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

I- Background
- Cloud computing providers (Amazon, Google, Rackspace, ...) serve virtual machines (VMs) that are hosted in large data centers.
- **Capacity Planning**: optimize resource 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 \(\rightarrow\) 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.

<table>
<thead>
<tr>
<th></th>
<th>Chord</th>
<th>LBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>0.496</td>
<td>0.863</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.378</td>
<td>0.798</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.629</td>
<td>0.935</td>
</tr>
</tbody>
</table>

**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.
Vers un Nuage responsable et sécurisé

**Parties prenantes**

MINES Nantes

**Contact**

Ronan-Alexandre Cherrueau, Mario Südholt
ASCOLA Research group
Mines Nantes
http://www.emn.fr/z-info/ascola

**Université de Nantes**

**Problématique**

Assurer les revendications des utilisateurs/fournisseurs sur l'utilisation des données pour un Nuage responsable 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**

- Requérir 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
- Décrire la violation d'une revendication

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

**Exemple**

Recommandation : OAuth Provider (OP)
- La revendication de la protection de l'identité des utilisateurs lors d'une authentification unique

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

**Conclusion & Perspectives**

- Langage de détection et de correction pour appliquer des politiques sur le Nuage
- Responsabiliser et sécuriser le Nuage

- Avoir une bibliothèque abstraite de solutions pour appliquer des politiques sur les données perso
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 l'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 voire 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.

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 figure 1 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 figure 2, l’algorithme K-MEANS partitionne un nombre important de petits groupes, assez homogènes en nombre de VM, finalement peu exploitables dans notre cas.

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, aucune méthode existante ne répondent à nos besoins. Nous avons alors développé notre propre algorithme de Clustering, insensible aux bruits, performants, multi-ressources et multicritères.
VM Scheduling for Capacity Planning in Distributed Clouds

I- Background

- Cloud computing providers (Amazon, Google, Rackspace, ...) serve virtual machines (VMs) that are hosted in large data centers.
- **Capacity Planning**: optimize resource usage within a cloud infrastructure.
- **VM Scheduling**: Overloaded VMs may be reallocated 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 reallocations
- 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 reallocations and minimize inter-site reallocations.

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.

<table>
<thead>
<tr>
<th></th>
<th>Chord</th>
<th>LBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>0.496</td>
<td>0.863</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.378</td>
<td>0.798</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.629</td>
<td>0.935</td>
</tr>
</tbody>
</table>

*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.

Contact
Jonathan Pastor, Mines Nantes, Ascola research team, Département Informatique Mines Nantes
jonathan.pastor@mines-nantes.fr

Site web http://beyondtheclouds.github.io/
**Organisation multi-agent pour la gouvernance du M2M**

- **Système multi-agent centré Organisation pour la gouvernance du M2M** :
  - **Agents**
    - Entité automatique pro-active qui gère le fonctionnement de l’infrastructure M2M.
  - **Artéfacts**
    - Composants de l’architecture M2M manipulables par les agents (ex. modèles de la plateforme, capteurs...).
  - **Spécifications Organisatoriales (SO) Framework**
    - Permet de garantir un fonctionnement global conforme aux spécifications du système.

- **Gouvernance Agile pour le passage à l’échelle de l’infrastructure M2M** :
  - Un ensemble de mesures de conduite stratégique et opérationnelle pour gérer la complexité de l’infrastructure M2M.

**Conclusion**

- **Contribution** :
  - Analyse du passage à l’échelle dans le M2M
  - Modèle de gouvernance des systèmes M2M basé sur une organisation Multi-Agents

- **Travaux futurs** :
  - Définir les comportements de réorganisation des agents : quand ? qui ? pourquoi ?
  - Garantir la cohérence des réorganisations
  - Déploiement/test sur l’infrastructure SensiCity

**Problématique du passage à l’échelle dans le M2M**

- **Définition : Scalability**
  - Un système S est scalable (déf. 1) sur un ensemble de dimensions (D) s’il peut être ajouté à un ensemble de propriétés (P), et s’il est capable de gérer les changements d’échelle en satisfaisant les spécifications (S) du système, ou si il est capable de faire un compromis sur l’un des paramètres (P), encore entre un ensemble satisfaisant parfaitement l’ensemble des spécifications.

- **États**
  - Dim. Infra : autonome, holistique, ouvert, optimisé, etc.
  - Dim. Prog : hiérarchisé, décentralisé, etc.

- **Définition de l’infrastructure M2M**
  - Les agents gèrent l’infrastructure M2M, dont les composants sont représentés par des artéfacts.
  - La spécification fonctionnelle (SF) définit les buts de l’infrastructure M2M, ainsi que les plans permettant d’atteindre ces objectifs.

- **Les normes ISO** assurent les missions aux rôles et définissent les conditions pour remplir ces missions.

**Contexte**

- Contraintes d’une architecture M2M en milieu urbain :
  - Masse de capteurs et actuateurs (M2M)
  - Utile-basse consommation et bas débit
  - Durée de vie (20 ans sur 1 batterie)

- Large échelle – milliers de capteurs, grande géographie, nombreuses applications, pour des volumes de données...
Système de recommandation de lieux basé sur l’intention, le contexte et le réseau social

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 :
  \[ \text{Règle } R : A \rightarrow B \]

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 \( A \subset (\text{session courante de visite}) \)
- Classement des règles \( R : A \rightarrow B \) selon une mesure de pertinence \( M_p \)

\[ M_p(R) = \frac{\text{Confiance}(R : A \rightarrow B) \times \text{Surprise}(R : A \rightarrow B)}{\text{P(A)}} \]

\( M_{\text{géographie}}(R) \)
\( M_{\text{intérêt}}(R) \)
\( M_{\text{social}}(R) \)

\[ M_p(R) = \alpha M_{\text{géographie}}(R) + (1 - \alpha) M_{\text{intérêt}}(R) \]

si intention = efficacité

si intention = découverte

Recour à des k lieux conséquents des k meilleures règles

APERÇU DU SYSTÈME DE RECOMMANDATION

APPLICATION SUR DONNEES REELLES
Recommandation de magasins, tests de montée en charge
- 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 réducteur unique (construction de l’arbre)
- Expérimentation et évaluation des recommandations sur des utilisateurs réels à venir
**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.
- **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].

Some References:

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

Multirate Digital Predistortion

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

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

Performance metrics simulated results

Future work

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

**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**

**Means**

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

**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
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.

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).

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.
**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 further add 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 ($v$) from all movement patterns.
- Determine population density from Census data.

**Model**

- Split the country into metapopulation [2,6] (subprefecture).
- Generate mobility between each metapopulation 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 across the country.

**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 manner (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 [5] 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 capital), Bouaké (the second largest city), Yopougon, (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 known to host mostly an agricultural region (Coffee, cocoa, 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 disparities 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".
Understanding the Evolution of Multimedia Content in the Internet through BitTorrent glasses

Reza Farahbaksh*, Angel Cuevas*, Ruben Cuevas**, Roberto Gonzalez**, Noel Crespi*
** 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
- The tool connects to the tracker and retrieves 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.

Analysis & Results

Content Evolution Analysis

- **Content Size Analysis**
  - BitTorrent content has doubled its size in a period of 2 years
  - In median from 223 MB to 458 MB

Content Popularity Evolution

- **Content Size per Category**
  - Box plot of content size per category for (a) pb09, (b) pb10, (c) pb11 and (d) pb12
    - 25th, 50th, 75th 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:
La Fibre: une révolution en marche
Les Infrastructures de réseaux d’accès
Les nouveaux usages

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
Un nouvel écosystème

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
**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.

**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.

**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 evolve a management system for content placement and delivery to mobile users by exploiting statistical patterns derived from mobility-, connectivity- and social information.

