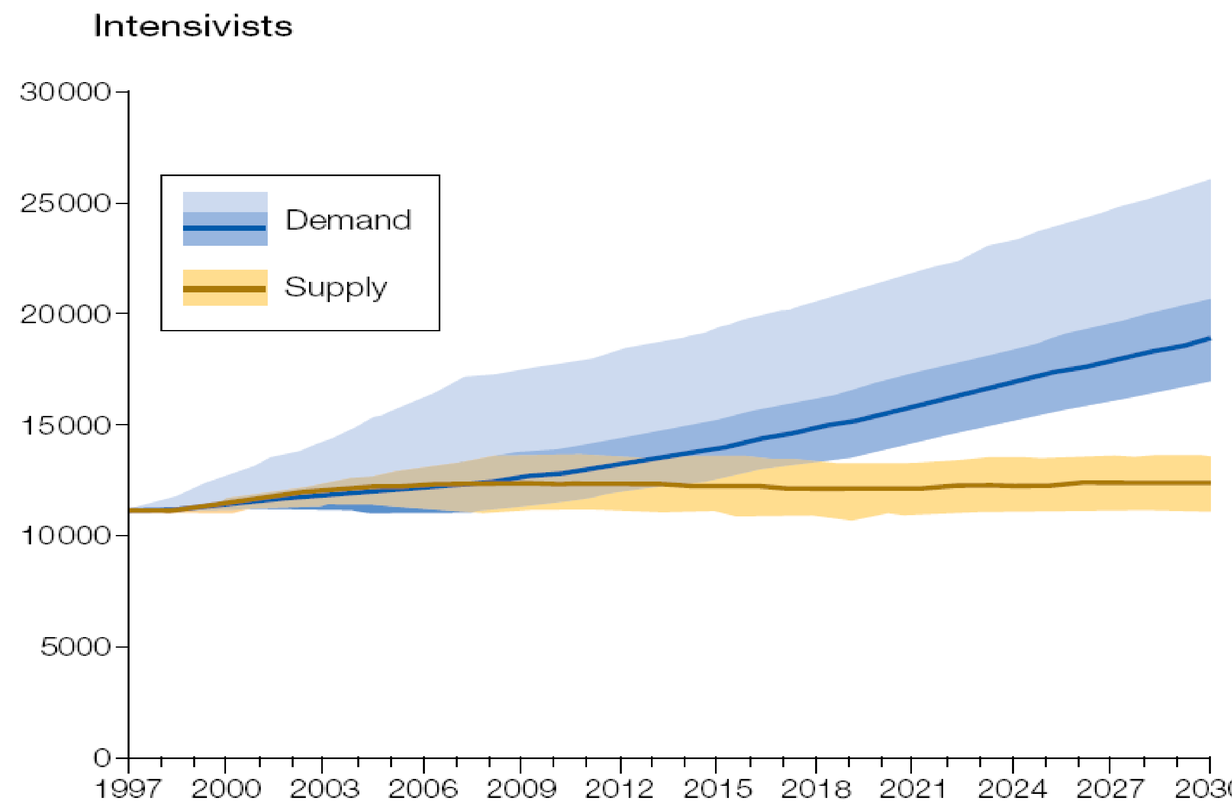
- Proposition of a new detection method for determined and underdetermined MIMO systems, based on sparse decomposition of the signal belonging to a finite constellation.
- The proposed detection method is solvable in polynomial time, and uses iterative algorithm such as primal dual interior point method.

Colloque « Numérique : Grande échelle & Complexité »
26-27 Mars 2014

Contact abdeldjalil.aissaelbey@telecom-bretagne.eu Site web www.telecom-bretagne.eu

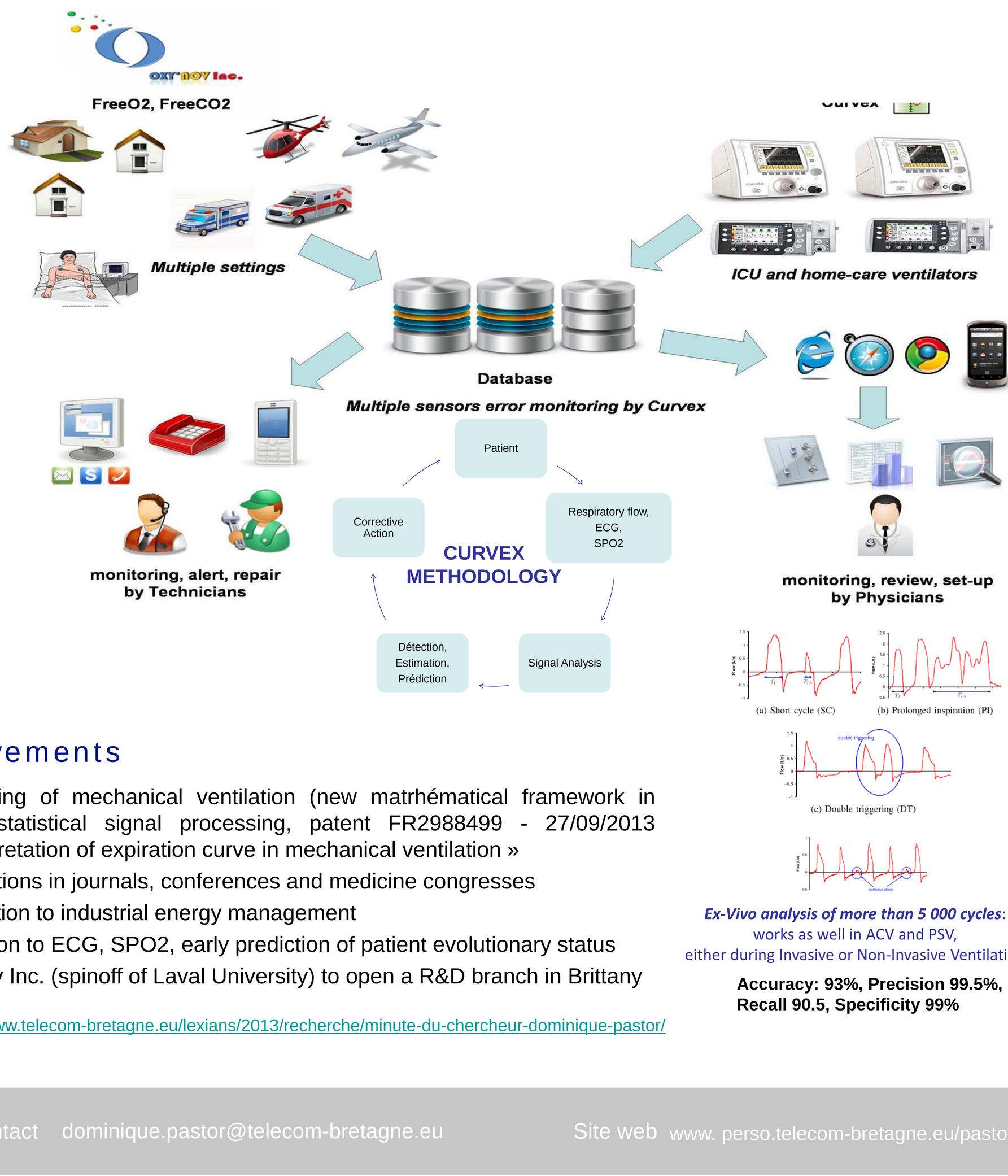
Plateforme CURVEX de surveillance des signaux physiologiques en médecine d'urgence et réanimation

Hospital System staffing under pressure



Source: Angus JAMA 2000

The CURVEX solution



Auteurs


Q.-T. Nguyen (TB/Lab-STICC)


S. Cherif (TB/Lab-STICC)


D. Pastor (TB/Lab-STICC)


E. L'Her (UBO/CHU/LATIM)

Partenaires

Achievements

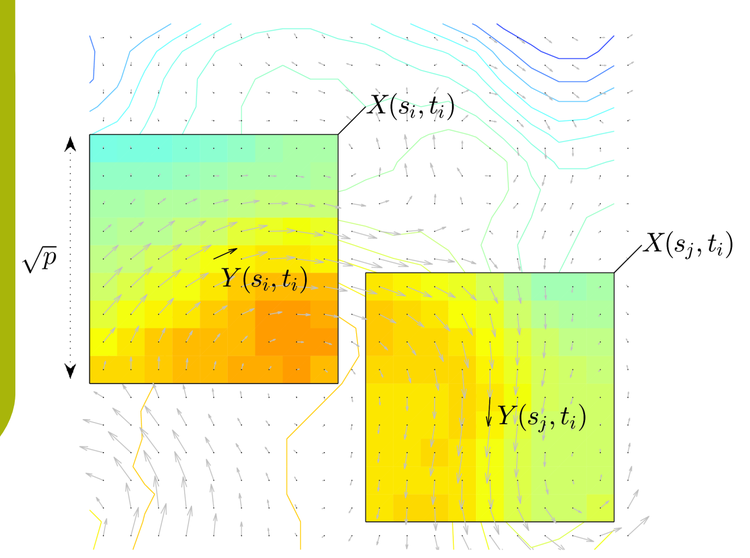
- Monitoring of mechanical ventilation (new matrhématical framework in robust statistical signal processing, patent FR2988499 - 27/09/2013 « interpretation of expiration curve in mechanical ventilation »
- Publications in journals, conferences and medicine congresses
- Application to industrial energy management
- Extension to ECG, SPO2, early prediction of patient evolutionary status
- Oxy'nov Inc. (spinoff of Laval University) to open a R&D branch in Brittany

http://www.telecom-bretagne.eu/lexians/2013/recherche/minute-du-chercheur-dominique-pastor/

Ex-Vivo analysis of more than 5 000 cycles:
works as well in ACV and PSV,
either during Invasive or Non-Invasive Ventilation

Accuracy: 93%, Precision 99.5%, Recall 90.5, Specificity 99%

Une approche statistique pour la caractérisation et le suivi des dynamiques superficielles des océans à partir d'images satellitaires



Auteurs

Pierre Tandeo
Sileye Ba
Ronan Fablet
Institut Mines-Telecom
Telecom Bretagne
LabSTICC – TOMS
Brest, France

Bertrand Chapron
Emmanuelle Autret
Ifremer
Laboratoire d'océanographie spatiale
CERSAT
Brest, France

Partenaires

Ifremer

Lab-STICC

Remerciements

AVISO

Remote Sensing Systems
www.remss.com

Références

[1] Isern-Fontanet et al. Potential use of microwave sea surface temperatures for the estimation of ocean currents, GRL, vol. 33, pp. L24608, 2006

[2] Lapeyre et al. Dynamics of the upper oceanic layers in terms of surface quasigeostrophy theory, JPO, vol. 36, pp. 165--176, 2006

[3] De Sarbo et al. A maximum likelihood methodology for clusterwise linear regression, Journal of Classification, vol. 5, pp. 249--282, 1988

Résumé

Les mesures satellitaires de courants (U,V) et de température de l'eau de surface (SST), fournissent une information sur les dynamiques de l'océan. Certaines études (cf. **[1]** et **[2]**) ont montré que les champs de température peuvent être considérés, dans certaines situations, comme des traceurs actifs de la dynamique de surface. Dans ce cas, de fortes corrélations existent entre les variations locales de SST et les courants (U,V). Existe-t-il d'autres relations entre la température et les courants ? Quand et où la SST peut-elle être considérée comme un traceur actif ou passif ? Dans cette étude, nous mettons en place une méthode statistique et explorons un historique d'observations satellitaires pour identifier et suivre des modes dynamiques cachés.

Méthode

- K fonctions de transfert cachées entre :
  - **Y** → courant (U,V) en un point
  - **X** → température SST au voisinage (patch)

$$\mathbf{Y}(s_i,t_i) = \sum_{k=1}^{K} \mathbf{H}_k(\mathbf{X}(s_i,t_i))$$

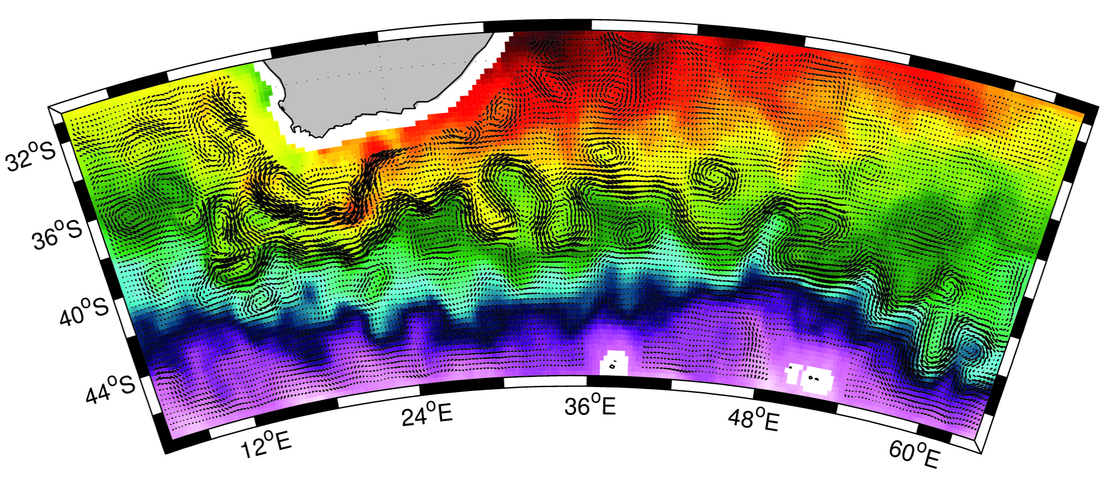
- Identification des K fonctions de transfert à partir d'un modèle de régressions linéaires latentes (cf. **[3]**) :

$$p\left(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta}\right) = \sum_{k=1}^{K} \lambda_k \mathcal{N}_k\left(\mathbf{Y};\mathbf{X}\boldsymbol{\beta}_k,\boldsymbol{\Sigma}_k\right)$$

- Estimation des paramètres par l'algorithme EM
- Suivi des modes dynamiques à partir des cartes de probabilités *a posteriori*

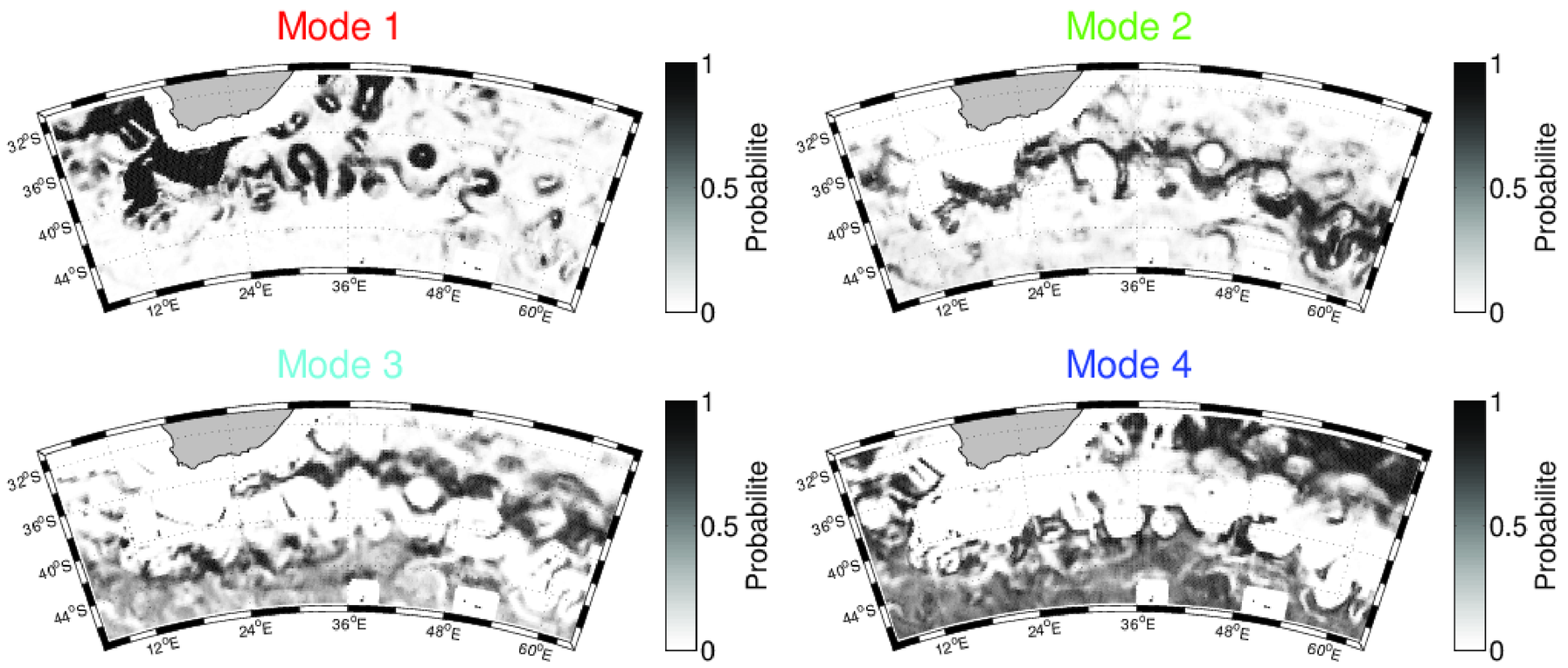
Données

- Température SST :
  Produit interpolé RSS
  Jounalier au 1/4°
- Courant (U,V) :
  Produit interpolé AVISO
  Journalière au 1/3°



Résultats

	Mode 1	Mode 2	Mode 3	Mode 4
Courant (intensité)	fort (principal Aiguilles + tourbillons)	fort (segondaire Aiguilles)	faible	faible
Courant (sens)	Nord-Sud	Ouest-Est	Est-Ouest	Ouest-Est
Température	élevée	élevée	moyenne	froide
SQG-like		√		√



Conclusion

La méthode proposée permet un apprentissage à l'aveugle, sans *a priori* physique, de fonctions de transfert cachées entre la SST et (U,V). Ces relations cachées correspondent à différents modes dynamiques dont certains s'apparentent à la théorie SQG. A partir du calcul des probabilités *a posteriori*, nous pouvons suivre l'évolution spatio-temporelle de ces modes dynamiques.

Perspectives

L'utilisation d'autres traceurs actifs tels que la salinité (SSH) ou la couleur de l'eau (Chl-a) ainsi que des données à haute résolution spatiale permettrait un rafinement du modèle. De plus, le suivi des probabilités *a posteriori* par un modèle stochastique et l'utilisation des K fonctions de transfert permet d'envisager une estimation des courants de surface à partir de la température.

GRETSI
Septembre 2013

Contact pierre.tandeo@telecom-bretagne.eu

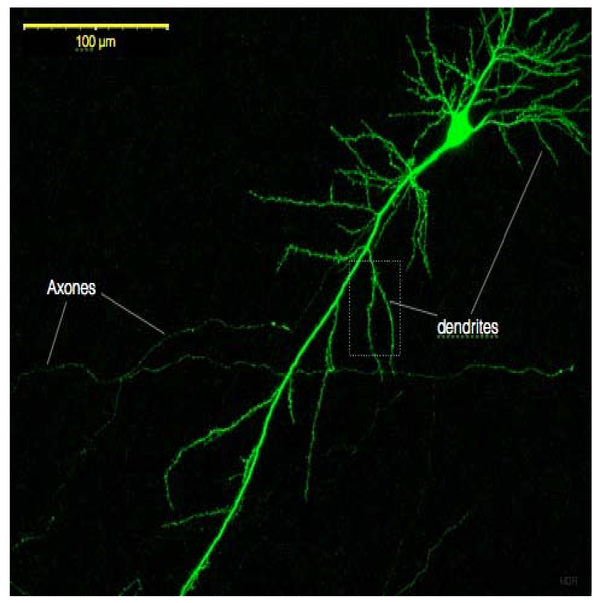
Site web http://tandeo.wordpress.com

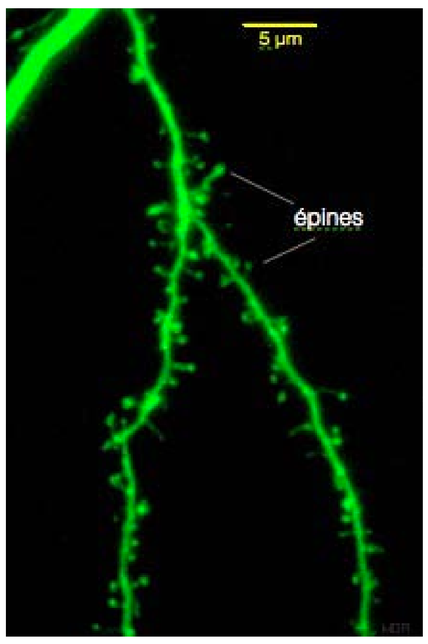
Nouveau modèle de neurone formel biologiquement plausible

DE NOUVEAUX RÔLES POUR UNE STRUCTURE ANCIENNEMENT CONNUE

Les clusters d'épines discrètes agissent comme des unités de calcul

- Les épines dendritiques ont été identifiées anatomiquement par Ramon Y. Cajal en 1911 qui les a qualifiées de "*espinas*" en raison de leur ressemblance avec des épines sur les tiges des fleurs.
- L'idée que les dendrites ne sont que des câbles passifs qui relaient les signaux entrants sur le corps de la cellule ne tient plus.
- Des études récentes révèlent que les sections dendritiques contenant des clusters d'épines discrètes agissent comme unités de calcul (Blom H, Rönnlund D, Scott L, Westin L, Widengren J, et al., *Nature*, 2013) .
- Cette clusterisation est influencée par les entrées sensorielles (Frost N. A., Shroff H., Kong H., Betzig E., Blanpied T. A., *Neuron*, Vol. 67, Issue 1, 15 July 2010).





Plasticité structurale et apprentissage

- Les épines dendritiques subissent en réalité d'une part des changements de leur forme et d'autre part un turn-over permanent (elles apparaissent et disparaissent)
- Lamprecht and Le Doux (2004) proposent une revue des mécanismes de plasticité structurelle associées aux épines dendritiques du neurone post-synaptique. Ces mécanismes sont généralement associées au renforcement et à la stabilisation de l'apprentissage synaptique.

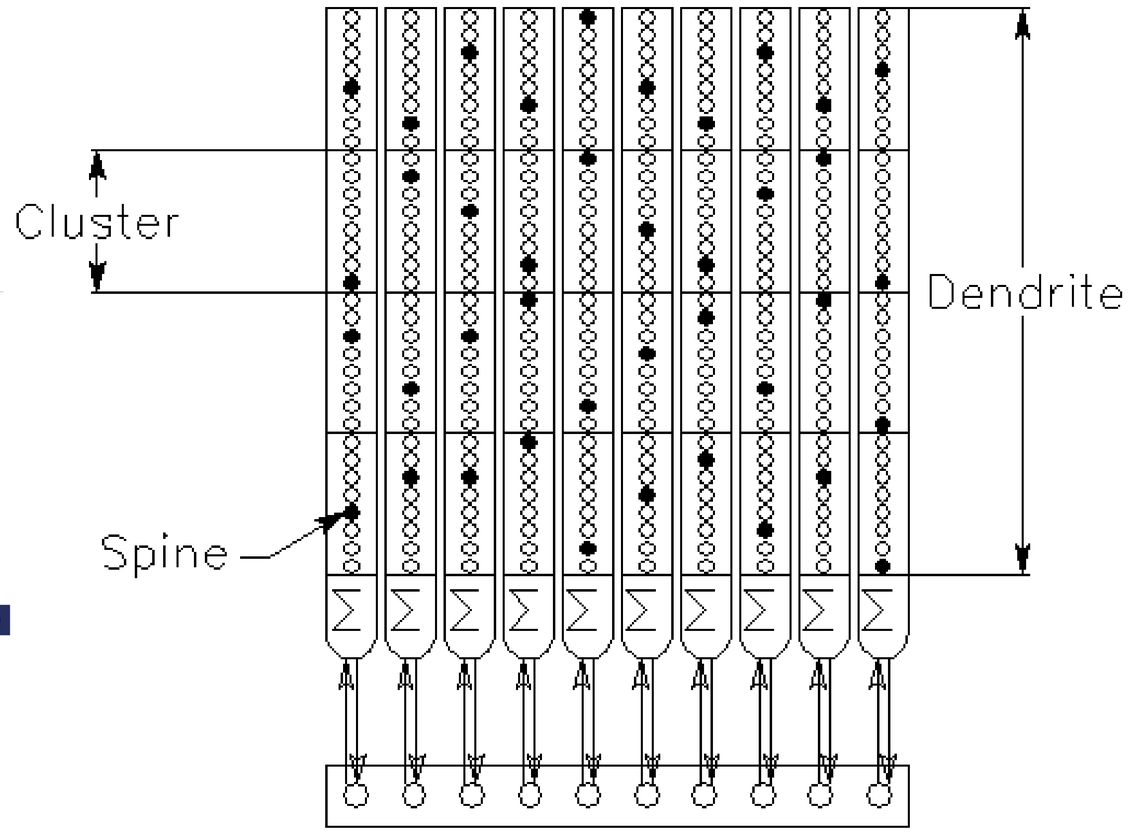
Auteurs

Ehsan Sedgh Gooya

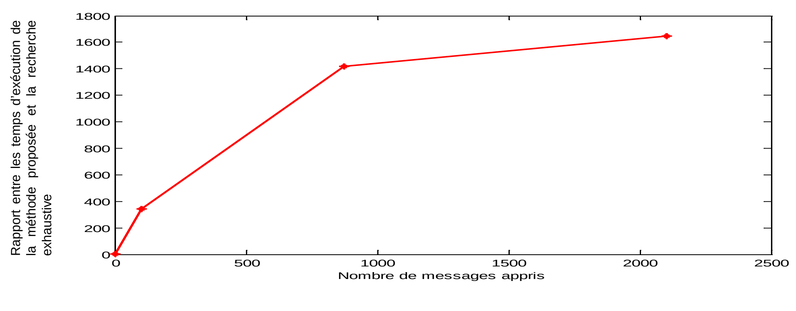
Dominique Pastor

Partenaires

MODELE FORMEL ET PROPRIETES



- Comme les réseaux à cliques de Berrou et Gripon, ce modèle n'emploie que des connexions binaires (pas de poids synaptiques)
- La taille de la mémoire (le stockage) n'est pas un réel problème. Par contre, il faut savoir retrouver très vite et sans erreur un message, ce que permet le neurone formel proposé
- La formalisation mathématique montre que ce neurone ne commet aucune erreur: il reconnaît tout ou une partie des messages qu'il a appris et ne reconnaît aucun message ou partie de message qu'il n'aurait pas appris
- Il est aussi performant qu'une recherche exhaustive, mais pour un coût de calcul largement inférieur.



- Les calculs de cette simulation sont faits en utilisant une programmation Matlab et des signaux générés aléatoirement
- Pour cette simulation, le dendrite est composé de 4 clusters et 8 épines par clusters. Chaque dendrite est associée de manière univoque à un message appris.
- Aucune itération de décodage
- On apprend en ajoutant du matériel, sans altérer ce qui a déjà été appris et en gardant les mêmes capacités de discrimination

PERSPECTIVES

- Introduction de la plasticité à l'aide d'un paramètre de tolérance afin de traiter le cas de messages distordus et/ou bruités
- Structure distribuée et/ou hiérarchique (réseau de neurones)
- Concept d'apprentissage attentif
- Vers des machines auto-apprenantes

26-27 mars 2014 Colloque IMT "Numérique, grande échalle et complexité"

Contact dominique.pastor@telecom-bretagne.eu
ehsan.sedghgooya@telecom-bretagne.eu

Site web www.perso.telecom-bretagne.eu/pastor/

Suivi d'objets multiples dans des scènes vidéo complexes

Parties prenantes

TELECOM Lille

Auteurs

Christelle Garnier
Mehdi Oulad Améziane
François Septier
Yves Delignon
Emmanuel Duflos

Partenaires

LAGIS
UMR CNRS 8219

CENTRALE LILLE

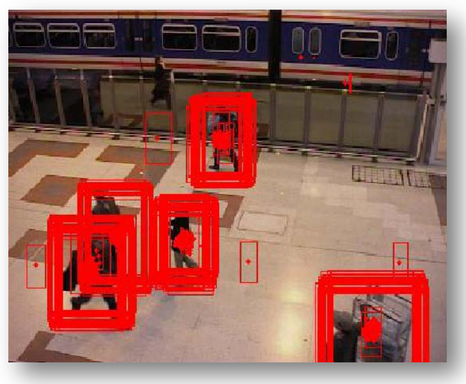
CONTEXTE

- Très grand volume de données vidéo à analyser dans de nombreux domaines d'application : vidéosurveillance, robotique, véhicules autonomes, interactions homme-machine, imagerie du vivant…
- Besoin de systèmes autonomes et intelligents, capables d'extraire automatiquement les informations utiles et de les interpréter.

OBJECTIF, PROBLEMATIQUE ET METHODES

Développer des modèles et des méthodes pour détecter et suivre de façon automatique, fiable et robuste des objets multiples dans des séquences vidéo issues d'une ou de plusieurs caméras.

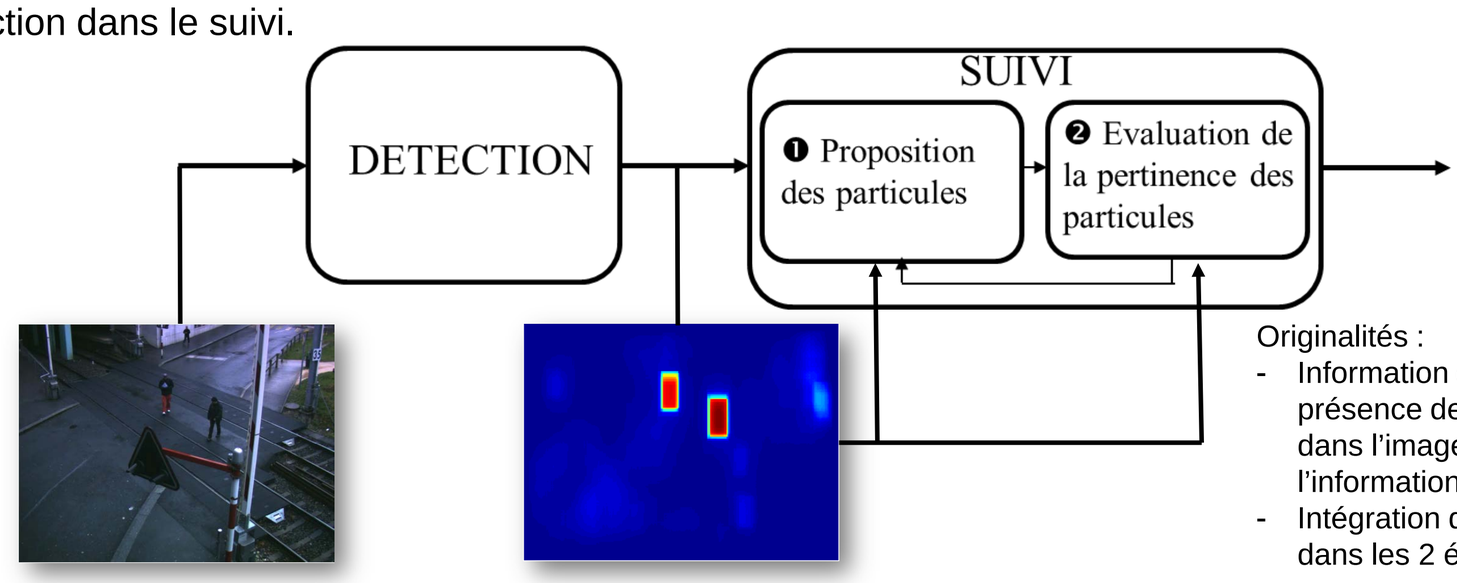
- **Principales difficultés** : nombre croissant d'objets (grande dimension), apparitions / disparitions d'objets, bruit, fausses détections, complexité de l'environnement, non-stationnarités (variations de l'environnement, du mouvement et de l'apparence des objets), occlusions…
- **Problème d'estimation séquentielle** : déterminer le nombre d'objets et leurs paramètres caractéristiques au cours du temps.
- **Outils méthodologiques** :
  Méthodes séquentielles de Monte-Carlo (filtrage particulaire),
  Méthodes de Monte-Carlo par chaînes de Markov (MCMC).



Exploration de l'espace d'état par les particules

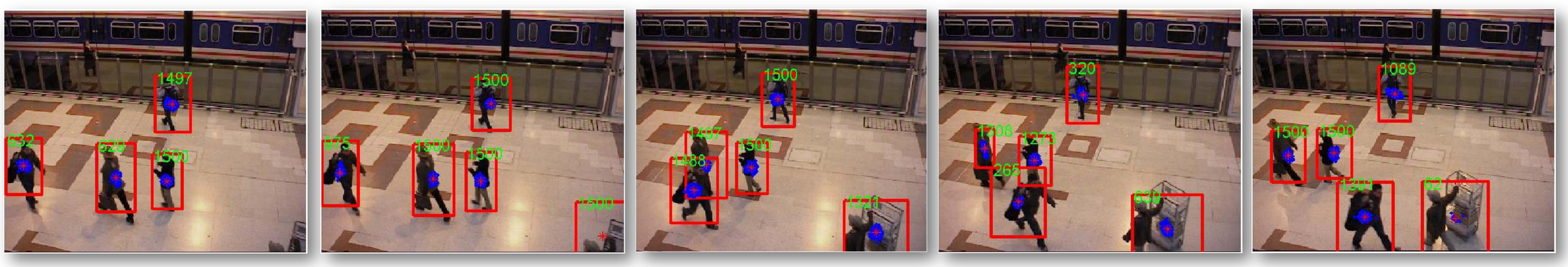
TRAVAUX ACTUELS

- **Optimisation des liens entre détection et suivi** : intégration de la sortie « soft » d'une méthode de détection dans le suivi.



Originalités :
- Information de probabilité sur la présence des objets d'intérêt dans l'image (plus riche que l'information binaire usuelle),
- Intégration de cette information dans les 2 étapes de suivi.

- **Gestion des variations de mouvement et d'apparence des objets** : nouveaux modèles dynamiques.
- **Amélioration du suivi en grande dimension** :
  - Versions séquentielles des méthodes MCMC plus performantes que les filtres particulaires lorsque la dimension augmente,
  - Lois de proposition plus efficaces pour explorer l'espace d'état et guider rapidement les algorithmes vers les zones à forte vraisemblance.



Suivi multi-objets : (1) suivi de 4 personnes, (2) détection automatique d'une 5ème personne, (3) 1ère occlusion partielle, (4) 2ème occlusion partielle, (5) suivi après les occlusions

Colloque Numérique Institut Mines Telecom
Mars 2014

Contact christelle.garnier@telecom-lille.fr
Site web www.splab.telecom-lille1.eu

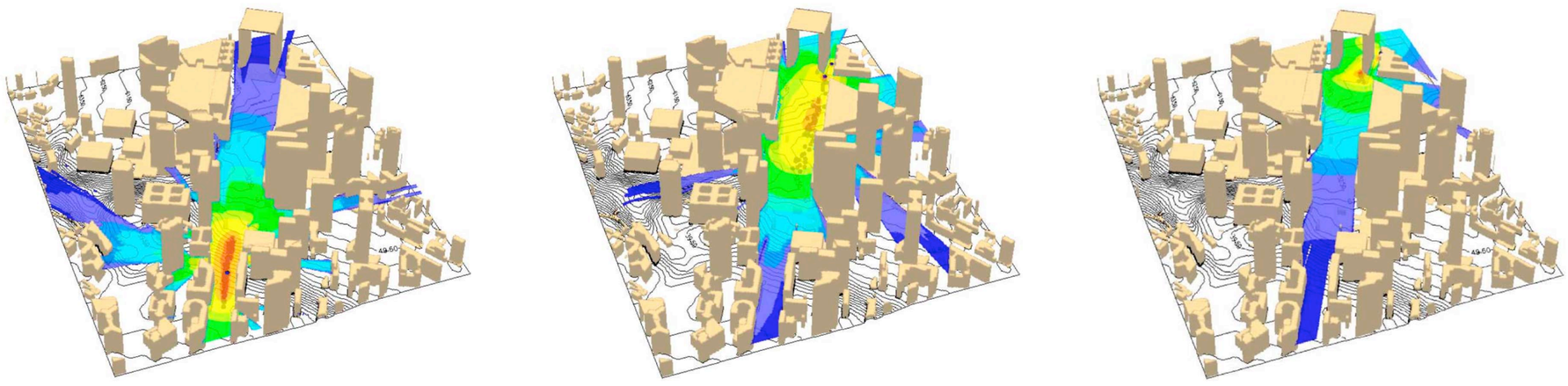
Localisation de sources de pollution depuis un réseau de capteurs

Contexte

- Menace grandissante de rejets délibérés ou accidentels d'agents nucléaires, radioactifs, biologiques ou chimiques (NRBC) ayant des conséquences dramatiques pour la population et l'environnement
- Mise en place d'un réseau mondial de capteurs dans le cadre du Traité d'Interdiction Complète des Essais Nucléaires (TICEN)

Objectifs

Développer une méthode de détection et de localisation de sources de rejets polluants depuis des mesures bruitées de concentration issues de multiples capteurs



Simulation d'un rejet d'agent toxique au coeur du quartier de la Défense à Paris

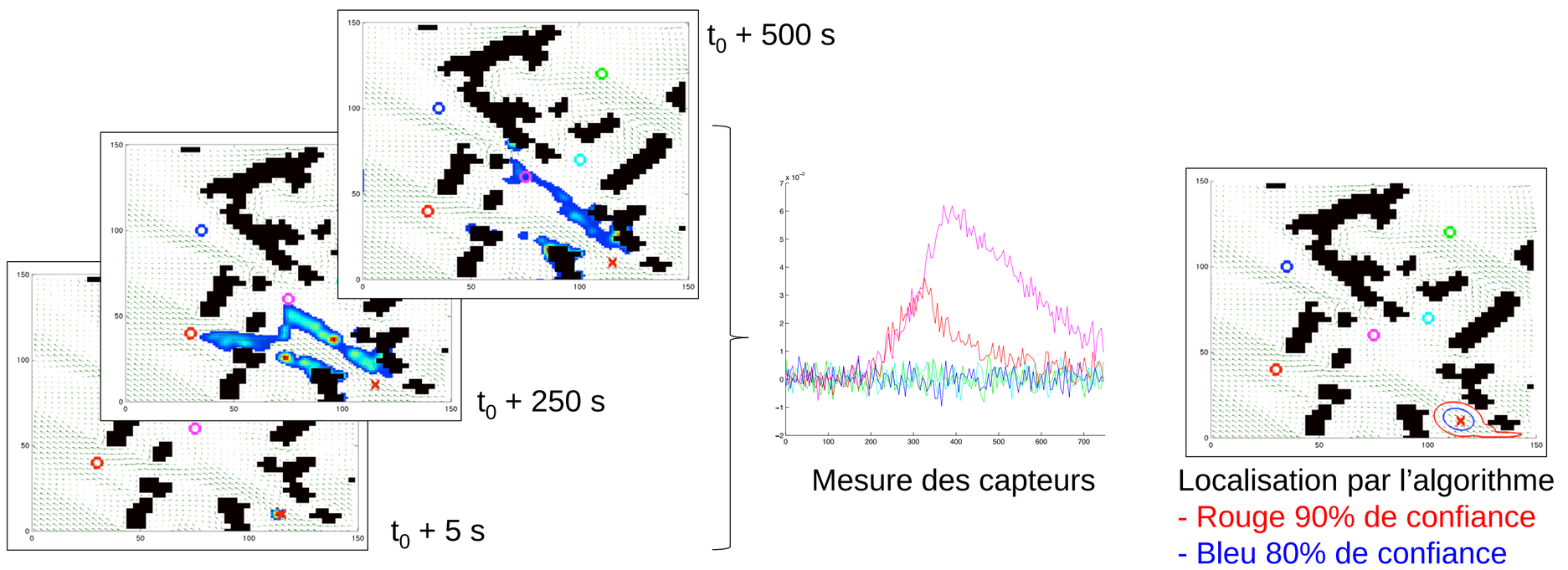
Problématiques

- Complexité des modèles météorologiques nécessaires pour la simulation réaliste de la dispersion atmosphérique d'agents toxiques.
- Imperfection et inhomogénéité des capteurs utilisés.
- Besoin d'une solution rapide et fiable afin de minimiser les conséquences d'un rejet.

Travaux actuels

Développement d'une méthode statistique de type Monte-Carlo permettant contrairement aux approches existantes de:

- Fournir un intervalle de confiance sur l'estimation fournie à l'utilisateur.
- Exploiter un modèle de dispersion atmosphérique de flux de turbulences complexes par des modèles Lagrangien (Parallel Micro-SWIFT-SPRAY, PMSS)
- Converger plus rapidement vers les zones de rejet les plus probables grâce à l'utilisation de techniques adaptatives (algorithmes PMC, AMIS).



