Privacy-Enhancing Technologies for the Internet

Yves Deswarte
deswarte@laas.fr

LAAS-CNRS, Toulouse

Security vs. Privacy

- "Privacy" ≈ confidentiality of personal data (and meta-data)
  PII : Personally Identifiable Information
- = subset of "scurity" (*CIA*)
- But…

… "*the devil is in the details*"

- Keep evidence, for possible future dispute
- Traceability to identify malicious actors
- Strong authentication
- … danger for privacy !!!

Outlines

- *"Privacy"* : Definition, Regulations
- Basic Principles
- PETs : Privacy Enhancing Technologies
  - Managing Multiple Identities
  - Anonymous Communications and Accesses
  - Privacy-Preserving Authorization
  - Personal Data Management
- The Prime Project

Privacy: definitions

- "*The state or condition of being free from being observed or disturbed by other people*"
- Common Criteria (ISO 15408) :
  Privacy = one functional class, with 4 requirements *to provide a user protection against discovery and misuse of identity by other users*:
  - Anonymity: ensures that a user may use a resource or service without disclosing the user's identity
  - Pseudonymity: ensures that a user may use a resource or service without disclosing its user identity, but can still be accountable for that use
  - Unlinkability: ensures that a user may make multiple uses of resources or services without others being able to link these uses together
  - Unobservability: ensures that a user may use a resource or service without others, especially third parties, being able to observe that the resource or service is being used

Pseudonymity < anonymity < unlinkability < unobservability

Regulations (1)

- Universal Declaration of Human Rights: Art. XII: "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks", UN General Assembly, December 10, 1948
- International: OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (September 3, 1980), UN guidelines concerning Computerized personal data files (December 14, 1990)
- European: Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Council of Europe, ETS-108, 26/01/81), directives 95/46/EC (free movement), 2002/58/EC (electronic communications sector, replacing directive 97/66/EC) + directive 2006-24-EC on Data Retention
- French: Protection des données nominatives -> à caractère personnel : loi "Informatique et Libertés" du 06/01/78, révisée par loi du 6 août 2004 + loi 94-548 (recherche médicale) http://www.cnil.fr
  - Article 1er : « *L'informatique doit être au service de chaque citoyen* […] *Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques* »

Regulations (2)

- Secret professionnel (N^eau Code Pénal, art. 226-13) et secret des correspondances (NCP art. 226-15)
  + code des postes et télécommunications
- + art. L-34-1, inséré par la "Loi relative à la sécurité quotidienne" du 15/11/2001, révisé par la "Loi pour la sécurité intérieure" du 18/03/2003, la "Loi sur l'économie numérique" du 21/06/2004, puis la "Loi relative aux communications électroniques et aux services de communication audiovisuelle" n°2004-669 du 9 juillet 2004, décret du 24/03/06.

Basic Principles

1st Principle to protect privacy:

- **Personal Data Minimization**
  personal information should be transmitted only to those who need it to achieve the task they have been entrusted with -> "*need-to-know*"
  then destroy/forget

- … on the Internet like in the real world

- … with limits: some personal data must be provided to judiciary authorities in case of dispute or investigation (e.g., against money laundering) : pseudonymity rather than total anonymity

  Links: minimization <--> proportionality and legitimate purpose

Example: electronic commerce (1)

- Involved parties:
  a client, a merchant, a delivery service, banks, credit card issuer, Internet service provider, …

- The merchant does not (generally) need to know the client's identity, but must be ensured the purchase will be paid.

- The delivery company does not need to know the purchaser's identity, nor what is purchased (except for the physical characteristics), but needs to know the deliveree 's identity and address.

Example: electronic commerce(2)

- The purchaser's bank does not need to know the merchant, nor what is purchased, but only the bank account and amount to be credited …
- The merchant's bank does not need to know the purchaser …
- The ISP does not need to know anything about the transaction, except the technical characteristics of the connection …

2nd Principle to protect privacy:

- "Sovereignty": the person shall maintain control on his/her personal [meta-]data

  -> stored on a personal device:
  (smartcard, PDA, PC…)

  -> if these data are disclosed to a third party, impose **obligations** on their use

  - Expiration dates
  - Notification in case of transfer or unexpected use
  - etc…

Privacy-Enhancing Technologies

PETs : *Privacy-Enhancing Technologies*

- Managing Multiple Identities
- Anonymous Communications and Accesses
- Privacy-Preserving Authorization
- Managing Personal Data