**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 Darmstadt, Germany;
- UNIVERSITY OF CAMBRIDGE, United Kingdom;
- UNIVERSIDAD CARLOS III DE MADRID, Spain

**Contact:**

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

http://www.ict-ecousin.eu
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

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

Additional Constraints
- Node mapping \[ \sum_{k \in V_R} x_{ik} = 1, \forall i \in V_P \]
- Limited storage \[ \sum_{k \in V_R} s_{k} \leq STO_{k}, \forall k \in V_P \backslash R \]
- Node & Link mapping \[ \sum_{k \in V_R} y_{ij,k,k_1,k_2} = 1, \forall (ij) \in E_P \]
- Latency \[ \sum_{k \in V_R, k \neq k_1} t_{ij,k} x_{ij} \leq t_{ij,k_1} \]
- Localisation \[ \sum_{k \in V_R} t_{ij,k} x_{ij} = \min \{ t_{ij,k_1} | k_1 \in V_R \} \]

Performance Results
- Convergence speed comparison (from hours to ms, \( ~10^4 \) or better improvement factor with heuristic close to optimal)
- Scalability of Heuristic Algorithm as a function of input and reference graph sizes
**Objective:** Design of a distributed context management framework for IoT context-aware applications

**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**

- Multiscaleability ≠ 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
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

Resource Analyzer

- Data extraction
- Push and Pull models
- Generic
- Granular
- Extendable

Plan

- Analysis rules
- Generic
- Extendable

Monitor

- Strategies
- Generic
- Extendable

Execute

- Actions
- Granular
- Generic
- Extendable

4. OCCI defined Entities and Mixins

Category

- Mix-in

Kind

- entity

Action

- Link

Alert

- Link

NotificationTool

SubscriptionLink

SubscriptionLink

ActionLink

AlertLink

5. Monitoring and Reconfiguration Infrastructure

OCCI Platform

- COAPS

Application

6. References


Contact: { Mohamed.Mohamed, Djamel.Belaid, Samir.Tata } @telecom-sudparis.eu
2. SYSTEMES INDUSTRIELS COMPLEXES
Interoperability requirements for collaborative process analysis

Auteurs
Nicolas Daclin
Sihem Mallek
Vincent Chapurlat

Interoperability requirements for collaborative process analysis

Overall context
- To verify interoperability requirements within collaborative processes models using formal verification techniques to detect and anticipate interoperability problems.

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?

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

Partenaires
Université Bordeaux 1

Interoperability requirements framework

Collaborative process model (BPMN 2.0)
**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 métamodelles, mais la création d’éditeurs graphiques est un point difficile).
- Un outillage compatible avec les environnements déjà existants dans l’écosystème Eclipse/EMF.

**Contexte**

**Contributions**

<table>
<thead>
<tr>
<th>Id</th>
<th>Aspect</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>open-source</td>
<td>Publié avec une licence libre et gratuite.</td>
</tr>
<tr>
<td>2</td>
<td>intégration des standards</td>
<td>Intégré dans l’écosystème Eclipse/EMF.</td>
</tr>
<tr>
<td>3</td>
<td>interaction utilisateur</td>
<td>Editeur de diagramme de classes étendu permettant la saisie des nœuds graphiques pour les entités du métamodèle.</td>
</tr>
<tr>
<td>4</td>
<td>indépendant de la plateforme</td>
<td>Les diagrammes créés sont indépendants de la plateforme spécifique.</td>
</tr>
<tr>
<td>5</td>
<td>abstraction</td>
<td>La structure d’un diagramme est définie par un métamodèle publié.</td>
</tr>
<tr>
<td>6</td>
<td>grammaire graphique</td>
<td>Le langage graphique et le système abstrait sont définis dans un langage de type DSL (Domain Specific Language).</td>
</tr>
<tr>
<td>7</td>
<td>composition</td>
<td>Un mécanisme de vues multiples, incluant la navigation entre les vues, fait partie intégrante de la syntaxe concrète.</td>
</tr>
<tr>
<td>8</td>
<td>héritage graphique</td>
<td>Les langages peuvent être composés, par un mécanisme de fusion.</td>
</tr>
<tr>
<td>9</td>
<td>support des hiérarchies</td>
<td>Un support natif des hiérarchies est fourni.</td>
</tr>
<tr>
<td>10</td>
<td>grammaire positionnelle</td>
<td>La position d’un nœud graphique est indépendante de la syntaxe concrète.</td>
</tr>
<tr>
<td>11</td>
<td>syntaxe statique</td>
<td>Les langages peuvent être composés, par un mécanisme de fusion.</td>
</tr>
<tr>
<td>12</td>
<td>héritage graphique</td>
<td>L’héritage graphique est implémenté par un mécanisme de fusion.</td>
</tr>
<tr>
<td>13</td>
<td>interface de construction</td>
<td>Des composants modèles de bases sur des motifs sont inférées.</td>
</tr>
</tbody>
</table>

**Un langage pour décrire des diagrammes**

*(sa syntaxe concrète)*

*Les mots-clés*

*Le métamodèle*

*(sa syntaxe abstraite)*

**Un DSML obtenu avec DIAGRAPH**

*sa syntaxe abstraite annotée avec les mots-clés DIAGRAPH*
Guide pour la modélisation et la vérification de la cohérence modes de fonctionnement / scénarios opérationnels en Ingénierie Système : GEMOS

**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)...

**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 UPPAAAL (TCTL et Automates à états temporisés) et/ou vers COGITANT (Graphes Conceptuels)
- **Simuler le comportement** : sémantique opérationnelle du GEMOS

**Processus opératoire actuel**

VEREINT : vétérinaire de recherche et d’intervention NRBC
**CONTRIBUTIONS** (Lô et al. 2013)

**Conceptuelle**
- Modèle conceptuel des données pour l’évaluation d’architectures en IS

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

**Technique**
- Enrichissement de l’atelier d’IS CORE pour l’évaluation d’alternatives de solutions

**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.
**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.

**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 ;

**Les projets (passés et présents) :**

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

---

**Parties prenantes**

- **Auteurs**
  - Frédéric BOYER
    - Professeur
    - Ecole des Mines de Nantes
  - Vincent LEBASTARD
    - Maître assistant
    - Ecole des Mines de Nantes
  - Mathieu POREZ
    - Maître assistant
    - Ecole des Mines de Nantes

**Partenaires**

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

---

2) Straightforward CP Model

\[
\text{minimise}(x) \quad \text{s.t.} \quad \bigvee_{i \in K} \quad \text{AtMostNValue}(x, z) \quad \forall z \in \mathbb{Z} \quad \text{Communication}
\]

\[
\begin{align*}
W \cup W' \cap \text{Dom}(x_i) & = \emptyset, \\
\text{Dom}(x_i) & = W_i, \\
W_i & = \{w_1, w_2, w_3\} \\
W' & = \{w_4, w_5\}
\end{align*}
\]

\[
\text{AtMostNValue}(x, z) = \bigvee_{i \in K} \quad \text{AtMostNValue}(x, z_i)
\]

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 \( G_i \):

- \( R_1: \bar{z} = \max([z_i]) \)
- \( R_2: A = \bar{z} \Rightarrow \forall i \in V, \text{Dom}(x_i) \leftarrow \text{Dom}(x_i) \cap \bigcup_{j \in A} \text{Dom}(x_j) \)

4) Improving filtering (Graph & Rules)

The constrained intersection graph \( G_{C2} \) of the set of variables \( X \) and the set of differences constraints \( C \) is deduced from \( G_i \) by removing edges \((i, j)\) whenever \( \text{neg}(i, j) \notin C \)

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

Results after 5 min with \( \text{AMNV} < G_{C2} | \text{MD}, R^k | R_{1,2} > \)

6) Results & Literature

- AMNV < \( G_{C1} | \text{MD} | R_{1,2} > \)
- AMNV < \( G_{C1} | \text{MD} | R_{1,3} > \)
- \( LB_a \)
- AMNV < \( G_{C2} | \text{MD}, R^k | R_{1,3} > \)
- Krishnamoorthy et al.
- MIP
- CP
- Smet et al.
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)

- Forecast Programme Gaspard
- Main Components of UCP-ModelSeeker
- Why, What, How?