Scénario d'un rejet à t_0 d'une source localisée à (115,10) dans un environnement avec bâtiments (noir) et 5 capteurs (cercles)

Auteurs

Harizo Rajoana (Aria Tech.)
François Septier
Yves Delignon
Patrick Armand (CEA)

Partenaires

UMR CNRS 8219

Contact françois.septier@telecom-lille.fr

Site web www.telecom-lille.fr/people/septier

Colloque Numérique Institut Mines Telecom

Mars 2014

Riemannian Geometry for 3D Human Video Retrieval

Rim Slama*, Hazem Wannous*, Mohamed Daoudi+;*
+ Institut TELECOM/TELECOM Lille 1, France.
* Laboratoire d'Informatique Fondamentale de Lille (UMR Lille1/CNRS 8022), Université de Lille 1, France
{rim.slama, hazem.wannous, mohamed.daoudi}@telecom-lille.fr

Context and Issues

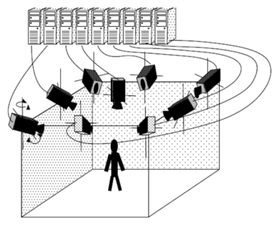
- Long sequence of 3D videos : massive amounts of data

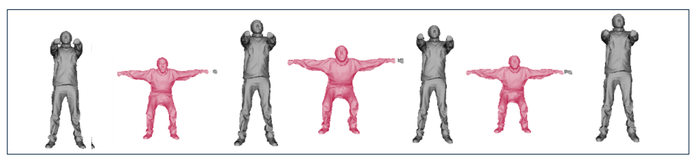
☠ browsing and searching for relevant information quickly become difficult

➔Need for 3D video segmentation system

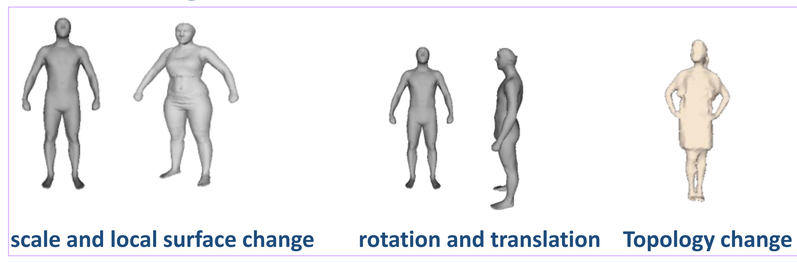
➔Pose/Motion retrieval

➔Video summarization



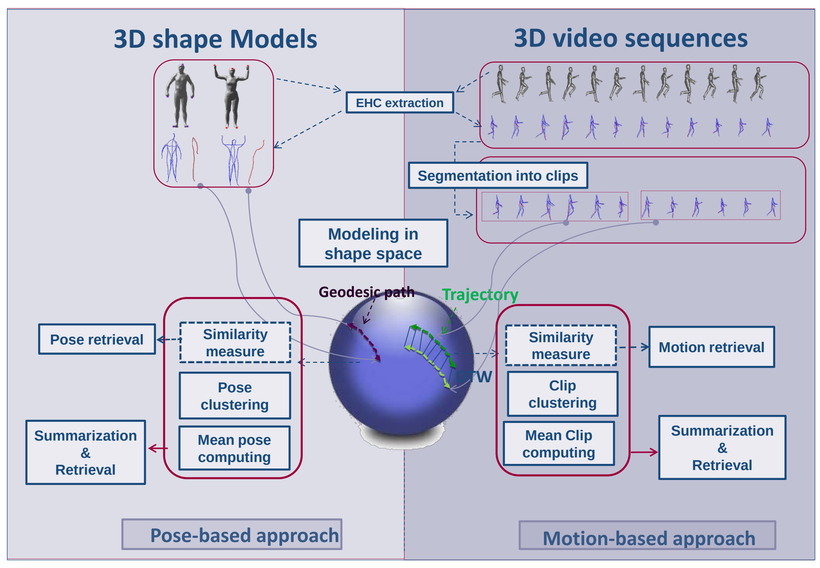


- Challenges:

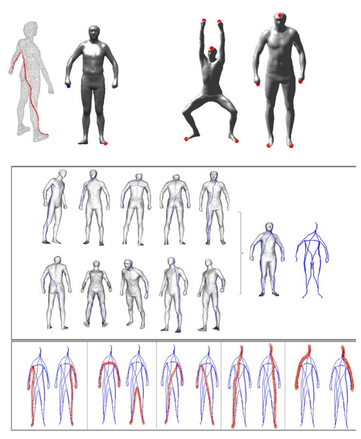


Approach: Riemannian Geometry

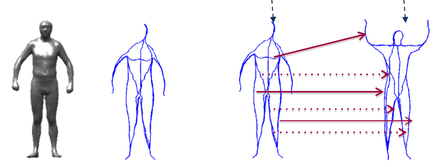
Overview



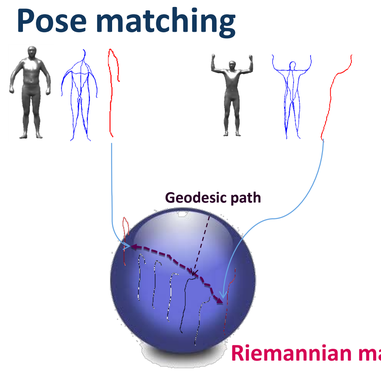
Extremal Human Curve extraction



EHC comparison



Pose matching

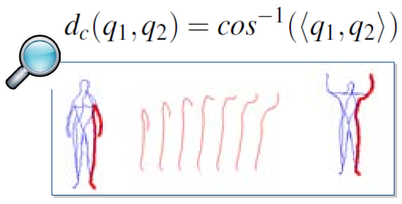


$$\beta \xrightarrow{\text{SRVF}} q : \mathbb{S}^1 \to \mathbb{R}^3, \quad q(t) \doteq \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}}$$

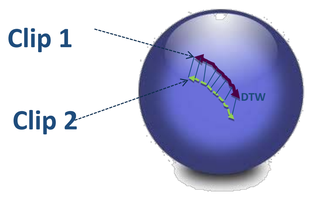
$$\mathcal{C} = \{q : I \to \mathbb{R}^3, \|q\| = 1\} \subset \mathbb{L}^2(I, \mathbb{R}^3)$$

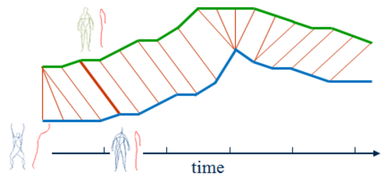
Geodesic distance between $q_1, q_2 \in \mathcal{C}$

$$d_c(q_1, q_2) = cos^{-1}(\langle q_1, q_2 \rangle)$$



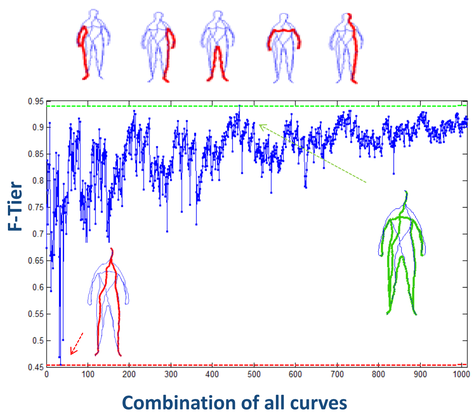
Clip matching



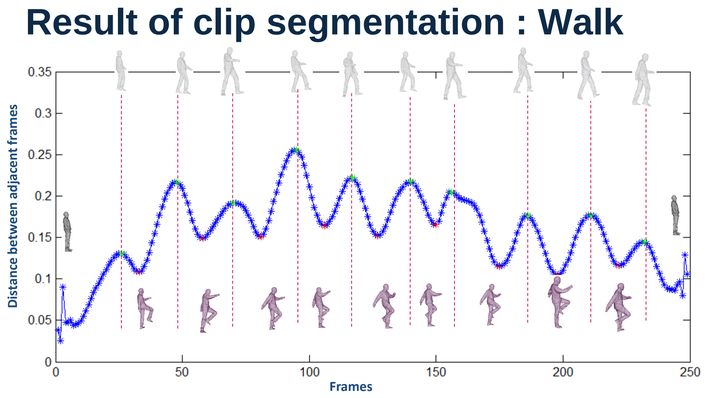


Results and Conclusion

Selected curves



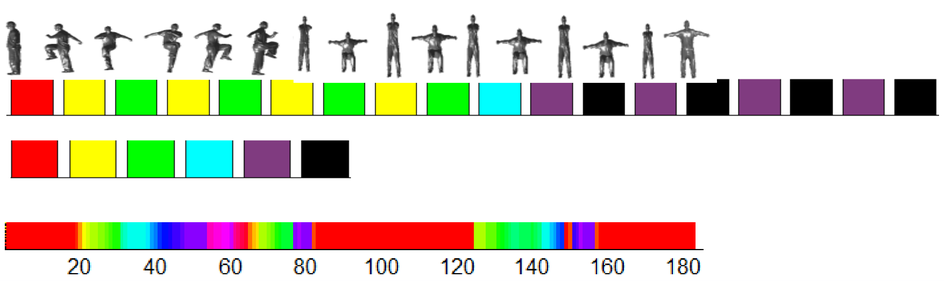
Result of clip segmentation : Walk



Conclusion

- Five curves are sufficient to represent at best the body pose
- Velocity curve is used to segment the long sequences into clips
- Clip matching using DTW on Riemannian manifold gives 93.44% of second tier rate and allows being invariant to speed
- Summarization by clustering is exploited in content-based motion retrieval

Videoclustering



Selected publications

- R. Slama, H. Wannous, M. Daoudi, 3D human motion analysis framework for shape similarity and retrieval. Image Vision Computing Journal 32(2): 131-154 (2014)

Office D106N, Telecom Lille1 -
Rue G. Marconi, Cité Scientifique
Villeneuve d'Ascq, 59650, FRANCE.
rim.slama@telecom-lille1.eu
(+33) (0)3 20 33 55 17
(+33) (0)3 20 33 55 99

Rim Slama, PhD. student,

Statistical Shape Analysis of a Large 3D Faces Dataset

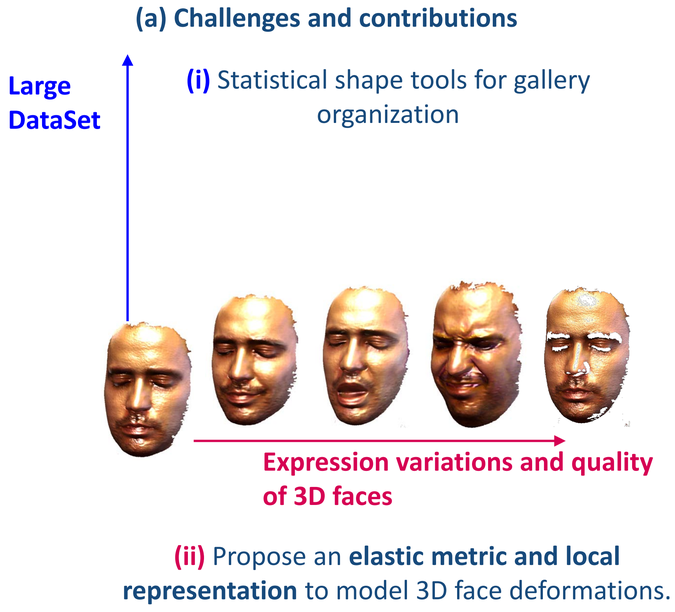
Hassen Drira*, Boulbaba Ben Amor*, Mohamed Daoudi* and Anuj Srivastava[T]

* Institut Mines-Télécom/Télécom Lille, France.
[T] Departement of Statistics, Florida State University, Tallahassee, FL 32306, USA.

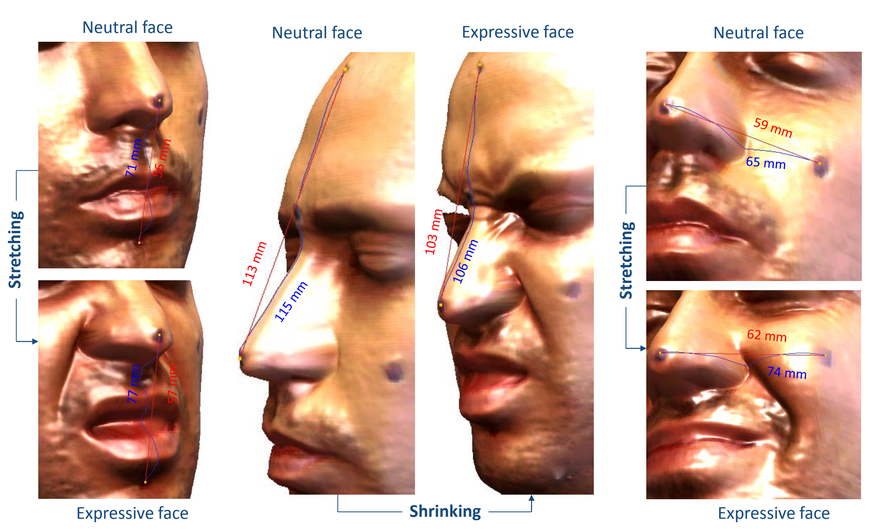
hassen.drira@telecom-lille.fr

MOTIVATIONS

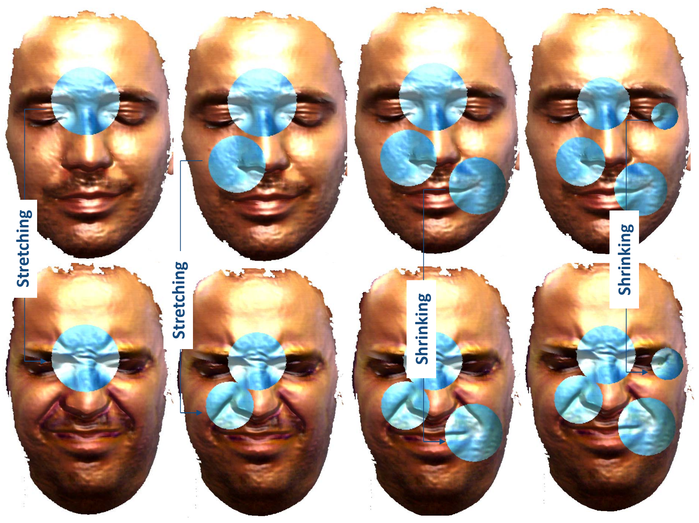
(a) Challenges and contributions



(b) Need for elastic representation of facial surfaces

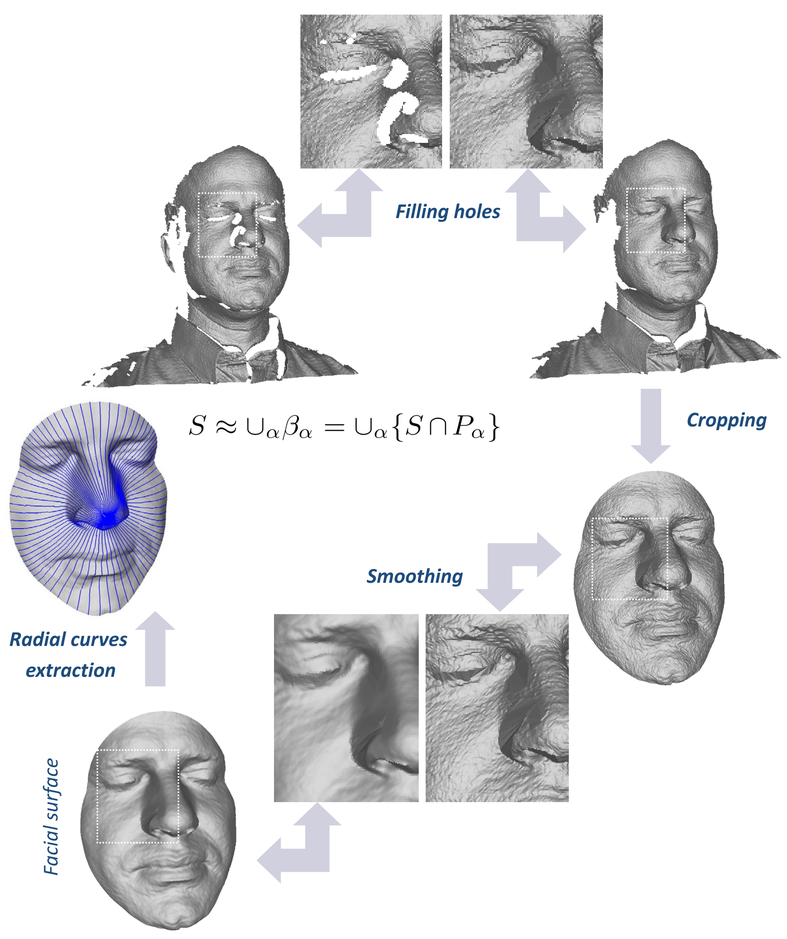


(c) Facial patches Stretching/Shrinking (elastic deformations)

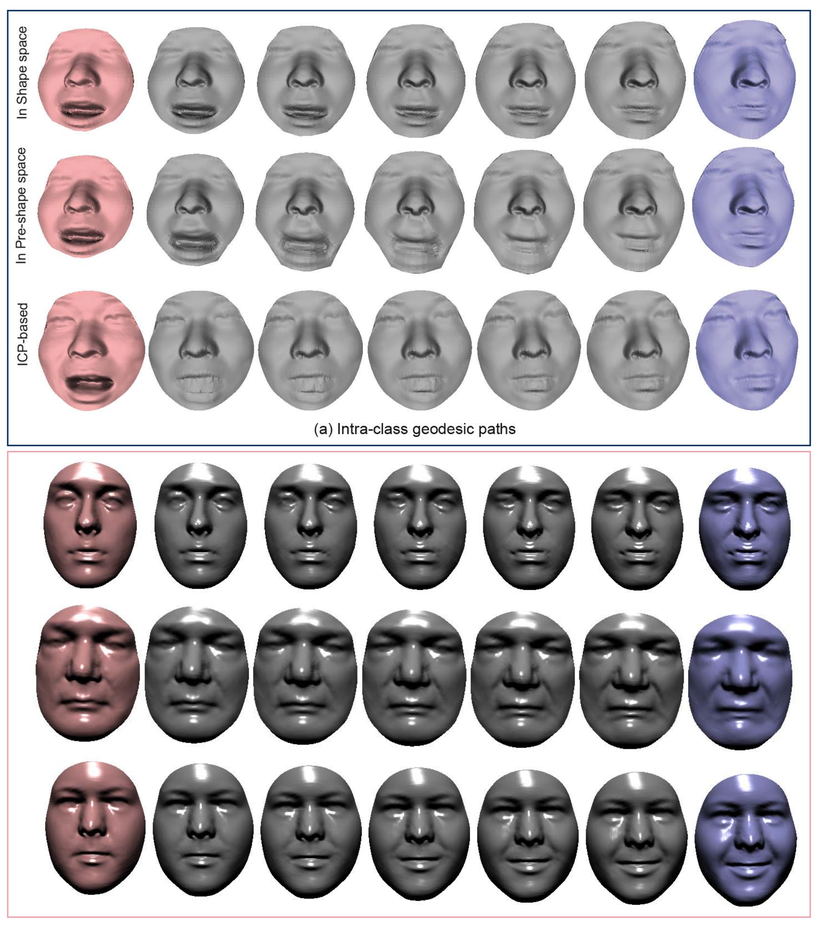


STATISTICAL SHAPE ANALYSIS of 3D FACE

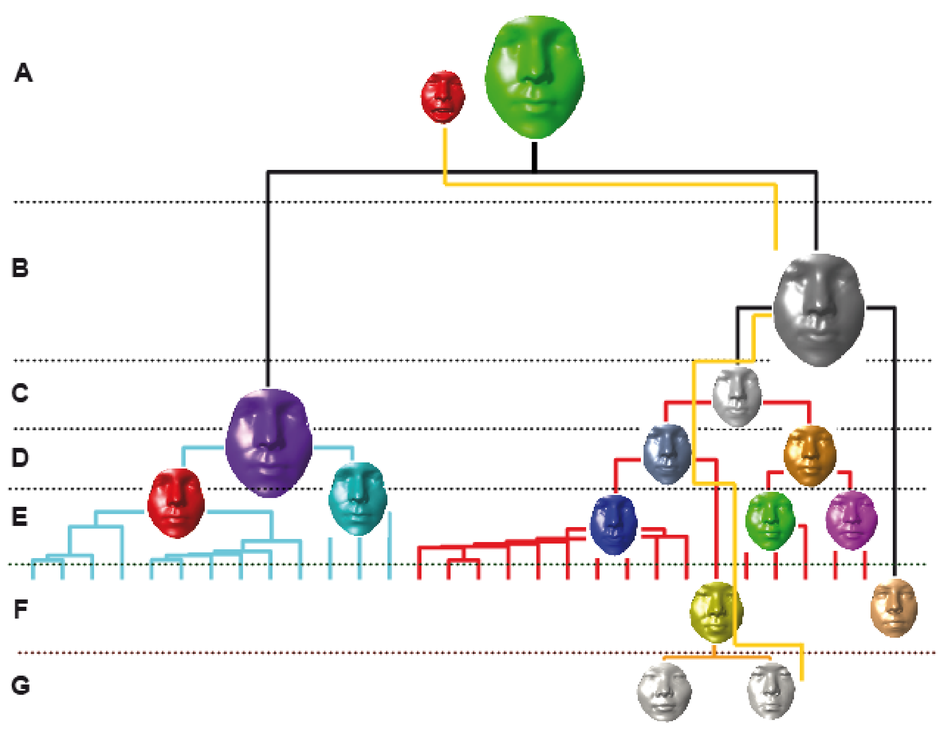
(a) Preprocessing of 3D scans and facial representation



(b) Examples of geodesics



(c) Gallery organization: Hierarchical retrieval of 3D faces Gallery



EXPERIMENTAL RESULTS

(a) Experimental protocol

- FRGCv2 datset: 4007 facial scans
- 466 subjects
- Gallery: neutral faces of 466 subjects
- Probe: remaining

(b) Experimental results

Comparison of rank-1 scores on the FRGCv2 dataset with the state-of-the-art results.

Spreeuwers [29]	Wang et al. [32]	Haar et al. [31]	Berretti et al. [2]	Queirolo et al. [26]	Faltemier et al. [9]	Kakadiaris et al. [13]	Our approach
99%	98.3%	97%	94.1%	98.4%	97.2%	97%	97%

Comparison of verification rates at FAR=0.1% on the FRGCv2 dataset with state-of-the-art results (the ROC III mask and the All vs. All scenario).

Approaches	Kakadiaris et al. [13]	Faltemier et al. [9]	Berretti et al. [2]	Queirolo et al. [26]	Spreeuwers [29]	Wang et al. [32]	Our approach
ROC III	97%	94.8%	-	96.6%	94.6%	98.4%	97.14%
All vs. All	-	93.2%	81.2%	96.5%	94.6%	98.13%	93.96%

Selected publications

1. Hassen Drira, Boulbaba Ben Amor, Anuj Srivastava, Mohamed Daoudi, Rim Slama: 3D Face Recognition under Expressions, Occlusions, and Pose Variations. IEEE Trans. Pattern Anal. Mach. Intell. 35(9): 2270-2283 (2013).

2. Hassen Drira, Boulbaba Ben Amor, Anuj Srivastava, Mohamed Daoudi: A Riemannian analysis of 3D nose shapes for partial human biometrics. ICCV 2009: 2050-2057

SUPPORTED BY ANR

Laboratoire d'Informatique Fondamentale de Lille
CNRS/Lille1 UMR 8022

Improved bandit algorithms: Go Bayesian!

Emilie Kaufmann, Telecom ParisTech & CNRS

In a nutshell

What is the performance of Bayesian bandit algorithms from a frequentist point of view? Bayes-UCB and Thompson Sampling appear to outperform frequentist algorithms on their own ground, which is supported by optimal regret bound for the Bernoulli case.

Bayesian vs. Frequentist Model for MAB

K independent arms. Arm a depends on parameter θ_a and has expectation μ_a; optimal arm is $a^* = \text{argmax } \mu_a$ and $\mu^* = \mu_{a^*}$ is the highest expectation of reward associated.

Two probabilistic modelings

Frequentist :	**Bayesian :**
• $\theta_1, \dots, \theta_K$ unknown parameters	• $\theta_a \overset{i.i.d.}{\sim} \pi_a$
• $(Y_{a,t})_t$ is i.i.d. with distribution ν_{θ_a}	• $(Y_{a,t})_t$ is i.i.d. conditionally to θ_a with distribution ν_{θ_a}

At time t, arm A_t is chosen and reward $X_t = Y_{A_t,t}$ is observed

Two measures of performance

• Minimize (classic) regret $R_n(\theta) = \mathbb{E}_\theta \left[\sum_{t=1}^{n} \theta^* - \theta_{A_t} \right]$	• Minimize "Bayesian" regret $R_n = \int R_n(\theta) d\pi(\theta)$

Optimal algorithms

• *Asymptotically optimal algorithms* satisfy, for $a : \mu_a < \mu^*$, $\limsup_{n\to\infty} \frac{\mathbb{E}_\theta[N_a(n)]}{\log(n)} \leq \frac{1}{\text{KL}(\nu_{\theta_a}, \nu_{\theta^*})}$ They are optimal in the sense of Lai and Robbins' lower bound (1985) on the number of draw of a sub-optimal arm	• An index policy inspired by that of Gittins (1979) adapted to non-discounted rewards minimizes Bayesian regret → no frequentist guarantees

⇒ Our goal: Design algorithms inspired by the Bayesian modeling that are asymptotically optimal in the frequentist setting.

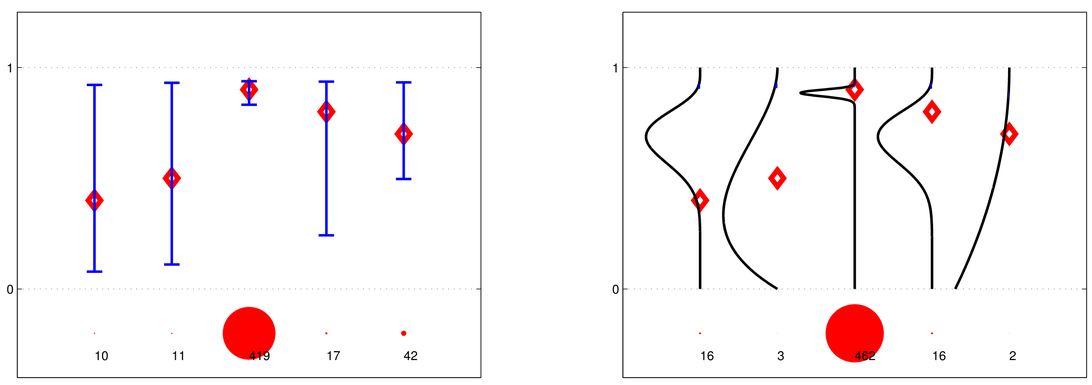
Bayesian versus frequentist algorithms

Some quantities that naturally arise in the Bayesian modeling are

- $\Pi_t = (\pi_1^t, \dots, \pi_K^t)$ the current posterior over $(\theta_1, ..., \theta_K)$
- $\Lambda_t = (\lambda_1^t, \dots, \lambda_K^t)$ the current posterior over the means $(\mu_1, ..., \mu_K)$

Successful algorithms inspired by the frequentist modeling use

- Upper Confidence Bound for the empirical mean... (UCB)
- ... built using KL-divergence (KL-UCB, asymptotically optimal)



Whereas a *Bayesian algorithm* uses Π_t to determine action A_t.

Bayes-UCB and Thompson Sampling

Bayes-UCB algorithm chooses $A_t = \underset{a=1...K}{\text{argmax}}\ q_a(t)$, with

$$q_a(t) = Q\left(1 - \frac{1}{t(\log t)^c}, \lambda_a^t\right)$$

Thompson Sampling is a *randomized* algorithm:

$$\begin{cases} \forall a \in \{1..K\}, \quad \theta_a(t) \sim \lambda_a^t \\ A_t = \text{argmax}_a\ \theta_a(t) \end{cases}$$

Parameters: c (in practice, take $c = 0$), initial prior Π_0

Bayes-UCB: Theoretical elements

ν_{θ_a} is the Bernoulli distribution $\mathcal{B}(\theta_a)$, π_a^0 the (conjugate) prior $\mathcal{U}([0,1])$

- **Bayes-UCB is asymptotically optimal for Bernoulli bandits**

Theorem 1 *Let $\epsilon > 0$; for the Bayes-UCB algorithm with parameter $c \geq 5$, the number of draws of a sub-optimal arm a is such that :*

$$\mathbb{E}_\theta[N_a(n)] \leq \frac{1+\epsilon}{KL\left(\mathcal{B}(\theta_a), \mathcal{B}(\theta^*)\right)} \log(n) + o_{\epsilon,c}\left(\log(n)\right)$$

- **Bayes-UCB is very close to a frequentist algorithm**

The Bayes-UCB index $q_a(t)$ is closely related to the one used by the KL-UCB algorithm (Cappé et al. 2013): $\tilde{u}_j(t) \leq q_j(t) \leq u_j(t)$ with:

$$u_a(t) = \underset{x > \frac{S_a(t)}{N_a(t)}}{\text{argmax}} \left\{ d\left(\frac{S_a(t)}{N_a(t)}, x\right) \leq \frac{\log(t) + c\log(\log(n))}{N_a(t)} \right\}$$

$$\tilde{u}_a(t) = \underset{x > \frac{S_a(t)}{N_a(t)+1}}{\text{argmax}} \left\{ d\left(\frac{S_a(t)}{N_a(t)+1}, x\right) \leq \frac{\log\left(\frac{t}{N_a(t)+2}\right) + c\log(\log(n))}{(N_a(t)+1)} \right\}$$

where $d(x,y) = KL\left(\mathcal{B}(x), \mathcal{B}(y)\right) = x\log\frac{x}{y} + (1-x)\log\frac{1-x}{1-y}$

Bayes-UCB appears to build automatically confidence intervals based on Kullback-Leibler divergence, that are adapted to the geometry of the problem in this specific case.

Thompson Sampling: Theoretical elements

- **TS is asymptotically optimal for Bernoulli bandits**

Theorem 2 *Let $\epsilon > 0$. With b defined below, for every sub-optimal arm a, there exists a constant $N(b, \epsilon, \theta_a, \theta^*)$ such that*

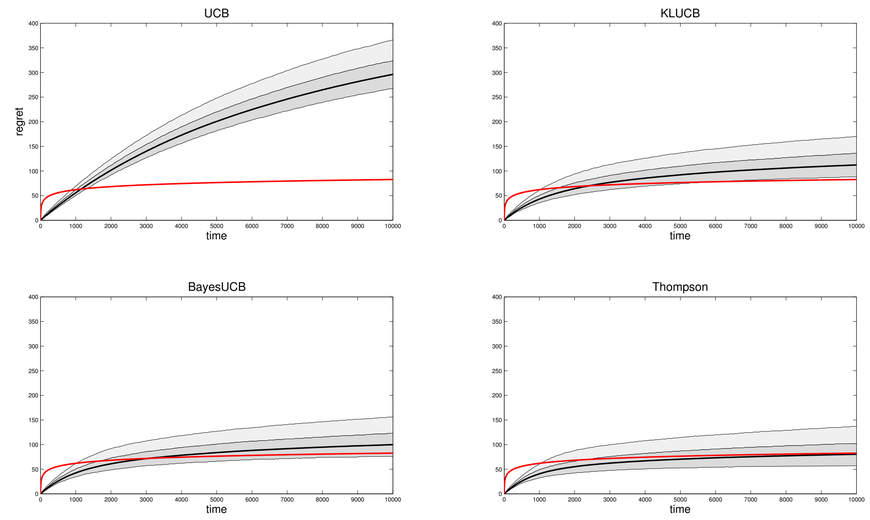
$$\mathbb{E}[N_a(n)] \leq (1+\epsilon)\frac{\ln(n) + \ln\ln(n)}{d(\theta_a, \theta^*)} + N(b, \epsilon, \theta_a, \theta^*) + 5 + 2C_b.$$

Proof Bottleneck: For some constants $b = b(\mu) \in (0,1)$ and $C_b < \infty$,

$$\sum_{t=1}^{\infty} \mathbb{P}\left(N_1(t) \leq t^b\right) \leq C_b.$$

Numerical experiments and beyond

- **Bayesian algorithms are practically as efficient as optimal frequentist algorithms** or even better!



Regret of the various algorithms as a function of time. The red curve show the lower bound, the black bold curve the mean regret and the dark and light shaded the central 99% and the upper 0.05%

- **They are easier to implement:** KL-UCB solves an optimization problem whereas Thompson Sampling only produces one sample!
- **They are easy to generalize:** general models where sampling from a posterior distribution is possible (using MCMC), sparse linear bandit, contextual bandit model...

References

[1] O. Cappé, A. Garivier, O. Maillard, R. Munos, G. Stoltz Kullback-leibler Upper Confidence Bounds for optimal sequential allocation *To appear in Annals of Statistics*, 2013

[2] E. Kaufmann, O. Cappé, A. Garivier, Bayesian Upper Confidence Bounds for bandit problems *AISTATS*, 2012

[3] E. Kaufmann, N. Korda, R. Munos Thompson Sampling: an asymptotically optimal finite-time analysis *ALT*, 2012

Numerical cost for time series prediction via aggregation

Andrés Sánchez Pérez
andres.sanchez-perez@telecom-paristech.fr

François Roueff
francois.roueff@telecom-paristech.fr

1. Summary

The aggregation technique provides an estimator with well-established and excellent theoretical properties that applies for a wide family of times series which includes the AR(d). However the numerical computation of this estimator relies on a Markov chain Monte Carlo method whose performances should be evaluated.

2. Stable AR(d) process

The $\boldsymbol{AR}(d)$ is the stationary solution of :

$$X_t \quad = \quad \sum_{j=1}^{d} \theta_j X_{t-j} + \sigma \xi_t \ , \tag{1}$$

where the innovations ξ_t are i.i.d. with $\mathbb{E}\xi_t = 0$.

Denote by $s_d(1) = \left\{ (\theta_1, \ldots, \theta_d) \ : \ \boldsymbol{\theta}(z) = 1 - \sum_k \theta_k z^k \neq 0 \text{ for } |z| < 1 \right\}$.

We assume that the parameter $\bar{\boldsymbol{\theta}} = (\theta_1, \ldots, \theta_d) \in s_d(1)$ and that (ξ_t) have compact support $\Rightarrow \exists \mathcal{B} \in \mathbb{R}_+^* : X_t \in [-\mathcal{B}, \mathcal{B}] \, \forall t$.

3. Forecasters

Let $(X_1, \ldots, X_n)$ observed values from this ***stationary time series*** $X = (X_t)_{t \in \mathbb{Z}}$. Consider a family of predictors $\{f_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$. For any $\boldsymbol{\theta} \in \Theta$, $f_{\boldsymbol{\theta}}$ is a function from which we obtain :

$$\hat{X}_t^{\boldsymbol{\theta}} \quad = \quad f_{\boldsymbol{\theta}}\left(X_{t-1}, \ldots, X_{t-d}\right) \ , \tag{2}$$

a possible forecasting of X_t according to $\boldsymbol{\theta}$.
Let ℓ be a loss function; we define the prediction risk as
$R(\boldsymbol{\theta}) \quad = \quad \mathbb{E}\left[\ell\left(\hat{X}_t^{\boldsymbol{\theta}}, X_t\right)\right]$ and the empirical version of the risk as
$r_n\left(\boldsymbol{\theta}; X_1, \ldots, X_n\right) = \dfrac{1}{n-d} \sum_{t=d+1}^{n} \ell\left(\hat{X}_t^{\boldsymbol{\theta}}, X_t\right)$.

4. Gibbs estimator

For a $\lambda > 0$ (temperature parameter), we define the ***Gibbs estimator*** as the expectation of a r.v. drawn under the Gibbs measure $\pi\{-\lambda r_n\}$:

$$\hat{\boldsymbol{\theta}}_{\lambda,n} \quad = \quad \pi\{-\lambda r_n\}[\mathrm{Id}] = \int_{\Theta} \boldsymbol{\theta} \pi\{-\lambda r_n(\cdot)\}(\mathrm{d}\boldsymbol{\theta}) \ , \tag{3}$$

where $\nu[h] = \int h \mathrm{d}\nu$ and $\nu\{h\}(\mathrm{d}\boldsymbol{\theta}) = \dfrac{\exp(h(\boldsymbol{\theta}))}{\nu[\exp(h)]} \nu(\mathrm{d}\boldsymbol{\theta})$.

5. A theoretical result

Theorem 1 *In the context of the AR(d), for a bounded $\Theta \subset \mathbb{R}^p$, a uniform* ***prior*** *π yields that $\exists$ a constant $\mathcal{E} : \forall \epsilon > 0$, with probability at least $1 - \epsilon$,*

$$R\left(\hat{\boldsymbol{\theta}}_{\sqrt{n},n}\right) \quad \leq \quad \inf_{\boldsymbol{\theta} \in \Theta} R(\boldsymbol{\theta}) + \mathcal{E} \frac{\log^2(n)}{\sqrt{n}} + \frac{2}{\sqrt{n}} \log\left(\frac{1}{\epsilon}\right) \ . \tag{4}$$

6. Numerical approximation [2]

The ***Metropolis-Hastings algorithm*** generates a Markov chain $\boldsymbol{\Phi} = \{\boldsymbol{\Phi}_i\}_{i \geq 0}$ with the target distribution ρ as a unique invariant measure, based on another Markov chain which serves as a proposal. We tested :

- The *independent Hastings algorithm* where the proposal is i.i.d. with density q such that $\dfrac{q(y)}{\rho(y)} \geq \beta, \forall y \in \Theta$ for some $\beta > 0$.
- The *Metropolis-Hastings algorithm* where the proposal is a Markov chain with conditional density kernel q on $\bar{\Theta} \times \bar{\Theta}$ such that $\beta = \inf_{x \in \bar{\Theta}, y \in \bar{\Theta}} \dfrac{\rho(y)}{\rho(x)} \inf_{x \in \bar{\Theta}, y \in \bar{\Theta}} q(x, y) > 0$.

$\bar{\boldsymbol{\theta}}_m = \dfrac{1}{m} \sum_{i=0}^{m-1} \boldsymbol{\Phi}_i$ is a numerical estimate of $\int x \, \rho(x) \, \mathrm{d}x$.

Theorem 2 *Note by* $\mathrm{diam}(\Theta) = \sup_{x,y \in \Theta} \|x - y\|$ *and define :*

$$M(\alpha, \gamma, \epsilon) = \frac{(2-\gamma)\,\mathrm{diam}(\Theta)}{2\alpha^2 \epsilon \gamma} + \frac{1}{2} \sqrt{\left(\frac{(2-\gamma)\,\mathrm{diam}(\Theta)}{\alpha^2 \epsilon \gamma}\right)^2 + \frac{4\mathrm{diam}(\Theta)}{\alpha^2 \epsilon \gamma}} \ , \tag{5}$$

In the two previous cases, $\forall m \geq M(\alpha, \beta, \epsilon)$, with probability at least $1 - \epsilon$,

$$\left| \bar{\boldsymbol{\theta}}_m - \int x \, \rho(x) \, \mathrm{d}x \right| \quad \leq \quad \alpha \, . \tag{6}$$

7. Application to the AR(d) process

Since $s_d(1) \subseteq B_d\left(2^d - 1\right)$, the prior π can be defined on $\Theta = s_d(1)$ or $B_d\left(2^d - 1\right)$. These two possibilities are combined with two different proposals in the Metropolis-Hasting algorithm.

- **Uniform prior on** $B_d\left(2^d - 1\right)$
  - *Uniform proposal :* $\beta_{\lambda,n} = \exp\left(-\lambda \mathcal{B}^2 \left(1 + \sqrt{d}\left(2^d - 1\right)\right)^2\right)$.
  - *Constrained random walk with Gaussian increment :*

$$\beta_{\lambda,n} = \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-2\left(2^d - 1\right)\left(\lambda 2^{d+1} \mathcal{B}^2 + \left(2^d - 1\right) n\right)\right) \ .$$

- **Pushforward measure on** $s_d(1)$

  A map from the reciprocal roots of $\boldsymbol{\theta}(z)$ into the coefficients $\theta_1, \ldots \theta_d$ and a measure on the first ones allow to define a prior on $s_d(1)$.
  - *Uniform proposal :* $\beta_{\lambda,n} = \exp\left(-\lambda \mathcal{B}^2 \left(1 + \sqrt{d}\left(2^d - 1\right)\right)^2\right)$.
  - *Constrained random walk with Gaussian increment :*

$$\beta_{\lambda,n} = \left(\frac{n}{2\pi}\right)^{1+p+2\lfloor \frac{d}{2} \rfloor} \exp\left(-2\left(\lambda 2^{d+1}\left(2^d - 1\right)\mathcal{B}^2 + \left(2^{1+d+2\lfloor \frac{d}{2} \rfloor} - 1\right)^2 n\right)\right).$$

Theorem 3 *$\exists$ a constant $\mathcal{F}$ such that $\forall m \geq M\left(\dfrac{\log(n)}{n}, \beta_{\sqrt{n},n}, \epsilon\right)$, with M defined as in Theorem 2, with probability at least $(1 - \epsilon)^2$,*

$$R\left(\bar{\boldsymbol{\theta}}_{\sqrt{n},n,m}\right) \quad \leq \quad \inf_{\boldsymbol{\theta} \in \Theta} R(\boldsymbol{\theta}) + \mathcal{F} \frac{\log^2(n)}{\sqrt{n}} + \frac{2}{\sqrt{n}} \log\left(\frac{1}{\epsilon}\right) \ . \tag{7}$$

This result remains true for the whole family of Causal Bernoulli Shifts (CBS) processes (see [1] for the definition).

8. Numerical results

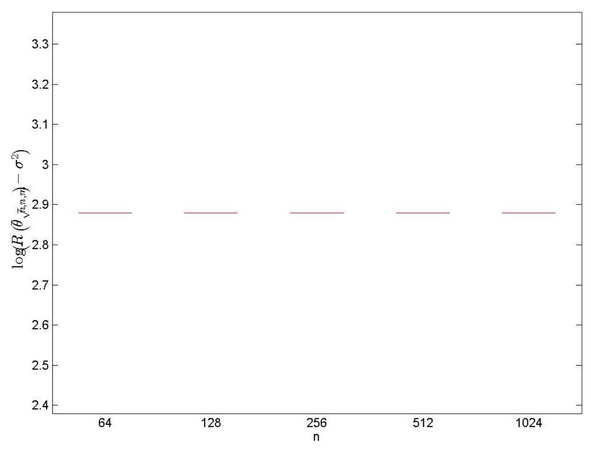


Figure 1: Uniform proposal, $d = 8$, $\Theta = B_8\left(2^8 - 1\right)$

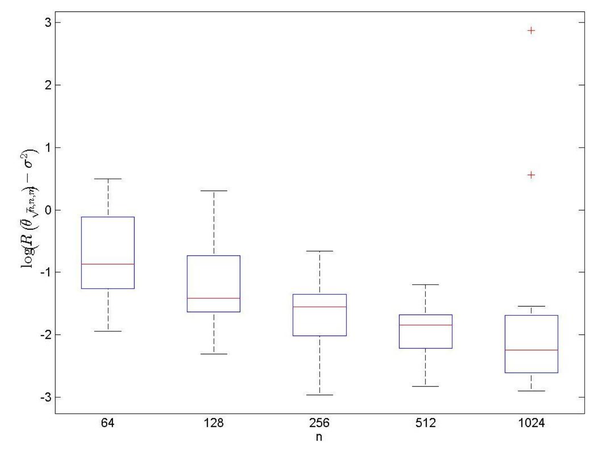


Figure 2: Gaussian proposal, $d = 8$, $\Theta = s_8(1)$

Figure 2 shows good results in contrast to Figure 1. However, using (5) and the obtained expressions of β yield to the following equivalence for the minimal number of iterations m guarantying a correct prediction error :

$$m \geq C_1(d) \frac{\log^2 n}{2n^2 \epsilon} \exp\left(C_2(d) \sqrt{n}\right) \ ,$$

where C_1 and C_2 are positive functions. Hence, to guaranty the Formula (7) we need a very high number of iterations, which can lead to a prohibitive computational cost.