1st PET: Managing Multiple Identities

- Identity = the representation of a physical person
- Reduce/control the links between the person and the personal data and meta-adata (control the *linkability*)
  - communications and accesses shall be unlinkable
- But: customized/privileged accesses: virtual identity = pseudonym
  - Preferences (ex: meteo) -> "*cookies*"
  - Different roles -> different pseudonyms
    - Ex: tax payer and elector
  - Authentication strength should be adapted to the risks of identity theft (and liablity)
  - Lifetime should be adapted to the needs of linkability -> throw-away pseudonyms
- Multiple virtual identities vs. "*single-sign-on*"
  Liberty Alliance <http://www.projectliberty.org>
  vs. Microsoft Passport

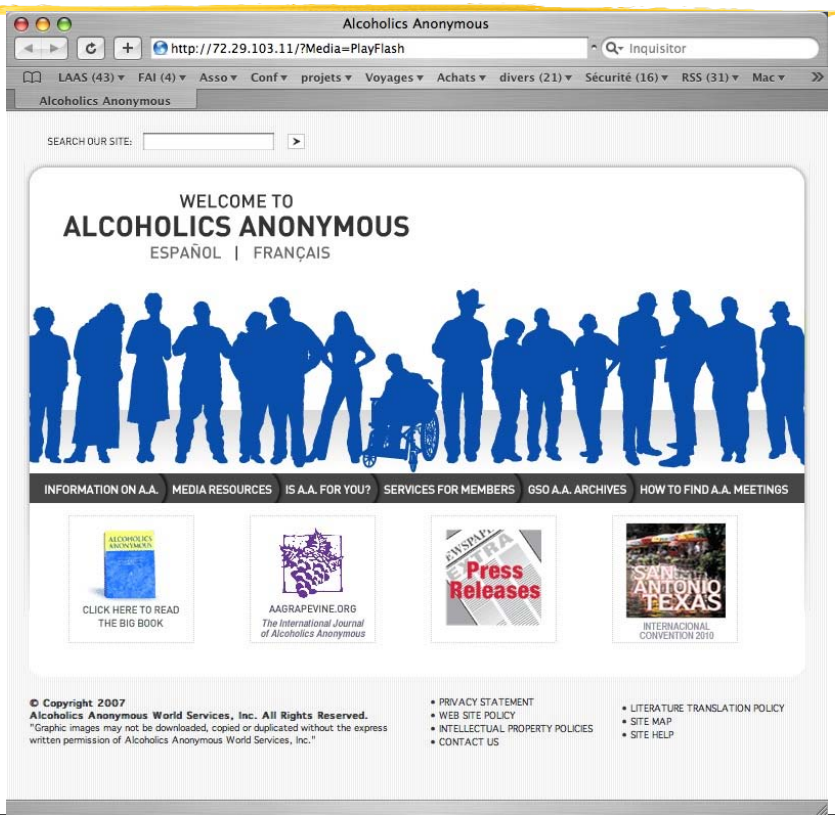
@IP= identifying data

Example :

```
Return-Path: <Yves.Deswarte@laas.fr>
Received: from laas.laas.fr (140.93.0.15) by mail.libertysurf.net
(6.5.026)
      id 3D518DEF00116A4D for yves.deswarte@libertysurf.fr; Tue, 13 Aug
2002 13:44:40 +0200
Received: from [140.93.21.6] (tsfyd [140.93.21.6])
    by laas.laas.fr (8.12.5/8.12.5) with ESMTP id g7DBid1D001531
    for <yves.deswarte@libertysurf.fr>; Tue, 13 Aug 2002 13:44:39 +0200
(CEST)
User-Agent: Microsoft-Entourage/10.1.0.2006
Date: Tue, 13 Aug 2002 13:44:38 +0200
Subject: test
From: Yves Deswarte <Yves.Deswarte@laas.fr>
To: <yves.deswarte@libertysurf.fr>
Message-ID: <B97EBDC6.2052%Yves.Deswarte@laas.fr>
Mime-version: 1.0
Content-type: text/plain; charset="US-ASCII"
Content-transfer-encoding: 7bit
```