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

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

<table>
<thead>
<tr>
<th>Variant</th>
<th>Search</th>
<th>Cluster</th>
<th>Split</th>
<th>MAL</th>
<th>MSE</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Frequent</td>
<td>All</td>
<td>no</td>
<td>452.30</td>
<td>87.53</td>
<td>1.10</td>
</tr>
<tr>
<td>B</td>
<td>Frequent</td>
<td>1</td>
<td>no</td>
<td>449.67</td>
<td>104.62</td>
<td>0.53</td>
</tr>
<tr>
<td>C</td>
<td>Frequent</td>
<td>3</td>
<td>no</td>
<td>206.43</td>
<td>70.90</td>
<td>0.87</td>
</tr>
<tr>
<td>D</td>
<td>Random</td>
<td>All</td>
<td>no</td>
<td>649.97</td>
<td>114.30</td>
<td>1.45</td>
</tr>
<tr>
<td>E</td>
<td>Random</td>
<td>1</td>
<td>no</td>
<td>492.90</td>
<td>106.68</td>
<td>0.54</td>
</tr>
<tr>
<td>F</td>
<td>Random</td>
<td>2</td>
<td>no</td>
<td>422.33</td>
<td>82.94</td>
<td>0.86</td>
</tr>
<tr>
<td>G</td>
<td>Frequent</td>
<td>All</td>
<td>yes</td>
<td>445.10</td>
<td>87.82</td>
<td>2.30</td>
</tr>
<tr>
<td>H</td>
<td>Frequent</td>
<td>1</td>
<td>yes</td>
<td>431.43</td>
<td>101.70</td>
<td>0.95</td>
</tr>
<tr>
<td>I</td>
<td>Frequent</td>
<td>2</td>
<td>yes</td>
<td>294.00</td>
<td>70.70</td>
<td>1.74</td>
</tr>
<tr>
<td>J</td>
<td>Random</td>
<td>All</td>
<td>yes</td>
<td>547.37</td>
<td>97.23</td>
<td>2.37</td>
</tr>
<tr>
<td>K</td>
<td>Random</td>
<td>1</td>
<td>yes</td>
<td>510.22</td>
<td>111.71</td>
<td>1.03</td>
</tr>
<tr>
<td>L</td>
<td>Random</td>
<td>2</td>
<td>yes</td>
<td>397.68</td>
<td>78.38</td>
<td>1.73</td>
</tr>
</tbody>
</table>

For More Information
http://4c.ucc.ie/~hsimonis/modelseeker.pdf
Real-Time Control of Metal Sheet Lamination Process

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.

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

Paved-Road Aging Evaluation

Detection and analysis of cracks and 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.


contact: Petr.Dokladal@mines-paristech.fr

Automated Visual Inspection of Electronic Cards

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
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.

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 an ecosystem. The similarity between the requirements of the railway and aviation safety equipment, make such tool chains a good basis for the project.
3. GRANDES MASSES DE DONNEES
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
- Désambiguiser, 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

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 par articles scientifiques biomédicaux

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
- Polarisations conceptuelles, extraction conceptuelle et évaluation de critères

Indexation conceptuelle

Indexation par propagation
- Interface interactive d’assistance à l’indexation
- Annotation conceptuelle semi-automatique

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

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

1Centre For Computational Biology, Mines ParisTech ; INSERM U1000 ; Institut Curie
2Paris 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 information 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 → 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 $x_i$ be a vector of descriptors of cell line $i$ and $y_j$ be a vector of descriptors of chemical $j$

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

$$t_{ij} = f(x_i, y_j) + b_i + c_{ij}$$

We estimate $f$ and $b$ by penalized least-square regression

$$\min_{f \in \mathcal{F}, b \in \mathbb{R}^n} \sum_{i=1}^{m} \left( t_{ij} - f(x_i, y_j) - b_i \right)^2 + \lambda \| f \|_2^2$$

To specify $\| f \|_2^2$, we just have to choose two kernels:

→ $K_{cell}(x_i, x_k)$ which measures the similarity between cell lines $x_i$ and $x_k$.
→ $K_{drug}(y_m, y_n)$ which measures the similarity between chemicals $y_m$ and $y_n$.

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

$$f(x, y) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} K_{cell}(x_i, x_k) K_{drug}(y_j, y_k)$$

where $c_{ij}$ depend only on:

→ $K_{cell}(x_i, x_k)$
→ $K_{drug}(y_m, y_n)$

The toxicity matrix.

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

References

Simulation directe à partir de la datamasse des images 3D

**STRATÉGIE GÉNÉRALE**

**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**
- « Mailler d’images » : génération avec un maillage 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, ε=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

**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 ENVIROMENTS 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

**Auteurs**

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

**Partenaires**

**GEOSCIENCES**

**MODELISATION DES ENVIRONNEMENTS URBAINS**

**MATERIAUX COMPOSITES**

**MICROSTRUCTURES METALLIQUES**

**STRATÉGIE GÉNÉRALE**

**Auteurs**

Luisa Silva
Rima Ghazal
Daniele Craciun
Jia-Yin Zhao
Min Quan Thai
Sébastien Nadler
et al.
MoGDIW, an integrated workflow for cell motility genes discovery in high-throughput time-lapse screening data

1,2,3Alice Schoenauer Sebag, 2Céline Raulet-Tomkiewicz, 2Robert Barouki, 1Jean-Philippe Vert, 1Thomas Walter
1Center For Computational Biology, Mines ParisTech ; INSEERM U900 ; Institut Curie
2INSEERM U904 ; Paris V
3Agro 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.

MoGDIW: an integrated workflow

Aim : quantitatively 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

Strategy : use of Machine Learning

Structured learning for cell tracking

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 :
  \[
  \hat{x}(t) = \arg \max_{x(t), \sigma, w} L(x(t), x; w) = \sum_{\text{frames}} \left< w^*, r_{ij}^* > x_{ij} \right> \text{ s.t. } \forall i \in \{0, \ldots, N(t)\} \sum_{j=1}^{n} x_{ij} = 1 \text{ and } \forall j \in \{0, \ldots, N(t+1)\} \sum_{i=1}^{m} x_{ij} = 1
  \]
- 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
  - Constrained nearest neighbour (Cell Cognition) : 72.7%, 62.8% [1]
  - Linear assignment 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>
<thead>
<tr>
<th>Feature type</th>
<th>Examples of feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Track characterization</td>
<td>Diffusion coefficient, persistence, track evenness</td>
</tr>
<tr>
<td>Static quantification</td>
<td>Convex hull area, largest move, total path length</td>
</tr>
<tr>
<td>Dynamic quantification</td>
<td>Mean acceleration, mean instantaneous speed</td>
</tr>
</tbody>
</table>

Experiment characterization :
- Clustering of all trajectories across experiments
- Characterization of an experiment by the cluster histogram of its trajectories
- Pearson's \chi^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, Israel)

- Method : MoGDIW with K-means, k=4

- Results in comparison with [5] :
  - Small overlap of selected gene lists
  - No enrichment 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

References


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

Table of feature types and examples of corresponding features

<table>
<thead>
<tr>
<th>Feature type</th>
<th>Examples of feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Track characterization</td>
<td></td>
</tr>
<tr>
<td>Static quantification</td>
<td></td>
</tr>
<tr>
<td>Dynamic quantification</td>
<td></td>
</tr>
</tbody>
</table>

Publication: Mines ParisTech
Adaptation de maillage anisotope, calcul parallèle et capture d’interface
Applications au matériaux et à la mise en forme, mais pas seulement...

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
  - Répartitionnement 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
Visual textures as realizations of multivariate log-Gaussian Cox processes
Huynh-Giao Nguyen, Ronan Fablet, Jean-Marc Boucher
Institut Telecom / Telecom Bretagne / LabSTIC (CS 83818 - 29238 Brest Cedex 3 - France)
(huyn.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 [6] 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[11] with three datasets: UIUC, KTH-Tips, Brodatz.

Descriptive statistics of multivariate point processes

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

\[
\mu_p(B \times \ldots \times B) = \mathbb{E} \left[ N(B_1) \ldots N(B_p) \right]
\]

For a marked spatial point process \( \{ s_1, m_j \} \) in given bounded regions \( B_j \), where \( m_j \) is a mark associated to point \( s_j \), the second-order moment is given by:

\[
\gamma_2(B \times B) = \mathbb{E} \sum_{s_j \in B} I(s_j) M_B(s_j)
\]

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 \( i \) :

\[
K(r) = (\lambda_i \lambda_j)^{-1} \gamma_2(r) = (\lambda_i \lambda_j)^{-1} \mathbb{E} \sum_{s_j \in B} \mathcal{D}(m_j) \mathcal{D}(m_i) I(s_j - s_i \leq r)
\]

Log-Gaussian Cox processes

Cox processes \( \{ X \} \) with random intensity functions \( \{ \lambda_x \} \) are point processes such that \( X \) is a Poisson process, where \( Z \) is \( \exp \{ Y \} \). \( \{ Y \} \) is a multivariate Gaussian field on \( S \) characterized with mean function \( \mu = \mathbb{E}(Y) \) and covariance function \( \Sigma(Y) = \text{Cov}(Y(s), Y(s')) \).

- **Intensity function**: Pair correlation - Ripley’s K function

\[
\lambda(x) = \mathbb{E}(\exp(\mu x)) \quad c(r) = \mathbb{E}(\exp(\mu x)) \quad K(r) = 2\pi \int_0^r \mathcal{D}(r') dr'
\]

- **The estimation of the pair correlation function is given by**:

\[
g(r) = (2\pi \lambda)^2 \sum_{l \in h} \mathcal{D}(m_l) \mathcal{D}(m_i) I(s_l - s_i \leq r)
\]

- **Given a parameterization \( L(\beta, \alpha) \)**, namely Exponential, Hyperbolic or C Cardinal sine, model parameters are estimated from the minimization of the following criterion:

\[
\int_0 \left[ \mathcal{L}_L(\beta, \alpha) - c_0(r) \right]^2 dr
\]

The proposed descriptor is formed by \( (\alpha, \beta_0) \).