9. References

References

[1] Pierre Alquier and Olivier Wintenberger. *Model selection for weakly dependent time series forecasting.* Bernoulli, 18(3) : 883-913, 2012.

[2] Krzysztof Łatuszyński and Wojciech Niemiro. *Rigorous confidence bounds for MCMC under a geometric drift condition.* J. Complexity, 27(1) : 23-38, 2011.

A NEW BROADCAST BASED DISTRIBUTED AVERAGING ALGORITHM OVER WIRELESS SENSOR NETWORKS

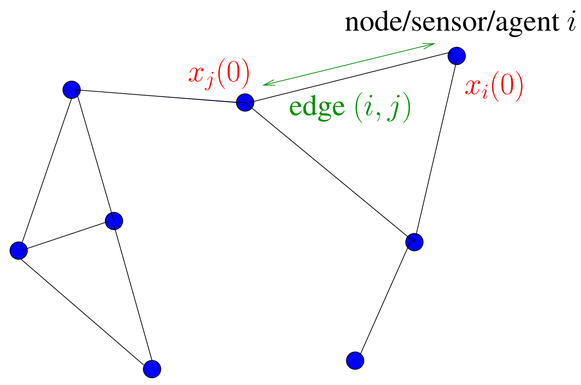
F. IUTZELER[(1)], P. CIBLAT [(1)] and W. HACHEM[(2)]

[(1)] Institut MINES-TELECOM ; TELECOM ParisTech ; CNRS LTCI | [(2)] CNRS LTCI ; TELECOM ParisTech

ABSTRACT

We study the distributed computation of the average value of initial measurements in a Wireless Sensor Network. Unlike existing works, we take benefit of the broadcast nature of the wireless channel to speed up the convergence speed

MODEL



- N nodes/sensors/agents
- $x_i(0)$ measurement at node i at time 0
- **Problem:** at each node, we want to compute

$$x_{\text{ave}} = \frac{1}{N}\sum_{i=1}^{N} x_i(0)$$

without fusion center and so with only local communications.

- **Applications:** practical measurements of temperature, gas pressure,
- At each time t, the sensors' updates can be rewritten in matrix form

$$\mathbf{x}(t+1) = \mathbf{K}(t)\mathbf{x}(t)$$

where $\mathbf{K}(t)$ is the update matrix for time t

Goal:

- We want $\mathbf{x}(t)$ to converge to the average consensus $x_{\text{ave}}\mathbf{1}$
- We want to wake up only one node at each time
- We want to use the broadcast nature of the wireless links

AVERAGING ALGORITHMS

Fundamental properties to ensure convergence to the average consensus:

	Property		Mathematical implication
1.	**Consensus preservation** $\mathbf{x}(t) = c\mathbf{1} \Rightarrow \mathbf{x}(t+1) = c\mathbf{1}$	$\Rightarrow$	**Row stochasticity** $\mathbf{K1} = \mathbf{1}$
2.	**Consensus value** $\mathbf{x}(t) = c\mathbf{1} \Rightarrow \mathbf{x}(t) = x_{\text{ave}}\mathbf{1}$	$\Rightarrow$	**Column stochasticity** $\mathbf{1}^T\mathbf{K} = \mathbf{1}^T$
3.	**Convergence** $\mathbb{E}\left[\|\mathbf{x}(t) - c\mathbf{1}\|^2\right] \to 0$	$\Rightarrow$	**Spectral radius** $\rho(\mathbb{E}[\mathbf{K}] - 1/N\mathbf{1}\mathbf{1}^T) < 1$

Consequences:

- Double-stochasticity needs feedback, so **no broadcast.**
- If only column-stochasticity, $\exists\ \mathbf{v}(t)$ non-negative vectors s.t.

$$\mathbf{K}(t)\cdots\mathbf{K}(2)\mathbf{K}(1) \sim \mathbf{v}(t)\mathbf{1}^{\top} \quad \text{and} \quad \mathbf{x}(t) \sim (Nx_{\text{ave}})\mathbf{v}(t)$$

Proposed solution:

- An other variable updated with the same matrix has to be considered to know $\mathbf{v}(t)$. Then $\mathbf{v}(t)$ can be removed by division.

$\Rightarrow$ **Variables do not converge to consensus, but the quotient does.**

PROPOSED ALGORITHM: *BWGossip*

Algorithm based on the *Sum-Weight* framework [Kempe2003, Bénézit2011] where sensors have **two** local variables jointly updated:

– a sum variable $\mathbf{s}(t)$
– a weight variable $\mathbf{w}(t)$

$$\begin{cases} \mathbf{s}(t+1) = \mathbf{K}(t)\mathbf{s}(t) \\ \mathbf{w}(t+1) = \mathbf{K}(t)\mathbf{w}(t) \end{cases} \quad \text{and} \quad \begin{matrix} \mathbf{s}(0) = \mathbf{x}(0) \\ \mathbf{w}(0) = \mathbf{1} \end{matrix}$$

BWGOSSIP

Assuming that at time t, the sensor i wakes up

▶ Sensor i broadcasts $\left(\frac{s_i(t)}{|\mathcal{N}_i|+1}; \frac{w_i(t)}{|\mathcal{N}_i|+1}\right)$

▶ At sensors in the neighborhood $\mathcal{N}_i$, we have:

$$\begin{cases} s_j(t+1) = s_j(t) + \frac{s_i(t)}{|\mathcal{N}_i|+1} \\ w_j(t+1) = w_j(t) + \frac{w_i(t)}{|\mathcal{N}_i|+1} \end{cases}, \forall j \in \mathcal{N}_i$$

▶ At sensor i, we have :

$$\begin{cases} s_i(t+1) = \frac{s_i(t)}{|\mathcal{N}_i|+1} \\ w_i(t+1) = \frac{w_i(t)}{|\mathcal{N}_i|+1} \end{cases}$$

▶ All other sensors stay idle.

RESULTS

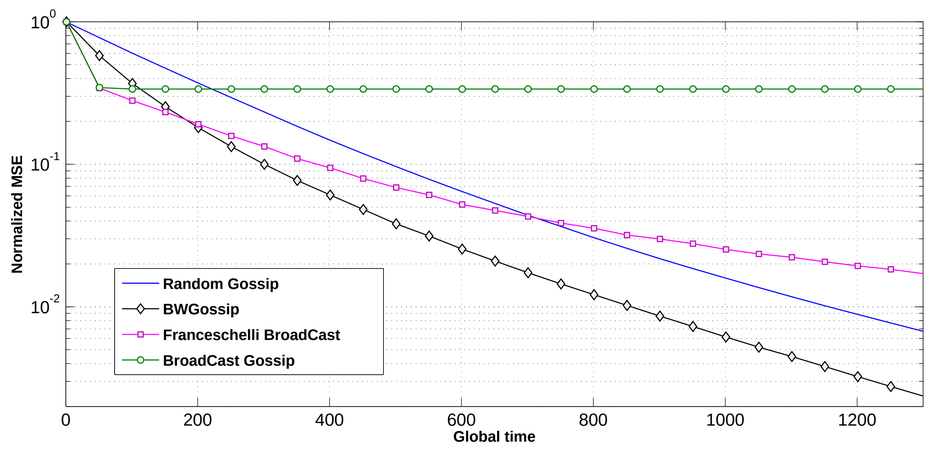
Theorem 1 CONVERGENCE

BWGossip *converges to the average consensus almost surely.*

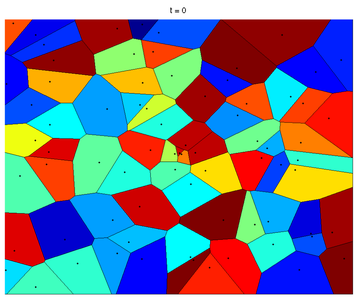
Theorem 2 CONVERGENCE SPEED **(main result)**

For the BWGossip *algorithm, we have:*

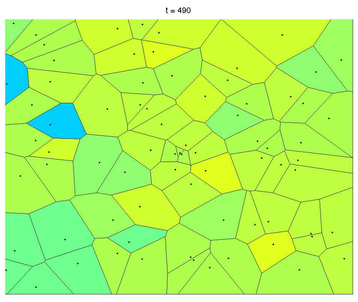
$$\forall \epsilon > 0, \quad \|\mathbf{x}(t) - x_{\text{ave}}\mathbf{1}\|_2^2 = \mathcal{O}_P\left((\Gamma+\epsilon)^t\right)$$
$$\textit{with } \Gamma = \rho\left((\mathbf{I}-\mathbf{J}) \otimes (\mathbf{I}-\mathbf{J}) . \mathbb{E}\left[\mathbf{K}\otimes\mathbf{K}\right]\right) < 1$$



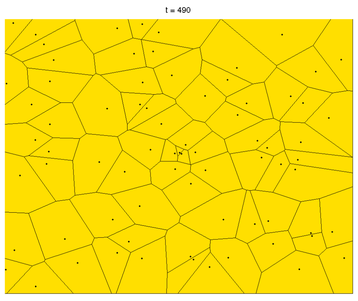
- ***BWGossip* outperforms the existing algorithms.**



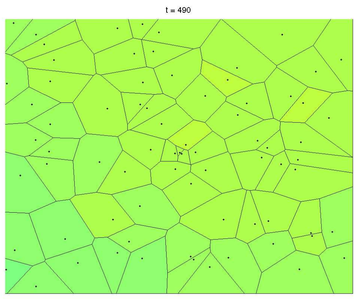
INITIALIZATION



RANDOM GOSSIP



BROADCAST GOSSIP

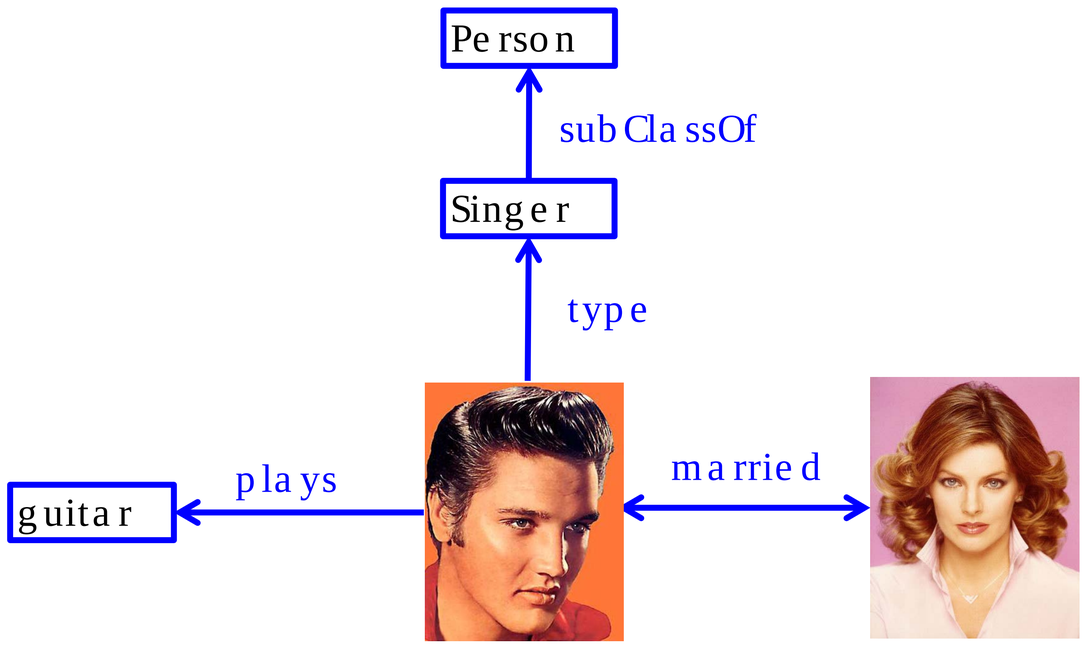


BW GOSSIP

The Independent Otto Hahn Research Group Ontologies

Fabian M. Suchanek

Ontology Construction



Ontologies serve, e.g., for disambiguation, translation, and question answering. We develop YAGO, the largest public ontology with a quality guarantee. YAGO is built automatically from Wikipedia and other sources.

http://yago-knowledge.org

YAGO – A Core of Semantic Knowledge
Fabian M. Suchanek, Gjergji Kasneci, Gerhard Weikum (WWW 2007)
+ follow-up publications in 2008, 2011, 2012, 2013

Rule Mining

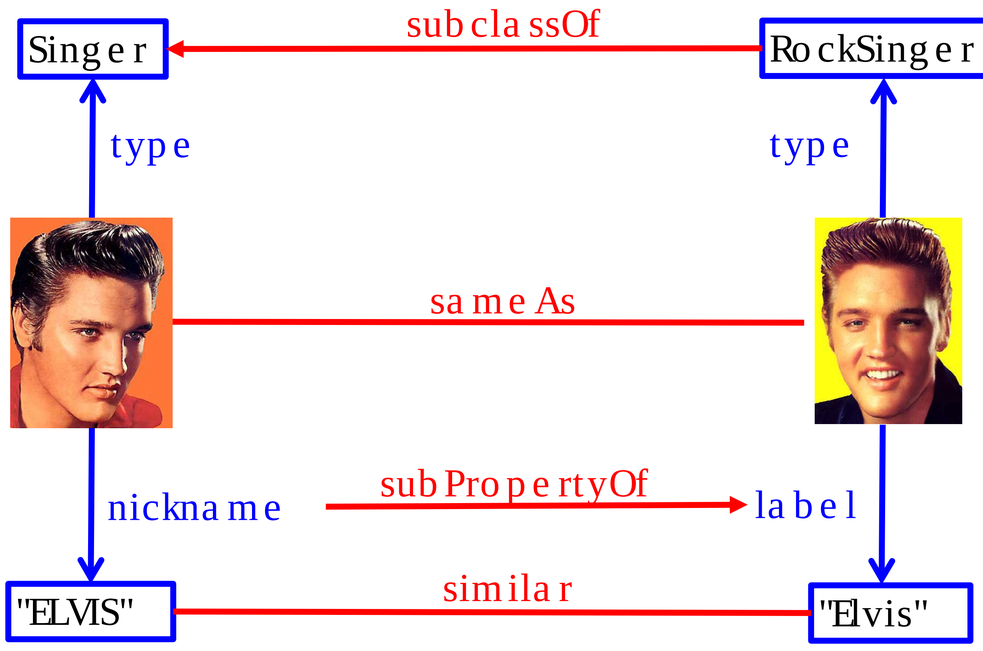
We develop techniques to mine rules, correlations, and schemas from an ontology. These serve to propose missing links, detect inconsistencies, reveal correlations, and make sense out of data. The semantics, incompleteness, and the scale of the data are different from classical settings.

Some rules we mined on YAGO are:

```
hasChild(x,y) & hasChild(z,y) => married(x,z)
wonAward(x,LeibnizPreis) => livesIn(x,Germany)
acadAdvisor(x,y) & almaMater(y,z) => worksAt(x,z)
```

AMIE: Association Rule Mining under Incomplete Evidence
Luis Galárraga, C. Teflioudi, K. Hose, F. M. Suchanek (WWW 2013)

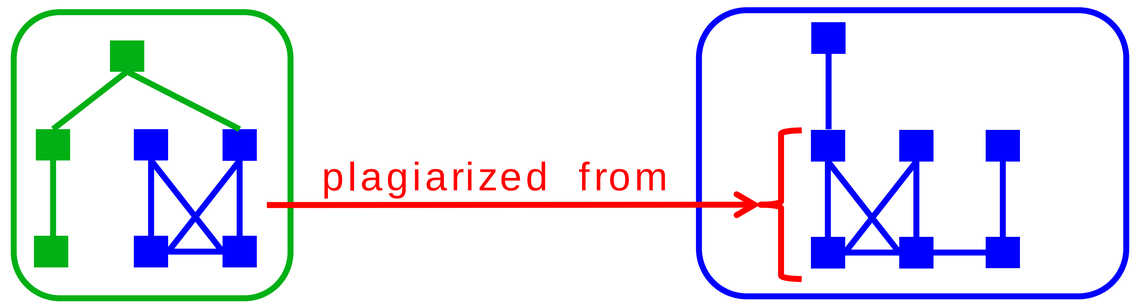
Ontology Matching



Many ontologies on the Semantic Web contain information about the same entities. To make use of complementary information, one has to determine which entities, classes, literals, and properties correspond. We develop statistical, logical, and probabilistic models and algorithms for this purpose.

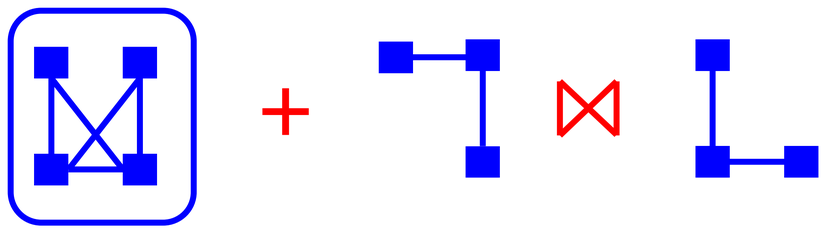
PARIS: Probabilistic Alignment of Relations, Instances and Schema
Fabian M. Suchanek, S. Abiteboul, P. Senellart (VLDB 2012)

Provenance



Statements can be illegally copied over from another ontology. We develop techniques to trace plagiarisms.

Watermarking for Ontologies (ISWC 2011)
Fabian M. Suchanek, David Gross-Amblard, Serge Abiteboul



We develop logic-based models to integrate Web services into ontologies.

SUSIE: Search using Services and Information Extraction
Nicoleta Preda, F. M. Suchanek, W. Yuan, G. Weikum (ICDE 2013)

MAX-PLANCK-GESELLSCHAFT

max planck institut informatik

20th international world wide web conference
28th march - 1st april 2011
hicc hyderabad, india

Casting a Web of Trust over Wikipedia: an Interaction-based Approach

Silviu Maniu, Talel Abdessalem, Bogdan Cautis; Télécom ParisTech – CNRS LTCI, Paris, France, {firstname.lastname@telecom-paristech.fr}

Our goal

Uncover a signed network over Wikipedia contributors from their interactions.

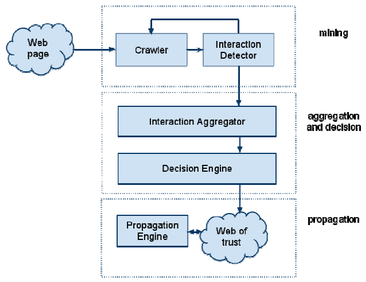
Signed link = a representation of the degree of trust/similarity or distrust/dissimilarity between two users

Several signed networks are already present in social media: Epinions (trust/distrust tags), Slashdot (friend/foe), Wikipedia Elections (support/oppose votes)

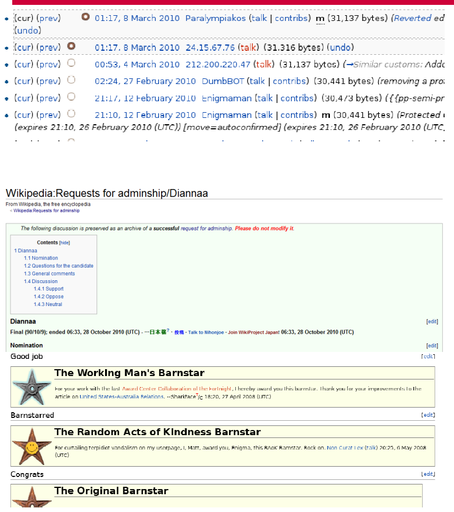
Motivation:

- Social applications can be enhanced by knowing such signed links (social search systems, recommender systems, trust and reputation, etc.)

General Architecture



Interactions in Wikipedia



Article editing:

- on text content (inserting, deleting and replacing text between the contributors)
- on the article revisions (reverting/discarding a version of an article and restoring another)

Adminship election:

- Contributors participate in so called requests for adminship, elections in which contributors can:
  - Support the candidate
  - Oppose the candidate

Interactions on user pages:

- Contributors can give each other prizes called barnstars (generally for good behavior)

Deciding link signs

+	−	−	+	−	+	−	+
insert	replace	delete	restore	revert	support	oppose	barnstars
operations on article text			operations on article revisions		adminship elections (RFAs)		user pages

The interaction vector = aggregation of all interactions for each pair of users

The decision process: -1 (negative) or +1 (positive) link

1. Annotate the atomic interactions with signs (positive or negative, as shown above)
2. Each interaction votes with its weight (measure) for the sign of the higher-order type
3. All types vote for the final link sign

WikiSigned - the resulting network

- 71,770 nodes and 463,312 edges, of which 85.93% positive

Dataset

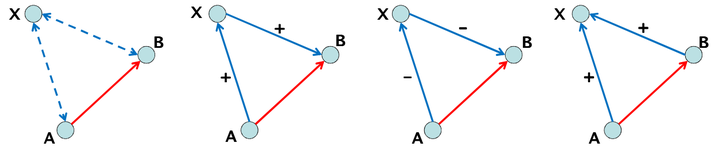
From the politics domain of the English Wikipedia:

- we extracted 320 articles, for a total of 442,297 revisions submitted by 105,177 contributors
- resulting in 800,057 total interactions, in which participate 42,631 adminship votes and 2,913 barnstars

Validation

Does this network represent a *plausible configuration of link signs*?

First approach: comparing our network with three existing explicit signed networks (Epinions, Slashdot, Wikipedia Elections)



the general case t_1 t_6 t_9

Using the concept of link triads and the predictions of two social theories: balance and status, for measuring:

The global properties of WikiSigned

- Our network has similar triad distribution as the explicit networks
- And it has the global structure of a network in which status theory holds (only one contradiction for triad signs)

The local properties of WikiSigned

- It can self-predict its link signs with 0.822 accuracy (AUC of 0.899)
- Good accuracy also in cross training-predicting (training on the row data and predicting on the column data)

	Epinions	Slashdot	Elections	WikiSigned
Epinions	0.906	0.905	0.787	**0.727**
Slashdot	0.929	0.806	0.792	**0.732**
Elections	0.922	0.895	0.814	**0.733**
WikiSigned	**0.889**	**0.844**	**0.784**	**0.822**

Second approach: application-level validation

- Predicting the importance or quality of articles by using the knowledge of link composition (number of positive and negative links) in training predictive models
- This knowledge of link signs helps the prediction when we predict the article importance

feature	Quality	Importance
Contributors	0.683	0.566
Contributors + normal links	0.740	0.779
Social links	0.807	0.750

Maui, Hawaii
CIKM 2012
Oct 28 - Nov 2, 2012

primates

MINES TELECOM
INSTITUT Mines-Télécom

A Privacy Management System for Social Networks

Imen BEN DHIA, Talel ABDESSALEM, Mauro SOZIO
Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, Paris, France
first.last@telecom-paristech.fr
http://dbweb.enst.fr/

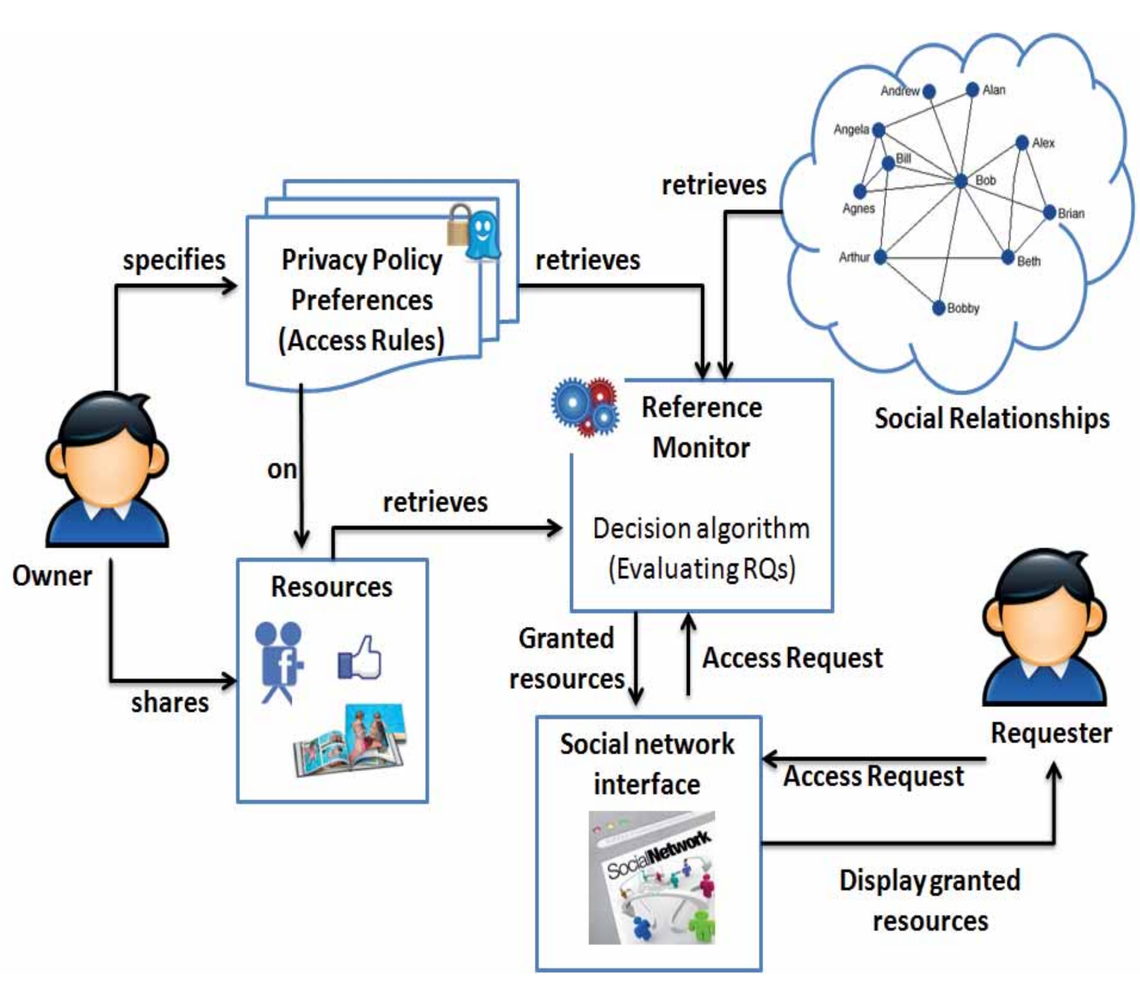
Context and Problem

- Social network users :
  - have different kinds of relationships (friend, colleague, etc.)
  - share content (personal data, photos, videos, etc.)
- Social networks usually grow quickly in terms of number of users, relationships established and pieces of shared information.

Problem: *Social network users have difficulty with specifying which information should be shared with whom.*

Solution : Enable users to specify their privacy policy preferences in a more flexible and efficient way than existing privacy management systems.

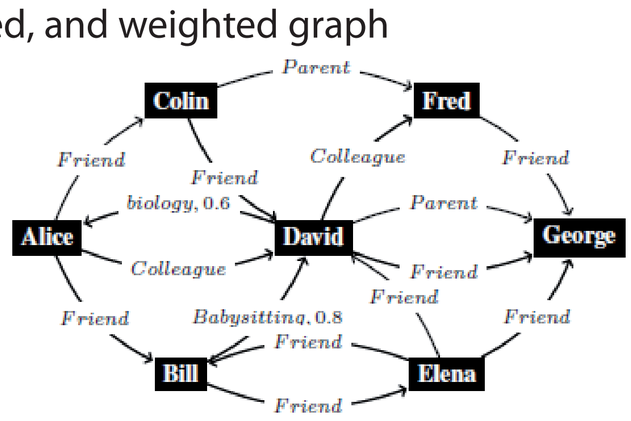
System Overview



Access Control Model

- **Social network model :**
  - Directed, edge-labeled, and weighted graph

(V, E, Σ, φ)



A Social Network Subgraph

- **Access rules specification based on reachability constraints :**
  - Semantics of the links, Links direction, Indirect relationships, Distance, User properties, Trust.

- **Access Rule (AR) :** Specification of the profiles of authorized users to access a given resource.

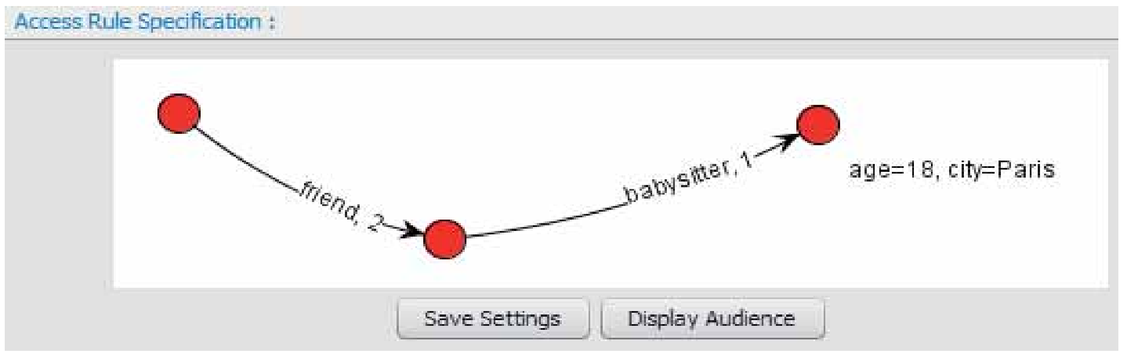
$$AR = (o, p, t_{min})$$

$$p = \{s_i\} \qquad s_i = (r, dir, l, C)$$

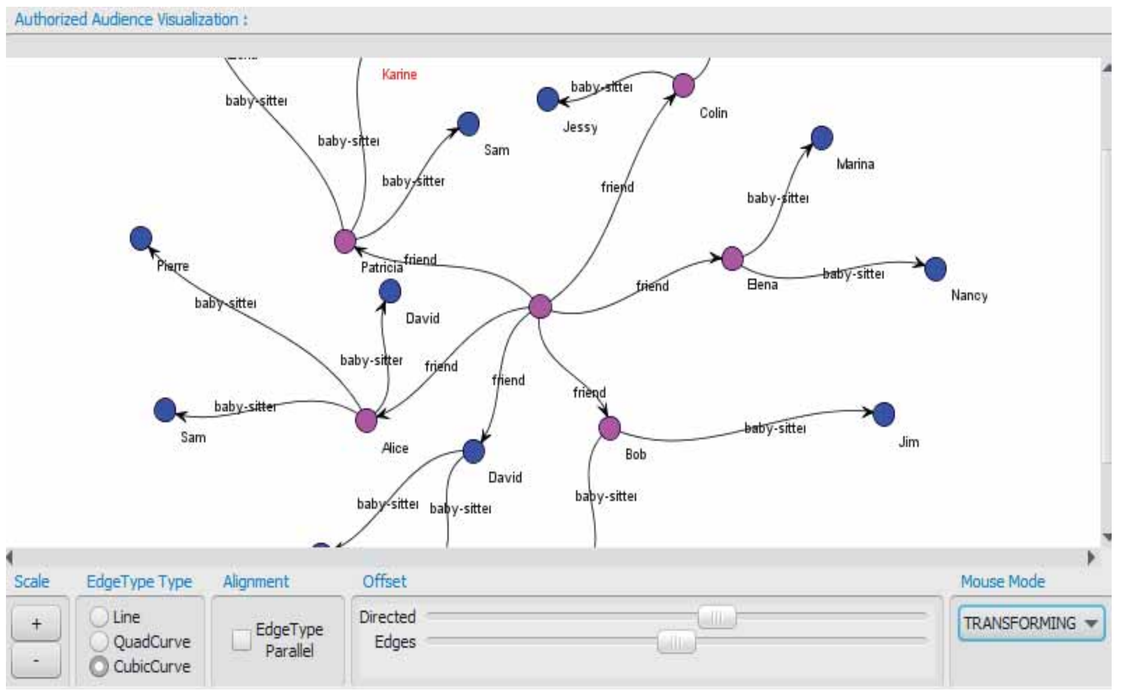
- **Example :**
  - p = Friend $^{+}$ [1, 2][city = Paris]/BabySitter $^{+}$ [1][city = Paris]
  - t_{min} = 0.8

Demonstration

- **Dataset :** LiveJournal social network, ~5 Million users, ~80 Million edges
- **Access Rule Specification :**



- **Authorized Audience Visualization :**



A Probabilistic XML Merging Tool

Talel Abdessalem
Télécom ParisTech
Paris, France

Mouhamadou Lamine BA
Université Cheikh Anta DIOP
Dakar, Senegal

Pierre Senellart
Télécom ParisTech
Paris, France

http://dbweb.enst.fr/

What this tool aim at...

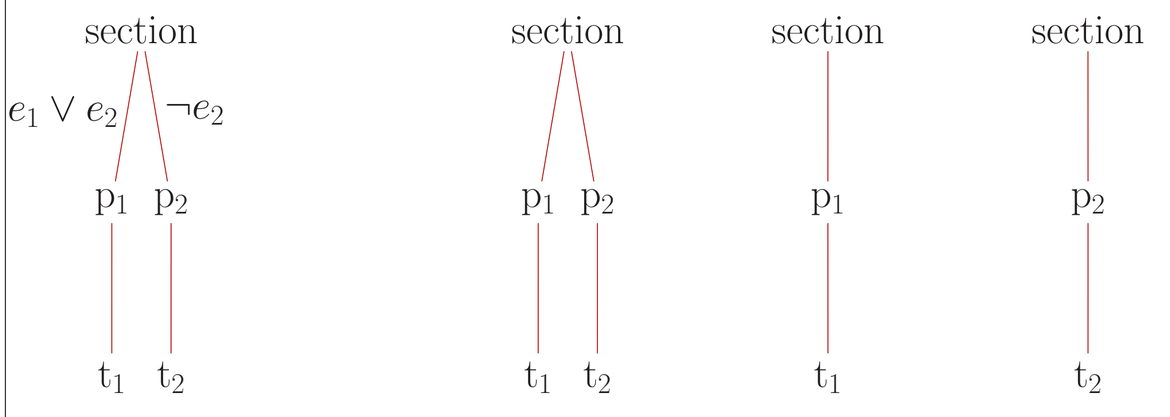
– Representing the outcome of semi-structured documents integration as a probabilistic tree

– Evaluating the uncertainty (modeled as probability values) of the result of the merge

– Querying the probabilistic repository with a subset of the XPath query language

Application domain: Wikipedia revisions
The tool enables merging the revisions of a given Wikiepdia page with:
→ an efficient evaluation of the uncertainty of the obtained result
→ an automatic management of conflicts.

Probabilistic Documents



$\Pr(e_1) = 0.7$
$\Pr(e_2) = 0.6$
P-Document

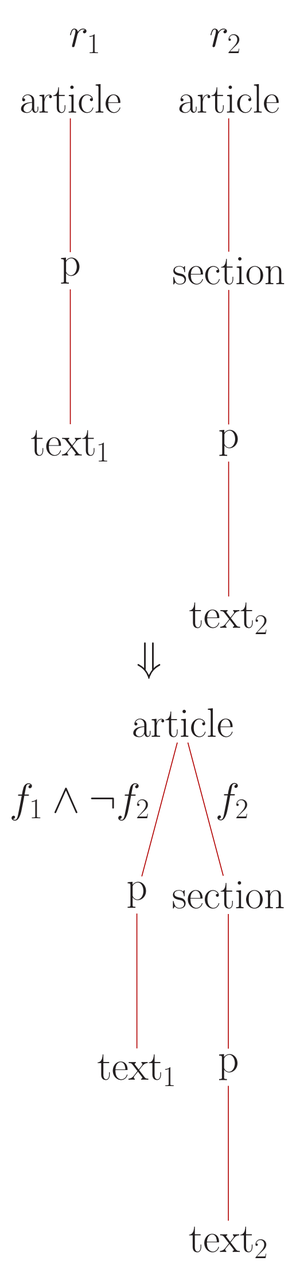
0.28 | 0.6 | 0.12

Corresponding possible documents and their probabilities

Merging of Wikipedia revisions

– A two-way tree merging technique for P-Documents

– Two steps: Matching of Revisions and Merging Matches



The result of the merge process

1. Matching of Revisions

Input: two revisions r_{k-1} and r_k and their associated event formula.
Output:
– **Deleted nodes x: $x \in r_{k-1}$ and x has no match in r_k.**

– **Added nodes x: $x \in r_k$ and x has no match in r_{k-1}.**

– **Matched couples (x, y): $x \in r_{k-1}$ and $y \in r_k$ match.**

2. Merging Matches

– **Deleted nodes:**

$$fie_{\text{new}}(x) = fie_{\text{old}}(x) \wedge (\neg f_k)$$

– **Matched couple:**

$$fie_{\text{new}}(x) = fie_{\text{old}}(x)$$

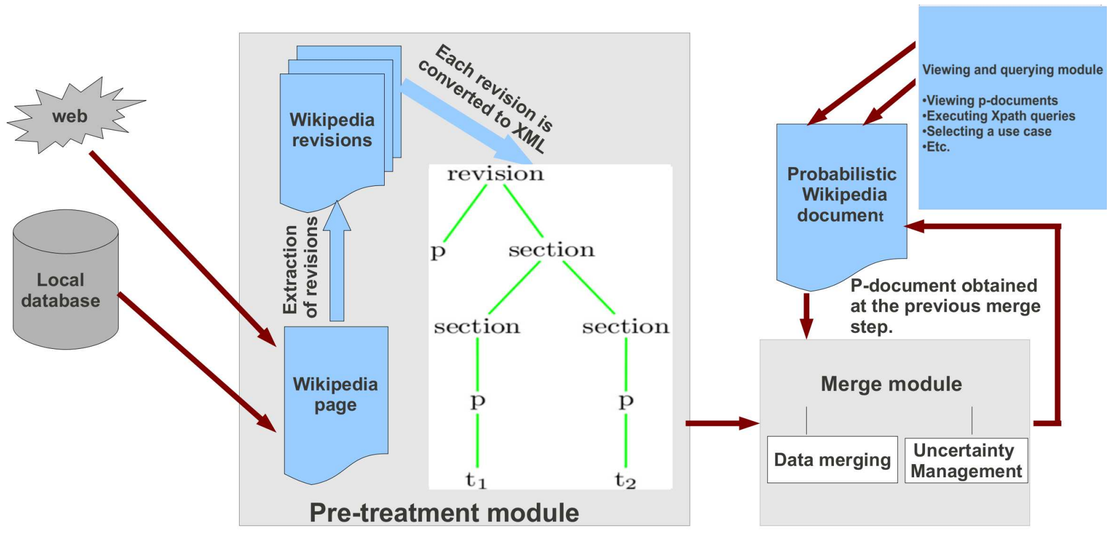
– **For added nodes:**

$$fie_{\text{new}}(x) = f_k$$

or

$$fie_{\text{new}}(x) = fie_{\text{old}}(x) \vee f_k$$

Architecture of the system



Description of the system

– System for managing Wikipedia documents.