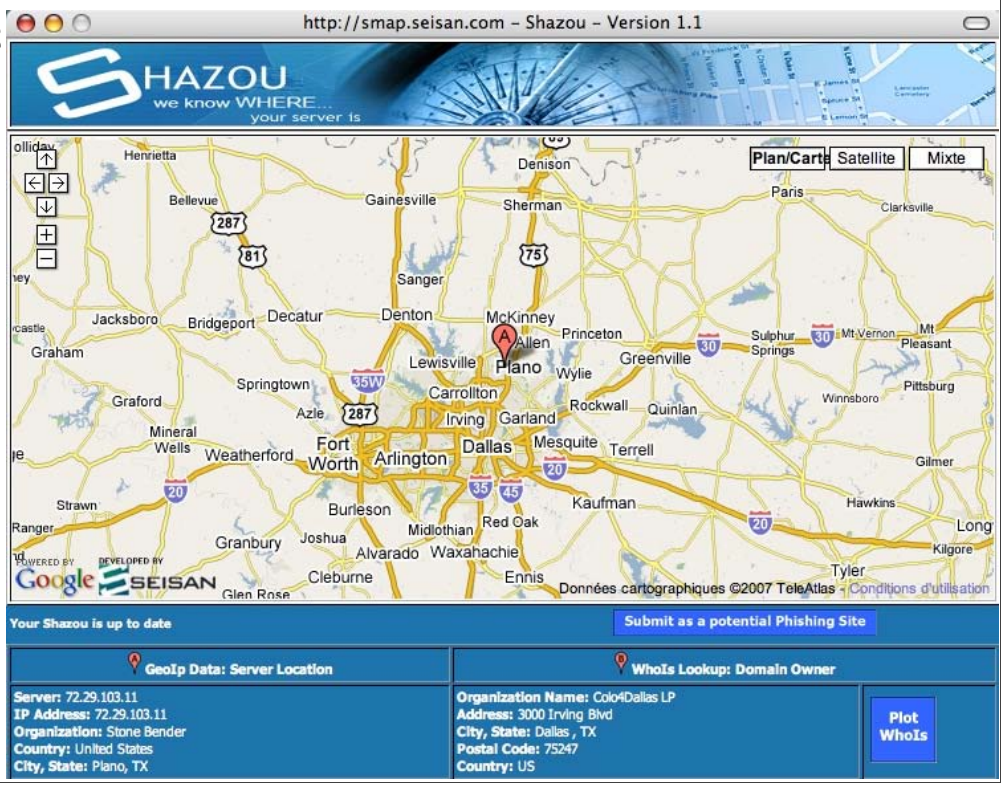
@IP= sensitive content

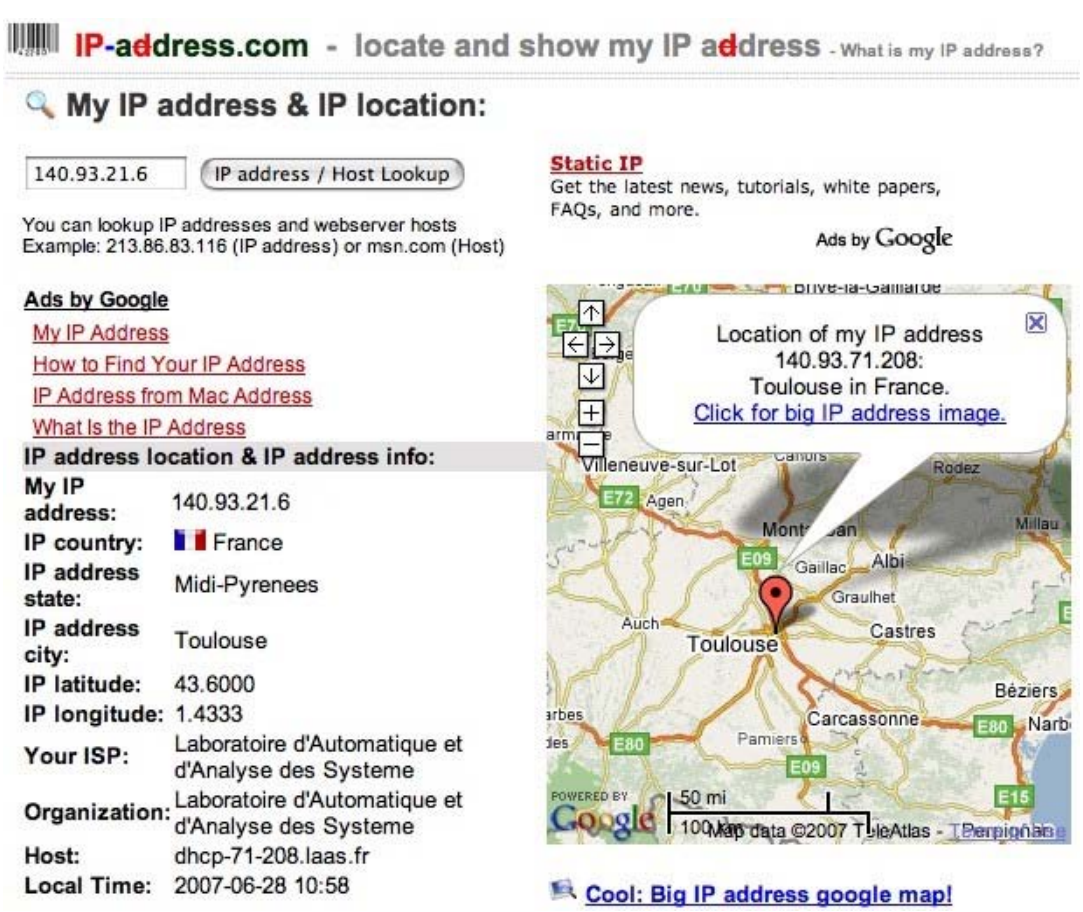
Example :
http://72.29.103.11/



@IP= location information

Example :