Edge-effect corrections:

Feature dimension reduction: A codebook of keypoint pairs \( u = (s, s') \) from two categorized keypoint \( s, s' \) is considered, such that:

\[
g(u) = (2\pi \lambda)^2 \sum_{l \in h} \mathcal{D}(M(s, s')) \mathcal{D}(M_l) I(s_l - s_i \leq r)
\]

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

\[
r_i = \frac{r_{ref}}{s_{ref}}
\]

Application to texture recognition

- Categories of visual keypoints \( k = \{60, 120, 150\} \), pairs of keypoints \( k = 60 \).
- \( r_{ref} = \min(w(h)/2 \text{ nx}) \) where \( \min(w(h)) \) the size of image, \( \exp(0.1 \exp(1)) \).
- Covariance function \( L(\beta, \alpha) \) is Gaussian function.
- Classifier: random forest.

Result:

<table>
<thead>
<tr>
<th>( k )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>UIUC</td>
<td>0.86</td>
<td>0.84</td>
<td>0.85</td>
<td>0.83</td>
<td>0.82</td>
<td>0.81</td>
<td>0.80</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>KTH-Tips</td>
<td>0.87</td>
<td>0.86</td>
<td>0.85</td>
<td>0.84</td>
<td>0.83</td>
<td>0.82</td>
<td>0.81</td>
<td>0.80</td>
<td>0.79</td>
</tr>
<tr>
<td>Brodatz</td>
<td>0.88</td>
<td>0.87</td>
<td>0.86</td>
<td>0.85</td>
<td>0.84</td>
<td>0.83</td>
<td>0.82</td>
<td>0.81</td>
<td>0.80</td>
</tr>
</tbody>
</table>

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

References:

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 gray-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 multi-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 [6] and fully characterize the geometry of an image.

Correlated random walk model [1]:

\[ d\theta(s) = -\gamma(\theta(s) - \theta_0) + \sigma dW(s) \]  

Directional drift - Brownian process

Parameters \( \gamma \) and \( \sigma \) 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-Planck representation, from which one can derive the stationary statistics of the random direction \( \theta \) and turning angle \( \Delta \theta \) [1]:

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

\[ p(\theta) \propto \exp\left(-\frac{\alpha}{2}(\theta - \theta_0)^2\right) \quad \text{with} \quad \alpha = \Delta \theta (1 - \exp(-\Delta \theta)) \]

Stochastic geometry-driven texture model

Stochastic geometry-driven texture model generalization of (1) for orientation fields:

\[ \nabla_{\theta} \theta = -\gamma (\theta - \theta_0) + \sigma \nabla_{\theta} W \]  

Reference orientation - 2D Brownian sheet

Where \( \nabla_{\theta} \) is the gradient in direction \( \theta \).

Image \( I \) such that its level-lines are everywhere tangent to random vector field \( \nabla I(\theta_0, \theta_0) = 0 \), \( \forall \theta \) (3).

Implementation:

1. Simulate an orientation field from a numerical integration of the stochastic equation,
2. Solve for \( \theta \) as a variational minimization of (3).

Application to texture-based super-resolution of geophysical fields

Problem statement

\( I_{\text{HR}} \): High-resolution image \( \text{NLM} \) grid

\( I_{\text{LR}} \): Low-resolution image \( \text{NLM} \) grid (here, \( K=2^4 \))

Stochastic super-resolution model

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

\[ \frac{\partial \theta(p)}{\partial p} = -\gamma(p) \left( \theta(p) - \theta_{LR}(p) \right) \quad \text{dp} + \sigma(p) dW(p) \]

\[ \hat{I} = \arg \min_{I} \int \| \nabla I(p, \theta(p)) - \| dp \quad \text{subject to} \quad I_{\text{LR}} = \mathcal{P} [\hat{I}] \]

with \( W \) a Brownian sheet and \( \theta_{LR} \) 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:

\[ \mathcal{P} I_{\text{LR}} = \mathcal{P} I_{\text{H}} \quad \text{and} \quad \mathcal{P} \left[I_{\text{H}} - \mathcal{P} I_{\text{H}}\right] \]

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

Parameter fields \( \gamma \) and \( \sigma \) are set according to the following observation:

1. Large gradients in \( I_{\text{LR}} \) result in more regular level-lines in \( I_{\text{H}} \).
2. Conversely, weak gradients in \( I_{\text{LR}} \) involve regular level-lines in \( I_{\text{H}} \).
3. \( \gamma(p) = \gamma_{0} | \nabla_{\theta} I_{\text{LR}}(p)| \), \( \sigma(p) = \sigma_{0} | \nabla_{\theta} I_{\text{LR}}(p)|^{\gamma} \)

(Here, we set empirically \( \gamma_{0}=2^{4}, \sigma_{0}=0.54, \gamma=0.14, \beta=0.13) \)

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 (D).
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

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

\[ y = H B_s + z \]

Sparse Decomposition

- Let \( Q = \{ q_1, \ldots, q_M \} \) the finite alphabet transmit constellation. Let \( q = \{ q_1, \ldots, q_M \} \) the vector space in which the finite alphabet vector can be cast, and \( B \) the decomposition matrix defined as

\[ B = \begin{pmatrix} q & 0 & \cdots & 0 \\ 0 & q & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q \end{pmatrix} \]

- The transmitted symbol vector \( x \) can be rewritten after symbol decomposition on \( q \) as \( x = B y \).

- The received signal is reformulated as \( y = H B_s + z \).

New MIMO Detector based on \( \ell_1 \)-norm minimization

- The new problem is the decoding of the binary source vector \( s \). To this end, we propose to solve it using the following minimization problem

\[ \arg \min_{s \in \{0,1\}^N} \| y - H B_s \|_1, \quad \text{subject to } s \in \{0,1\}^N \]

where \( \epsilon \) is a constant defined later, and \( B = \begin{pmatrix} 1_0 & 0_0 & \cdots & 0_0 \\ 0_0 & 1_0 & \cdots & 0_0 \\ \vdots & \vdots & \ddots & \vdots \\ 0_0 & 0_0 & \cdots & 1_0 \end{pmatrix} \).

- In the literature of sparse reconstruction, the \( \ell_2 \)-norm can be relaxed by the \( \ell_1 \)-norm minimization, and the problem is resolved using

\[ \arg \min_{s \in \{0,1\}^N} \| y - H B_s \|_1, \quad \text{subject to } s \in \{0,1\}^N \]

Applications

- Large MIMO systems
  - In a noiseless channel, the equivalence between the \( \ell_2 \)-norm and the \( \ell_1 \)-norm hold for large dimensions of \( s \).
- MIMO frequency selective channel
  - The received signal can be written as

\[ y = \begin{pmatrix} y(1) \\ y(T_f + L) \\ \vdots \end{pmatrix} = H_{11} x(1) + H_{12} x(T_f) + \cdots + H_{1L} x(T_f + L) \]

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.
Hospital System staffing under pressure

![Graph showing demand and supply over time](https://via.placeholder.com/150)

Source: Angus JAMA 2000

The CURVEX solution

- FreeO2, FreeCO2
- ICU and home-care ventilators
- Multiple settings
- Database
- Corrective Action
- Monitoring, alert, repair by Technicians
- Respiratory flow, ECG, SPO2
- Signal Analysis
- Detection, Estimation, Prediction
- Multiple sensors error monitoring by Curvex
- CURVEX METHODOLOGY
- Monitoring, review, set-up by Physicians

Achievements

- Monitoring of mechanical ventilation (new mathematical 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

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

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.

Données

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

Méthode

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

- Identification des K fonctions de transfert à partir d’un modèle de régressions linéaires latentes (cf. [3]) :
  \[ p(Y|X, \theta) = \sum_{k=1}^{K} \lambda_k N_k(Y; X \beta_k, \Sigma_k) \]

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

Résultats

<table>
<thead>
<tr>
<th>Mode 1</th>
<th>Mode 2</th>
<th>Mode 3</th>
<th>Mode 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courant (intensité)</td>
<td>fort (principal Aiguilles + tourbillons)</td>
<td>fort (secondaire Aiguilles)</td>
<td>faible</td>
</tr>
<tr>
<td>Courant (sens)</td>
<td>Nord-Sud</td>
<td>Ouest-Est</td>
<td>Est-Ouest</td>
</tr>
<tr>
<td>Température</td>
<td>élevée</td>
<td>élevée</td>
<td>moyenne</td>
</tr>
<tr>
<td>SQG-like</td>
<td>√</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 raffinement 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.