Features

– A keyword-based search engine for Wikipedia pages
– Extracting the revisions of a given page
– Selecting the list of revisions to merge
– Building one's own Wikipedia article
– Displaying the result of the merge
– Demonstrating a certain number of use cases
– Using a subset of XPath query language

On-Line Learning Gossip Algorithm in Multi-Agents Systems with Local Decision Rules

Parties prenantes

MINISTÈRE DE LA DÉFENSE DGA

TELECOM ParisTech

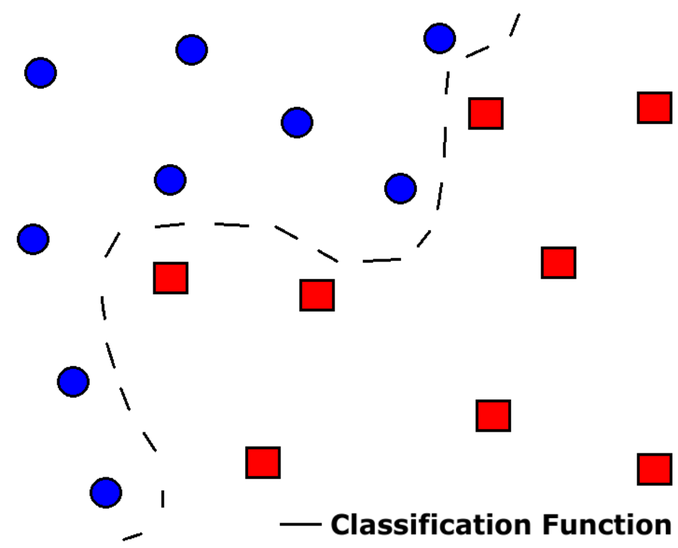
Auteurs

S. Clémençon, P. Bianchi, G. Morral et J. Jakubowicz

BACKGROUND

Motivation and applications

- Problem : investigate the **binary classification**
- Context : processing **BigData** for **statistical learning**
- Solution : implement in an **on-line** and **distributed** fashion



Learning problem

Given a training data set $(\boldsymbol{X},\boldsymbol{Y}) = (X_i, Y_i)_{i\geq 1}$, find the best prediction rule $sign(H^*)$ such the classifier function $H(X, \boldsymbol{\theta})$:

$\boldsymbol{\theta}^* = \min_{\boldsymbol{\theta}} R_\varphi(H(X, \boldsymbol{\theta}))$ → *minimizes the Risk function* R_φ

Particular case :

i) quadratic cost φ : $R_\varphi(H(X, \boldsymbol{\theta})) = ½\, E[\, (1- \boldsymbol{Y}H(X, \boldsymbol{\theta}))^2 \,]$

ii) mixture of experts : $H(X, \boldsymbol{\theta}) = \Sigma_j \theta_j\ h_j(X)$

PROPOSED DISTRIBUTED LEARNING

On-line Learning Gossip Algorithm (OLGA)

- A **distributed stochastic gradient descent** approach where the estimated parameter sequence $(\boldsymbol{\theta}_n)_{n\geq 1}$ is performed in 2 steps :

  [Gossip step] At iteration n, each agent i transmits $X_{n,i}$ to all randomly selected neighbours j with *probabiliy p* and obtain $h_j(X_{n,i}, \theta_{n-1,j})$

  [Local descent step] each agent i update its estimated parameter $\theta_{n,i}$ as follows :

$$\theta_{n,i} = \theta_{n-1,i} + \gamma_n \nabla_i h_i(X_{n,i}, \theta_{n-1,i})(Y_{n,i} - Y'^{(V)}_{n,i})$$

where

- $(\gamma_n)_{n\geq 1}$ is a decreasing step size sequence
- the r.v. $Y'^{(V)}_{n,i}$ *is an unbiased estimate of the global decision* $H(X_{n,i}, \theta_{n-1,j})$ *given by*

$$Y'^{(V)}_{n,i} = h_i(X_{n,i}, \theta_{n-1,i}) + 1/p\ \Sigma_j \delta^j_{n,i} h_j(X_{n,i}, \theta_{n-1,j})$$ *and* $(\delta^j_{n,i})$ are independent Bernoulli r.v 's $B(p)$

Asymptotic behaviour

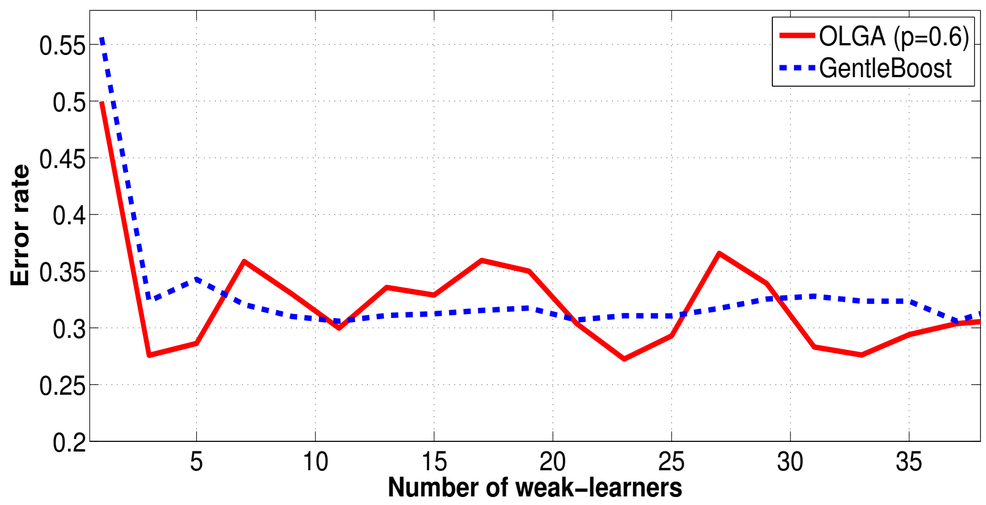
Theoretical results (under suitable assumptions)

- ✓ Consistency : $(\boldsymbol{\theta}_n)_{n\geq 1}$ **convergence *a.s.*** to the set of stationary points of R_φ
- ✓ Conditional Central Limit Theorem : qualify the error variance excess

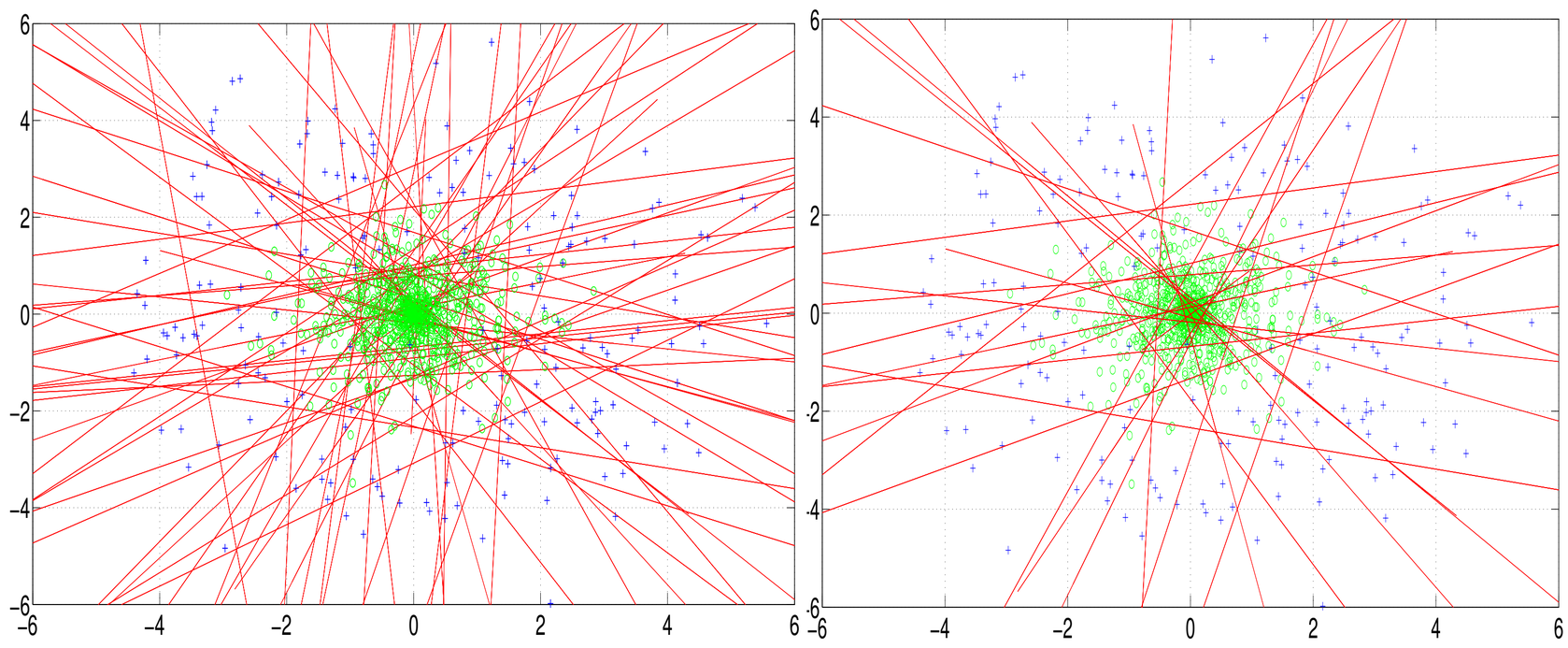
$$\gamma_n^{-1/2} (\theta_n - \theta^*) \rightarrow N(0 , \Sigma(\Gamma^*))$$

Γ^* : error in a centralized case **+ error excess induced by sparsification**

- ✓ The average network throughput is **reduced** by a factor **(1-p)**



Performance comparison between the centralized GentleBoost and OLGA for a benchmark dataset



Result classification with OLGA (-) on a simulated binary dataset (+ and o) using weak classifiers

Left : OLGA
Right : OLGA with **agent selection**

at each iteration n, each agent i of V declares idle or **active** under a suitable criterion → time-varying agents set $\boldsymbol{V_n}$

✓ Reduce redudancy classifiers and keep the relevants

26-27 Mars 2014 3ème colloque de l'Institut Mines-Télécom, « Numérique : Grande éc

Contact stephan.clemencon@telecom-paristech.fr

Presented at the IEEE Internation Conference on BigData 2013

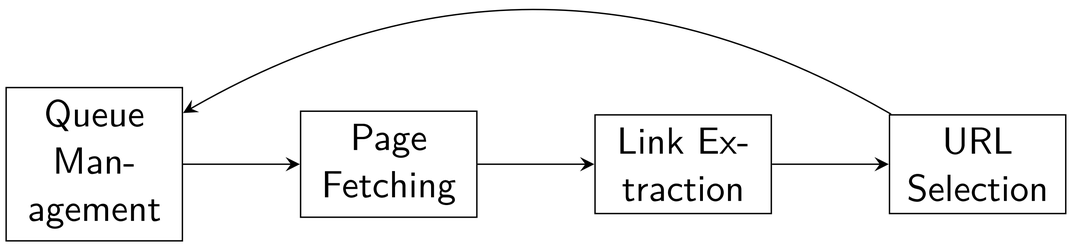
Demonstrating Intelligent Crawling and Archiving of Web Applications

Muhammad Faheem
Institut Mines–Télécom
Télécom ParisTech; CNRS LTCI
Paris, France
muhammad.faheem@telecom-paristech.fr

Pierre Senellart
Télécom ParisTech
& The University of Hong Kong
Hong Kong
pierre.senellart@telecom-paristech.fr

Traditional crawler

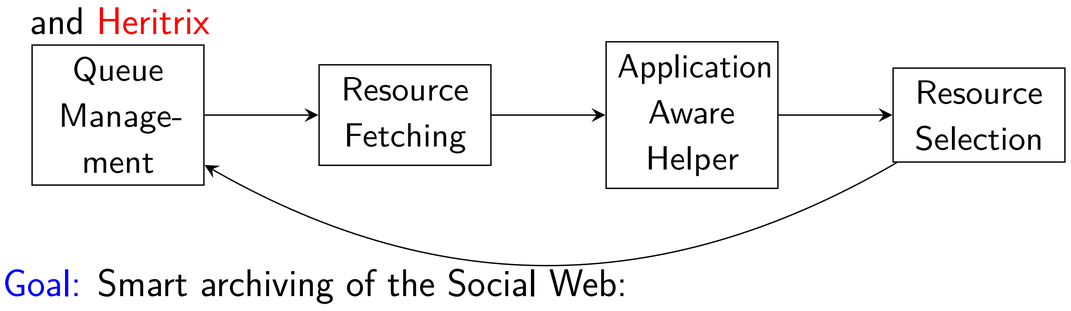
Traditional crawling: independent of the nature of the sites and their content management system



$\Longrightarrow$ Many HTTP requests, no guarantee of content quality

Application-aware helper

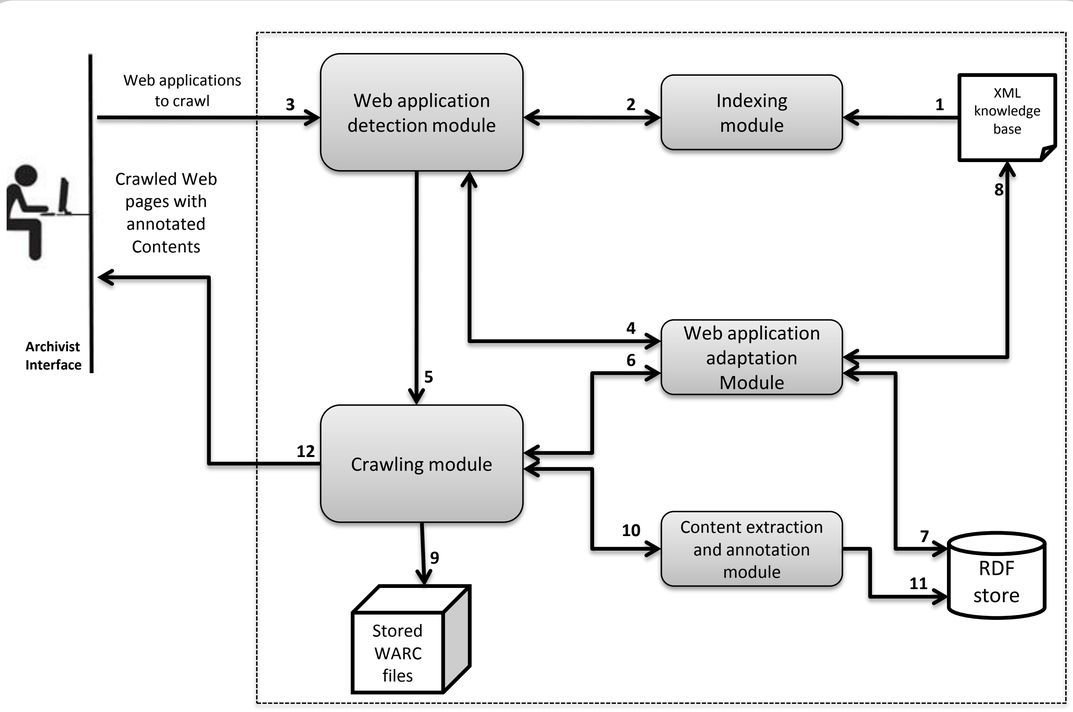
- Different crawling techniques for different Web sites
- Detect the type of Web application, kind of Web pages inside this Web application, and decide crawling actions accordingly
- Directly targets useful content-rich areas, avoids archive redundancy, and enriches the archive with semantic description of the content
- Implemented in 2 Web crawlers: Internet Memory Foundation crawler and Heritrix



Goal: Smart archiving of the Social Web:

1. Performing intelligent Crawling
2. Archiving Web objects

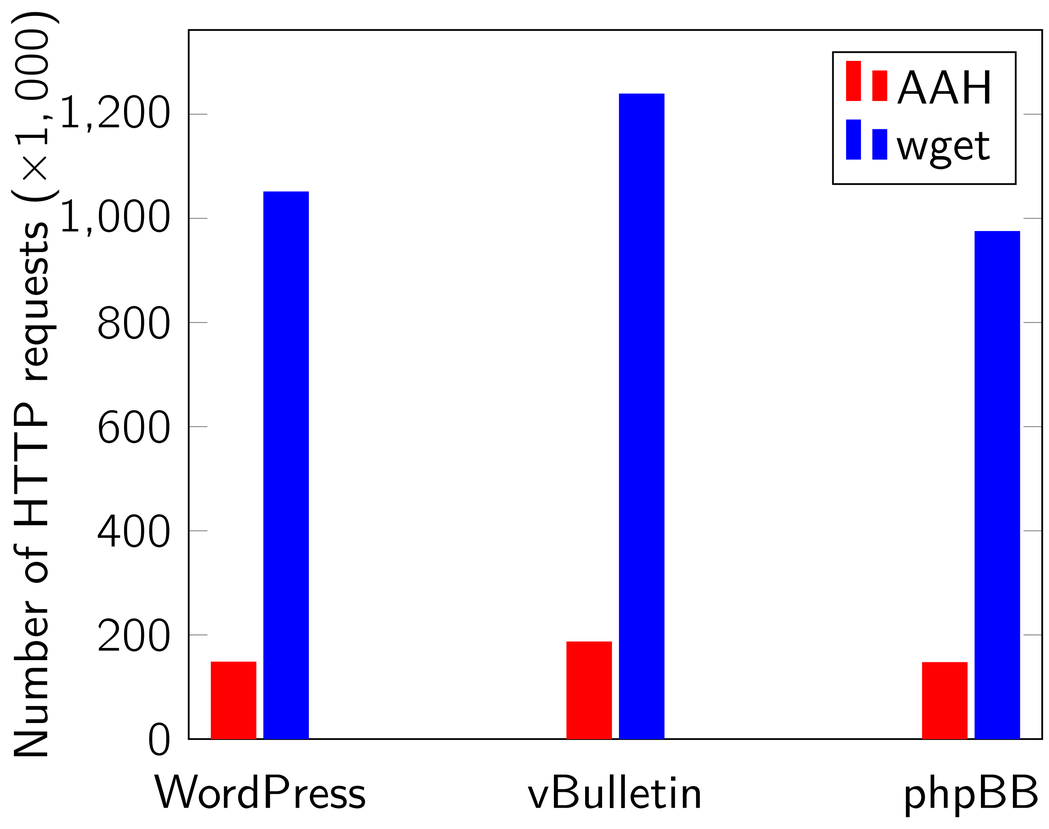
Architecture



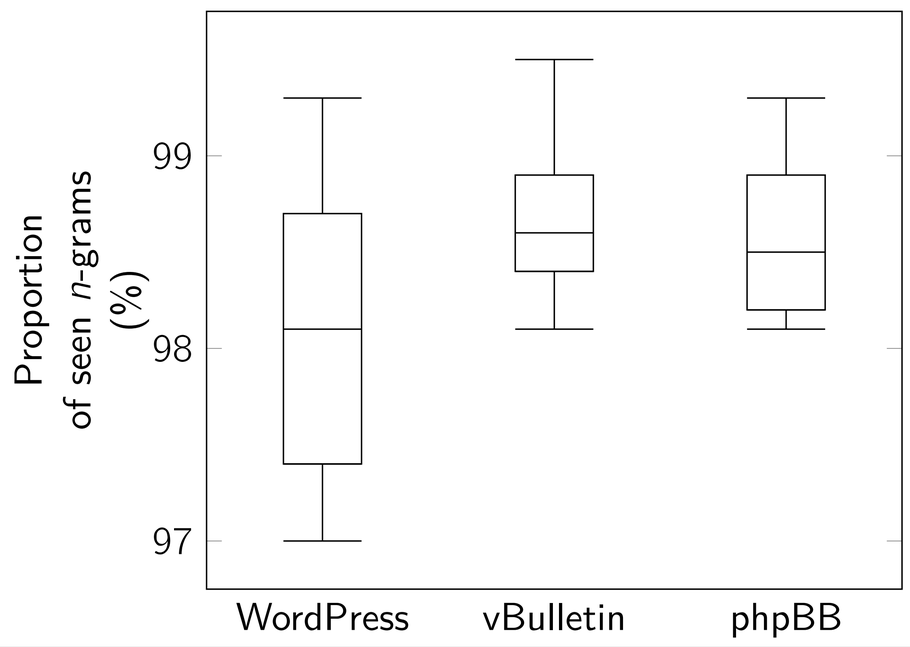
Methodology

- Knowledge base of known Web application types, algorithms for flexible and adaptive matching of Web applications to these types
  - Declarative, XML-based format
  - Integrated with YFilter for efficient indexing of KB.
- Type detected using URL patterns, HTTP metadata, textual content, XPath patterns, etc. E.g., vBulletin Web forum: `contains(//script/@src,'vbulletin_global.js')`
- Different crawling actions for different kinds of Web pages under a specific Web application
- Crawling action: not just a list of URLs; can be any action that uses REST API, complicated interaction with AJAX-based application, and extracts semantic Web objects

Crawl efficiency



Crawl effectiveness



CrowdMiner: Mining association rules from the crowd

Yael Amsterdamer Yael Grossman Tova Milo Pierre Senellart

TEL AVIV UNIVERSITY

TELECOM ParisTech

Introduction

- ***Crowd data sourcing*** collects data from the crowd, often by asking questions
- We want to learn about new domains from the crowd
  - E.g., health-related habits in some population
- Data is not recorded anywhere
- The contents of the domain are unknown
  - Discover what is **interesting** about this domain

What should we ask the crowd?

Data mining for the crowd?

- The discovery of data patterns in databases is done by **data mining.**
- Not suitable for our case
  - People do not remember enough details!

For example, it is unrealistic to expect people to remember every activity they did in the past, everything they have eaten, etc.

- They are far more likely to remember **personally prominent patterns**

"I drink red wine about once a week"

The model

We learn *association rules* of the form $a,b \rightarrow c,d$

- E.g. , *"heartburn"* $\rightarrow$ *"baking soda", "lemon"*

The answers contain

- **Rule support** – frequency of a,b,c,d
- **Rule confidence** – frequency of c,d given a,b
- **Items** (for an open question)

- **Significant rules** – average user support and confidence exceed fixed thresholds
- Users treated as random samples

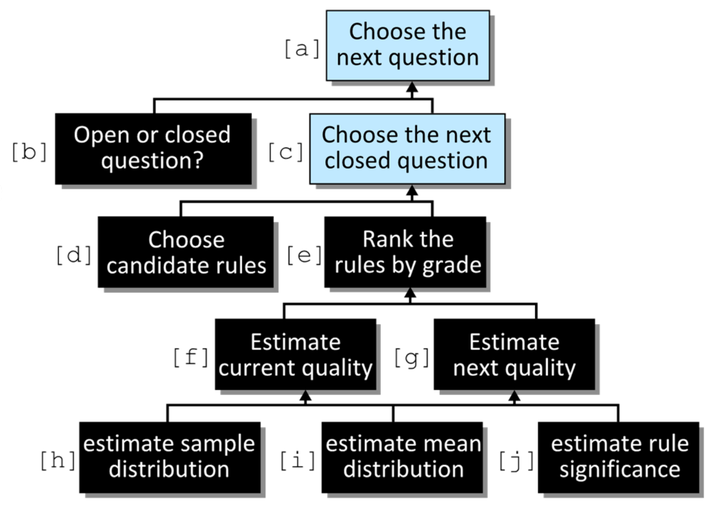
Our approach

- Use **personal summaries** to learn about **general trends**
- Treat individual answers as samples
- Combine two types of questions
  - **Open questions** *"Complete: When I feel* tired*, I usually* go for a walk*"*
  - **Closed questions** *"When you have a heartburn, do you take baking soda and lemon?"*
- Easier for users to answer
- Help digging deeper into their memories

We develop a system prototype *CrowdMiner* that interactively decides what to ask in order to discover significant data patterns

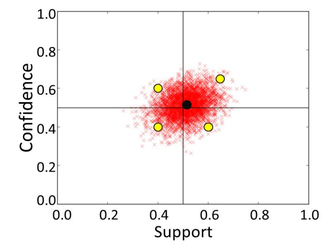
Choosing the Questions

A hierarchy of components that allow estimating the effect of the next question and choosing accordingly



Error Estimations

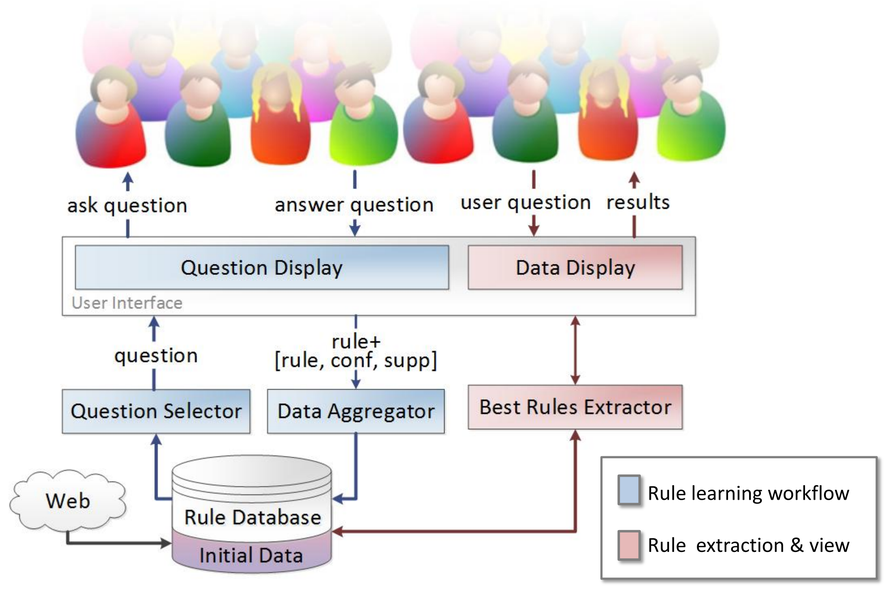
- Not all the users can be asked about every rule
- We want to estimate the probability of making an error – given the current knowledge
  - We learn a distribution of the answer support and confidence
  - **Significance estimation** – by the position of >0.5 of the distribution mass
  - **Error probability** – for the true mean to be on the other side of the thresholds
- The next question is the one expected to minimize the overall error



Well-Being Portal

- Learn about the **health habits** of others – by browsing the portal
  - Sports activities, eating habits, natural treatments
  - …
- Portal users are occasionally prompted with **questions**
  - About their personal habits
  - Computed by our algorithm
- User **answers** are processed to deduce rules (associations) between well-being concepts in the portal
- The portal allows browsing the learned rules

System Architecture



Opinion Dynamics and Community Detection

Institutions

INSTITUT Mines-Télécom

TELECOM SudParis

s@movar UMR 5157

Authors

Julio Cesar Louzada Pinto

Tijani Chahed

Jérémie Jakubowicz

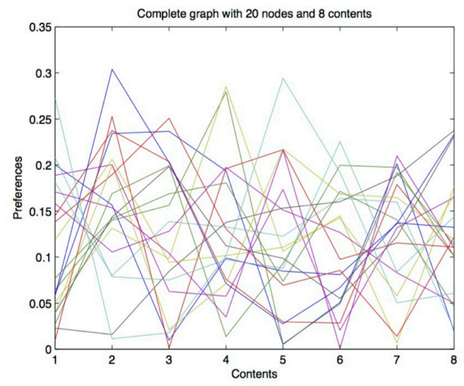
Partners

FONDATION TELECOM

INSTITUT CARNOT Télécom & Société numérique

Objectives

- Develop a stochastic opinion dynamics model with multiple contents and study the asymptotic behavior in simple cases.
- Develop a community detection (graph clustering) algorithm.



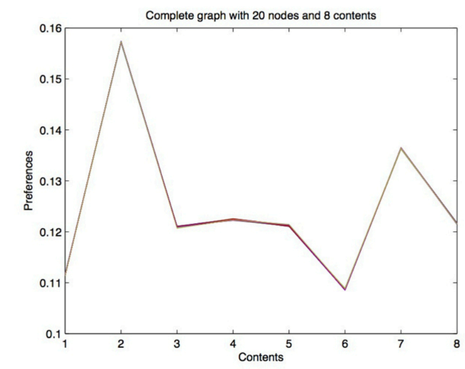
Model

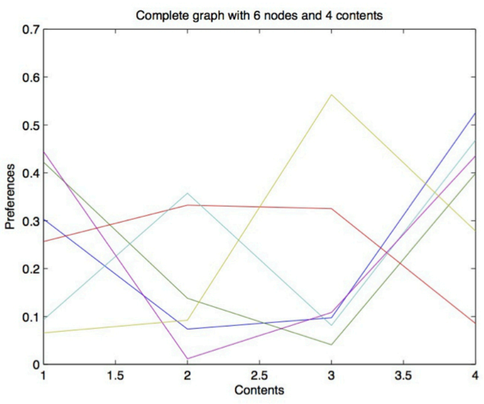
- *N* agents communicate about *K* contents via a graph *G = (V,E)* with inward adjacency matrix *A*.
- Agent *i* has score $X_t^{i,k}$ for content *k* at time *t.*
- *Preferences* $P_t^{i,k}$ are normalized scores, i.e., $P_t^{i,k} \propto X_t^{i,k}$.
- Agents update their scores linearly as

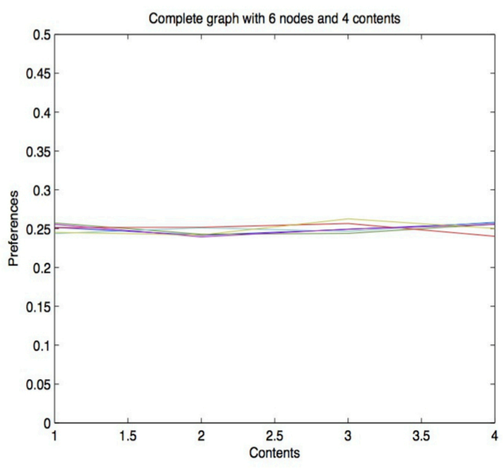
$$X_{t+1}^{i,k} = X_t^{i,k} + \sum_{j \to i} 1_{u_{t+1}^j = k}$$

with $P\left(u_{t+1}^j = k \mid F_t\right) = \left(f\left(P_t\right)\right)_{jk}$ the probability of agent *j* broadcasting content *k.*

- Function $f : \left(\Delta_K\right)^N \to \left(\Delta_K\right)^N$ models the way agents choose the contents to broadcast, where $\left(\Delta_K\right)^N$ is the set of the *NxK* stochastic matrices.

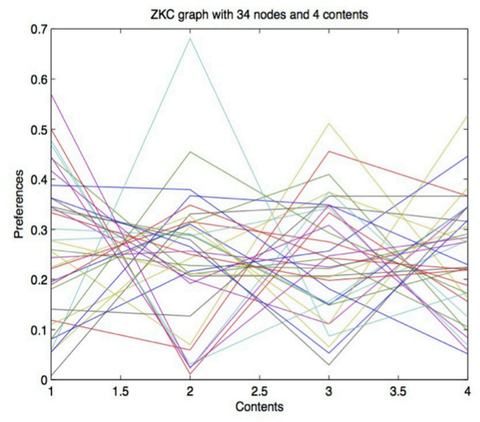


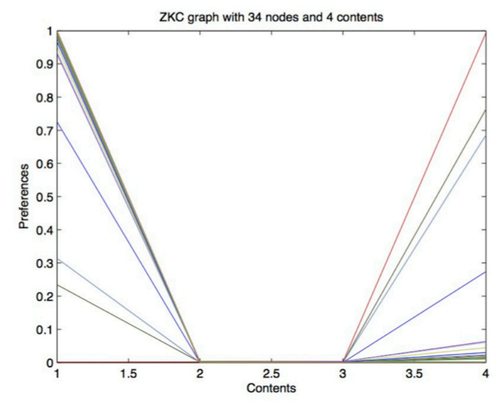




Types of function f

- f(x) = x – identity function.
- $\left(f\left(x\right)\right)_{ik} = \dfrac{e^{\beta x_{ik}}}{\sum_l e^{\beta x_{il}}}$ - soft-max function.
- $\beta \ll 1 \to \left(f\left(x\right)\right)_{ik} \sim \dfrac{1}{K}$
- $\beta \gg 1 \to \left(f\left(x\right)\right)_{ik} \sim 1_{k = argmax_j x_{ij}}$



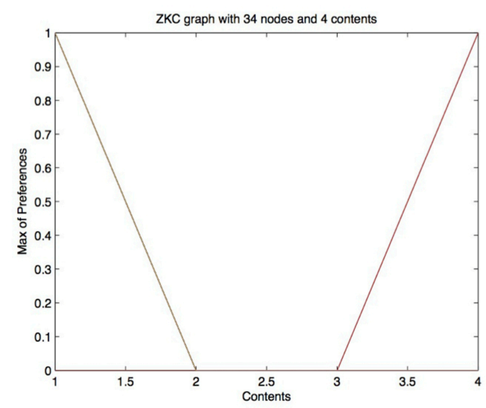


Results

- For f(x) = x: there exists a random variable $P_\infty \in \left(\Delta_K\right)^N$ such that $P_t \to P_\infty$ almost surely.

If *G* is strongly connected, then there exists a random variable $\pi \in \Delta_K$ such that $P_t \to 1\pi^T$ almost surely .

- For soft-max exponential with $\beta \ll 1$: If $inf_i \sum_j A_{ij} > 0$, then there exists a $\beta_{min} > 0$ such that for all $\beta \in \left[0, \beta_{min}\right]$ we have that $P_t \to \dfrac{11^T}{K}$ almost surely.
- If $\beta \gg 1$, then we expect the graph to be clusterized in communities which broadcast the same content; the one with the maximum preference. This creates an algorithm for community detection.
- Publication: A stochastic opinion dynamics model with multiple contents, CDC, Firenze, December 2013.



Journée Futur et Ruptures

January 2014

Contact julio.louzada_pinto@telecom-sudparis.eu Site web www-public.it-sudparis.eu/~louzada

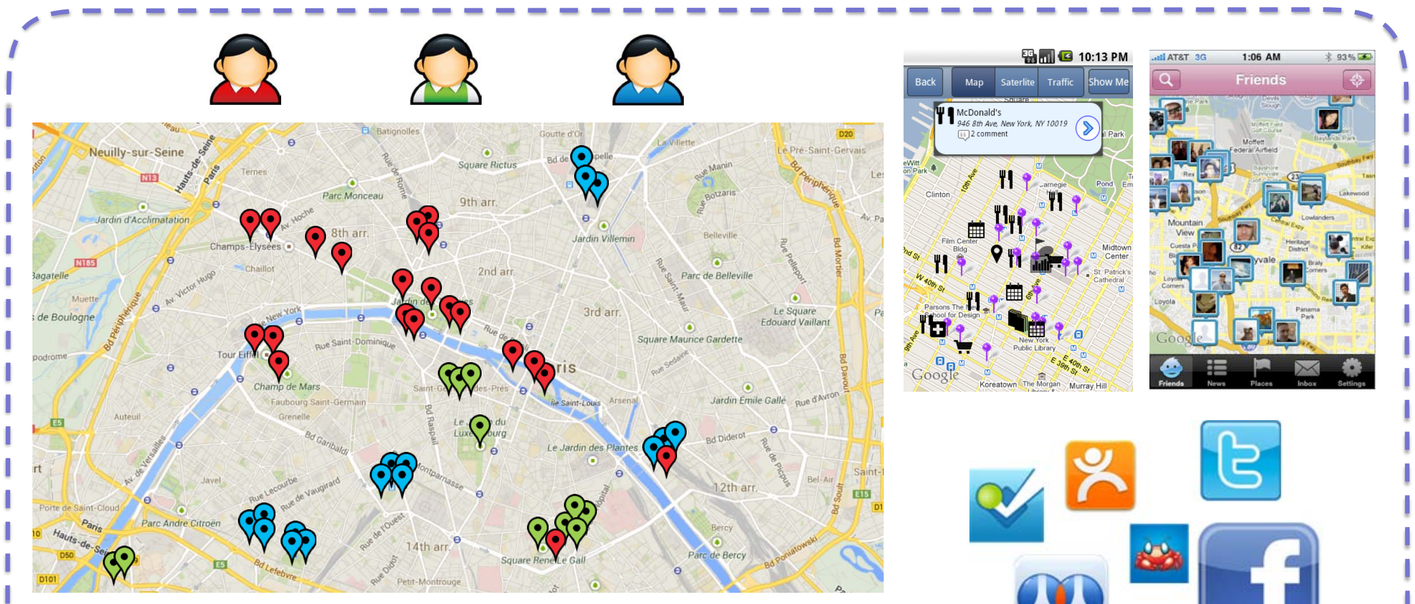
From Big Social Media Data to Service Personalization

Authors : Daqing Zhang, Dingqi Yang, Zhu Wang, Zhiyong Yu

Location Based Social Network Data Analytics

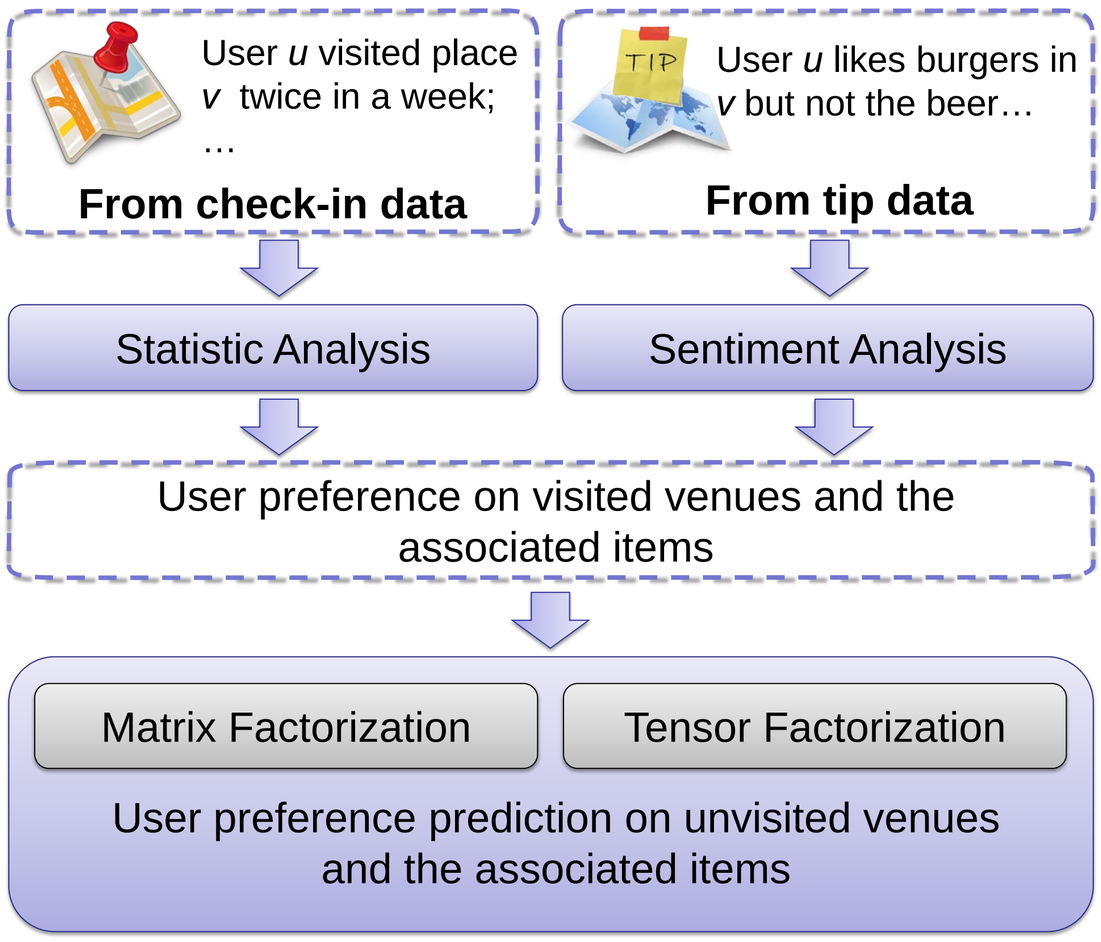
- With the increasing popularity of location based social networks, users generated significant volume of heterogeneous social media, e.g.,
  - Texts
  - Photos
  - Videos
  - Presences
  - …

- These digital footprints massively contain users' fine-grained preference
- Understanding this user preference can enable ubiquitous, personalized location based services.



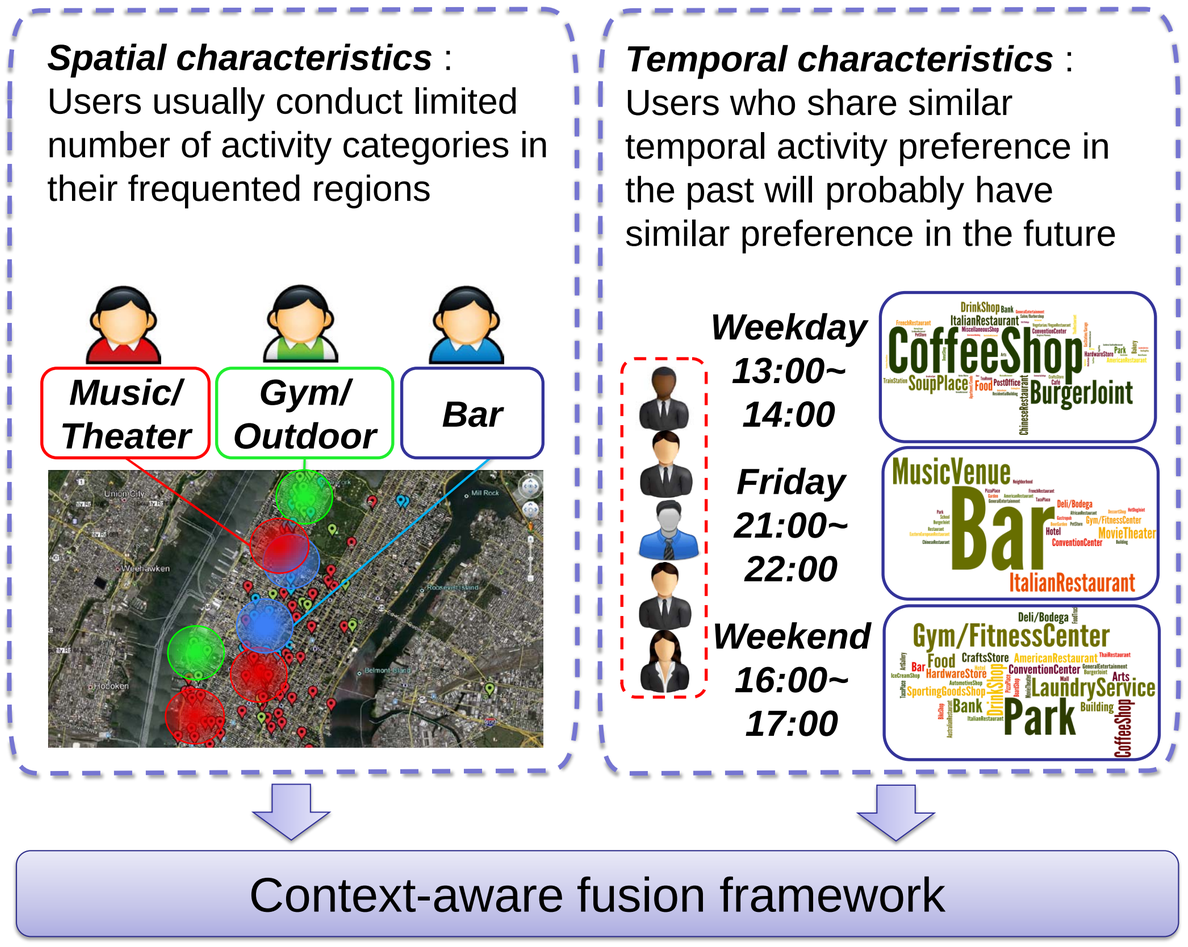
Preference Awareness

- Extracting fine-grained user preference on venues from heterogeneous data.
- Predicting user preference on unvisited venues.



Context Awareness

- Studying and modeling spatial temporal characteristics of user activity.
- Inferring a user's interest according to his current context.



An Example of Personalized Location Based Services

Contact Daqing.Zhang@telecom-sudparis.eu
Site web http://www-public.it-sudparis.eu/~zhang_da/

A Context-Risk-Aware Recommender System

Institut Mines-Télécom

TELECOM SudParis

s@movar UMR 5157

cnrs

Authors

Djallel Bouneffouf, Amel Bouzeghoub and Alda Gançarski

Context

s@movar UMR 5157 — TELECOM SudParis — cnrs — Mines-Télécom Institut

Nomalys

Objectives and Issues

Motivations

- Provide personalized and context-aware recommendations in mobile environments
- Consider content dynamicity and user's situations risk level

Key Challenges

- Infer higher-level goals from low-level observed operations
- Handle cold start and sparseness effect
  - Requires a large amount of information in order to make accurate recommendations
- Exploration vs. exploitation tradeoff
  - How to sacrifice a short term small reward to privilege larger rewards in the long term?
  - How to associate the situations risk level to the exploration/exploitation tradeoff?