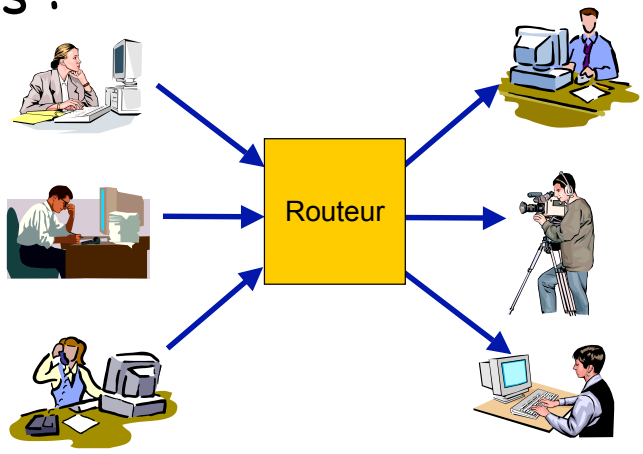


IP V6, ad hoc networks, …

- Tomorrow : IP everywhere (*pervasive/ubiquitous computing, ambiant intelligence, sensor networks, RFID, 4G convergence …*)
- every device will have an implicit IP@ *unique and permanent* (based on a manufacturing serial number)
- Every person will own several devices …
- … that will connect to other close devices (ad hoc)
- … that will identify each other, route their communications, provide contextual information, etc.

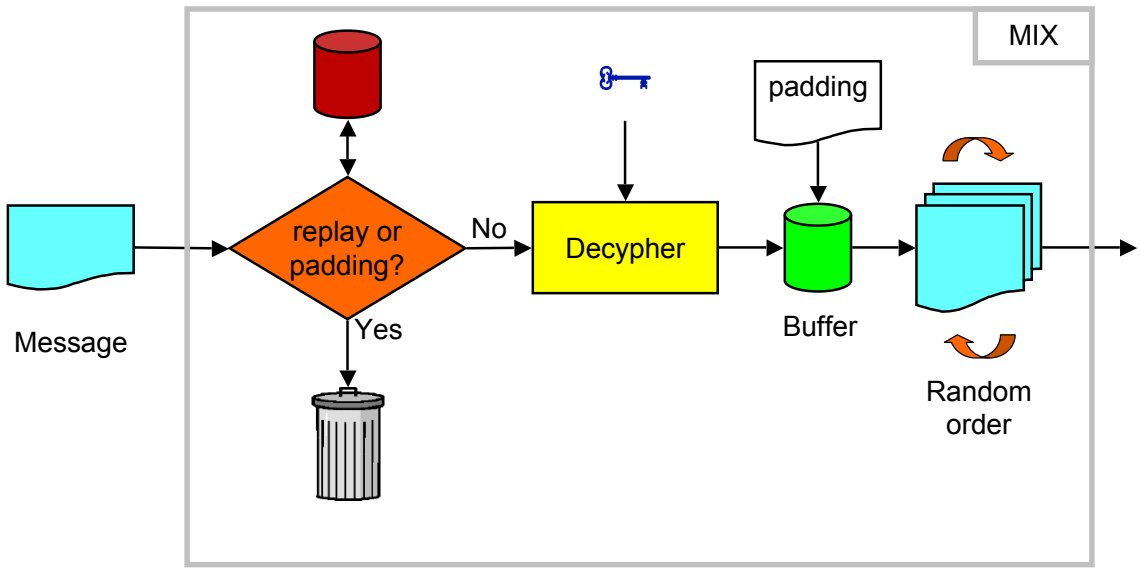
2nd PET: Anonymous Communications

- To protect IP@:
  dynamic assignment of IP addresses:
  (DHCP, PPP, NAT, …)
- Anonymity routers :
  - MIX
  - Onion Routing
  - Crowds