Références


Partenaires

Ifremer
Lab-STICC – TOMS
Brest, France

Remerciements

Contact
pierre.tandeo@telecom-bretagne.eu

Site web
http://tandeo.wordpress.com
DE NOUVEAUX RÔLES POUR UNE STRUCTURE ANCIENEMENT CONNU

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.

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)

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.

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
- 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
Suivi d’objets multiples dans des scènes vidéo complexes

**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, PROBLÉMATIQUE ET MÉTHODES**

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).

**TRAVAUX ACTUELS**

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

**PARTENAIRES**

- Christelle Garnier
- Mehdi Oulad Améziane
- François Septier
- Yves Delignon
- Emmanuel Duflos

**AUTEURS**

- Christelle Garnier
- Mehdi Oulad Améziane
- François Septier
- Yves Delignon
- Emmanuel Duflos

**DEFINITIONS**

- MCMC : Méthodes de Monte-Carlo par chaînes de Markov.
- Particle Filter : Filtre particulaire.

**SUJETS ACTUELS**

- 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.

**PARTENAIRES**

- UMR CNRS 8219
- CENTRALE LILLE
- TELECOM 100 Lille
- LACIS

**CONTACT**

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

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).

Partenaires

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

Parties prenantes

Auteurs

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

Simulation d’un rejet d’agent toxique au cœur du quartier de la Défense à Paris

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

Approach: Riemannian Geometry

Overview

Extremal Human Curve extraction

Pose matching

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

Selected publications

Statistical Shape Analysis of a Large 3D Faces Dataset

Hassen Drira*, Boulbaba Ben Amor*, Mohamed Daoudi* and Anuj Srivastava†

* Institut Mines-Télécom/Télécom Lille, France.
† Department of Statistics, Florida State University, Tallahassee, FL 32306, USA.

hassen.drira@telecom-lille.fr

Statistical Shape Analysis of a Large 3D Faces Dataset

MOTIVATIONS

(a) Challenges and contributions
(i) Statistical shape tools for gallery organization
(ii) Propose an elastic metric and local representation to model 3D face deformations.

(b) Need for elastic representation of facial surfaces

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

Neutral face
Expression face
Stretching
Neutral face
Expression face
Shrinking

Expression variations and quality of 3D faces

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

Galley

Expression variations and quality of 3D faces

(iii) Statistical shape tools for gallery organization

Neutral face
Expression face
Expressive face

Distance along line (Euclidean)
Distance along surface (Geodesic)

EXPERIMENTAL RESULTS

(a) Experimental protocol
(b) Experimental results

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>99%</td>
<td>98.3%</td>
<td>97%</td>
<td>94.1%</td>
<td>98.4%</td>
<td>97.2%</td>
<td>97%</td>
</tr>
</tbody>
</table>

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).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All vs. All</td>
<td>-</td>
<td>95.2%</td>
<td>81.2%</td>
<td>96.5%</td>
<td>94.6%</td>
<td>98.1%</td>
</tr>
</tbody>
</table>

Selected publications


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 \( \theta_a \) and has expectation \( \mu_a \): optimal arm is \( a^* = \text{argmax} \mu_a \) and \( \mu^* = \mu_{a^*} \) is the highest expectation of reward associated.

Two probabilistic modelings

**Frequentist:**
- \( \theta_1, \ldots, \theta_K \) unknown parameters
- \( (Y_{a,t})_t \) i.i.d. with distribution \( \nu_{\theta_a} \)
- At time \( t \), arm \( A_t \) is chosen and reward \( X_t = Y_{A_t,t} \) is observed

**Bayesian:**
- \( \theta_a \) i.i.d. \( \pi_a \)
- \( (Y_{a,t})_t \) i.i.d. conditionally to \( \theta_a \) with distribution \( \nu_{\theta_a} \)

Two measures of performance

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

**Optimal algorithms**

- Asymptotically optimal algorithms, for \( a : \mu_a < \mu^* \),
  \[ \lim_{n \to \infty} \frac{\mathbb{E}[N_a(n)]}{\log(n)} \leq \frac{1}{KL(\nu_{\theta_a}, \nu_{\theta^*})} \]
- An index policy inspired by that of Gittins (1979) adapted to non-discounted rewards minimizes Bayesian regret
  \[ \Rightarrow \text{Our goal: Design algorithms inspired by the Bayesian modeling that are asymptotically optimal in the frequentist setting.} \]

Bayesian versus Frequentist Algorithms

Some quantities that naturally arise in the Bayesian modeling are

- \( \Pi_l = \{\pi_1^l, \ldots, \pi_K^l \} \) the current posterior over \( (\theta_1, \ldots, \theta_K) \)
- \( \lambda_t = \{\lambda_1^t, \ldots, \lambda_K^t \} \) the current posterior over the means \( (\mu_1, \ldots, \mu_K) \)

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 \( \Pi_l \) to determine action \( A_t \).

Bayes-UCB and Thompson Sampling

**Bayes-UCB algorithm** chooses \( A_t = \text{argmax}_a q_a(t) \), with \( q_a(t) = Q \left( 1 - \frac{1}{t \log(1/t)}, \lambda_a^t \right) \)

**Thompson Sampling** is a randomized algorithm:

\[
\{ \forall a \in [1..K], \theta_a(t) \sim \lambda_a^t \}
\]

Parameters: \( c \) (in practice, take \( c = 0 \)), initial prior \( \Pi_0 \)

Bayes-UCB: Theoretical Elements

- \( \nu_{\theta_a} \) is the Bernoulli distribution \( B(\theta_a), \pi_a^0 \) the (conjugate) prior \( U((0, 1]) \)
- Bayes-UCB is asymptotically optimal for Bernoulli bandits

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

\[ \mathbb{E}[N_a(n)] \leq \frac{KL(B(\theta_a), B(\theta^*))}{1 + c} \log(n) + \alpha_{\varepsilon}(\log(n)) \]

- 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): \( \tilde{u}_a(t) \leq q_a(t) \leq u_a(t) \) with:

\[
u_a(t) = \text{argmax}_{a > \frac{q_a(t)}{N_a(t)}} \left\{ d \left( \frac{S_a(t)}{N_a(t)}, \frac{1}{N_a(t)} \log \left( \frac{N_a(t)}{N_a(t) + 1} \right) \right) \right\}
\]

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

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

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

\[ \sum_{t=1}^\infty P \left( N_i(t) < \varepsilon t \right) \leq C_b \]

Numerical Experiments and Beyond

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

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 \( \{f_\theta, \theta \in \Theta \} \). For any \( \theta \in \Theta \), \( f_\theta \) is a function from which we obtain:

\[
\hat{X}_t^\theta = f_\theta(X_{t-1}, \ldots, X_{t-d}) \tag{2}
\]

A possible forecasting of \( X_t \) according to \( \theta \).

Let \( \epsilon \) be a loss function; we define the prediction risk as

\[
R(\theta) = \mathbb{E} \left[ (\hat{X}_t^\theta - X_t)^2 \right]
\]

and the empirical version of the risk as

\[
r_n(\theta) = \frac{1}{n-d} \sum_{t=d+1}^{n} (\hat{X}_t^\theta - X_t)^2
\]

3. A THEORETICAL RESULT

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

\[
R(\theta) \leq \inf_{\theta \in \Theta} R(\theta) + \mathbb{E} \left[ \frac{\log^2(n)}{\sqrt{n}} + \frac{2}{\sqrt{n}} \log \left( \frac{1}{\epsilon} \right) \right] \tag{4}
\]

7. APPLICATION TO THE AR(d) PROCESS

Since \( s_d(1) \subseteq B_d(2^d - 1) \), the prior \( \pi \) can be defined on \( \Theta = s_d(1) \) or \( B_d(2^d - 1) \). These two possibilities are combined with two different processes in the Metropolis-Hastings algorithm.

- Uniform prior on \( B_d(2^d - 1) \)
  - Uniform proposal : \( \beta_{\lambda,n} = \exp \left( -\lambda \mathbb{E} \left[ (\hat{X}_t^\theta - X_t)^2 \right] \right) \)
  - Constrained random walk with Gaussian increment:
    \[
    \beta_{\lambda,n} = \left( \frac{n}{2\pi} \right)^{\frac{1}{2}} \exp \left( -2 \left( 2^d - 1 \right) \right)
    \]

- Pushforward measure on \( s_d(1) \)

8. NUMERICAL RESULTS

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

\[
m \geq C_1 \left( \frac{\log^2(n)}{\sqrt{n}} \mathbb{E} \left[ \frac{\log(n)}{\sqrt{n}} \right] \right)
\]

where \( C_1 = \frac{1}{\alpha} \) and \( C_2 = \frac{1}{\alpha^2} \) are positive constants. Hence, to guaranty the Formula (7) we need a very high number of iterations, which can lead to a prohibitive computational cost.