Key Words & Key Technologies

- **Context-Aware Recommender Systems** (CARS) combine characteristics from context-aware systems and recommender systems in order to provide personalized recommendations to users in ubiquitous environments.
- **Machine learning** algorithms can be used to learn models and predict documents
  - **Reinforcement learning** is learning what to do: how to map situations to actions
- In probability theory, the **multi-armed bandit problem** models an agent that simultaneously attempts to acquire new knowledge and optimize her decisions based on existing knowledge
  - In each round, a learner takes an **action** (or **arm**) and in return receives a numerical **reward**
  - The goal is to optimize action-selection policy to maximise the total reward received
  - The learner needs to **explore** (try) the different actions and **exploit** the seemingly most rewarding arms
  - In practice, the learner has access to **contextual information** in each round to infer which action leads to the highest rewards

Contributions

Models

- Modeling the context-aware recommender system as a bandit algorithm
- Modeling user, context, situation and risk

Algorithms

- A new semi-uniform strategy: contextual-epsilon-greedy strategy
  - Combining content-based filtering and reinforcement learning
- An algorithm R-UCB
  - Computes the probability of exploration by using the situation risk level $R(S^t)$
  - Three methods of risk computing
    - Using situations similarity (R_m)
    - Using situations concepts (R_c)
    - Using a Gaussian distribution (R_v)

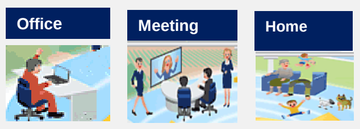
General Approach

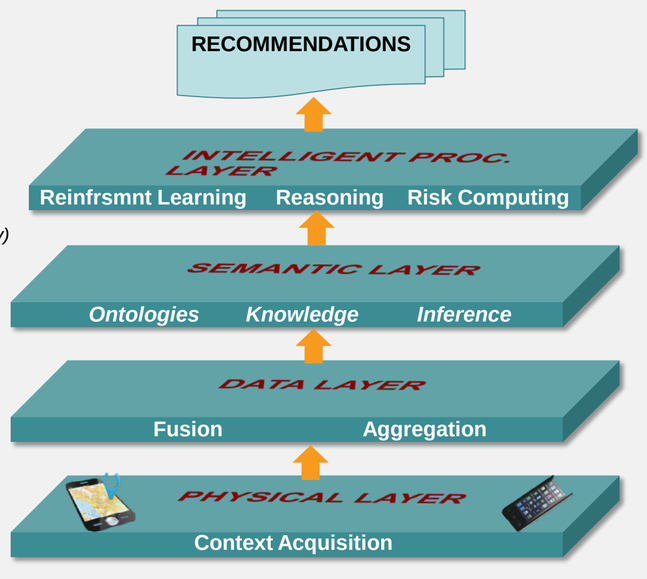
High level user's situation

$O_{Location}.Evry$, $O_{Temps}.afternoon(awd)$, $O_{sociale}.Company)$

Low level user's situation

("2.86, 1.34 ", Mon Oct-3-13 :10 :00-2013, "Arval")





Evaluation Results

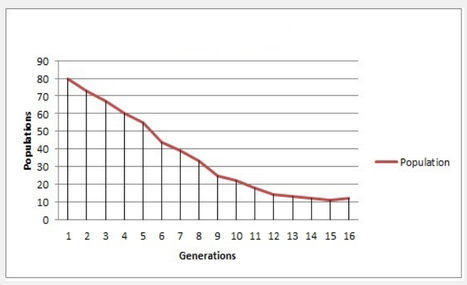
Data Set and Parameters

Nomalys Data Set

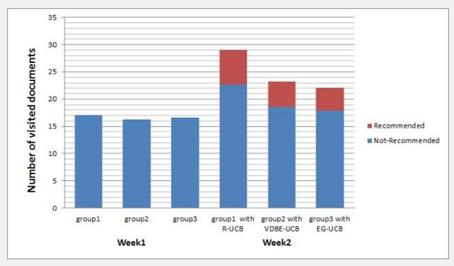
- 356 738 situations
- 5 518 566 navigations data
- 3500 users

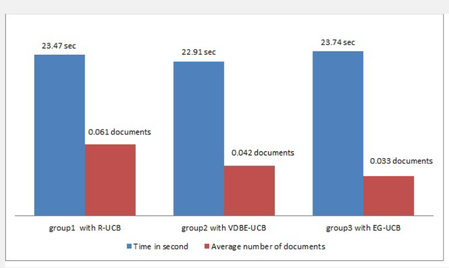
Genetic Algorithm

- Population: 80 Chromosomes
- Chromosomes: ϵ_{min} , ϵ_{max}, Threshold
- Results: $\epsilon_{min} \in [0.05; 0.13]$, $\epsilon_{max} \in [0.47; 0.56]$, Threshold $\in [0.7; 0.82]$

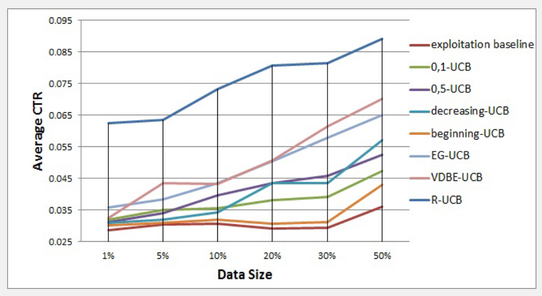


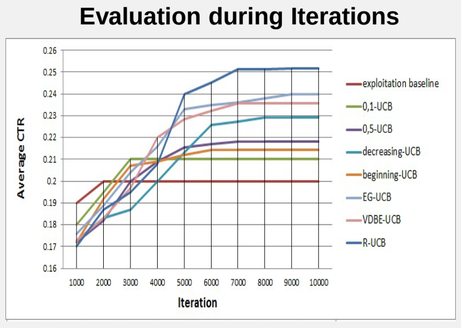
Online Evaluation

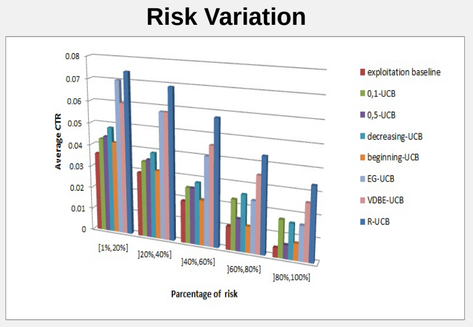




Offline Evaluation







4. SECURITE, SURETE ET RISQUES

POSEIDON (2012-2015) integrated security policies for outsourced data

Parties prenantes

Auteurs

Caroline Fontaine (Lab-STICC, CNRS + Télécom Bretagne)

Frédéric Cuppens (Lab-STICC, Télécom Bretagne)

Nora Cuppens-Boulahia (Lab-STICC)

Gouenou Coatrieux (LaTIM, Télécom Bretagne)

David Gross-Amblard (IRISA, Univ. Rennes 1)

Sébastien Gambs (IRISA,Univ. Rennes 1-INRIA)

Nicolas Prigent (IRISA, Supélec)

Partenaires

Context

Outsourced data / Cloud:

- **Various contexts**: outsourced storage and omputation, multimedia content distribution (e.g., Video on Demand), personal data management.
- **Various protagonists**: industry, administration, citizens, social networks .

All of end users are concerned with the same security issues: confidentiality, integrity, authentication, copyright protection, privacy and anonymity. These issues are traditionnally addressed with the help of security and cryptographic mechanisms, e.g. encryption, signature, watermarking, etc.

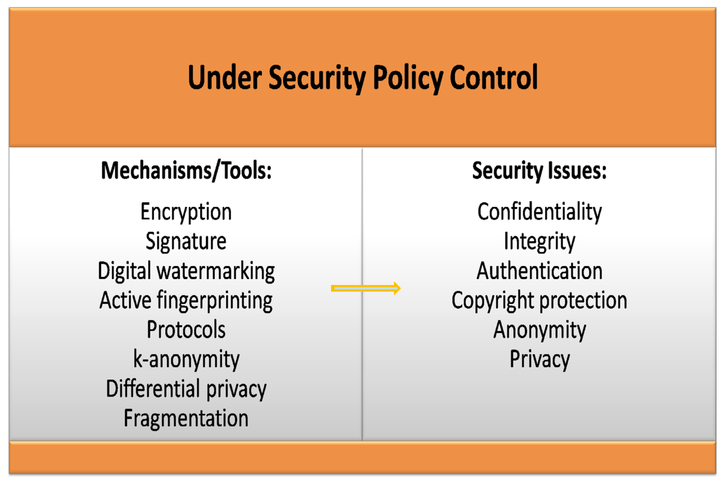
A security policy formalizes the security expectation with respect to the system. It specifies the involved entities, the data and services to protect, the threats. It conditions the actions choices and deployments of the security mechanisms.

While these mechanisms are known to be efficient when used independently, they often have to be combined.

Hence, to ensure a good security level of outsourced data, we need:

1. A **formal expression of the Security Policy**.
2. **Adapted security mechanisms** and a formal expression of the security properties each of them may guarantee.
3. An extension of this formalism to properly state the consequences of the **combination of several such mechanisms**. This is essential to enable an **automated deployment** of the policy: automated selection of the mechanisms depending on the context and related security priorities, automated analysis of possible incompatibilities.

Of course all these formalisms must be compliant with each other.



First Results

- The first track concerns the **study and improvement of security mecanisms** related to data or request privacy in the Cloud. In particular, we focused on the following ones:

  Fully Homomorphic Encryption schemes.

  Anonymous delivery protocol for multimedia content, which enables both privacy and traceability of malicious users.

- The second track focused on the design of a support tool allowing, **for a given security policy, selection of the best mechanism or combination of mechanisms** to enforce this security policy.

- The third track concerns the **adaptation of security solutions to the particular contexts of Cloud and peer-to-peer networks**

Selected publications (more on www.poseidon.cominlabs.ueb.eu)

Recent advances in homomorphic encryption: a possible future for signal processing in the encrypted domain, C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat, R. Sirdey. IEEE Signal Processing Magazine, Number 2, Volume 30, pp. 108-117 (2013).

Preserving Multi-relational Outsourced databases Confidentiality using Fragmentation and Encryption, Bkakria, A., Cuppens, F., Cuppens-Boulahia, N., Fernandez, J.M., Gross-Amblard, D. Journal of Wireless Mobile Networks, UbiquitousComputing, and Dependable Applications (JoWUA) (2013).

Robust Lossless Watermarking of Relational Databases Based on Circular Histogram Modulation, J. Franco-Contreras, G. Coatrieux, F. Cuppens, N. Cuppens-Boulahia, C. Roux, IEEE Transactions on Information Forensics and Security 9(3): 397-410 (2014)

Contact caroline.fontaine@telecom-bretagne.eu

Site web www.poseidon.cominlabs.ueb.eu

A Joint Watermarking/Encryption Algorithm for Verifying Medical Image Reliability in Both Encrypted and Spatial Domains

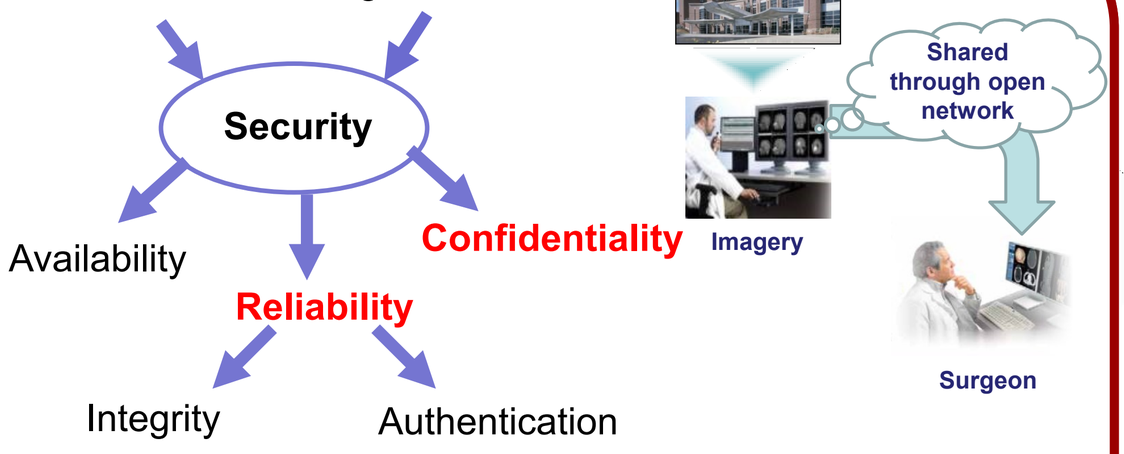
Partners

Authors

Dalel BOUSLIMI
Gouenou COATRIEUX
Michel COZIC
Christian ROUX

MEDICAL DATA PROTECTION

Rules of ethics Legislatives rules



Reliability: *a degree of confidence/trust in medical data*

Encryption (“a priori” protection) Watermarking (“a posteriori” protection)

Combination

Joint Watermarking/Encryption Algorithm

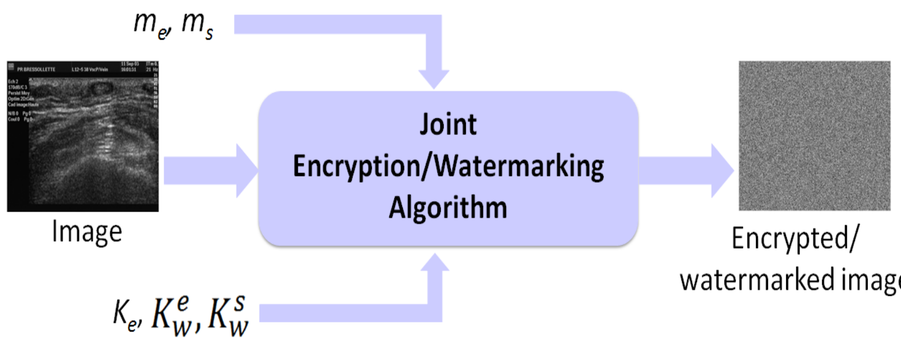
Constraint: interoperability with/transparent to the DICOM standard

⇨ *Use of the AES block cipher in CBC mode.*

SYSTEM ARCHITECTURE

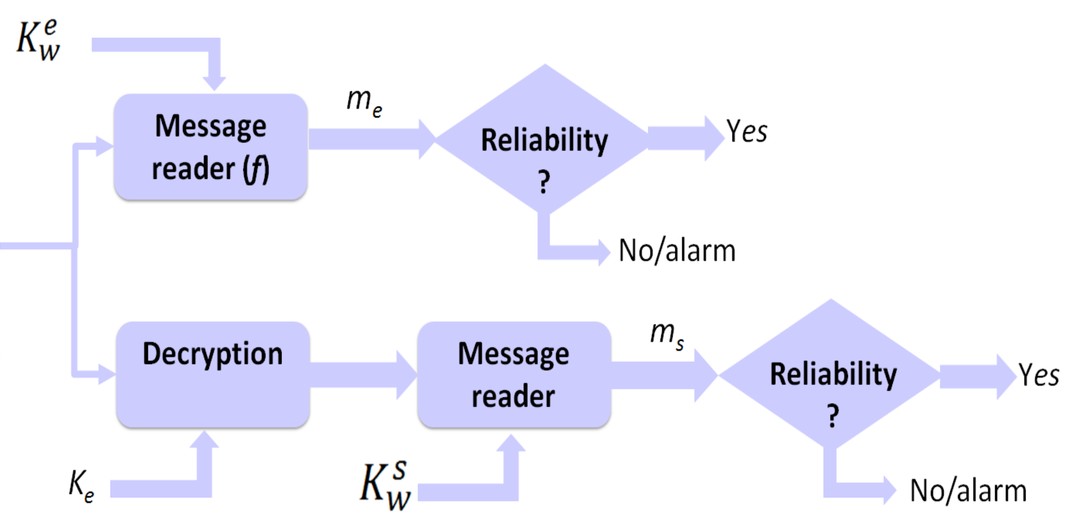
Protection

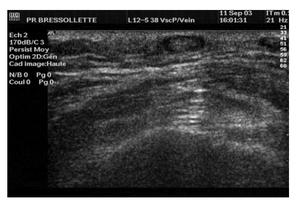
• m_s and m_e : messages available in the spatial and the encrypted domains, respectively.



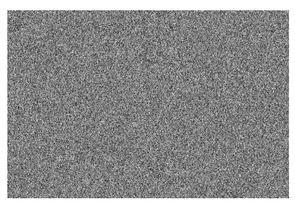
K_w^e, K_w^s : watermarking keys, K_e: encryption key

Verification

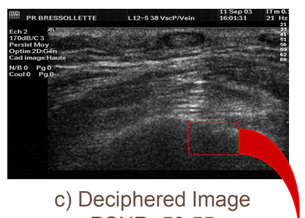




a) Original Image, Entropy=6,76 bits/pixel



b) Joint Watermarked/ ciphered Image



c) Deciphered Image PSNR=53,55

d) Zoom in image difference between (a) and (c)- gray levels values mapped in the range 0-255.

JOINT WATERMARKING/ENCRYPTION

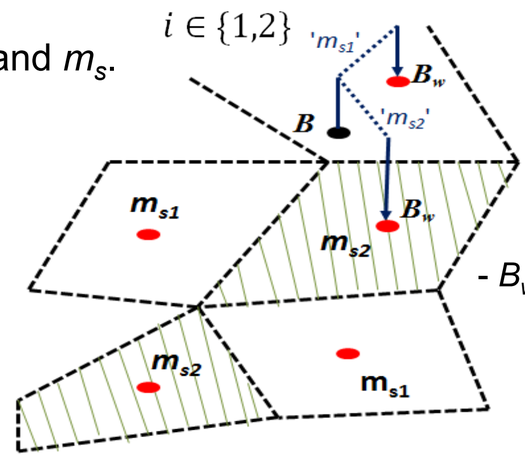
- Modification of the quantization index modulation (QIM):
Disrupt/modulate the image pixels to encode simultaneously m_e and m_s.

-**QIM** : insertion based on codebooks C_{ms_i}, which represents the message m_{si}.

- **QIM/chiffrement** Constitution of sub-codebooks $C_{ms_i me_j}$ according to the AES.

$$C_{m_{si}} = \bigcup_{j=1}^{q} C_{m_{si}m_{ej}} \text{ et } C_{m_{si}m_{ej}} \cap C_{m_{si}m_{ek}} = \emptyset, j \neq k$$

$$C_{ms_i me_j} = \{Y \in C_{ms_i} / \ f(AES(Y, K_e), K_w) = me_j\}$$



$$C_{m_{si}} = \bigcup \emptyset_{m_{si}},$$
$$C_{m_{si}} \cap C_{m_{sj}} = \emptyset, i \neq j$$
$$B_w = \{B_w \in C_{m_{si}} / \|B - B_w\| = \min(\|B - Y\|), Y \in C_{m_{si}}\}$$

- B_w : the watermarked vector B.

EXPERIMENTAL RESULTS

- **Performance Indicators**
  - Image distortion measure: PSNR (dB).
  - Capacity rate (bpp: Bit Per Pixel).
- **100 ultrasound images- 576 × 688 pixels, 8-bit depth.**
  - Capacity rate: 1/16 bpp in each domain.
  - PSNR: greater than 60dB.

CONCLUSION

-The proposed joint encryption/watermarking algorithm guarantees *a priori* as well as *a posteriori* protection.
-The use of the AES in CBC mode makes our method transparent and compliant with the DICOM Standard

- Message insertion introduces very low image distortion.
- **Future works** will focus on making our scheme more robust to attacks like lossy image compression (ex. JPEG),

Contact Gouenou.Coatrieux@telecom-bretagne.eu Site web http://latim.univ-brest.fr/

Lossless Watermarking based on Circular Interpretation of Bijective Transformations for the Protection of Medical Databases

Objectives/Solution/Results: Verify the reliability (authenticity, integrity) of **medical relational databases** / A fragile/robust lossless watermarking algorithm based on a circular interpretation of bijective transformations embedding a message within **numerical data** of a relational database / Our method preserves the value of the database while allowing the embedding of a digital signature or an authentication code for verifying the database integrity and origins (even if the database is modified – traitor tracing).

Parties prenantes

TELECOM Bretagne

Auteurs

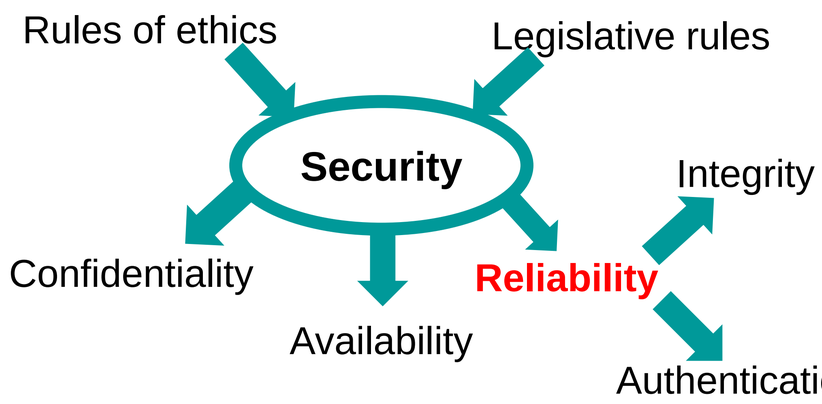
J. Franco-Contreras
G. Coatrieux
N. Cuppens-Boulahia
F. Cuppens
C. Roux

Partenaires

Région Bretagne

1. MEDICAL DATA PROTECTION

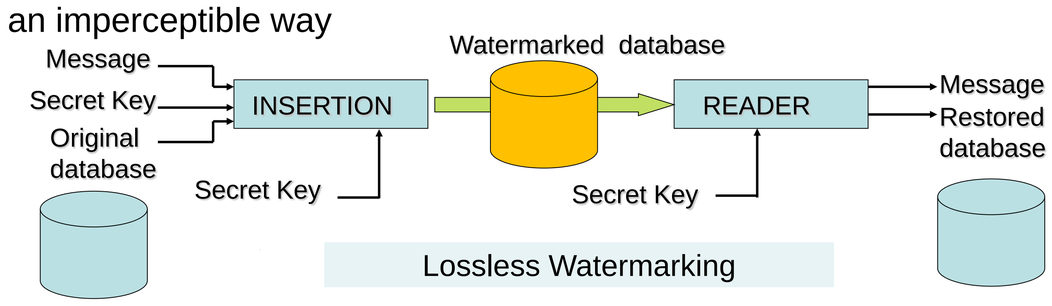
Security Objectives



Reliability: *a degree of confidence/trust in medical data*

Watermarking

Allows the embedding of a message within a content by modifying its values in an imperceptible way

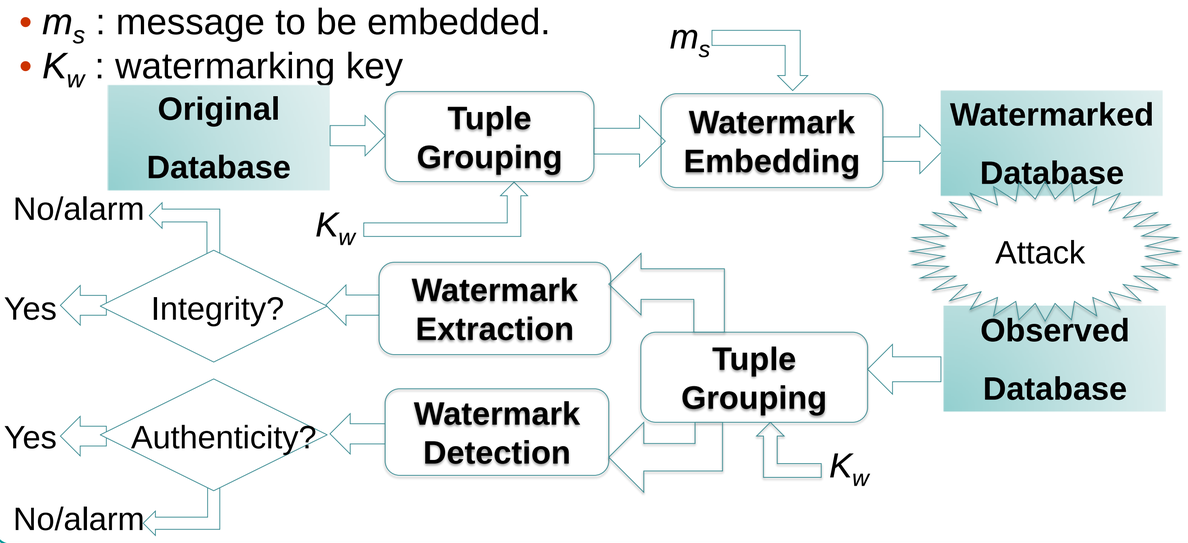


Constraint: Do not perturb the normal interpretation of data

Lossless or reversible property allows watermark removal and exact data restoration.

2. A COMMON DATABASE WATERMARKING CHAIN

- m_s : message to be embedded.
- K_w : watermarking key



Tuple Grouping → Make embedding independent from database storage

- Reorganization of tuples $\{t_u\}_{u=1...Mu}$ in N_g groups depending on their primary key ($t_u.PK$) and a secret watermarking key K_w.
- Each tuple is assigned to the group

$$n_u = H(K_w \mid H(K_w \mid t_u.PK)) \bmod N_g.$$

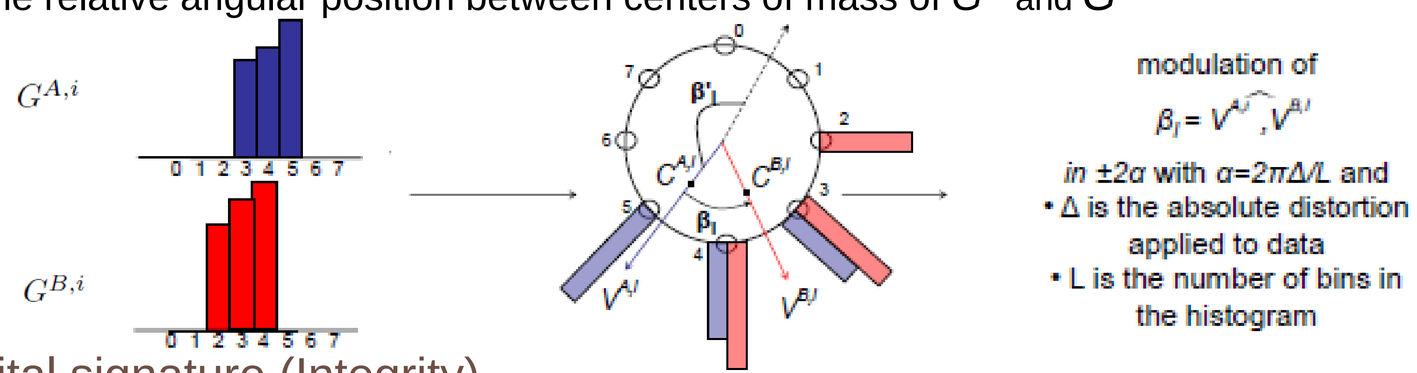
  - H : Cryptographic HASH operation
  - $\mid$: concatenation operation

- One bit is ideally embedded in each group.

3. PROPOSED METHOD

Modulation Principles

- One group is uniformly split into two subgroups ($G^{A,i}$, $G^{B,i}$).
- Histograms for a **numerical attribute** in each subgroup are mapped into a circle
- Modification of the relative angular position between centers of mass of $G^{A,i}$ and $G^{B,i}$



Fragile Embedding → Insertion of a digital signature (Integrity)

$$\beta_i^w = \begin{cases} \beta_i + 2\alpha \text{ if } b=0 \longrightarrow +\Delta \text{ to the values in } G^{A,i} \text{ and } -\Delta \text{ to the values in } G^{B,i} \\ \beta_i - 2\alpha \text{ if } b=1 \longrightarrow -\Delta \text{ to the values in } G^{A,i} \text{ and } +\Delta \text{ to the values in } G^{B,i} \end{cases}$$

- Digital Signature is extracted from the observed database and compared to the one computed from the data.

Robust Embedding → Insertion of an Authentication Pattern (Reliability)

- Same modulation principle but with the insertion of a secret pattern.
- Detection based on correlation (origin) and/or extraction of the pattern (integrity).

*** Special case (Non carriers)**

Groups where $|\beta_i| > 2\alpha$ (as in β'_i) cannot carry information (**non-carriers**) and

$$\beta_i^w = \begin{cases} \beta_i + 2\alpha \text{ if } \beta_i > 0 \\ \beta_i - 2\alpha \text{ if } \beta_i < 0 \end{cases}$$

4. EXPERIMENTAL RESULTS

❖ Experimental Database

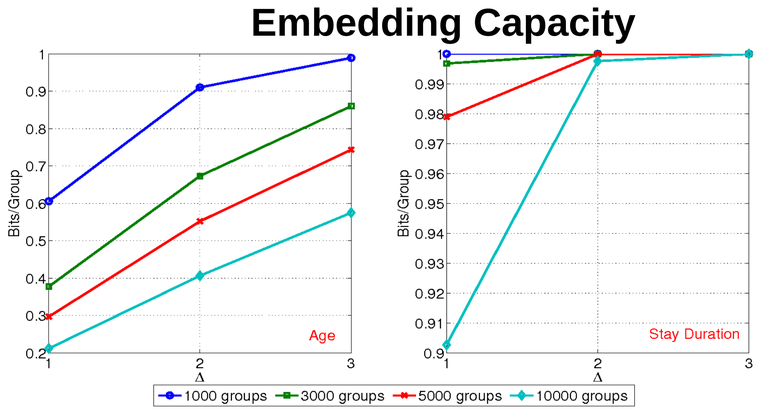
Real medical database related to inpatient stays at the hospital with 1048575 tuples

Id_hospital	Id_patient	age	Stay duration
601433878	7892	29	13
601484325	28653	40	31
601527723	14552	65	4

❖ Fragile Scheme → How much can be embedded

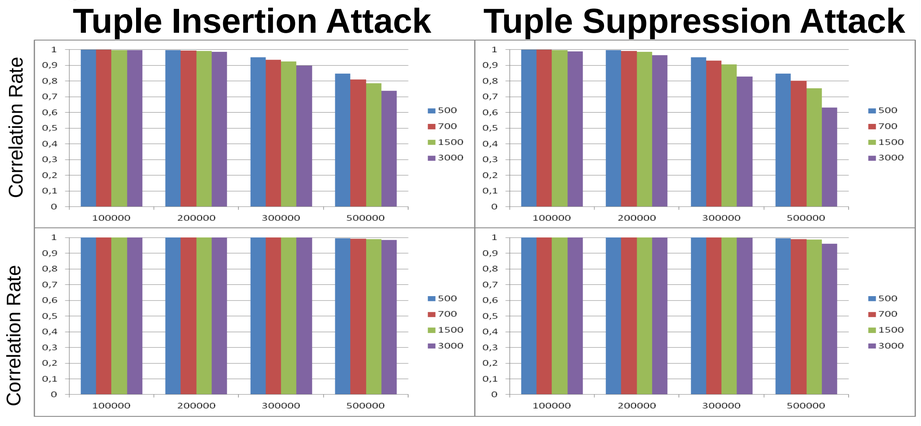
- Three values of Δ=[1,2,3] considered

Embedding Capacity



❖ Robust Scheme

- Att. Age with Δ=3; Att. Stay Duration with Δ=1

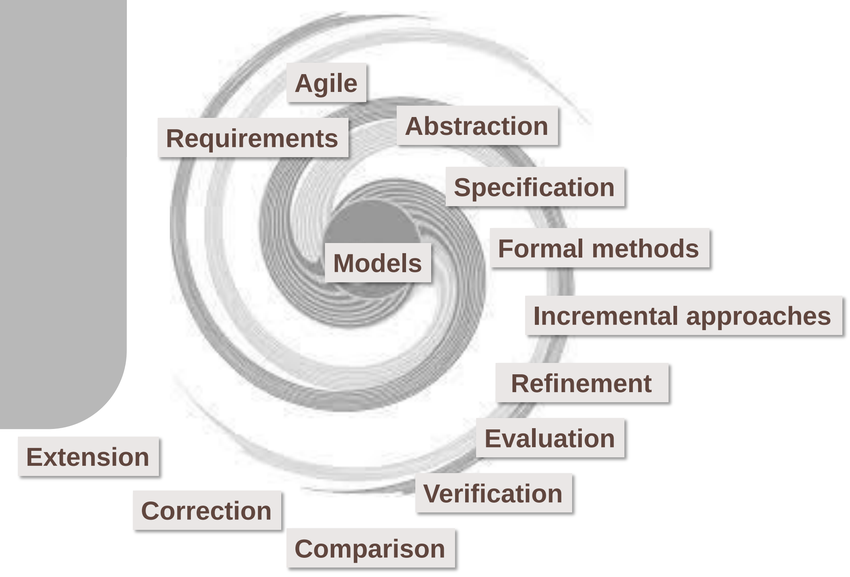


In both cases, results depend on the statistical properties of the attributes, more specifically in their standard deviations.

Contact gouenou.coatrieux@telecom-bretagne.eu
Site web http://latim.univ-brest.fr/

Mars 2014

Incremental development and behavioural verification of reactive systems



Problem

How to assist model development of critical reactive systems?

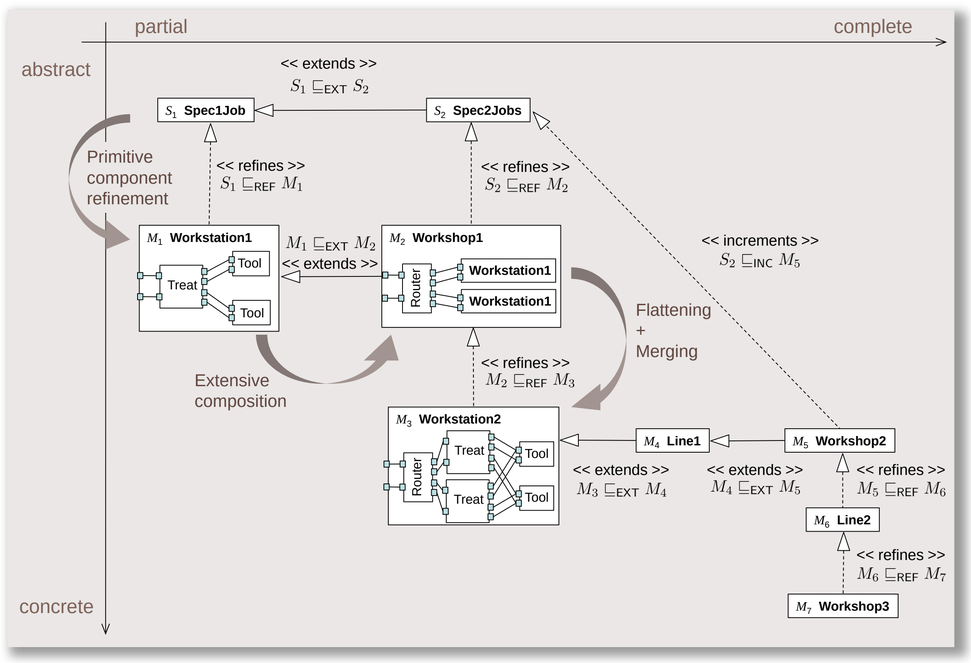
Model engineering is not mature.

- “Engineers don’t know why their system works. [...] They can not be sure a critical system is free of critical errors.” J. Sifakis
- “Today for most software systems, the analogy of building something like a cathedral is no longer a good choice. [...] Requirements change all the time, we need a short time-to-market, we need feed back all the time...” M. Lippert
- “If you want to get it right, be ready to start over at least once.” E.S. Raymond

Need to develop and **verify several** model versions:
from abstract and partial ones,
to detailed and completed ones.

IDF – Incremental Development Framework

Combining model refinements and extensions



Two sets of techniques … to support:

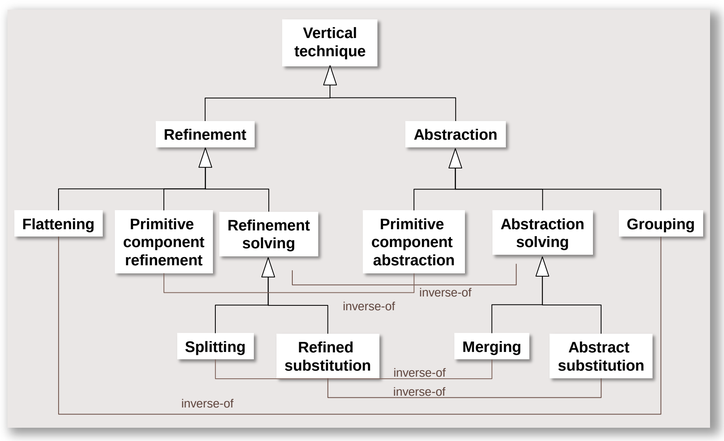
- Construction techniques
- Evaluation techniques

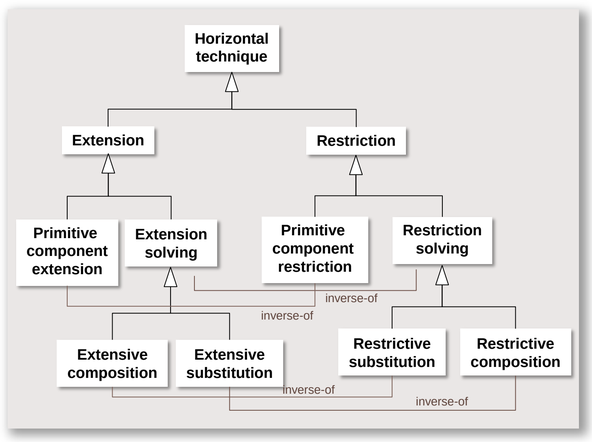
⇒ Incremental development processes

Two axis:

- abstraction level (vertically)
- completion level (horizontally)

Construction techniques





Evaluation techniques

M_2 conf M_1 M_2 is a correct implementation of M_1: M_2 preserves liveness properties of M_1.

$M_1 \sqsubseteq_{INC} M_2$ M_2 increments M_1: any implementation of M_2 is an implementation of M_1.

$M_1 \sqsubseteq_{EXT} M_2$ M_2 extends M_1: M_2 preserves liveness properties of M_1 and has more behaviours.

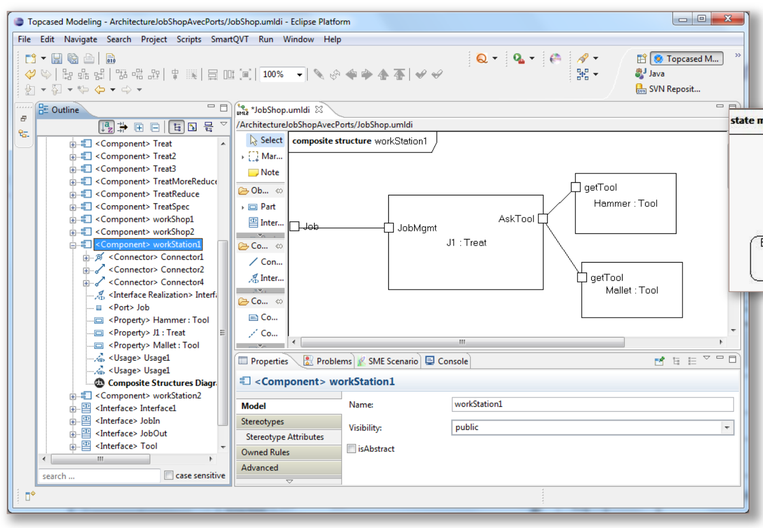
$M_1 \sqsubseteq_{REF} M_2$ M_2 refines M_1: M_2 preserves liveness and safety properties of M_1.

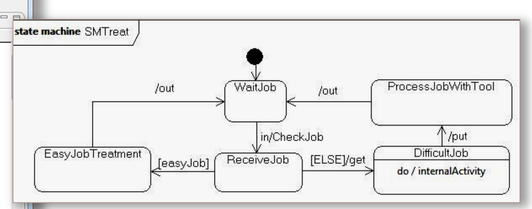
$M_1 \sqsubseteq_{SUB} M_2$ M_2 can substitute M_1: M_2 refines M_1 and can safely replace M_1.

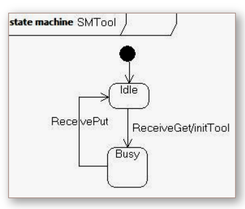
IDCM – Incremental Development of Compliant Models

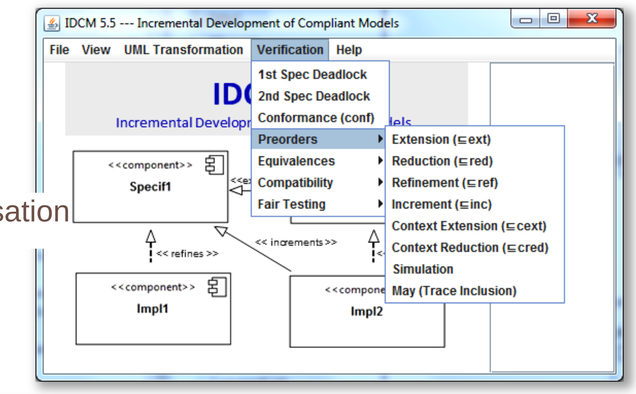
A tool to support IDF

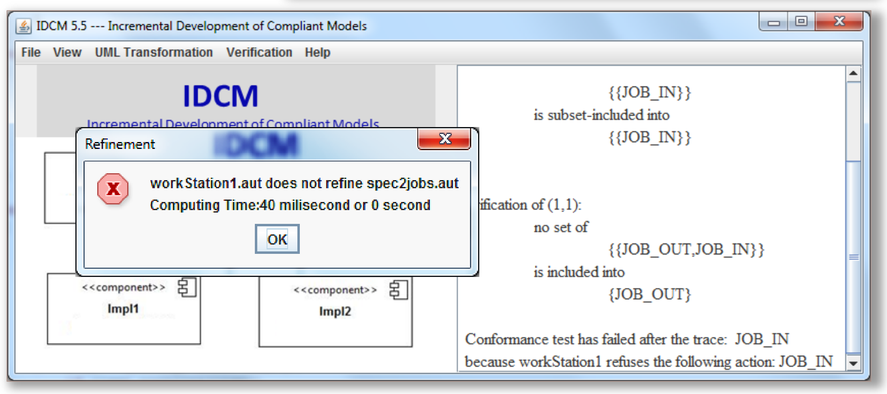
- Transformation of UML models into LTS (Labelled Transition Systems):
  - UML primary components (state machines) and
  - architectures (composite structures).
- Use of CADP (Construction and Analysis of Distributed Processes) features for LTS composition and minimisation
- Implementation of conformance, increment, extension, refinement and substitution relations
- Analysis of models pointing out traces of failure and denied actions whenever relations are not satisfied











Institution

MINES Alès

Authors

Anne-Lise Courbis
Thomas Lambolais
Hong-Viet Luong
Thanh-Liem Phan

Partners

Institut de Recherche en Informatique de Toulouse

Christian Percebois

UNIVERSITÉ MONTPELLIER 2 SCIENCES ET TECHNIQUES

Thérèse Libourel

Contact thomas.lambolais@mines-ales.fr Web site http://lgi2p.mines-ales.fr

Methodology of system interdependency analysis after a major disruption

Parties prenantes

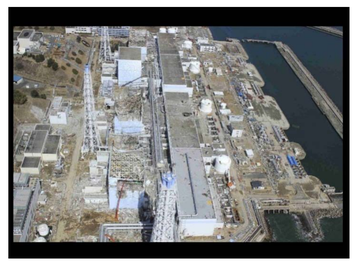
Auteurs

REY Benjamin (EMA)
TIXIER Jérôme (EMA)
DANDRIEUX Aurélia (EMA)
DUSSERRE Gilles (EMA)
LAPEBIE Emmanuel (CEA)

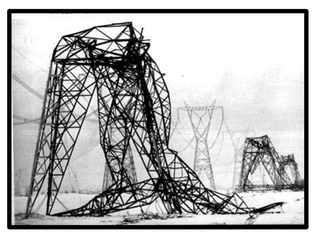
Context

■ Beyond the dramatic deaths, injured people, evacuated families due to the Japan earthquake and tsunami in 2011, another consequence was the destruction of 30% of the electricity production plants. Because of physic interdependencies, many essential activities have been affected by this production disruption (for instances chemical and petrochemical industries) and indirectly the whole country and its population.
Attacks on the World Trade Center in New York in 2001, ice storm in Canada in 1998 are other examples of cascading failures.
It enhances the need for research on the functional and spatial interdependencies in a system (territory, industrial site, organization,…). Project problematic can be formulated in order to answer to this question :

How to assess a major disruption impact in a system (organization, territory,…) composed of several interdependent elements ?



Fukushima Daiichi nuclear plant, 2011, Japan
(origin : SIPA/Ap)



Ice Storm 1998, Canada
(origin : radio-canada website)

Conceptual tools to analyze system interdependencies

What are risk management methodology characteristics about independencies between elements of a system ?

Empirical approach

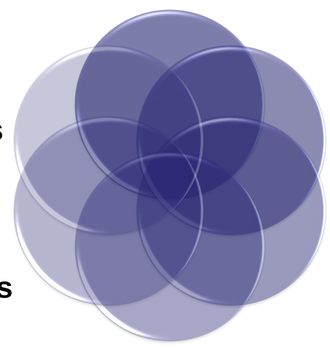
Databases,…
(Pederson and *al*, 2006)
(Yusta et *al.*, 2010)

Physic, geographic, cybernetic, logic
(Rinaldi et *al*., 2001)

Natural hazard, industrial accident, threat

Modeling interdependencies and cascading effects tools

4 Types of Interdependencies



Major disruptions

Predictive approach

Multi-agent system, dynamic system, network theory, petri network, input-output model ,…
(Pederson et *al*, 2006) (Yusta et *al.*, 2010)

Activity sectors Energy, health, food, water, transport, defense, government, telecommunication,…
(SGDN, 2008)

4 Scale levels International, national, regional, local

Physical security, security management,…
(Petit et *al.*, 2011)

Protection

Protection and resiliency measures

Resiliency

Redundancy, stocks, new resources,…
(Petit et *al.*, 2012)

Methodology

Approach

System representation

- Create system typologies of : processes, functions, physic elements, services and constraints

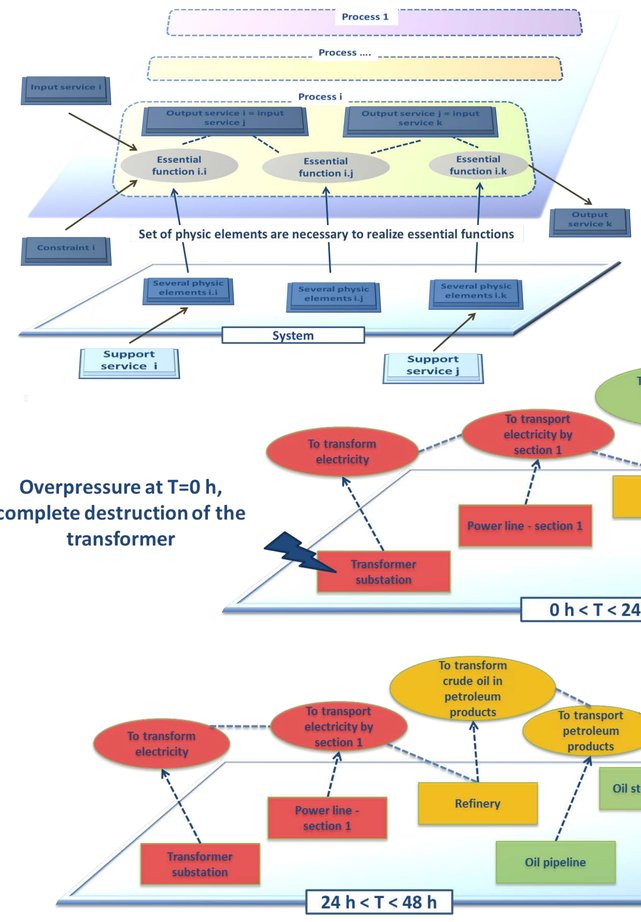
Scenarios building

- Create disruption typology
- Identify fault physic elements and failure functions with their characteristics (maximum capacity, dynamic propagation of a disruption…)

System vulnerability assessment

- Scenario simulation and results :
  1/ service disruption assessment by means of different indexes to define : the number of affected end users, the injured people, …
  2/ to highlight vulnerable elements of the system and define major scenarios

Results



Papers

■ REY B., TIXIER J., BONY-DANDRIEUX A., DUSSERRE G., MUNIER L., LAPEBIE E., (2013), Interdependencies between industrial infrastructures: Territorial vulnerability assessment, Chemical Engineering transactions, vol. 31, 2013

■ M PETIT F., ROBERT B., REY B., 2010, Protection des infrastructures critiques, Les Techniques de l'Ingénieur, Paris: Édition T.I. p.146-152.

Contact benjamin.rey@mines-ales.fr Site web www.mines-ales.fr/

Numérique : Grande échelle et complexité

March 2014

Crisis management training and debriefing

« Improvement of the organizational training and the psychosocial approaches' contribution»

Parties prenantes

MINES Alès

Auteurs

LAPIERRE Dimitri (EMA)
WEISS Karine (Unîmes)
DUSSERRE Gilles (EMA)
BONY-DANDRIEUX Aurélia (EMA)
TENA-CHOLLET Florian (EMA)
TIXIER Jérôme (EMA)

Partenaires

FONDATION TELECOM

INSTITUT CARNOT Télécom & Société numérique

Context

- Feedbacks from the nuclear power plant accident in Fukushima (Japan, 2011) and the explosion of the chemical plant AZF in Toulouse (France, 2001) underline that strategic decision-making is taken in a complex and dynamic environment, characterized by emergency. Improving crisis management of disasters, requires more effective training sessions (i.e. by using simulation game) and methodologies allowing evaluation and debriefing.

Collective failures :

- **Cognitive:** misrepresentation of the situation, *sensemaking* collapse, loss of structuring frame
- **Behavioral:** feelings, lack of understanding, block for acting, non-critical group think, disorientation
- **Organizational:** wrong execution of decisions, coordination collapse, wrong tasks' repartition, leadership deletion, lack of communication, blind support of the procedures

Crisis management's limits

Emergency

- *Collective failures*
- *Maladaptive training for actors*
- *Inexperienced decision makers*

Doubt

Training objectives
- **Facilitate the implementation of learning in real situations**
- **Promote expertise and improve teamwork skills**

Stress

Crisis management improvement

- *Development of non technical skills*
- *Decision-making improvement*
- *Higher organizational reliability*

Ambiguity

Dynamic nature of the event

Time pressure

Resilience capacity

Lack of immersion	Non immersive situation, low feeling of stress, low mediatic pressure, difficulties in mobilizing all actors
Procedural decision-making	Low awareness of the decision-making's models, rigidity of procedural decisions, poor preparation to cope with elements of surprise
Wrong consideration of needs	Few upstream studies for identifying the real needs of participants, similarities between exercises
Psychosocial factors	Low consideration of human and organizational factors, no-promotion of the collective representation of the situation
Debriefing	No clear structuration of debriefing, no match between debriefing animation and trainees' performance and reactions (orchestration level of debriefing)

Methodological elements to improve crisis management training

1. Improve training scenarios

Diversity of crisis situations

CTT
EBAT
Check list
Level adjustment
TADMUS approach
Educational training
Needs adjustments via upstream survey

2. Activate learning levers

Experiential learning

Strength of thinking quickly
Educational leader
CRM Training's tools
Decision making model
Shared Mental Models

Trainings
- **EBAT : Event-Based-Approach Training**
- **CTT : Critical Thinking Training**
- **CRM : Crew Resource Manager**

Simulation exercise

Technical skills

Non technical skills

Improvement decision making

3. Assess a group in real time

Skills mobilized

Observation grid
Questionnaire
Check list
Case test

EXTRACT

Phases	Assessment tools
During the exercise	**Check-lists** ▪ Grid of observations adapted from NOTECHS (Non Technical Skills) (objective achievement with the behavior/attitude for each skill) ▪ Grids for group behavior compared to the scenario's' events ▪ Grid of actions to carry out (technical assessment) **Assessment model of the group performance** ▪ Adapted from CRM
After the exercise	**Feedback form** ▪ Self-assessment ▪ Identification of encountered difficulties

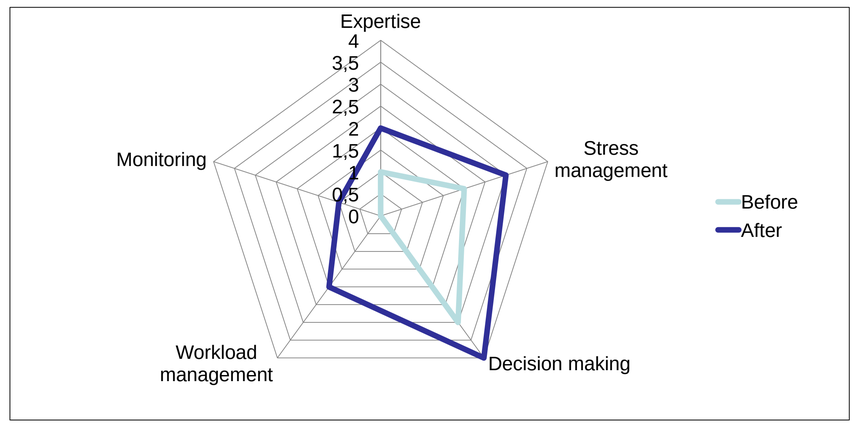
4. Organize the debriefing

Continue learning

Judgment method
Sort of debriefing
Steps'structuring
Orchestration
Feedback

Perspectives

- Index of trainee's ability to cope with crisis situation based on preliminary questionnaire highlighting needs and skills before the training session
- Indexes of monitoring of trainees during the training session to collect real time data (non technical and technical skills, group dynamic, behavior) in order to promote the animation ability
- Assessment tool dedicated to the improvement of debriefing



Collective radar « before / after » training session

Le projet GEPET-Eau

Gestion Efficiente Prédictive ET adaptative de la ressource en Eau des voies navigables dans un contexte de changement climatique

Contexte : Plan National d'Adaptation au Changement Climatique

Constats

- Diminution des ressources en eau
- Augmentation des températures
- Accroissement en fréquence et amplitude des extrêmes

APR 2012 GICC – Projet 2013-2015

- Méthodes d'évaluation des effets directs et indirects
- Réduction de la vulnérabilité aux variations climatiques
- Etude de la résilience aux évènements extrêmes
- Adaptation au changement climatique
- Financement CGDD, DGEC, DGITM

Objectifs

Contributions

- Déterminer les conséquences du changement climatique sur la navigation
- Prédire les conditions exceptionnelles potentielles à partir d'étude sur l'impact du changement climatique
- Disposer d'un modèle générique de la dynamique des voies navigables
- Estimer la résilience des voies navigable - bief Cuinchy-Fontinettes
- Pouvoir disposer, à terme, d'un outil d'aide à la décision

Verrous scientifiques

- Caractériser des scénarios caractéristiques du changement climatique
- Disposer d'un modèle de voies navigables
- Disposer de modèles d'actionneurs (écluses/vannes)
- Etudier la résilience des voies navigables
- Concevoir des stratégies de gestion prédictive
- Proposer une architecture de conduite

Premiers résultats

Architecture de conduite

- Gestion prédictive et adaptative
- Simulation de scénarios extrêmes
- Conception de stratégies de conduite

Modélisation des voies navigables

- « Boîte grise »
- IDZ (Integral Zero Delay)
- IR - Modèle de Résonnance
- Multi-échelle (débit & volume)

Contrôle des voies navigables

- Commande prédictive MPC

Colloque sur le Numérique
Mars 2014

GICC Le programme Gestion et Impacts du Changement Climatique

Parties prenantes

ARMINES

vnf voies navigables de France

Auteurs

K. Horvath
E. Duviella
J. Blesa
L. Rajaoarisoa
S. Lecoeuche
D. Juge-Hubert
K. Chuquet
E. Sauquet
F. Guibert
N. Gaffet

Partenaires

UPC UNIVERSITAT POLITÈCNICA DE CATALUNYA

irstea

AGENCE DE L'EAU ARTOIS PICARDIE

Liberté • Égalité • Fraternité RÉPUBLIQUE FRANÇAISE
Direction régionale de l'Environnement, de l'Aménagement et du Logement NORD-PAS-DE-CALAIS

Liberté • Égalité • Fraternité RÉPUBLIQUE FRANÇAISE
Ministère de l'Écologie, du Développement durable et de l'Énergie

Contact eric.duviella@mines-douai.fr

Site web http://gepeteau.wordpress.com/
ia.ur.mines-douai.fr

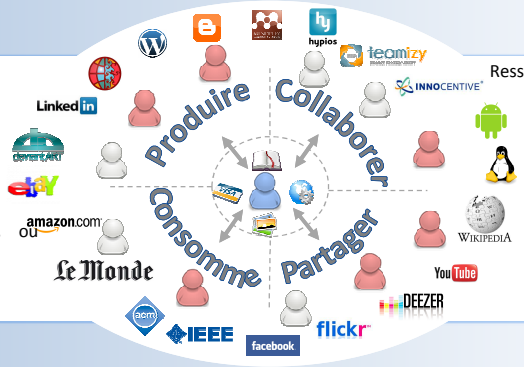
Gestion de la confiance au sein de communautés virtuelles

Réda Yaich, Gauthier Picard, Philippe Jaillon, Olivier Boissier
ISCOD - Institut Henri Fayol
Ecole Nationale Supérieure des Mines de Saint-Etienne, France
Contact : Reda Yaich* (yaich@emse.fr)

Contexte

Les communautés virtuelles :
- Groupes d'entités
- Interagissant via internet
- Partageant des **pratiques, intérêts, valeurs, principes** communs

Leurs objectifs **: production**, **consommation**, **partage, collaboration** autours de ressources (Informations, services, idées, etc.).

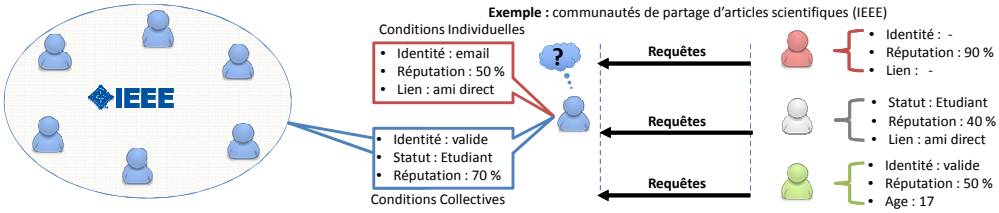


Ressources **sensibles** dont la manipulation comporte un **risque**.

Prise de décision rendue difficile par un contexte :
- **large**
- **distribué**
- **hétérogène**
- **ouvert**
- **décentralisé**
- **dynamique**

La **confiance** est nécessaire pour :
- **maîtriser** le risque
- **réduire** la complexité et l'incertitude

Motivation



Méthode

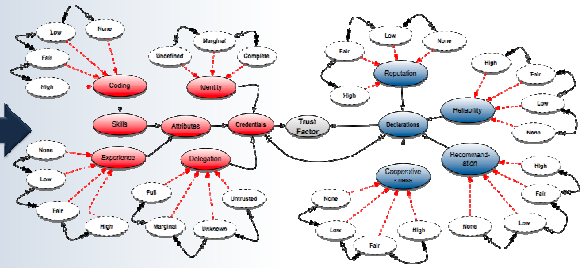
1. Définir un langage de spécification de politiques de confiance **sémantique** et **flexible**.

- **Sémantique** : intelligible par les humains et les agents.
- **Flexible** : dont l'évaluation n'est pas binaire et dont les règles peuvent être modifiées à la volée.

Principe

- ***Une ontologie*** *répertorie les critères de confiance utilisés dans la communauté (ex. Identité, propriétés, et réputation) ainsi que leur domaine de valeurs.*
- *Une politique est spécifiée à partir d'un* ***ensemble*** *de* ***critères de confiance <Type, Valeur, Poids>,*** *Ex. <Réputation, 0.6, 2>.*
- ***La politique est adaptée*** *par l'ajout, la suppression et /ou la modification des critères de confiance.*

Solution

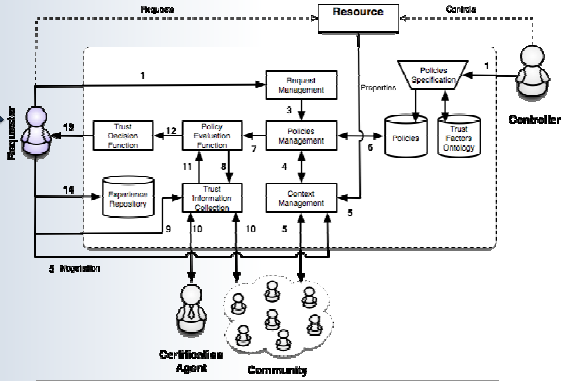


Approche

2. Concevoir un système de gestion de la confiance adaptatif et social

- **Adaptatif** : ajuster au mieux les politiques au contexte métier (risques, opportunités, menaces, etc.)
- **Social :** par articulation des politiques individuelles et collectives.

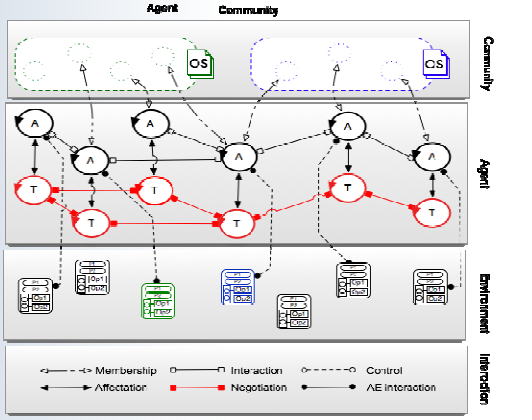
- ***Usage de politiques (expressions en logique pondérée)*** *pour représenter à la fois des politiques individuelles et collectives.*
- ***Usage de métapolitiques (règles ECA)*** *pour adapter et combiner les politiques .*
  - ***E*** *: Evènement du contexte*
  - ***C*** *: Condition de garde*
  - ***A*** *: Une liste d'actions*



3. Implémenter Système de Gestion de la confiance **intelligent** et **autonome** pour adapter, combiner/intégrer et vérifier les politiques de confiance.

- **Intelligent** afin de raisonner sur son contexte (métier et social) et ses politiques.
- **Autonome** afin de prendre des décisions d'adaptation quand c'est nécessaire.

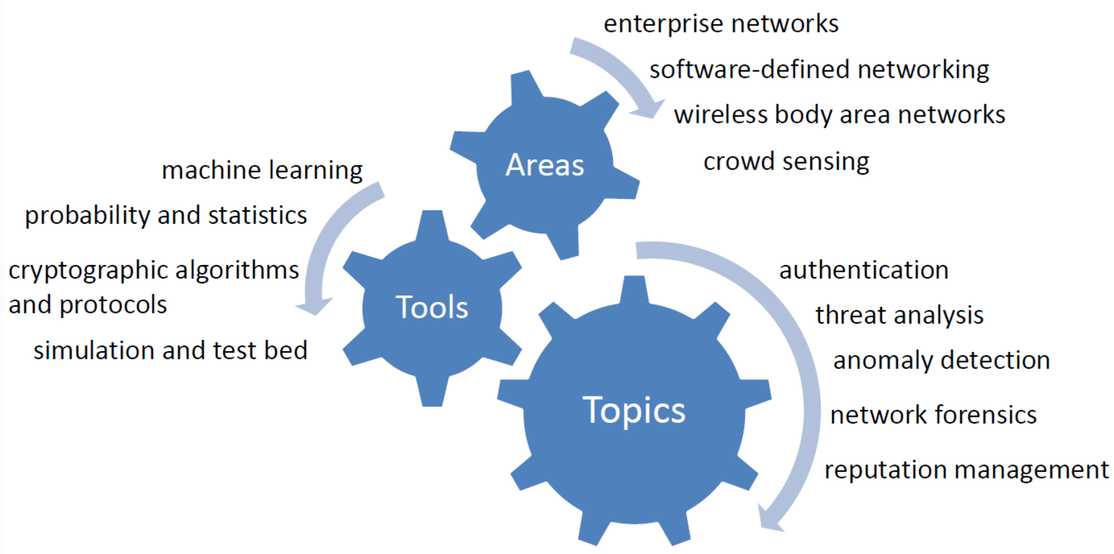
- *Utilisation d'****agents assistants*** *dédiés à la gestion de la confiance.*
- *Pour chaque interaction, l'agent* ***évalue*** *un degré de confiance et* ***recommande*** *à l'utilisateur une décision.*
- *L'agent peut* ***assouplir*** *ou* ***durcir*** *les politiques qu'il utilise en fonction du contexte (métier et social).*



Network Performance and Security

Zonghua Zhang, and Ahmed Meddahi

Research Outline



Challenges

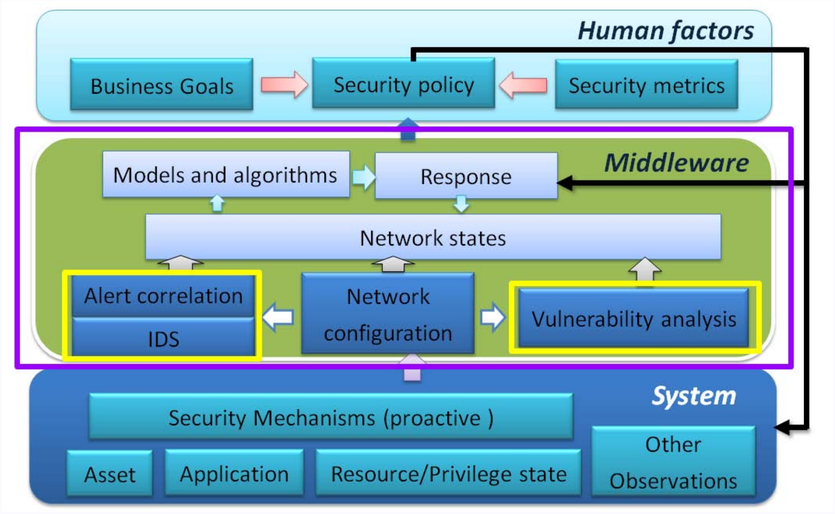
- Building a perfecturly secure system in practice is mission impossible
- The increasing complexity of today's computer and communication systems, and the diversity of networking applications and services lead to the rapid emergence of zero-day vulnerabilities and attacks
- The interactions between system, human, and organization are too complicated to be characterized, modelled, and analyzed
- A prefect in-depth defense line is not available
- ....

We are interested in investigating the significant yet implicit relations between *network performance* (or quality of services) and *security*, designing and developing efficient protocols, models, and algorithms to achieve the best trade-off between expected performance goals and specified security metrics

Selected Research Topics

● Cost-effective security management

Developing adaptive and scalable middleware to enhance *usability*, *effectiveness* and *interoperability* of legacy security mechanisms in enterprise networks. The objective is to assist security administrators in taking optimal security hardening, ranging from vulnerability patching to security mechanism re-configuration and policy enforcement, by leveraging network failure cost resulting from attacks and maintenance cost incurred by defenses. The advent of Software-Defined Network and Cloud Computing has significantly reformed the battlefield between attackers and defenders.



Reference

✓ Shuzhen Wang, Zonghua Zhang, Youki Kadobayashi: Exploring attack graph for cost-benefit security hardening: A probabilistic approach, *Computers & Security* 32: 158-169 (2013) (Details are given below)

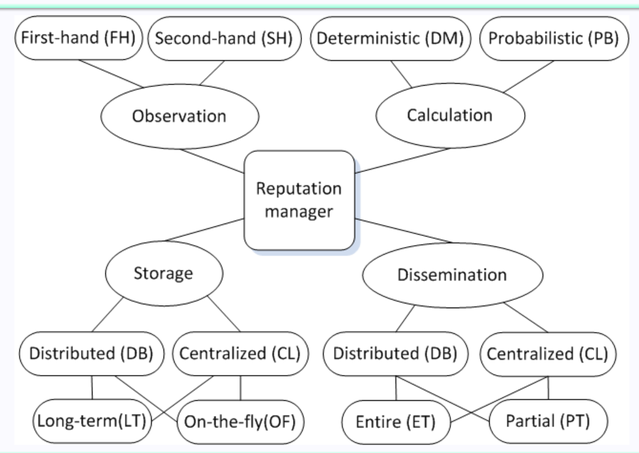
✓ Zonghua Zhang, Farid Naït-Abdesselam, Pin-Han Ho, Youki Kadobayashi: Toward cost-sensitive self-optimizing anomaly detection and response in autonomic networks. *Computers & Security* 30(6-7): 525-537 (2011)

✓ Zonghua Zhang, Pin-Han Ho, Liwen He: Measuring IDS-estimated attack impacts for rational incident response: A decision theoretic approach. *Computers & Security* 28(7): 605-614 (2009)

✓ Zonghua Zhang, Hong Shen: M-AID: An adaptive middleware built upon anomaly detectors for intrusion detection and rational response. *ACM Trans. on Autonomic and Adaptive Systems* 4(4) (2009)

● Reputation and trust management

Designing *privacy-preserving*, *robust* and *light-weight* reputation systems to enhance the quality of services in wireless networks, mobile social networks, and smart city oriented crowd sensing. One of the major aims is to encourage the network entities, varying from static sensors to mobile phone users, to actively contribute their local information for global data processing, knowledge discovery and decision-making.



Reference

✓ Zonghua Zhang, Pin-Han Ho, Farid Naït-Abdesselam: RADAR: A reputation-driven anomaly detection system for wireless mesh networks. *ACM Wireless Networks* 16(8): 2221-2236 (2010)

✓ Juan Li, Zonghua Zhang, Weiyi Zhang: MobiTrust: Trust Management System in Mobile Social Computing. in *Proceeding of CIT 2010*: 954-959

✓ Zonghua Zhang, Jingwei Liu, Youki Kadobayashi: STARS: A Simple and Efficient Scheme for Providing Transparent Traceability and Anonymity to Reputation Systems. in *Proceeding of DPM/SETOP 2010*: 170-187

● Privacy-preserving network forensics

Designing *efficient* and *reliable* privacy-preserving methods, algorithms and protocols for forensic analysis on threat data of interest. The purpose is to integrate cross-site encrypted footprints associated with multi-layer observations for manifesting and characterizing attack behavior.

Reference

✓ NECOMA Project (Nippon-European Cyberdefense-Oriented Multilayer threat Analysis, EU FP7) http://www.necoma-project.eu/

✓ Zonghua Zhang, Hong Shen: Constructing Multi-Layered Boundary to Defend Against Intrusive Anomalies: An Autonomic Detection Coordinator. in *Proceedings of IEEE DSN.*

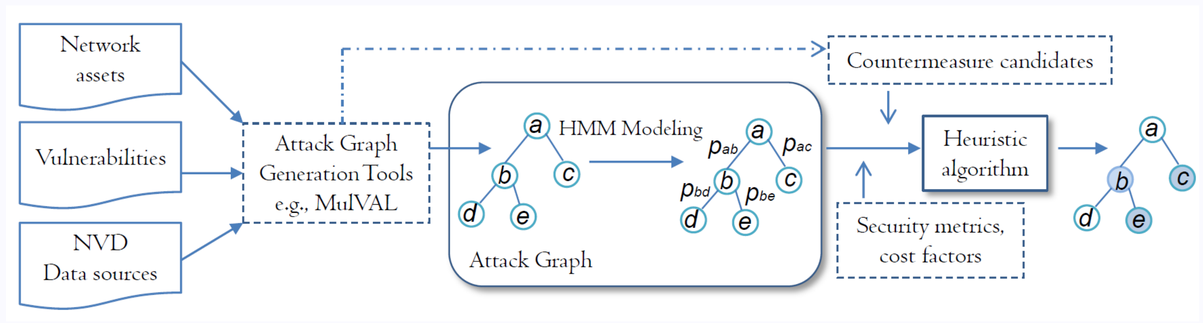
Exploring attack graph for cost-benefit security hardening: a probabilistic approach

● Design Goal

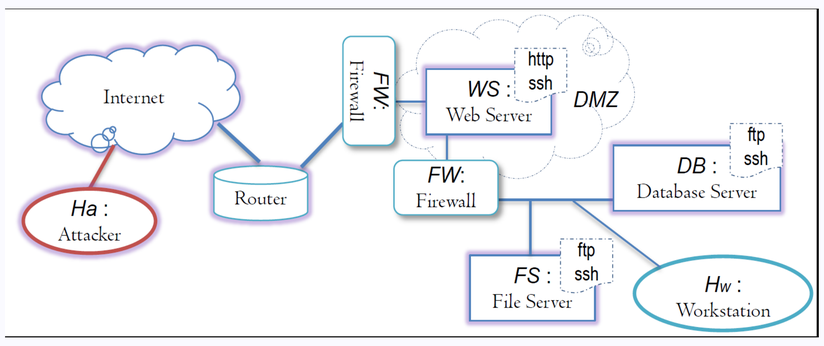
Given a target network and a set of collected observations, how to effectively estimate and predict the implicit system states, subsequently to identify the root causes that may lead to the significant loss of security properties in terms of a set of security metrics, which are specified as security demand and business goal of an organiztaion.

● Design Rationale: Our Approach AG-HMM

- ✓ Commercial/Open source tools are used to identify network vulnerabilities
- ✓ Network assests, vulnerabilities, user priviledges are collected as observations
- ✓ Dependency Attack Graph (AG) is modified and applied to represent observations
- ✓ Hidden Markov Model (HMM) is used to estimate implicit system states based on AG-represented observations
- ✓ System states is quantified as pre-specified security metrics according to a well-defined cost function
- ✓ The observations associated with key system states are removed

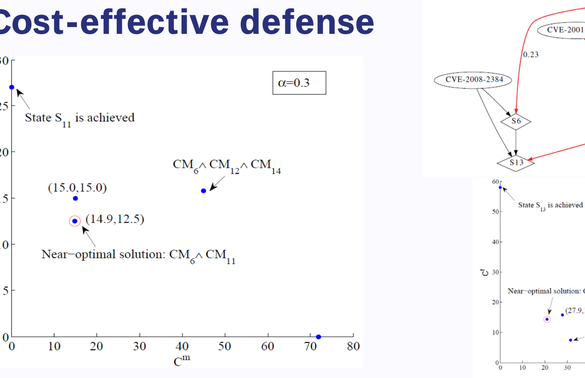


● Example network



Notations	*Network asset*	*Applications*	*Role*
H_a	Attacker	Any	Conducting attacks
H_w	Workstation	Any	Normal activities
WS	Web Server	http, ftp	WWW, database queries
FS	File Server	ftp, ssh	Storing confidential info., etc.
DB	Database Server	ftp, ssh	Storing data accessed via *WS*
FW	Firewall	Loose policies	Traffic filtering and control
IDS	IDS	Signature-based	Intrusion detection and alerts

● Generation of dependency attack graph



● Attack state estimate using Hidden Markov Model

Notation	Countermeasure
CM_1	Disconnect(DB, Internet)
CM_2	Disconnect(FS, Internet)
CM_3	Disconnect(WS, Internet)
CM_4	Patch(DB, CVE-2008-2384)
CM_5	Patch(FS, CVE-2001-0755)
CM_6	Patch(WS, CVE-2006-3747)
CM_7	Patch(H_w, all)
CM_8	Disable(DB, ssh/ftp)
CM_9	Disable(FS, ssh/ftp)
CM_{10}	Disable(WS, ftp)
CM_{11}	Disable(H_w, all)
CM_{12}	Configure(FW)
CM_{13}	Configure(IDS)
CM_{14}	Configure(H_w)

● Cost-effective defense

GRAPHICAL CODE FOR AUTHENTICATION
INFORMATION THEORETICAL ANALYSIS

School

Authors

Anh Thu PHAN HO
Wadih SAWAYA
Patrick BAS (CNRS)

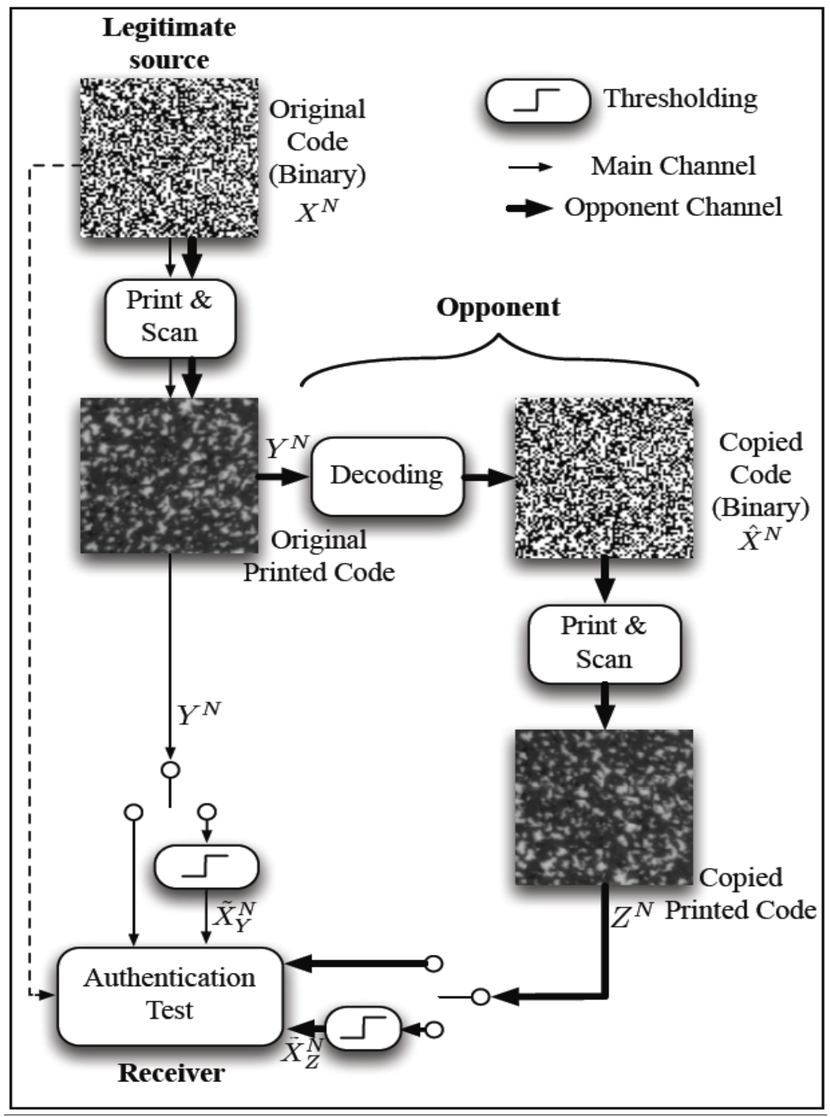
Partner

CONTEXT

- Counterfeiting is rising rapidly in many areas such as food, medicines, cosmetics…
- Fighting against counterfeit by printing a 2D barcode on package of products
- Assuming that printing and acquisition are stochastics and irreversible processes
- Opponent's strategy: generating $\hat{X}^N$ such that Z^N is considered as authentic

OBJECTIVES

- Develop a theoretical authentication model using information theoretic tools
- Extract bounds on the success probabilities of the opponent
- Define the parameters of optimal codes for authentication

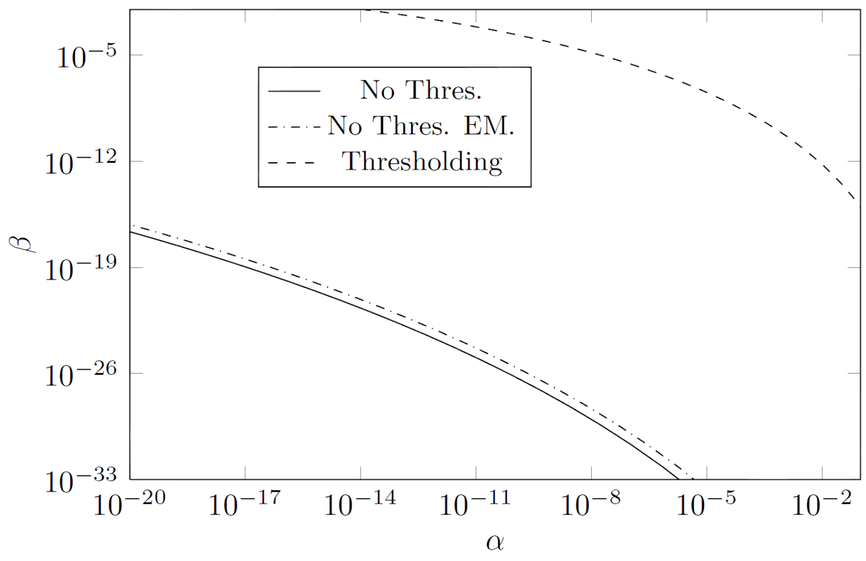


RESULTS

- Gray level observation strategy better than binary thresholding for authentication
- Assuming the models of processes known, using Neyman-Pearson test

$$L = \log \frac{P\left(o^N/x^N,\, H_1\right)}{P\left(o^N/x^N,\, H_0\right)} \underset{H0}{\overset{H1}{\gtrless}} \lambda \text{ where } o^N|H_0 = y^N;\; o^N|H_1 = z^N$$

- Computing false alarm and non detection probabilities
  - ➢ Gaussian aproximation
  - ➢ Chernoff bounds (threshold far away the mean)



PERSPECTIVES

- Authentication for structured codes
- Studying the model of the broadcast channel

REFERENCES

[1] A-T.Phan Ho, B-A. Hoang Mai, W.Sawaya, P.Bas. Document Authentication Using Graphical Codes : Impacts of the Channel Models. In ACM Workshop on Information Hiding and Multimedia Security, Montpellier, France, June 2013

[2] R. G. Gallager. Information theory and reliable communication, volume 15. Willey 1968

Le colloque "Numérique : Grande échelle et complexité"
Mars 2014

Contact phanho@telecom-lille.fr

Site web

RFID, une technologie controversée : entre usages et perception du risque

Objectifs

Un programme de recherche sociologique sur la construction sociale du risque RFID

Analyse des controverses et perceptions du risque en situation

- Analyse de la presse écrite (généraliste et spécialisée, anglophone/francophone, 1990-2010, 100000 articles)
  - Cartographie et chronologie du débat public
  - Identification des acteurs
  - Qualification des risques
- Enquête dans la R&D et production de la RFID
- Ethnographie d'une expérimentation d'usage (santé, DASRI)

Partenaires

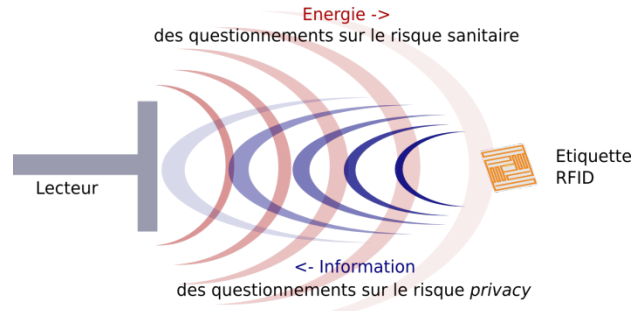
Télécom ParisTech, Dép. SES, DEIXIS-Sophia (leader)
Mines Saint-Etienne, Centre Microélectronique de Provence
Université de Coimbra, OSIRIS, Observatoire sur les Risques

Projets et soutien

Risc - Radiofréquences : Identification des Sources de Controverse. Le cas de la RFID

Trace-De-TIC

Objet de la recherche



Valorisations

Conférence-débat
Télécom ParisTech
14 mars 2014

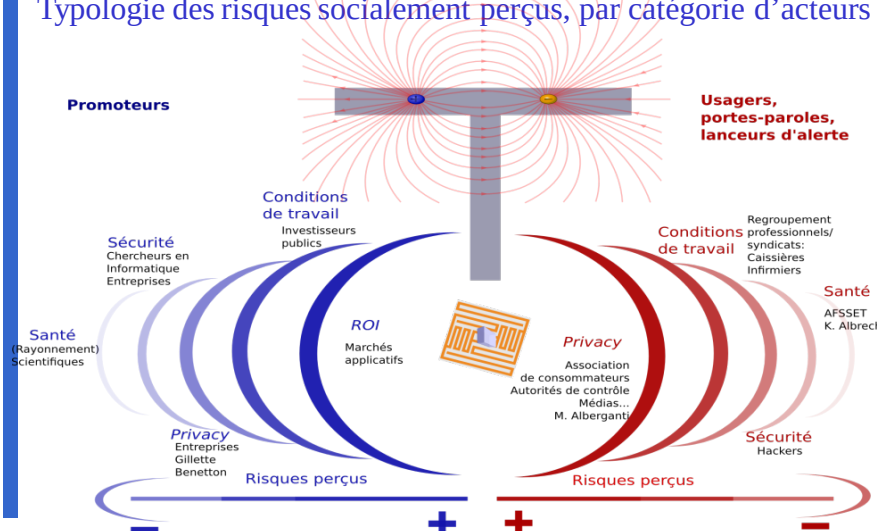
La RFID à l'épreuve de l'innovation responsable

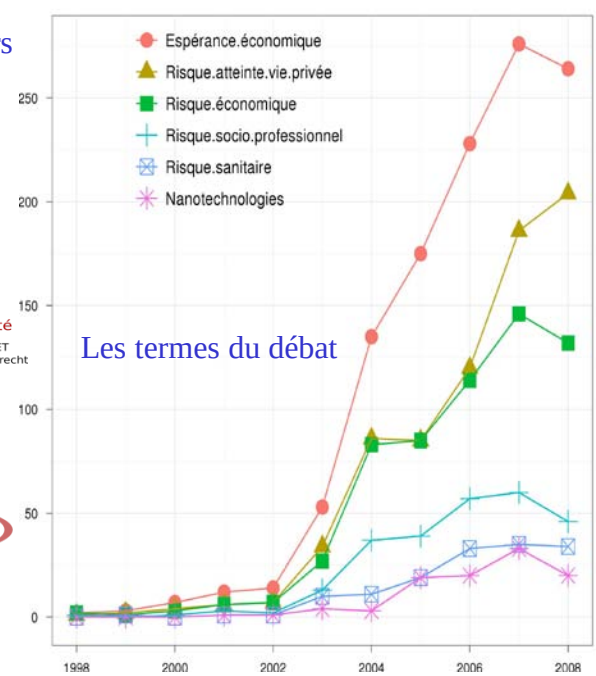
Évaluation d'impact sur la vie privée
gestion de l'exposition humaine
éco-conception et recyclage

Quelles exigences et opportunités pour une approche responsable de l'innovation ?