9. References


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.

Algorithm based on the Sum-Weight framework [Kempe2003, Bénétiz2011] where sensors have two local variables jointly updated:
- a sum variable \( s(t) \) \( \{ s(t+1) = K(t)s(t) \text{ and } s(0) = x(0) \} \)
- a weight variable \( w(t) \) \( \{ w(t+1) = K(t)w(t) \text{ and } w(0) = 1 \} \)

**BWGossip**

 Assuming that at time \( t \), the sensor \( i \) wakes up

- Sensor \( i \) broadcasts \( (\frac{1}{|N|}I^T)w(t) \)
- At sensors in the neighborhood \( \mathcal{N}_i \), we have:
  \[
  s(t+1) = s(t) + \frac{s(t)}{|N|}w(t), \quad \forall j \in \mathcal{N}_i
  \]
- At sensor \( i \), we have :
  \[
  s(t+1) = \frac{s(t)}{|N|}w(t)
  \]
- All other sensors stay idle.

**RESULTS**

**Theorem 1 Convergence**

BWGossip converges to the average consensus almost surely.

**Theorem 2 Convergence speed**

(\text{main result})

For the BWGossip algorithm, we have:

\[
\forall \epsilon > 0, \quad \|x(t) - x_{ave}\|^2 = O_p \left( (I + \epsilon)^3 \right)
\]

\text{with} \quad \Gamma = \rho \left( (I - J) \otimes (I - J) \cdot E[K \otimes K] \right) < 1

\[
\rho(E[K] - 1/|N|I)^2 < 1
\]

- **BWGossip** outperforms the existing algorithms.
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](http://yago-knowledge.org)

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.

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)`

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

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

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

AMIE: Association Rule Mining under Incomplete Evidence
Luis Galárraga, C.Teflioudi, KHae, FMSuchanek (WWW 2013)
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)

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)

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)

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

<table>
<thead>
<tr>
<th>feature</th>
<th>Quality</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributors</td>
<td>0.683</td>
<td>0.566</td>
</tr>
<tr>
<td>Contributors + normal links</td>
<td>0.740</td>
<td>0.779</td>
</tr>
<tr>
<td>Social links</td>
<td>0.807</td>
<td>0.750</td>
</tr>
</tbody>
</table>

AA AAAA AAAA AA AA AA
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.

Access Control Model

- Social network model:
  - Directed, edge-labeled, and weighted graph
  \[(V, E, \Sigma, \varphi)\]

- 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.
  \[AR = (o, p, t_{\min})\]

  \[p = \{s_i\}\quad s_i = (r, \text{dir}, l, C)\]

  Example:
  \[p = \text{Friend} \times \{[1, 2][\text{city} = \text{Paris}/\text{BabySitter} \times [1][\text{city} = \text{Paris}]\] \quad t_{\min} = 0.8\]

System Overview

- Dataset: LiveJournal social network, ~5 Million users, ~80 Million edges
- Access Rule Specification:
- Authorized Audience Visualization:
A Probabilistic XML Merging Tool

Talel Abdessalem Mouhamadou Lamine BA Pierre Senellart
Télecom ParisTech Université Cheikh Anta DIOP Télecom ParisTech
Paris, France Dakar, Senegal 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 Wikipedia page with:
→ an efficient evaluation of the uncertainty of the obtained result
→ an automatic management of conflicts.

Merging of Wikipedia revisions

- A two-way tree merging technique for P-Documents
- Two steps: Matching of Revisions and Merging Matches

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:
  \( f_{IE_{old}}(x) = f_{IE_{old}}(x) \land \neg f_k \)
- Matched couple:
  \( f_{IE_{old}}(x) = f_{IE_{old}}(x) \)
- For added nodes:
  \( f_{IE_{old}}(x) = f_k \) or
  \( f_{IE_{old}}(x) = f_{IE_{old}}(x) \lor f_k \)

The result of the merge process

Probabilistic Documents

\( e_1 \lor e_2 \)  \( \neg e_2 \)
\( p_1 \quad p_2 \)  \( t_1 \quad t_2 \)
\( t_1 \quad t_2 \)
\( t_1 \quad t_2 \)
\( t_1 \quad t_2 \)

Pr(\( e_1 \)) = 0.7
Pr(\( e_2 \)) = 0.6

Corresponding possible documents and their probabilities

Architecture 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
**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 \((X, Y) = (X_j, Y_j)_{j=1}^{n}\), find the best prediction rule \(\text{sign}(H^*)\) such that the classifier function \(H(X, \theta)\):

\[ \theta^* = \min_{\theta} R_p(H(X, \theta)) \rightarrow \text{minimizes the Risk function } R_p \]

Particular case:

1. Quadratic cost \(\varphi\): \(R_p(H(X, \theta)) = \frac{1}{2} E[ (1 - YH(X, \theta))^2 ] \)
2. Mixture of experts: \(H(X, \theta) = \sum h_i(x)\)

### PROPOSED DISTRIBUTED LEARNING

On-line Learning Gossip Algorithm (OLGA)

- A distributed stochastic gradient descent approach where the estimated parameter sequence \((\theta_n)_{n \geq 1}\) is performed in 2 steps:
  - **[Gossip step]** At iteration \(n\), each agent \(i\) transmits \(X_{n,i}\) to all randomly selected neighbours \(j\) with probability \(p\) and obtains \(h_j(X_{n,i}, \theta_{n-1,j})\)
  - **[Local descent step]** Each agent \(i\) updates its estimated parameter \(\theta_{n,i}\) as follows:
    \[
    \theta_{n,i} = \theta_{n-1,i} + \gamma_{n,i} \cdot Y_{n,i} \cdot h(X_{n,i}, \theta_{n-1,i})
    \]

where

- \((\gamma_{n,i})_n\) is a decreasing step size sequence
- The r.v. \(Y_{n,i}\) is an unbiased estimate of the global decision \(H(X_{n,i}, \theta_{n-1,i})\) given by

\[
Y_{n,i} = h(X_{n,i}, \theta_{n-1,i}) + \frac{1}{p} \sum h(X_{n,i}, \theta_{n-1,i})\]

- \((\delta_{n,i})_n\) are independent Bernoulli r.v. ‘s \(B(p)\)

### Asymptotic behaviour

- Consistency: \((\theta_n)_{n \geq 1}\) convergence a.s. to the set of stationary points of \(R_p\)
- Conditional Central Limit Theorem: quantify the error variance excess

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

- \(\Gamma\): error in a centralized case + error excess induced by sparsification
- The average network throughput is reduced by a factor \((1-p)\)

---

**Theoretical results** (under suitable assumptions)

- Consistency: \((\theta_n)_{n \geq 1}\) convergence a.s. to the set of stationary points of \(R_p\)
- Conditional Central Limit Theorem: quantify the error variance excess

- 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**

- OLGA (p=0.6)

- GentLeBoost

**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 \(V_p\)

- Reduce redundancy classifiers and keep the relevant
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

- Queue Management
- Page Fetching
- Link Extraction
- URL Selection

⇒ 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

Architecture

Goal: Smart archiving of the Social Web:
1. Performing intelligent Crawling
2. Archiving Web objects

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

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?

The model
We learn association rules of the form \( a, b \rightarrow c, d \)
- E.g., “heartburn” → “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

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

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”

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

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
- **Ask question** → **answer question** → **user question results**
- **Data Display**
- **Question Display**
- **User Interface**
- **Question Selector**
- **Best Rules Extractor**
- **Rule Database**
- **Initial Data**
**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 \( X_i^k \) for content \( k \) at time \( t \).
- Preferences \( P_{ik}^t \) are normalized scores, i.e., \( P_{ik}^t \propto X_i^k \).
- Agents update their scores linearly as
  \[
  X_i^{k+1} = X_i^k + \sum_{j \neq i} P_{ij}^t \cdot 1_{u_j^t = k}
  \]
  with \( P_{ij}^t = k \mid P_i^t = \{ f(P_i^t) \}_k \) the probability of agent \( j \) broadcasting content \( k \).
- Function \( f : |\mathcal{A}_K|^N \rightarrow |\mathcal{A}_K|^N \) models the way agents choose the contents to broadcast, where \(|\mathcal{A}_K|^N\) is the set of the \( NK \) stochastic matrices.