Quelques résultats de recherche

Typologie des risques socialement perçus, par catégorie d'acteurs





Les termes du débat

Contact : Laura Draetta, TELECOM ParisTech, LTCI/CNRS
+33 (0)4 93 00 84 09 laura.draetta@telecom-parsitech.fr

A Simple Model for Evaluating Secure Key Generation from Random Radio Channels

Institut Mines-Télécom

Auteurs

Taghrid MAZLOUM

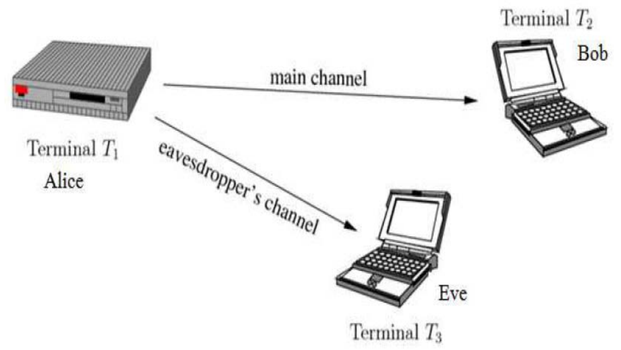
Alain SIBILLE

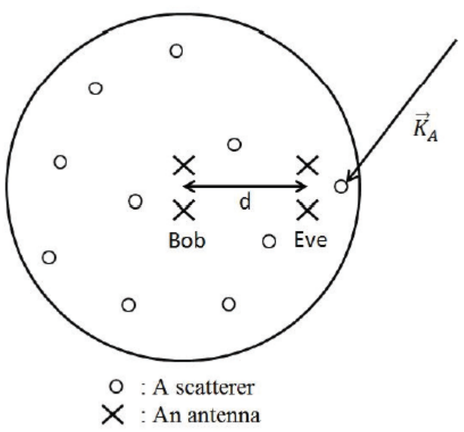
Partenaires

PHYLAWS

Introduction

- Security is a significant challenge in wireless communications. Owing to the public nature of information wireless transmission, an eavesdropper can easily access to the information exchanged between legitimate terminals.
- The widely used method to ensure security is to encrypt and decrypt messages using secret keys. However conventional techniques of generating and distributing such keys may suffer from complexity and high computational cost.
- An alternative solution is physical layer security (PhySec) that designs all kind of security methods which take advantage of the inherent properties of the propagation channel, e.g. noise, interference, and the time-varying nature of fading channels, to provide secure communications.
- We are particularly interested in secret key generation (SKG) achieved by exploiting and invoking radio channel properties, e.g. reciprocity and spatial decorrelation. Legitimate terminals generate independently an identical secret key by observing the same propagation channel considered as a source of randomness. Hence we propose a simple stochastic channel model for which we evaluate the security.





SKG through a simple channel model

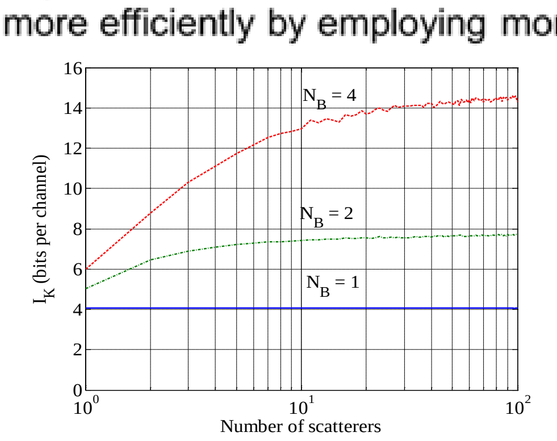
Description of the channel model

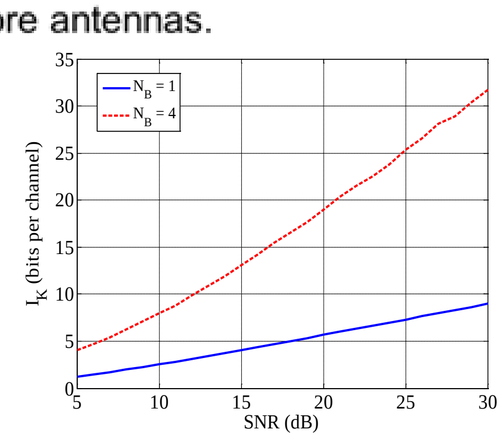
- In the literature, the security is evaluated for a worst-case scenario where one legitimate user, i.e. Alice or Bob, and the eavesdropper Eve are sufficiently close to each other. However, in order to explore more realistic scenarios, we propose a simple 2-D geometry-based stochastic channel model to study the effect of the lack of spatial stationarity between Bob and Eve on the SKG.
- We model a macroscopic environment by uniformly distributing scatterers within a disc around Bob. Eve is located at a distance *d* from Bob. They both are equipped with either single or multiple omnidirectional antennas. The transmitter Alice is assumed far away from the disc.
- The complex channel gain connecting Alice antenna with the *m*th antenna at Bob/Eve side can be defined as:

$$h_m = \sum_{l=1}^{N_S} \frac{\beta_l}{d_l} \exp\left\{j\left(K.d_{ml} + \overrightarrow{K_A}.\overrightarrow{r_l}\right)\right\}$$

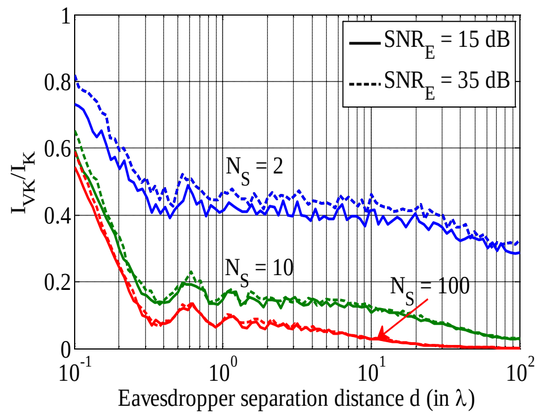
SKG evaluation

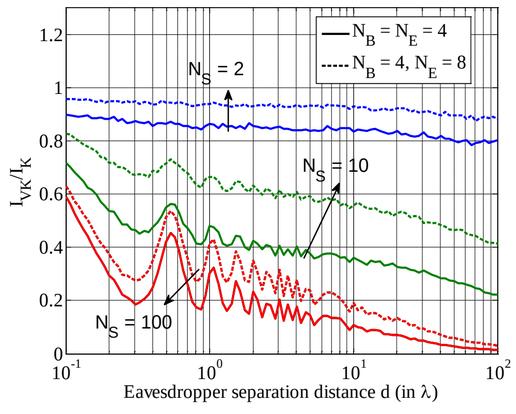
- I_K is the maximum number of bits that can be extracted from the reciprocal propagation channel. By increasing the spatial diversity of the channel, i.e. by employing multiple antennas, I_K increases especially for rich scattered channels and for higher SNR.
- Eve is able to know some bits of I_K. Therefore we define the vulnerable key bits I_{VK} as the number of bits leaked to Eve. The security is improved for rich scattered channels and for large separation distance Bob/Eve. Eve degrades the security by either increasing her SNR or more efficiently by employing more antennas.





The improvement of I_K with respect to the channel richness in scatterers, to the number of antennas and also to the SNR





The relative vulnerable key bits I_{VK}/I_K vs. Eve separation distance for either larger SNR or employing more antennas

Conclusion

- A simple channel model exploring the macroscopic variation between Bob and Eve is investigated in SKG context.
- The confidentiality is achieved owing to the spatial decorrelation between the legitimate channel and that measured by Eve, especially in rich scattered channels for large separation distance Bob/Eve where they both do not share a common stationarity region.
- Exploiting the degrees of freedom of multiple antennas improve the security by increasing I_K leading to generate more secure key bits. Unfortunately, it also helps Eve to gather more information about the shared key.

Contact {Taghrid.mazloum, alain.sibille}@telecom-paristech.fr Site web www.telecom-paristech.fr

Pseudonym update for privacy in cooperative V2X systems

Parties prenantes

TELECOM ParisTech

RENAULT

Auteurs

Rim Moalla, TPT.
Houda Labiod, TPT.
Brigitte Lonc, Renault.

Partenaires

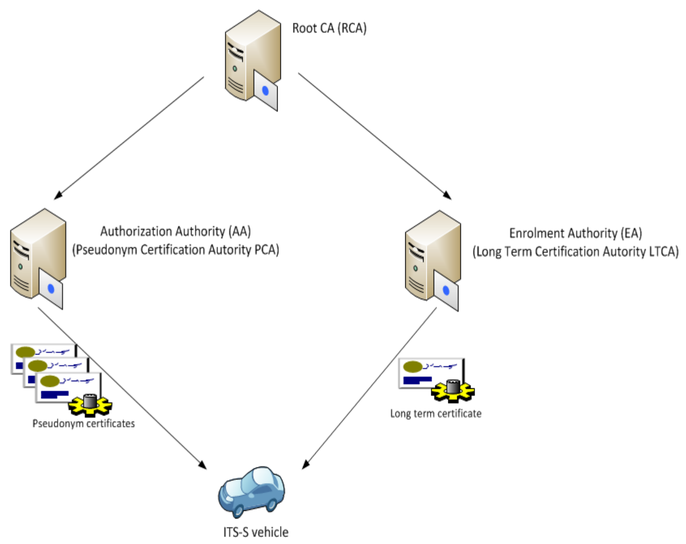
FONDATION TELECOM

INSTITUT CARNOT Télécom & Société numérique

Problem statement

Context

- Privacy is a complex requirement in V2X systems as we have to identify stations and to protect personal data.
- Several solutions are proposed to provide privacy: Anonymous certificates, group signature and pseudonyms certificates.
- Car2Car consortium and standardization organizations choose pseudonyms solution where ITS-S vehicle has two types of certificates: short term certificates named pseudonyms certificates and long term certificate.
- Pseudonyms certificates change frequently and consequently have to be updated.
- We consider pseudonyms certificates update over-the-air using the unsecure G5 (IEEE 802.11p) media.

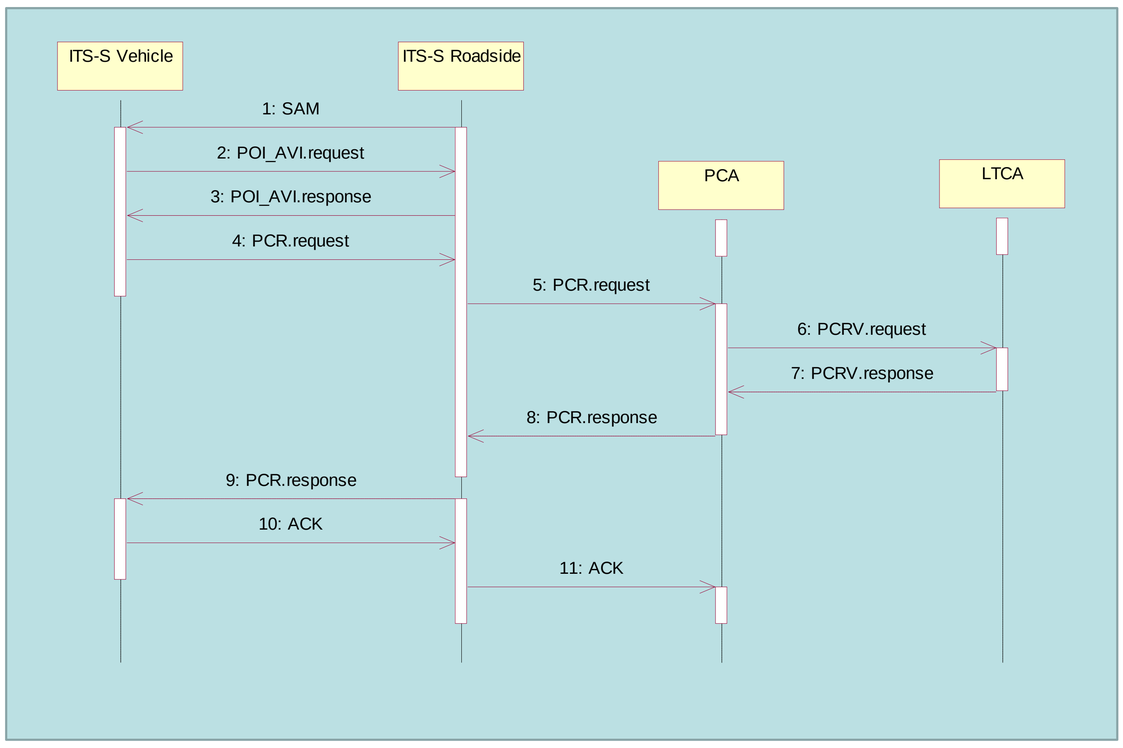


Requirements

- ITS-S vehicle has to transmit sensitive data to the PCA without conveying them to an intermediary node such as ITS-S roadside.
- ITS-S vehicle has to prove to the intermediary node that is authorized to establish communication with PCA.
- ITS-S roadside is required to prove its legitimacy to the PCA by providing that it is authorized to act on behalf of the ITS-S vehicle.

Proposed protocol

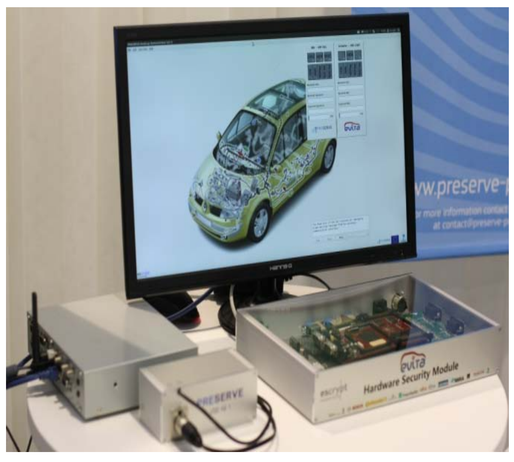
- Is composed by two phase: phase I covers service discovery and session key establishment and phase II for certificate update.
- Enables vehicle to securely update its certificates from a roadside unit (over G5).
- Considers performance, scalability and cost issues.



Pattent submitted in July 2013

Implementation

- We integrate security services into Score@F platform.
- Cryptographic operations are based first on Java crypto library and then on FP7 PRESERVE security solution.
- We evaluated during last tests session on September 2013:
  - Signature generation duration
  - Signature verification duration
  - Pseudonyms change strategies



Validation

- Score@F/ PRESERVE Workshop , NRIA-Rocquencourt, September2013
- Journée Mobilité 2.0, Satory, February 2014

Contact rim.moalla@telecom-paristech.fr Site web

Modeling and Characterization of Physically Unclonable Functions

*

Parties prenantes

TELECOM ParisTech

LABORATOIRE Hubert CURIEN UMR • CNRS • 5516 • ST-ETIENNE

Authors

PhD student :
Zouha Cherif Jouini
Supervisors:
Jean-Luc Danger
Lilian Bossuet

Partners

orange™

Physically Unclonable Functions

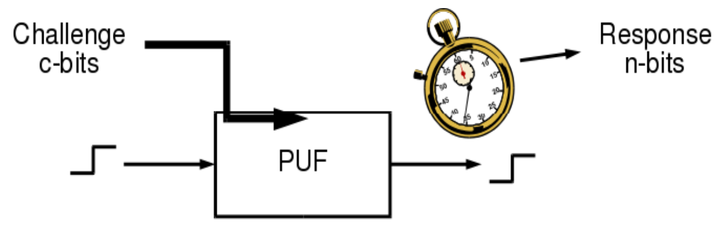
- It returns a signature intrinsic to a circuit (a fingerprint).

Applications

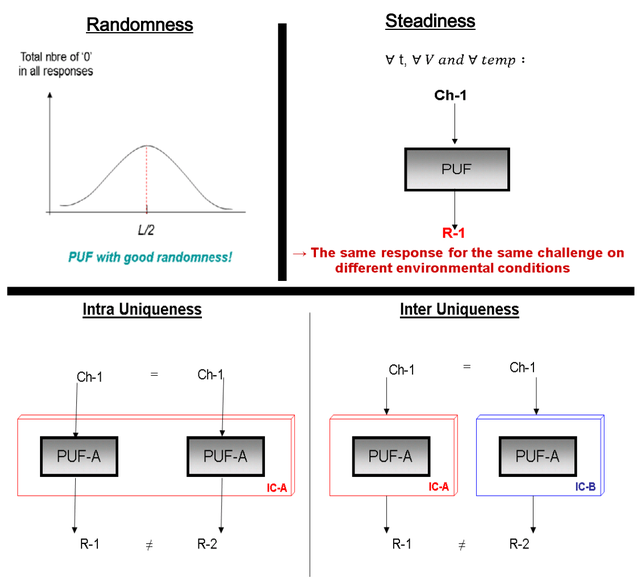
- Authentication of Integrated Circuits.
- Generation of cryptographic keys.

Types

- Silicon: Easy to implement : Arbiter PUF, Ring-Oscillator PUF, SRAM PUF, etc.
- Non silicon: Coating PUF, Optical PUF, etc.



Our Contributions



1- PUF characterization method

Principle

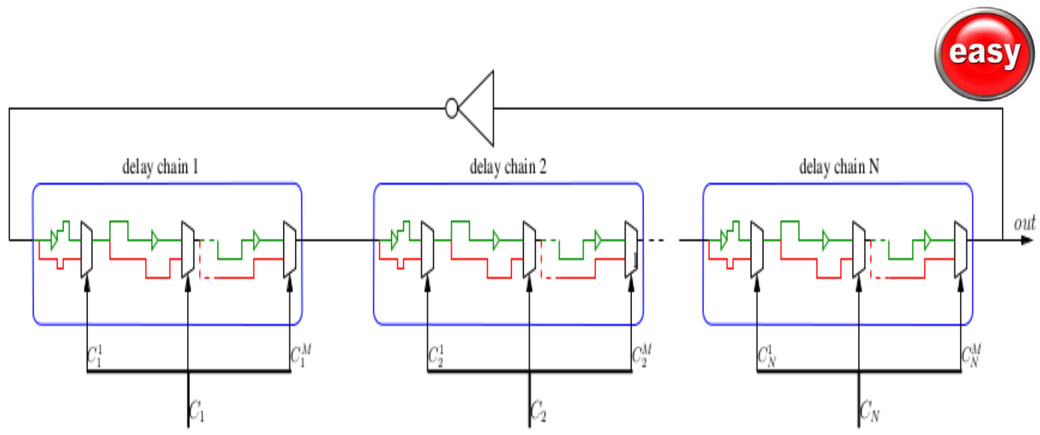
- Used to evaluate silicon PUFs (specially delay PUFs).
- Takes advantage of the physical characteristics of the PUF structures.

Performance Indicators

- Randomness
- Uniqueness
- Steadiness

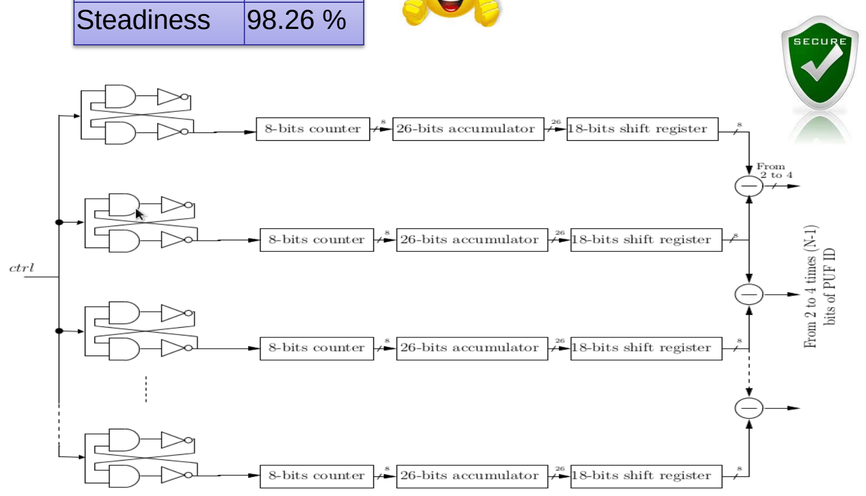
2- Loop PUF

- Silicon delay based PUF.
- Easy to implement :
  - ➢ No hard routing and placement constraints.



Performance Results

Randomness	98.97 %
Uniqueness	89.21 %
Steadiness	98.26 %



3- TERO PUF

- Silicon Ring-Oscillator based PUF.
- Not sensitive to locking phenomenon.

Performance Results

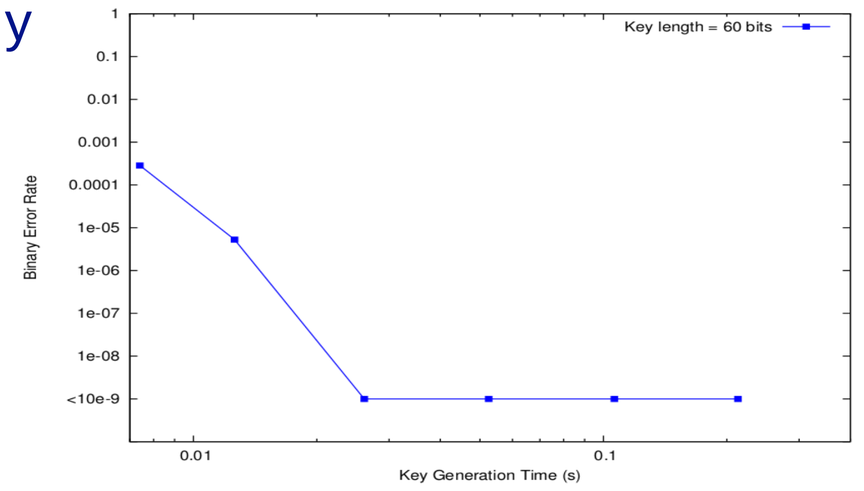
Randomness	98.61 %
Uniqueness	98.54 %
Steadiness	97.25 %

4- Loop PUF-based cryptographic key

Principle

- Smart selection of challenges.
- Increasing the number of tests.
- Unreliable bit identification.
- Key correction procedure.

➡ BER = 10-9 - 10ms - 101 slices in Xilinx FPGA.



* http://www.newelectronics.co.uk/electronics-news/puffin-aims-to-protect-against-identity-theft/45139

Contact Zouha.cherif@telecom-paristech.fr

Drone for Humanitarian Operations

Visit our website and have a look at our videos at, http://drone4u.eurecom.fr

Institutions

Authors

Tullio Joseph Tanzi (Télécom ParisTech)

tullio.tanzi@telecom-paristech.fr

Ludovic Apvrille (Télécom ParisTech)

Jean-Luc Dugelay (EURECOM)

Claire Migliaccio (LEAT)

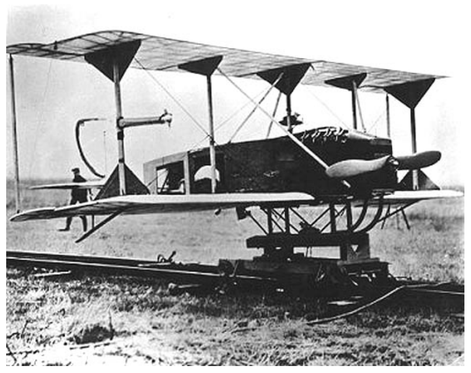
Julien Morel (Télécom ParisTech)

Franck Guarnieri (Mines ParisTech)

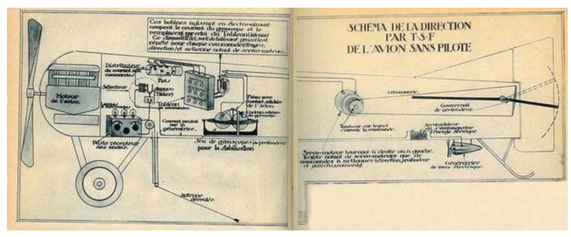
Partners

Main characteristics of drones

- **Unmanned** Aerial Vehicle: Drones can perform some specific missions with no on-board pilot;
- **Self-flying**: Autonomous fly is usually limited to reach a specific location given by GPS. More advanced autonomous functionalities can help in a decision process to react against unpredicted situations;
- **Reusable** and **reconfigurable**: Drones can be used for diverse missions, and can be appropriately customized.

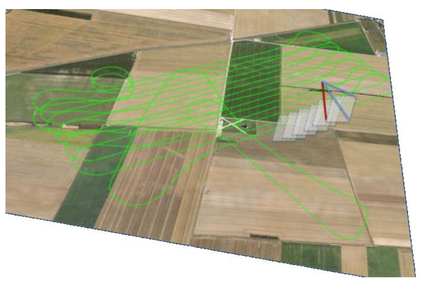


Hewitt-Sperry Automatic Airplane 1910



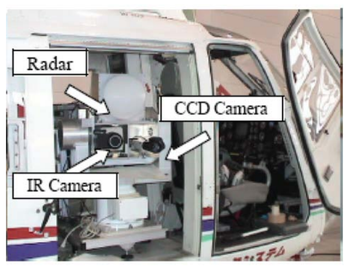
"Avion sans pilote« , by Maurice Percheron [Lectures pour tous, février 1923].

Drones for assisting disasters

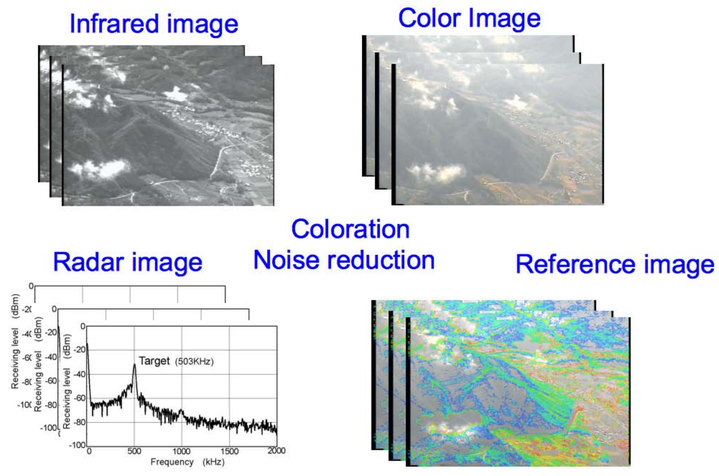


3D Reconstruction

- Spatial **coverage**:
  - Scanning a given area to establish an overview map of emergency;
- **Image processing**:
  - Detecting groups with a fast classification (e.g., adults vs children)
  - 3D reconstruction to allow drones to navigate autonomously with cameras
- **Specialized** on-board devices and **sensors**:
  - Detecting signals attached to wireless networks (e.g., mobile phones) so as to drive rescuers to areas where they are more likely to find persons.
- etc.

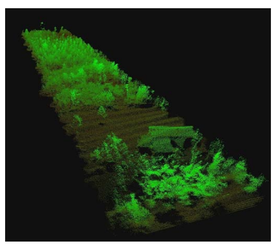


Non conventional payload



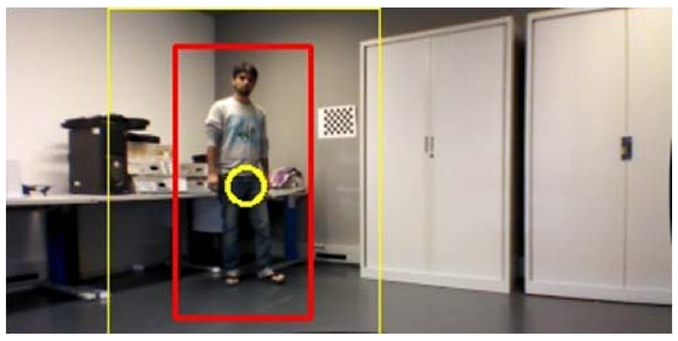
Embedded electronic and software architecture

- Cameras,
- Lidars,
- Low cost and efficient processing units (e.g., parallella ...)



LIDAR Image

UAV cluster



People detection and tracking

Summary

- Designing a civil drone to assist disasters;
- "Smart drone": Autonomous drone with some standalone capacities to make decisions;
- Safety and security are taken into account at design stage.
- Integration of complex sensors;
- Handling complex national and international rules and policies;
- Societal impacts, including privacy preservation.

TTool — An open source toolkit provided by

Colloque l'Institut Mines-Télécom , "Numérique : Grande échelle et complexité"

26 et 27 mars 2014

Formally Proved Security of Assembly Code Against Leakage

Pablo RAUZY
Sylvain GUILLEY

Institut MINES-TELECOM,
TELECOM-ParisTech,
CNRS LTCI (UMR 5141).
Paris, France

Context: countermeasures

	Hardware	Software
Masking	***	***
Hiding (dual-rail)	***	*few works!*

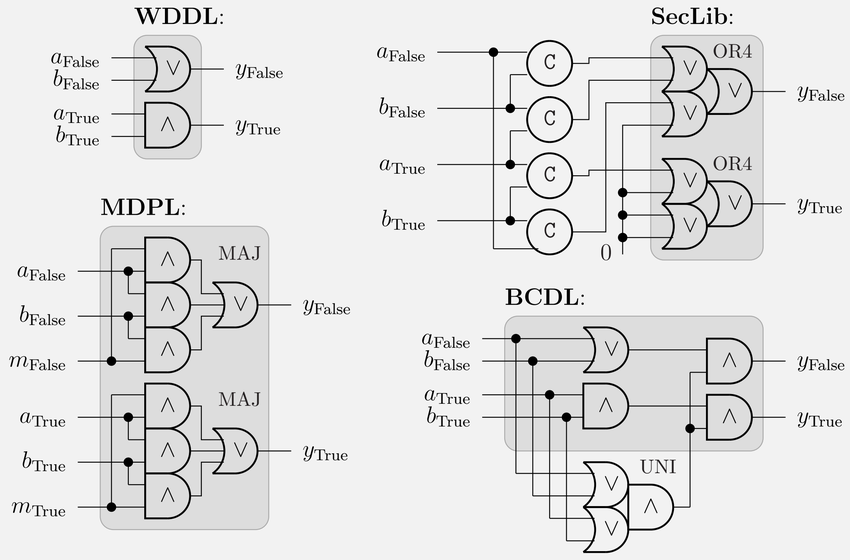
Problems of masking in software

- Lots of entropy (*not available on resource-constrained devices*)
- Structural vulnerability: existence *high-order* attacks

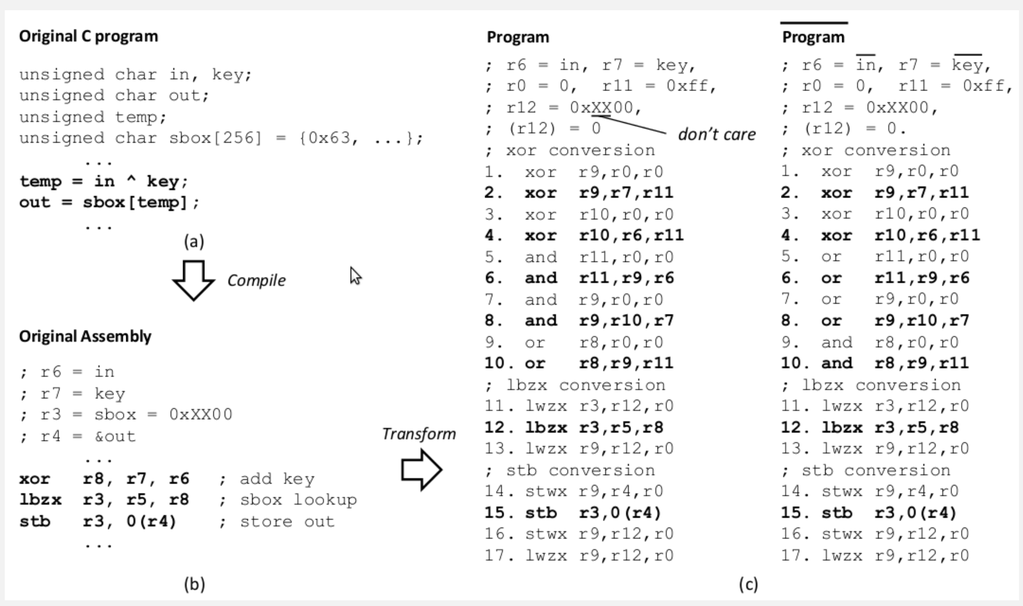
Dual-rail in software: opportunities

- No need for entropy
- Provable correction of leakage-free (with a finite number of *physical* hypotheses, to do by pre-characterization)

State-of-the-art about dual-rail in hardware [DGBN09]



State-of-the-art: mixed HW/SW, *e.g.*, dual-rail instruction set [CSS13]

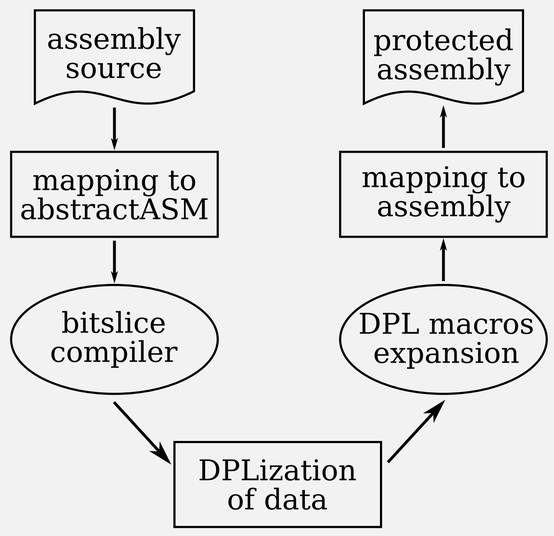


Courtesy of Zhimin Chen and Patrick Schaumont ECE Department, Virginia Tech Blacksburg VA 24061, USA

An example of Virtual Secure Circuit (VSC):

(a) `KeyAddition` and `SubBytes` operations in C code;
(b) Compiled assembly code;
(c) Converted VSC assembly code.

Pure software dual-rail: design flow



DPL: **D**ual-**R**ail with **P**recharge

Macro for Boolean operation *op*

```
r1 ← r0          mov r1 r0
r1 ← a           mov r1 a
r1 ← r1 ∧ 3      and r1 r1 #3
r1 ← r1 ≪ 1      shl r1 r1 #1
r1 ← r1 ≪ 1      shl r1 r1 #1
r2 ← r0          mov r2 r0
r2 ← b           mov r2 b
r2 ← r2 ∧ 3      and r2 r2 #3
r1 ← r1 ∨ r2     orr r1 r1 r2
r3 ← r0          mov r3 r0
r3 ← op[r1]      mov r3 !r1,op
 d ← r0          mov d r0
 d ← r3          mov d r3
```

Cost on PRESENT [BKL+07] case-study

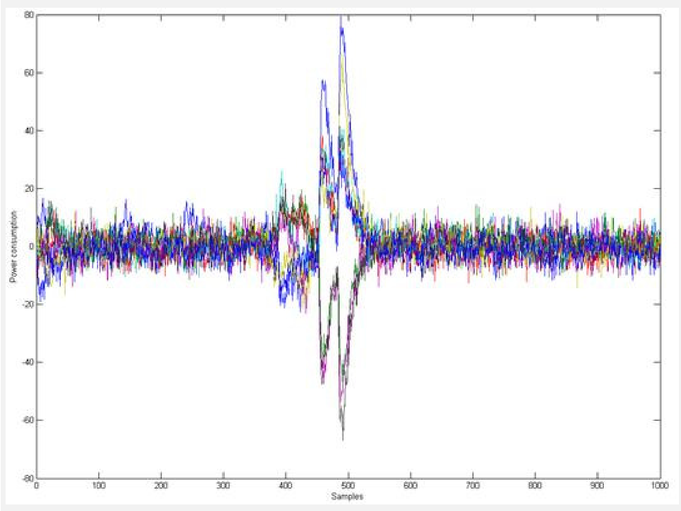
	cycle count	code size*	RAM words*
state-of-the-art	11342	1000	18
bitsliced	6473	1194	144
DPL protected	182572	2674	192

* The state-of-the-art code size and RAM words are given for encryption + decryption, while ours are for encryption only. Code size and RAM words are given in bytes.

Optimizations (still with formal proof of correction)

- The existence of non-sensitive signals (*e.g.*, the selection of key size); or loop counters;
- The limited data range of some variables, that makes some parts of the code use constant variables;
- The possibility to go from one macro to the other through register, thereby saving time from the memory transfers;
- The possibility to merge instructions given certain patterns;
- The use of architecture-specific intructions not included in our abstractASM.

Leakage analysis (physical part)

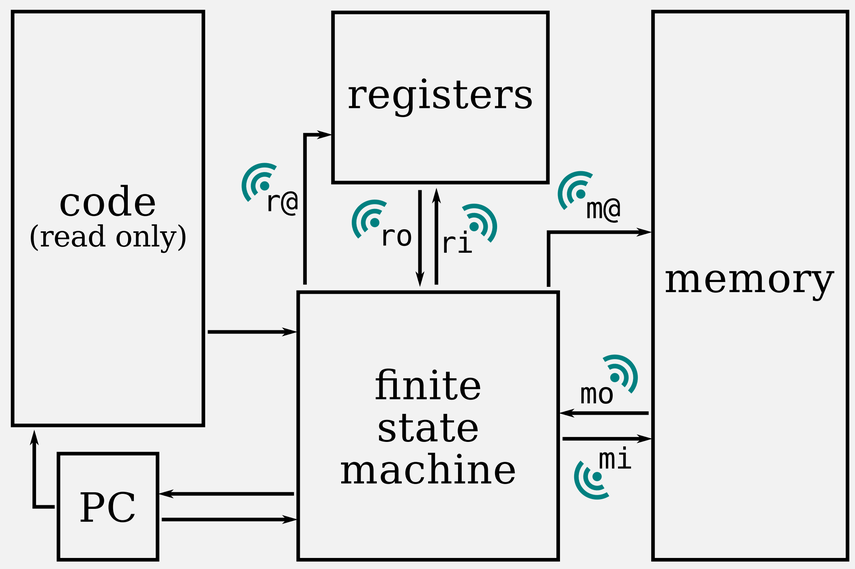


Verifications:
- indistinguishable resources
- for data and addresses

Tools:
- profiling
- linear regression

Stochastic characterization [SLP05] of every bit in a general purpose CPU.

Leakage analysis (formal part)



Verification:

- Leakage: **Hamming distance** of values, **should be constant**
- **Symbolic execution** to check this constantness property

References

[BKL+07] Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew J. B. Robshaw, Yannick Seurin, and Charlotte Vikkelsoe. PRESENT: An Ultra-Lightweight Block Cipher. In *CHES*, volume 4727 of *LNCS*, pages 450–466. Springer, September 10-13 2007. Vienna, Austria.

[CSS13] Zhimin Chen, Ambuj Sinha, and Patrick Schaumont. Using Virtual Secure Circuit to Protect Embedded Software from Side-Channel Attacks. *IEEE Trans. Computers*, 62(1):124–136, 2013.

[DGBN09] Jean-Luc Danger, Sylvain Guilley, Shivam Bhasin, and Maxime Nassar. Overview of Dual Rail with Precharge Logic Styles to Thwart Implementation-Level Attacks on Hardware Cryptoprocessors, — *New Attacks and Improved Counter-Measures* —. In *SCS*, IEEE, pages 1–8, November 6–8 2009. Jerba, Tunisia. DOI: 10.1109/ICSCS.2009.5412599.

[SLP05] Werner Schindler, Kerstin Lemke, and Christof Paar. A Stochastic Model for Differential Side Channel Cryptanalysis. In LNCS, editor, *CHES*, volume 3659 of *LNCS*, pages 30–46. Springer, Sept 2005. Edinburgh, Scotland, UK.

Maximizing the success of a side-channel attack

Institut Mines-Télécom

Authors

Annelie Heuser

Sylvain Guilley

Olivier Rioul

Télécom ParisTech / Institut Mines-Télécom / CNRS LTCI

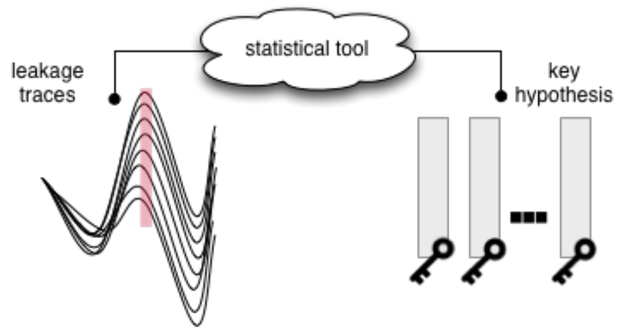
Annelie Heuser is a recipient of the Google Europe Fellowship in Privacy.

State of the Art

✧ What distinguishes known distinguishers, in terms of distinctive features?

✧ Given a side-channel context, what is the best distinguisher amongst all known ones?

- Distinguishers were chosen as (arbitrary) **statistical tools** (correlation, difference of means, linear regression, etc.)
- [1] highlights that proposed distinguishers behave **equivalent** when using the same leakage model, only "statistical artifacts" can explain different behavior [2]
- The **estimation** of the statistical tools (esp. mutual information) is very crucial and effective on the success [3]

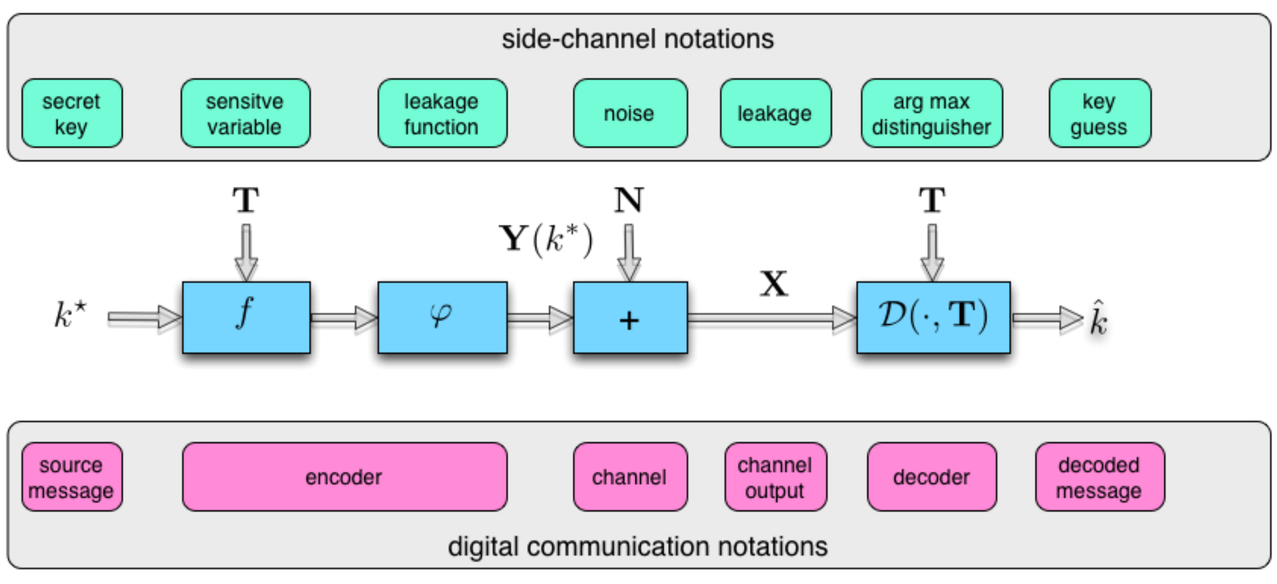


[1] Doget, Prouff, Rivain, and Standaert, JCEN, 2011
[2] Mangard, Oswald, and Standaert. IET, 2011
[3] Prouff and Rivain, IJACT, 2010.
[4] Heuser, Rioul, and Guilley, under submission

Side-channel analysis as a communication problem [4]

✧ Given a side-channel scenario, what is the best distinguisher, amongst all possible ones?