**Types of function \( f \)**

- \( f(x) = x \) - identity function.
- \( |f(x)|_{ik} = \frac{e^{Rx}}{\sum e^{Rx}} \) - soft-max function.
- \( \beta \ll 1 \Rightarrow |f(x)|_{ik} \sim \frac{1}{K} \)
- \( \beta \gg 1 \Rightarrow |f(x)|_{ik} \sim 1_{k=\arg\max_j s_{ij}} \)

**Results**

- For \( f(x) = x \): there exists a random variable \( P_x \in |\mathcal{A}_K|^N \) such that \( P_x \rightarrow P_{\infty} \) almost surely.
  If \( G \) is strongly connected, then there exists a random variable \( \pi \in \mathcal{A}_K \) such that \( \pi \rightarrow \pi^T \) almost surely.
- For soft-max exponential with \( \beta < 1 \): if \( \inf \sum A_{ik} > 0 \), then there exists a \( \beta_{\text{min}} > 0 \) such that for all \( \beta \in [0,\beta_{\text{min}}] \) we have that \( P \rightarrow \frac{1}{K} \) almost surely.
- If \( \beta \gg 1 \), then we expect the graph to be clustered 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.
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.

Statistic Analysis

Sentiment Analysis

User preference on visited venues and the associated items

Matrix Factorization

Tensor Factorization

User preference prediction on unvisited venues and the associated items

Context Awareness

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

Spatial characteristics:
Users usually conduct limited number of activity categories in their frequented regions

Temporal characteristics:
Users who share similar temporal activity preference in the past will probably have similar preference in the future

An Example of Personalized Location Based Services

He is probably interested in going to a bar.

Context-awareness

Time: 20:15 Friday
GPS: 48.8525,2.3344

He will like the x bar and the large-screen TV and sportive environment there.

Preference-awareness
A Context-Risk-Aware Recommender System

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 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 \( R(s) \)
  - Three methods of risk computing
    - Using situations similarity \( R_s \)
    - Using situations concepts \( R_c \)
    - Using a Gaussian distribution \( R_v \)

General Approach

Models and Algorithms
- Nomalys Data Set
  - 356,738 situations
  - 5,518,566 navigations data
  - 3500 users

Genetic Algorithm
- Population: 80 Chromosomes
- Chromosomes: \( \epsilon_{\min}, \epsilon_{\max}, \text{Threshold} \)
- Results: \( \epsilon_{\min} \in [0.05; 0.13], \epsilon_{\max} \in [0.47; 0.56], \text{Threshold} \in [0.7; 0.82] \)

Evaluation Results

Data Set and Parameters
- Nomalys Data Set
- Number of visited documents
- Time spent in documents
- Data Size Variation
- Evaluation during Iterations
- Risk Variation

Office
Meeting
Home

Evaluation during Iterations

Data Size Variation

Risk Variation
4. SECURITE, SURETE ET RISQUES
Context

Outsourced data / Cloud:
• Various contexts: outsourced storage and computation, 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 traditionally 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 mechanisms 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)


A Joint Watermarking/Encryption Algorithm for Verifying Medical Image Reliability in Both Encrypted and Spatial Domains

**MEDICAL DATA PROTECTION**

- Rules of ethics
- Legislatives rules
- Security
- Confidentiality
- Availability
- Reliability
- Integrity
- Authentication

**Reliability**: a degree of confidence/trust in medical data

**SYSTEM ARCHITECTURE**

**Protection**
- $m_x$ and $m_y$: messages available in the spatial and the encrypted domains, respectively.

**Verification**
- $m_x$, $K^e_w$, $K^a_y$: watermarking keys, $K_c$: encryption key

**EXPERIMENTAL RESULTS**
- Capacity rate: 1/16 bpp in each domain.
- PSNR: greater than 60dB.
- Image distortion measure: PSNR (dB).
- Capacity rate (bpp: Bit Per Pixel).

**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).

**JOINT WATERMARKING/ENCRYPTION**

- Modification of the quantization index modulation (QIM):
  - Disrupt/modulate the image pixels to encode simultaneously $m_x$ and $m_y$.
  - QIM: insertion based on codebooks $C_{m_{SI}}$, which represents the message $m_x$.
  - QIM/chiffrement: Constitution of sub-codebooks $C_{m_{SI}m_{ej}}$ according to the AES.

**EXPERIMENTAL RESULTS**

- 100 ultrasound images- 576 x 688 pixels, 8-bit depth.
- Capacity rate: 1/16 bpp in each domain.
- PSNR: greater than 60dB.

**CONTACT**

Dalel BOUSLIMI
Gouenou COATRIEUX
Michel COZIC
Christian ROUX

Gouenou.Coatrieux@telecom-bretagne.eu
www.mines-telecom.fr
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).

1. MEDICAL DATA PROTECTION

Security Objectives
- Rules of ethics
- Legislative rules

Security
- Confidentiality
- Availability
- Reliability: a degree of confidence/trust in medical data
- Integrity
- Authentication

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

2. A COMMON DATABASE WATERMARKING CHAIN

Tuple Grouping ➔ Make embedding independent from database storage
- Reorganization of tuples \( t \) in \( N \) groups depending on their primary key \( t.PK \) and a secret watermarking key \( K_w \)
- Each tuple is assigned to the group

\[
\nu = H(K_w | H(K_w | t.PK)) \mod N
\]
- \( H \): Cryptographic HASH operation
- \( \nu \): concatenation operation
- One bit is ideally embedded in each group.

3. PROPOSED METHOD

Modulation Principles
- One group is uniformly split into two subgroups \( (G^{A}, G^{B}) \).
- Histograms for a numerical attribute in each subgroup are mapped into a circle
- Modification of the relative angular position between centers of mass of \( G^{A} \) and \( G^{B} \)

Fragile Embedding ➔ Insertion of a digital signature (Integrity)

\[
\beta_i^w = \begin{cases} 
\beta_i + 2a & \text{if } \beta_i < 0 \\
\beta_i - 2a & \text{if } \beta_i > 0 
\end{cases}
\]
- Digital Signature is extracted from the observed database and compared to the one computed from the data.
- Same modulation principle but with the insertion of a secret pattern.
- Detection based on correlation (origin) and/or extraction of the pattern (integrity).

Robust Embedding ➔ Insertion of an Authentication Pattern (Reliability)
- Special case (Non carriers)

\[
\beta_i^w = \begin{cases} 
\beta_i + 2a & \text{if } \beta_i > 0 \\
\beta_i - 2a & \text{if } \beta_i < 0 
\end{cases}
\]

Groups where \(|\beta_i| > 2a\) (as in \( \beta_i \)) cannot carry information (non-carriers) and

4. EXPERIMENTAL RESULTS

- Three values of \( \Delta = \{1, 2, 3\} \) considered
- Att. Age with \( \Delta = 3 \); Att. Stay Duration with \( \Delta = 1 \)

In both cases, results depend on the statistical properties of the attributes, more specifically in their standard deviations.
Problem

How to assist model development of critical reactive systems?

- “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

IDF – Incremental Development Framework

Combining model refinements and extensions

Two sets of techniques to support:

- Construction techniques
- Evaluation techniques

Two axis:

- abstraction level (vertically)
- completion level (horizontally)

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.

http://lgi2p.mines-ales.fr
Methodology of system interdependency analysis after a major disruption

**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?**

**Conceptual tools to analyze system interdependencies**

**What are risk management methodology characteristics about independencies between elements of a system?**

- Empirical approach
- Predictive approach
- Multi-agent system, dynamic system, network theory, petri network, input-output model,…

**4 Types of Interdependencies**

- Activity sectors: Energy, health, food, water, transport, defense, government, telecommunication,…
- (SGDN, 2008)

**4 Scale levels**

- International, national, regional, local

**Protection and resiliency measures**

- Protection
- Resiliency
- Redundancy, stocks, new resources,…

**Approach**

1. Create system typologies of processes, functions, physic elements, services and constraints
2. Create disruption typology: Identify fault physic elements and failure functions with their characteristics (maximum capacity, dynamic propagation of a disruption,…)
3. 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**

**Building and simulation of failure scenarios**

**Papers**

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.