- Idea: Translate the problem of side-channel analysis into a problem of communication theory ➔ derive **optimal distinguisher**: maximize the success rate
- Leakage model is known to the attacker (**Theorem 1**)
  - Only statistical noise
  - Optimal decoding rule $\arg\max_k \Big(\mathbb{P}\{k\} \cdot p(\mathbf{x}|\mathbf{y}(k))\Big)$ (template attack, profiling is possible)
  - The optimal distinguisher only depends on the noise distribution (e.g., Laplacian, uniform, Gaussian)
- Leakage model is partially unknown to the attacker (**Theorem 2**)
  - Statistical and epistemic noise
  - Leakage arises due to a weighted sum of bits, where the weights follow a normal distribution



Theorem 1: optimal distinguisher when the leakage model is known

If the leakage arises from $X = Y(k^\star) + N$ *with known leakage model* $Y(k) = \varphi(f(k, T))$ *then the optimal distinguishing rule are*

- *Gaussian noise distribution:* $\mathcal{D}^{M,G}_{opt}(\mathbf{x}, \mathbf{t}) = \arg\max_k \ \langle \mathbf{x}|\mathbf{y}(k)\rangle - \frac{1}{2}\|\mathbf{y}(k)\|_2^2$,
- *Uniform noise distribution:* $\mathcal{D}^{M,U}_{opt}(\mathbf{x}, \mathbf{t}) = \arg\max_k \ -\|\mathbf{x} - \mathbf{y}(k)\|_\infty$,
- *Laplace noise distribution:* $\mathcal{D}^{M,L}_{opt}(\mathbf{x}, \mathbf{t}) = \arg\max_k \ -\|\mathbf{x} - \mathbf{y}(k)\|_1$.

Theorem 2: optimal distinguisher when the leakage model is partially unknown

Let $\mathbf{Y}_{\boldsymbol{\alpha}}(k) = \sum_{j=1}^{n} \alpha_j [f(\mathbf{T}, k)]_j$, $\mathbf{Y}_j(k) = [f(\mathbf{T}, k)]_j$ *and* $\mathbf{X} = \sum_{j=1}^{n} \alpha_j [f(T, k^\star)]_j + N$ *with* $N \sim \mathcal{N}(0, \sigma^2)$. *Assuming weights are independently deviating normally from the Hamming weight model, then the optimal distinguishing rule is*

$$\mathcal{D}^{\alpha,G}(\mathbf{x}, \mathbf{t}) = \arg\max_k \ (\gamma\langle \mathbf{x}|\mathbf{y}(k)\rangle + \mathbf{1})^t \cdot (\gamma Z(k) + I)^{-1} \cdot (\gamma\langle \mathbf{x}|\mathbf{y}(k)\rangle + \mathbf{1}) - \sigma_\alpha^2 \ln\det(\gamma Z(k) + I) \ ,$$

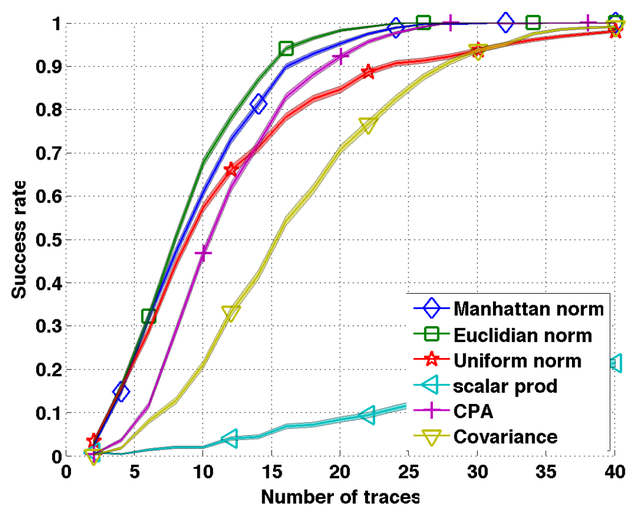
where $\gamma = \frac{\sigma_\alpha^2}{\sigma^2}$ *is the* **epistemic-to-stochastic-noise-ratio** *(ESNR).*

Correlation

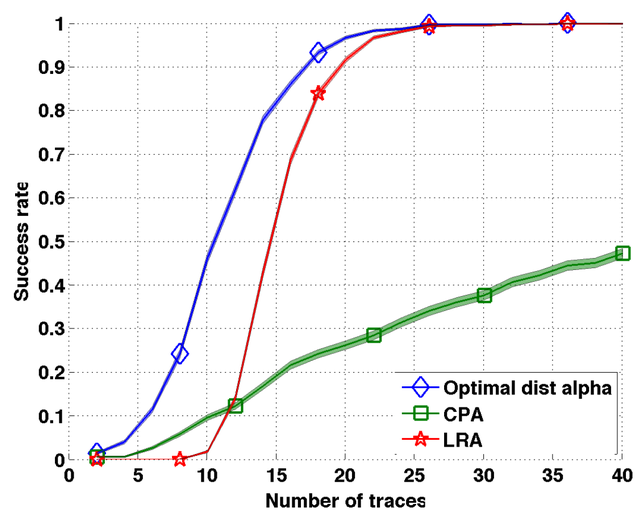
Covariance

Linear regression

Our novel **optimal** distinguishers **outperform** all state-of-the-art distinguishers depending on statistical tools in terms of the **success rate**!



Known model



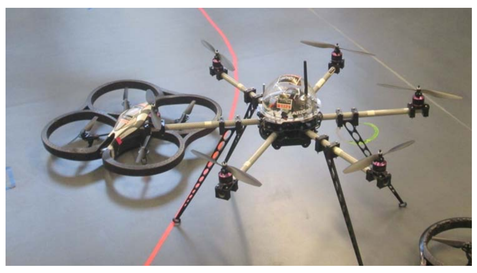
Partially unknown model

Contact annelie.heuser@telecom-paristech.fr Site web perso.telecom-paristech.fr/heuser

SysML-Sec: Capturing and Formally Proving Safety and Security Properties

Handling risk in safety-critical systems

- Automotive systems, avionics systems, nuclear power plants . . .
  - Digital car:
    - Security of over-the-air firmware updates, car control by malware [Koscher 2010], Autonomous vehicle safety (e.g., Google car)
    - Car navigation data spoofing [Andrea et al. 07]
  - Drones:
    - Sensitive data protection and communications security and safety
    - Autonomous support system: security and safety (hijacking, secure data fusion and interpretation, fault-tolerant attitude self-control)



Our proposal for security: SysML-Sec ...

- Objective: bring together system engineers and security experts
- Model-Driven Engineering from requirements to code generation
- Centered around a security-aware HW/SW partitioning
- Formal safety and security proofs
- Free software (TTool)

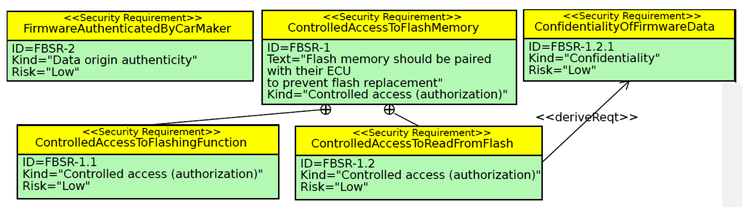
... Integration with safety models ongoing

TTool

An open source toolkit provided by

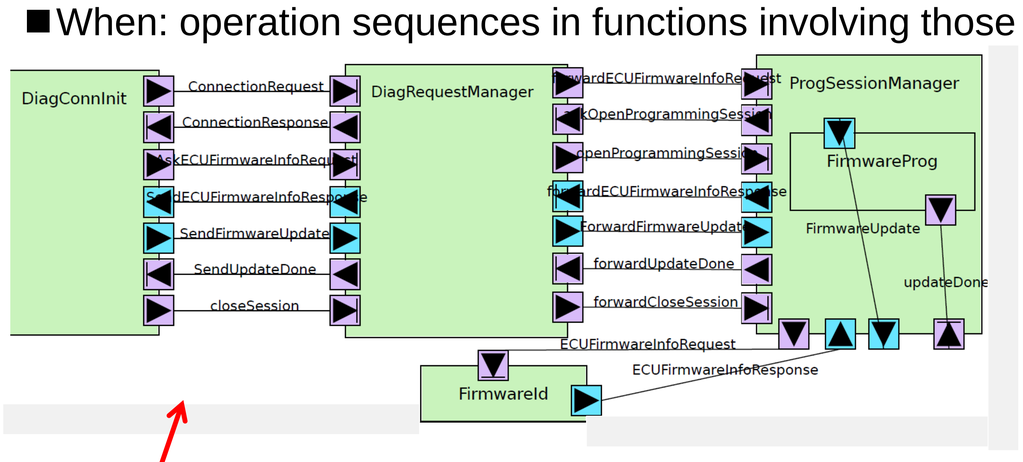
Requirements

- Who and why: stakeholders and security goals



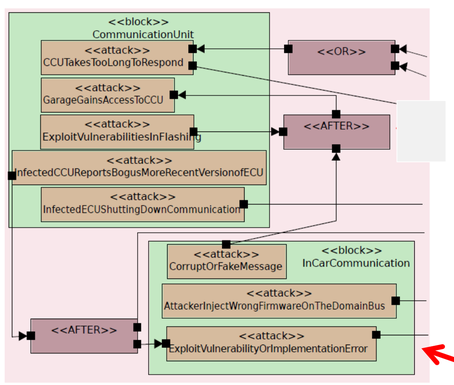
Application

- When: operation sequences in functions involving those assets



Attacks

- Who and Why: attackers, their capabilities, and objectives (risk analysis)

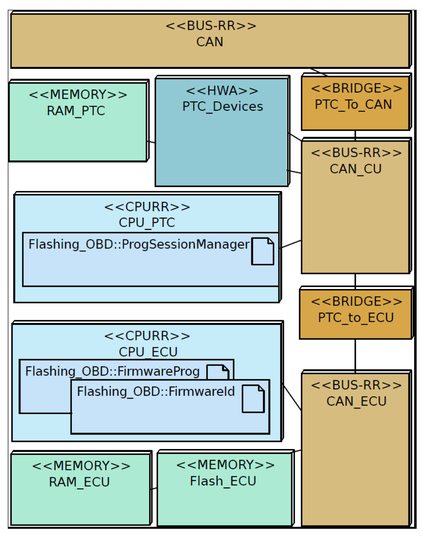


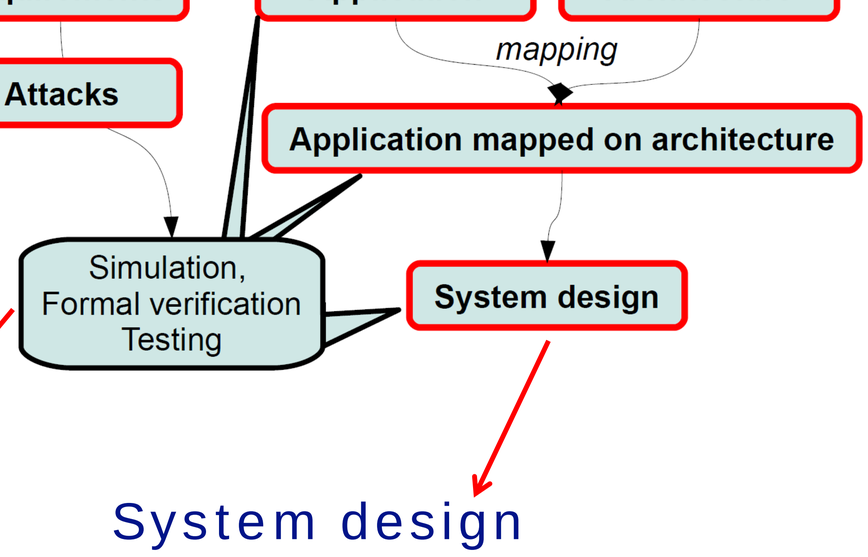
Architecture

- What: assets to be protected

Mapping

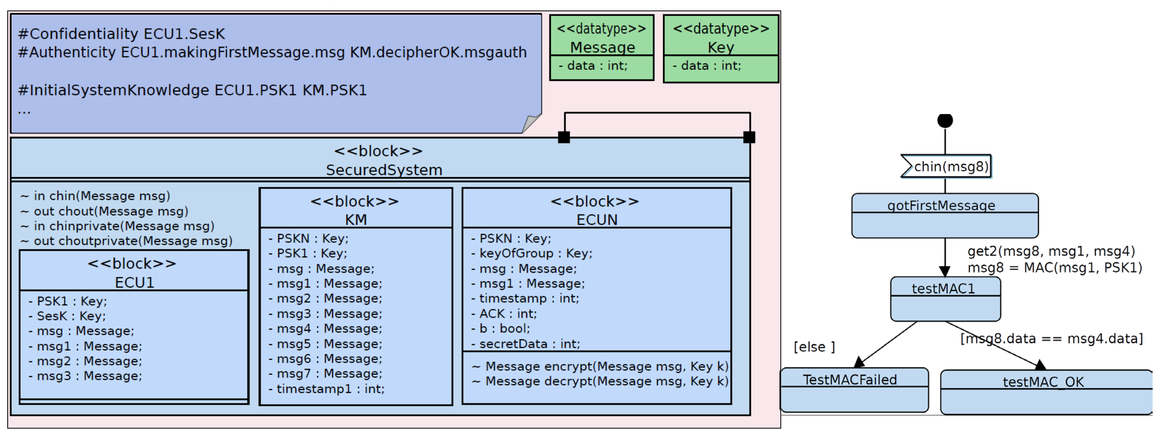
- Where: mapping of functions over architecture assets





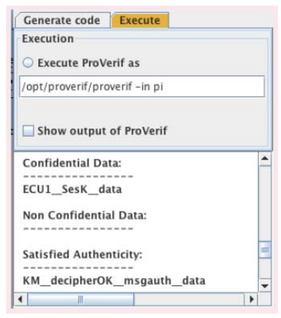
System design

- How: security objectives due to architecture (e.g., network topology, process isolation, etc.)



Formal verification

- Proof based on ProVerif
  - Authenticity, confidentiality
- Press-button approach from TTool



Authors

Ludovic Apvrille (Télécom ParisTech)

Ludovic.Apvrille@telecom-paristech.fr

Yves Roudier (EURECOM)

Tullio Joseph Tanzi (Télécom ParisTech)

Franck Guarnieri (Mines ParisTech)

Partners

Institutions: Télécom ParisTech, EURECOM, Mines ParisTech

Partners: Fondation Télécom, Institut Carnot Télécom & Société numérique

Circuits pour la Sûreté de Fonctionnement

Parties prenantes

Auteurs

Lirida Naviner, Jean-François Naviner, Hervé Petit

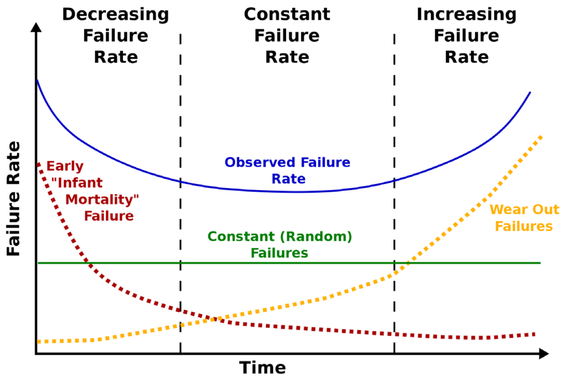
Doctorants : A. Ben Dhia, T. An, K. Liu, S. Sarrazin, C. Bottoni, B. Coeffic, N. Jovanovic, Y. Wang

Post-docs : M. Slimani, H. Cai, M. Costa

Partenaires

SURETE DE FONCTIONNEMENT

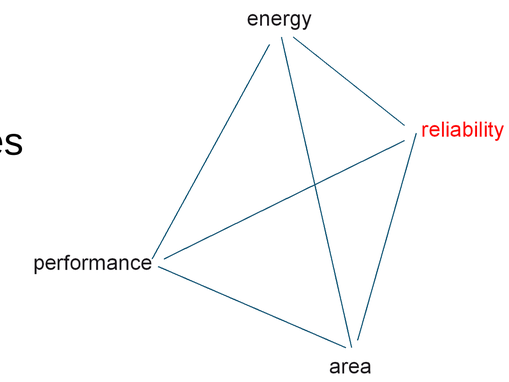
Contexte



- Technologie nanométrique et forte densité d'intégration (Loi de Moore)
- Circuits complexes et performants, mais vulnérables
- Augmentation du nombre de fautes
- Baisse du rendement de fabrication et de la fiabilité
- Industrie électronique « fabless »

Enjeux

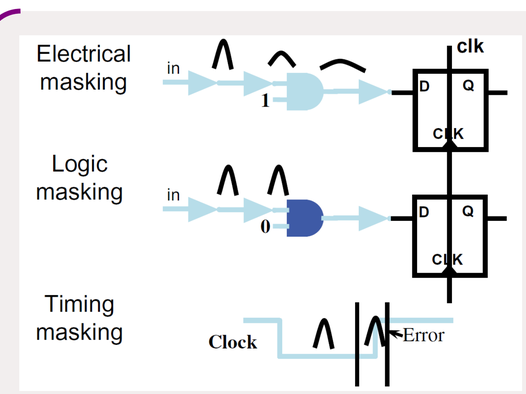
- Conception de systèmes électroniques sûrs et économiquement viables
- Intégration de la fiabilité dans le flot de conception
- Analyse et amélioration de la tolérance aux fautes

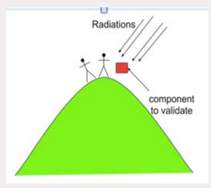


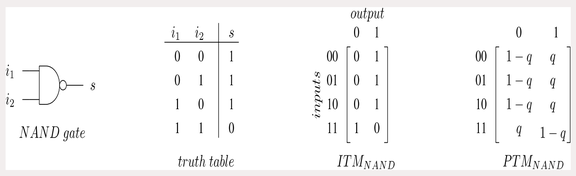
FAUTES TRANSITOIRES ET INTERMITTENTES

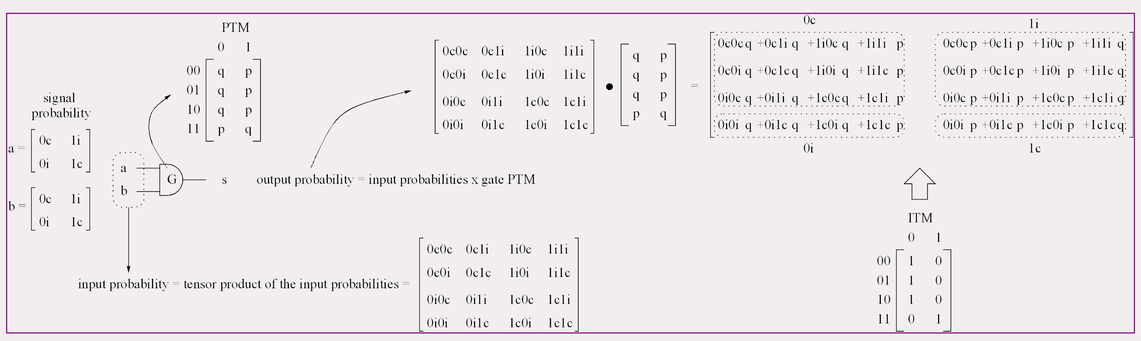
Rayonnement, Variabilité

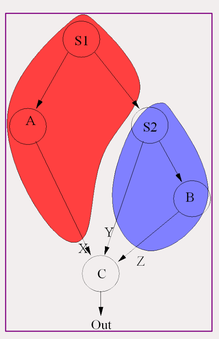
- Analyse de masquage logique
- Test en ligne (fautes de délai)
- Injection de fautes (FIFA)
- Analyse et compensation du bruit
- Durcissement sélectif
- Processeurs tolérants

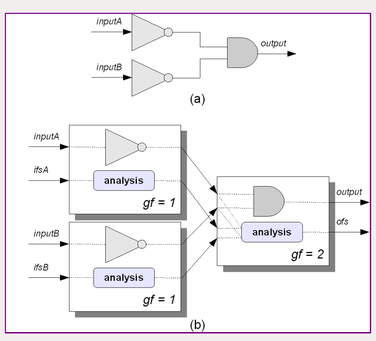


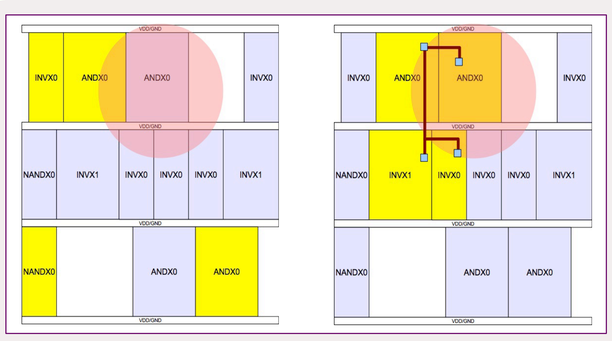


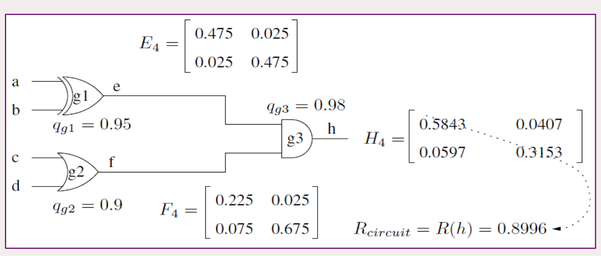


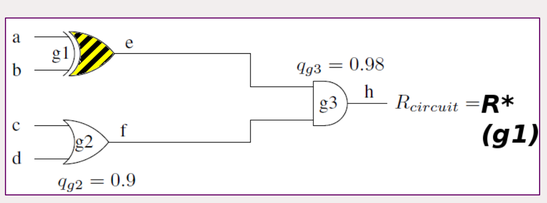








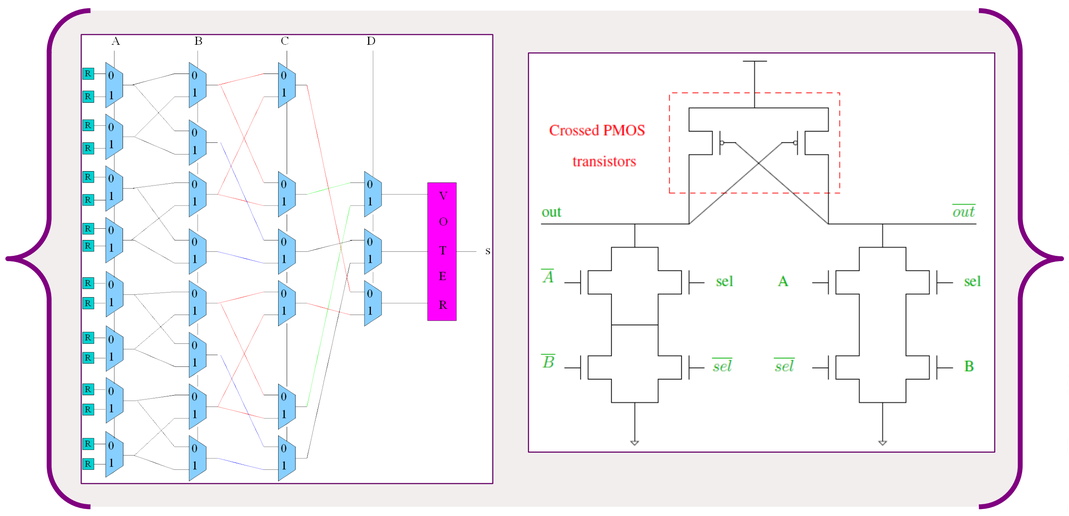




FAUTES PERMANENTES

Défauts de fabrication, Vieillissement

- Durcissement des blocs de base du FPGA
- Architectures robustes: Cross logic, DCVS
- Emulation/injection de défauts et analyse du taux de masquage



Numérique : Grande échelle et complexité

Paris, le 27 mars 2014

Contact lirida.naviner@telecom-paristech.fr

Site web nanoelec.wp.mines-telecom.fr

A geometrical model for selecting optimal countermeasures against multiple attack scenarios

Institut Mines-Télécom

Samovar UMR 5157

Authors

- Gustavo GONZALEZ GRANADILLO
- Hervé DEBAR
- Grégoire JACOB

RST Department

MOTIVATION

- Cyber-attacks are more sophisticated and complex.
- Challenges in the detection and reaction process.
- Huge amount of information from different sources.
- Current solutions do not provide a comprehensive impact analysis of attacks and countermeasures.
- Need of a model to evaluate complex and multiple attack scenarios.

RESULTS

- Quantitative model for evaluating, ranking, and selecting optimal countermeasures.
- Process to evaluate combinations of countermeasures, and select the one with the highest index.
- Deployment of the cost sensitive model over real attack scenarios provided by industrial partners.
- Geometrical model that represents the volume of systems, attacks, and countermeasures based on user accounts, channels, and resources.
- Impact evaluation and graphical representation of multiple attacks and countermeasures.

Context/funding

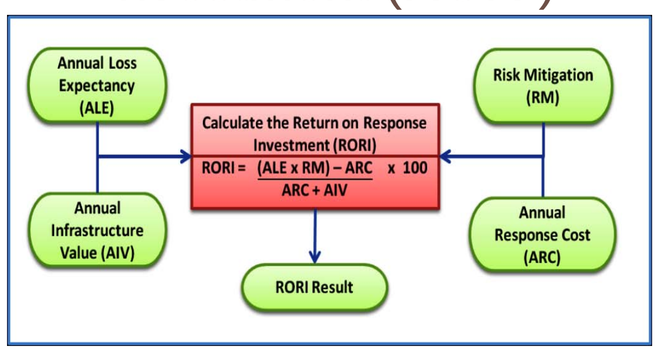
MASTF

SEVENTH FRAMEWORK PROGRAMME

ADAX

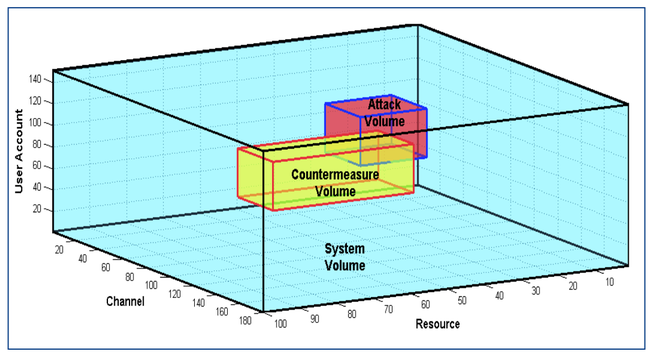
PROPOSAL

Return On Response Investment (RORI)



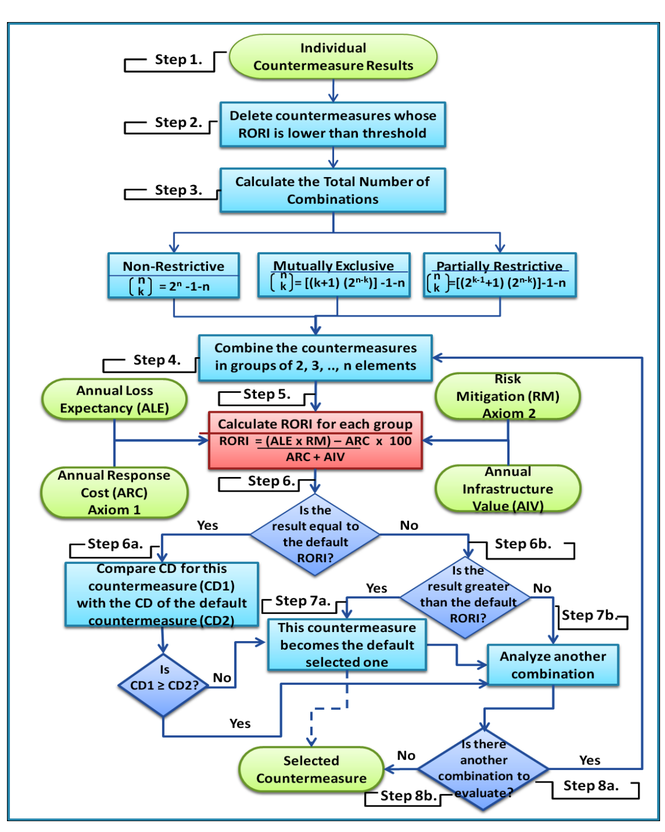
- **ALE:** Annual impact cost obtained in the absence of security countermeasures.
- **AIV:** Fixed costs expected on the system due to services, renting, and equipment maintenance.
- **RM:** Risk mitigation level associated to a particular countermeasure.
- **ARC:** Annual response cost incurred by implementing a new security action.

Attack Volume Model



- **System Volume:** Maximal space a given system is exposed to attackers.
- **Attack Volume:** Portion of the system that is targeted by a given attack based on the vulnerabilities it can exploit.
- **Countermeasure Volume:** Level of action a security solution has on a system over a given attack.

Countermeasure Selection Process

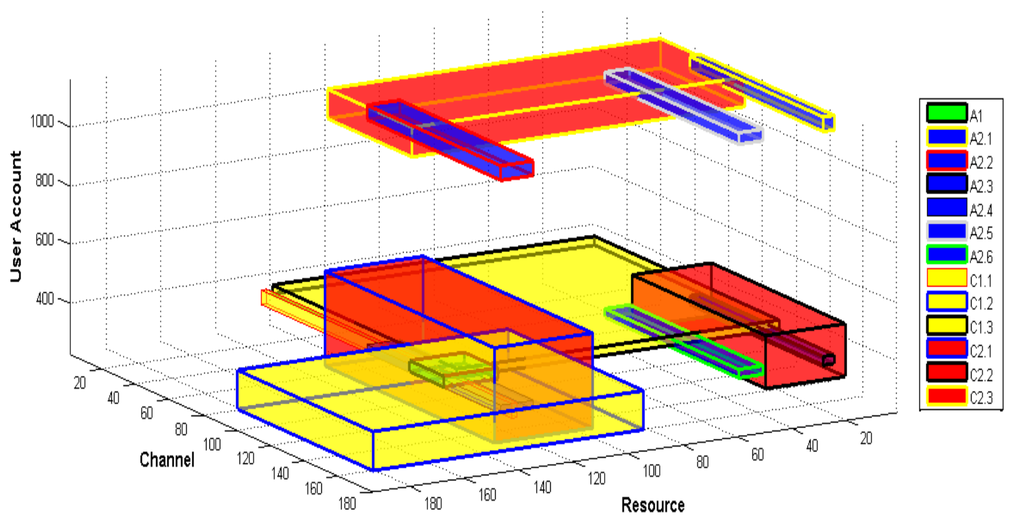


EXAMPLE

Multiple attacks detected at Telecom SudParis, France

Element	User Account	Channel	Resource	Volume (units^3)
System	U1:U3691	Ch1:Ch4512	R1:R993	430,106,901,440
Attack 1	U340:U377	Ch100:Ch120	R110:R130	904,932
Attack 2	U320:U349& U1110:U1159	Ch70:Ch149	R5:R9& R31:R40& R115:R127	8,380,800
CM 1.1	U300: U349	Ch1: Ch149	R121: R123	1,206,900
CM 1.2	U301: U433	Ch100: Ch179	R94: R193	57,456,000
CM 1.3	U330: U360	Ch1: Ch110	R1: R119	25,411,320
CM 2.1	U229: U550	Ch50: Ch110	R94: R130	35,124,840
CM 2.2	U270: U449	Ch70: Ch149	R1: R30	56,052,000
CM 2.3	U1030: U1130	Ch40: Ch90	R1: R123	14,551,218

Graphical representation of attacks and countermeasures



ITEA 2

INFORMATION TECHNOLOGY FOR EUROPEAN ADVANCEMENT

Contact gustavo.gonzalez_granadillo@ telecom-sudparis.eu

Site web www.telecom-sudparis.eu

Preuve formelle de la sécurité d'une application ferroviaire de signalisation en modes nominal et dégradés

Parties prenantes

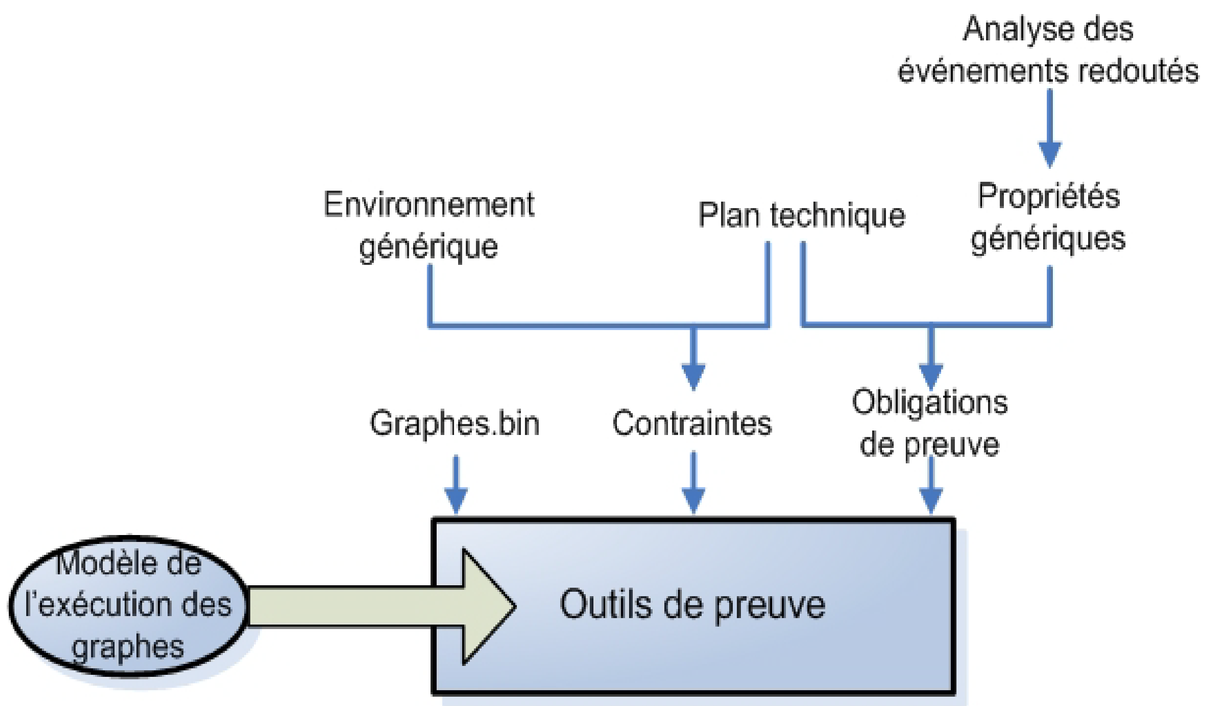
Auteurs

Amel Mammar (TSP)
Jean-Marc Mota (Thalès)

Approche générale par la preuve

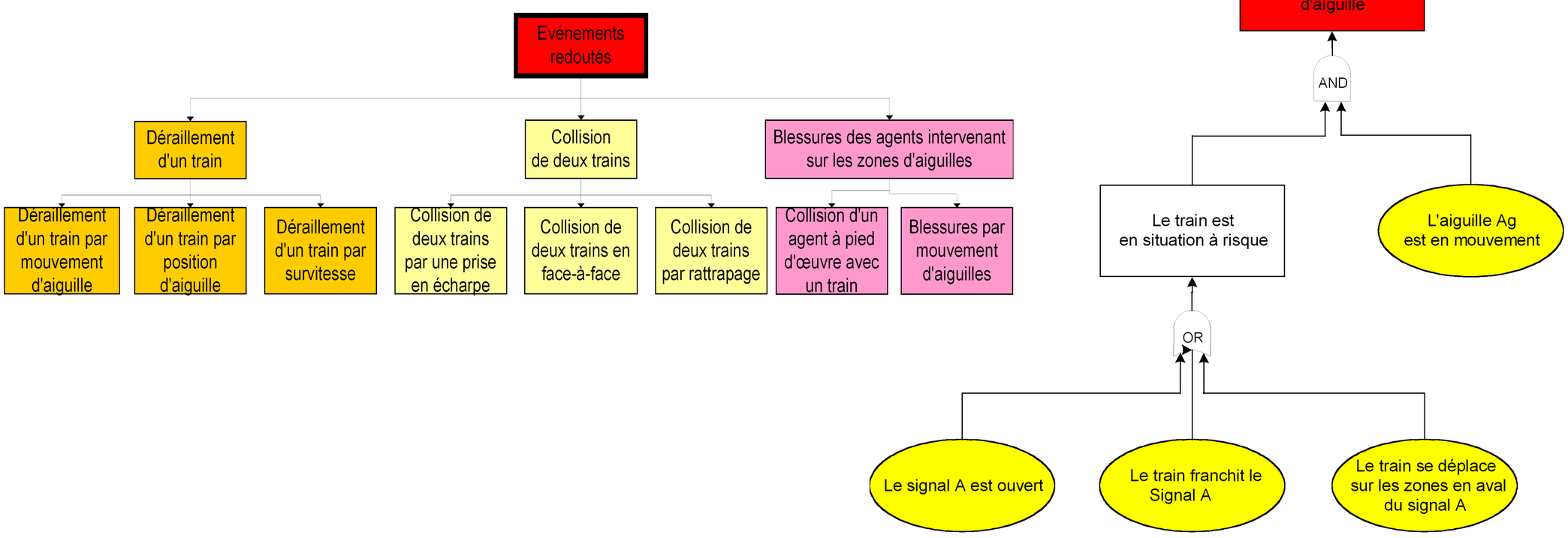
Objectifs

- Définir les différents scénarios menant aux situations redoutées
- Prendre en compte les défaillances possibles des éléments matériels
- Modéliser les scénarios par des propriétés de sécurité

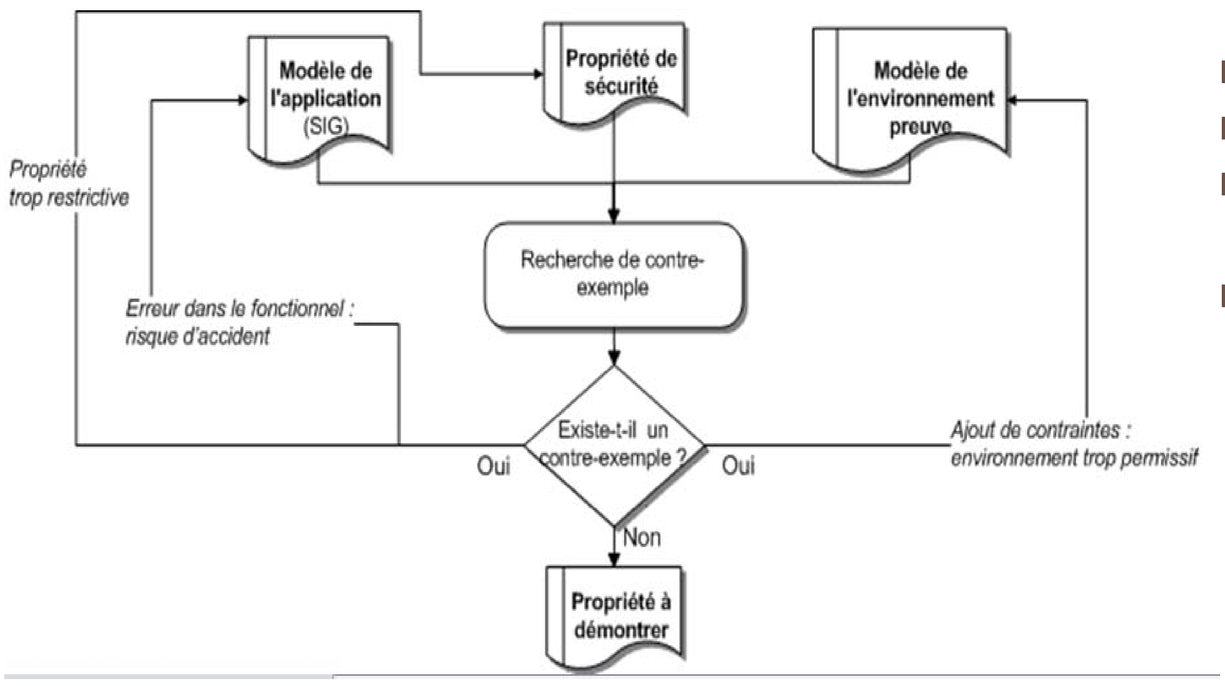


Analyse des événements redoutés

- Utilisation des arbres de fautes
- Démarche par raffinement d'événements



Preuve et modélisation de l'environnement par recherche de contre-exemples



- Une aiguille ne peut être simultanément à droite et à gauche
- Une aiguille ne change pas de position sans être commandée
- Une aiguille a besoin d'au moins deux cycles pour passer d'une position droite (resp. gauche) à une position gauche (resp. droite)
- Les actions de l'agent à pied d'œuvre sont conformes à la sécurité (modes dégradés)

colloque Institut Mines-Télécom "Numérique : Grande échelle et complexité"

Mars 2014

Contact amel.mammar@telecom-sudparis.eu Site web www. -public.it-sudparis.eu/~mammar_a/