Crisis management’s limits

- 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, breakdown of decisions, coordination failure, leadership deletion, lack of communication, blind support of the procedures

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

- Strength of thinking quickly
- Educational leader
- CRM Training’s tools
- Decision making model
- Shared Mental Models

3. Assess a group in real time

- Observation grid
- Questionnaire
- Check list
- Case test
- Feedback form
- Self-assessment
- Identification of encountered difficulties

4. Organize the debriefing

- 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
Le projet GEPET-Eau
Gestion Efficiente Prédictive ET adaptative de la ressource en Eau des voies navigables dans un contexte de 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 événements extrêmes
- Adaptation au changement climatique
- Financement CGDD, DGEC, DGITM

Objectifs

Connaissances
- Caractériser des scénarios caractéristiques du changement climatique
- Disposer d’un modèle générique de la dynamique des voies navigables
- Estimer la résilience des voies navigables - bief Cuinchy-Fontinettes
- Pouvoir disposer, à terme, d’un outil d’aide à la décision

Contributions scientifiques
- 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 navigables
- Conception de stratégies 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

Partenaires

Auteurs
K. Horvath
E. Duviella
J. Blesa
L. Rajaorisoa
S. Lecoeuche
D. Juge-Hubert
K. Chuquet
E. Sauquet
F. Guibert
N. Gaffet

Parties prenantes
Gestion de la confiance au sein de communautés virtuelles

Les communautés virtuelles :
- Groupes d'entités
- Interagissant via internet
- Partageant des pratiques, intérêts, valeurs, principes communs

Leurs objectifs : production, consommation, partage, collaboration autour 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

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.

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.

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.

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

- 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 : Événement du contexte
  - C : Condition de garde
  - A : Une liste d’actions

- 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

- **Cost-effective security mechanism development**
  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
- Zonghua Zhang, Hong Shen: Constructing Multi-Layered Boundary to Defend Against Intrusive Anomalies: An Autonomic Detection Coordinator. in Proceedings of IEEE DSN.
- Zonghua Zhang, and Ahmed Meddahi: 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.

- **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

- **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 use case scenarios for monitoring and characterizing attack behavior.

Reference
- Zonghua Zhang, Hong Shen: Constructing Multi-Layered Boundary to Defend Against Intrusive Anomalies: An Autonomic Detection Coordinator. in Proceedings of IEEE DSN.

Selected Research Topics

- **Example network**

- **Generation of dependency attack graph**

- **Attack state estimate using Hidden Markov Model**

- **Cost-effective defense**

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.
CONTRAT

- 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 $\tilde{X}^N$ such that $\tilde{Z}^N$ 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

$$L = \log \frac{P(o^N/x^N, H_1)}{P(o^N/x^N, H_0)} \geq \lambda$$

- Computing false alarm and non detection probabilities
  - Gaussian approximation
  - Chernoff bounds (threshold far away the mean)

PERSPECTIVES

- Authentication for structured codes
- Studying the model of the broadcast channel

REFERENCES


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-Étienne, 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

**Valorisations**

**Objet de la recherche**

Energie -> des questionnements sur le risque sanitaire

Lecteur

< - Information

des questionnements sur le risque privacy

**Partenaires**

Télécom ParisTech, Dép. SES, DEIXIS-Sophia (leader)  
Mines Saint-Étienne, Centre Microélectronique de Provence  
Université de Coimbra, OSIRIS, Observatoire sur les Risques

**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-paristech.fr
A Simple Model for Evaluating Secure Key Generation from Random Radio Channels

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 security 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 mth antenna at Bob/Eve side can be defined as:
  \[ h_m = \sum_{i=1}^{N_s} \frac{\beta_i}{d_i} \exp\{j(K_d m + K_A f_i)\} \]

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.

![Graphs showing SKG evaluation](image)

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.
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.

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, September 2013
- Journée Mobilité 2.0, Satory, February 2014
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.

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.

\[ \text{BER} = 10^{-9} - 10 \text{ms} - 101 \text{ slices in Xilinx FPGA.} \]
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.

Drones for assisting disasters

- **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.

Embedded electronic and software architecture

- Cameras,
- Lidars,
- Low cost and efficient processing units (e.g., parallella ...)

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.
Formally Proved Security of Assembly Code Against Leakage

Context: countermeasures

<table>
<thead>
<tr>
<th>Hardware</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masking</td>
<td>True</td>
</tr>
<tr>
<td>Hiding</td>
<td>True</td>
</tr>
<tr>
<td>dual-rail</td>
<td>True</td>
</tr>
</tbody>
</table>

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 forentropy
- Provable correction of leakage-free (with a finite number of physical hypotheses, to be 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]

Pure software dual-rail design flow

- assembly source
- protected assembly
- mapping to abstractASM
- mapping to assembly
- bit/slice compiler
- DPL macro expansion
- DPLization of data

Macro for Boolean operation $op$

<table>
<thead>
<tr>
<th>$r_1$</th>
<th>$r_2$</th>
<th>$r_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_1$</td>
<td>$r_2$</td>
<td>$r_3$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>$r_2$</td>
<td>$r_3$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>$r_2$</td>
<td>$r_3$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>$r_2$</td>
<td>$r_3$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>$r_2$</td>
<td>$r_3$</td>
</tr>
</tbody>
</table>

Cost on PRESENT [BKL07] case-study

<table>
<thead>
<tr>
<th>Operation</th>
<th>Code size</th>
<th>RAM words</th>
</tr>
</thead>
<tbody>
<tr>
<td>state-of-the-art</td>
<td>11382</td>
<td>192</td>
</tr>
<tr>
<td>Compiled</td>
<td>6473</td>
<td>144</td>
</tr>
<tr>
<td>DPL protected</td>
<td>182572</td>
<td>2674</td>
</tr>
</tbody>
</table>

Optimizations (still with formal proof of correction)

- The existence of non-negative 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 instructions not included in our abstractASM.

References


Maximizing the success of a side-channel attack

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]

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: \( \arg \max_k \{ P(k) \cdot p(x|y(k)) \} \) (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 \( X = Y(k^*) + N \) with known leakage model \( Y(k) = \psi(f(k, T)) \) then the optimal distinguishing rule are

- Gaussian noise distribution: \( D_{opt}^{G}(X, t) = \arg \max_k \{ x|y(k) \} - \frac{1}{2} \|y(k)\|^2 \)
- Uniform noise distribution: \( D_{opt}^{U}(X, t) = \arg \max_k \{-|x - y(k)|_\infty \}
- Laplace noise distribution: \( D_{opt}^{L}(X, t) = \arg \max_k \{-|x - y(k)|_1 \}

Theorem 2: optimal distinguisher when the leakage model is partially unknown

Let \( X_{\alpha}(k) = \sum_{j=1}^{n} \alpha_j f(T, k) + N \), \( Y_j(k) = f(T, k) \mod 2 \) and \( X = \sum_{j=1}^{n} \alpha_j f(T, k^*) + N \) 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

\[
D_{opt}^{G}(X, t) = \arg \max_k \{ \gamma(x|y(k)) + 1 \} \cdot (\gamma Z(k) + I)^{-1} \cdot (\gamma |x|y(k)) + 1 \} - \sigma^2 \ln \det(\gamma Z(k) + I),
\]

where \( \gamma = \frac{\sigma^2}{\sigma^2} \) is the epistemic-to-stochastic-noise-ratio (ESNR).

Our novel optimal distinguishers outperform all state-of-the-art distinguishers depending on statistical tools in terms of the success rate!
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)

Requirements
- Who and why: stakeholders and security goals

Attacks
- Who and Why: attackers, their capabilities, and objectives (risk analysis)

Application
- When: operation sequences in functions involving those assets

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

Institutions
Authors
Ludovic Apvrille (Télécom ParisTech)
Ludovic.Apvrille@telecom-paristech.fr
Yves Roudier (EURECOM)
Tullio Joseph Tanzi (Télécom ParisTech)
Franck Guarnieri (Mines ParisTech)

Partners

Contact   Ludovic.Apvrille@telecom-paristech.fr   Site web   http://www.telecom-paristech.fr.
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
A geometrical model for selecting optimal countermeasures against multiple attack scenarios

**Authors**
- Gustavo GONZALEZ GRANADILLO
- Hervé DEBAR
- Grégoire JACOB

**RST Department**

**Context/funding**

**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.

**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.

**EXAMPLE**

Multiple attacks detected at Telecom SudParis, France

**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.

**Graphical representation of attacks and countermeasures**

**Countermeasure Selection Process**

1. Calculate the Total Number of Countermeasures
2. Calculate the Annual Impact Cost in the absence of security countermeasures (ALE)
3. Calculate the Total Annual Cost of Security Countermeasures (TAC)
4. Select the countermeasure with the lowest TAC

**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.
**Preuve formelle de la sécurité d'une application ferroviaire de signalisation en modes nominal et dégradés**

**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)