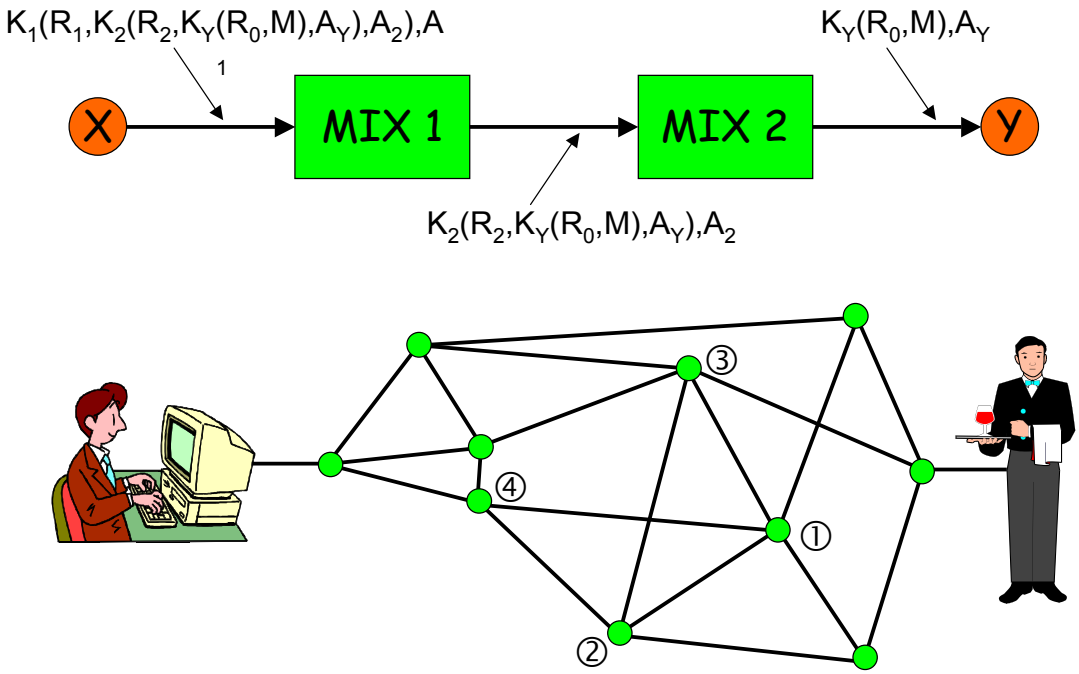


MIX: how does it work?

http://www.inf.tu-dresden.de/

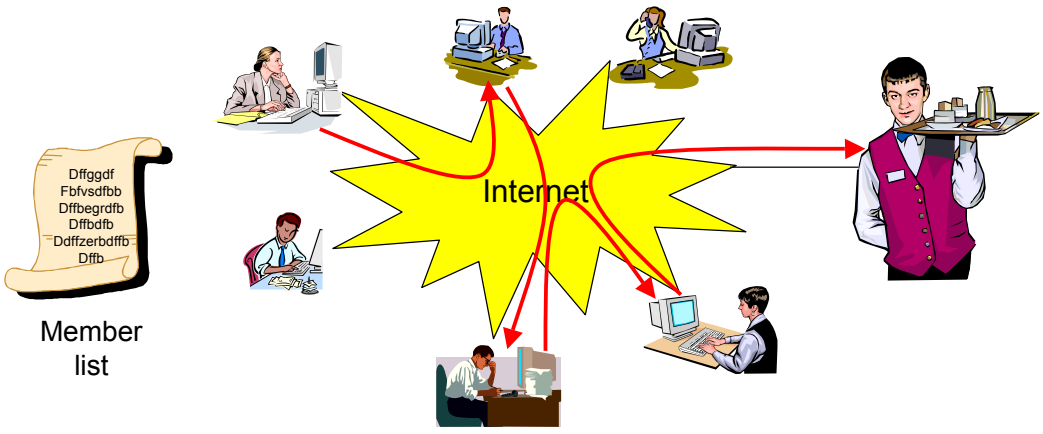


MIX / Onion Routing / Crowds



Crowds/Hords: peer-to-peer

- Each member is a MIX for the others
- Probability *p* to send the message to the destination (1-*p*) to send it to a randomly chosen member



MIX drawbacks

- Cost (# of messages, ciphering, latency, …)
- OK for mail, Web, … not for VoIP, …
- Vulnerable to collusion between MIXes
  --> independence between MIXes?
- Vulnerable to a global observer (statistical analysis)
  --> distribution over the Internet?
- Interactivity: return channel + relationship anonymity
- Inefficient on LANs …

Le dîner des cryptographes

- Comment savoir si quelqu'un a payé, sans pouvoir savoir qui ?

DC-network

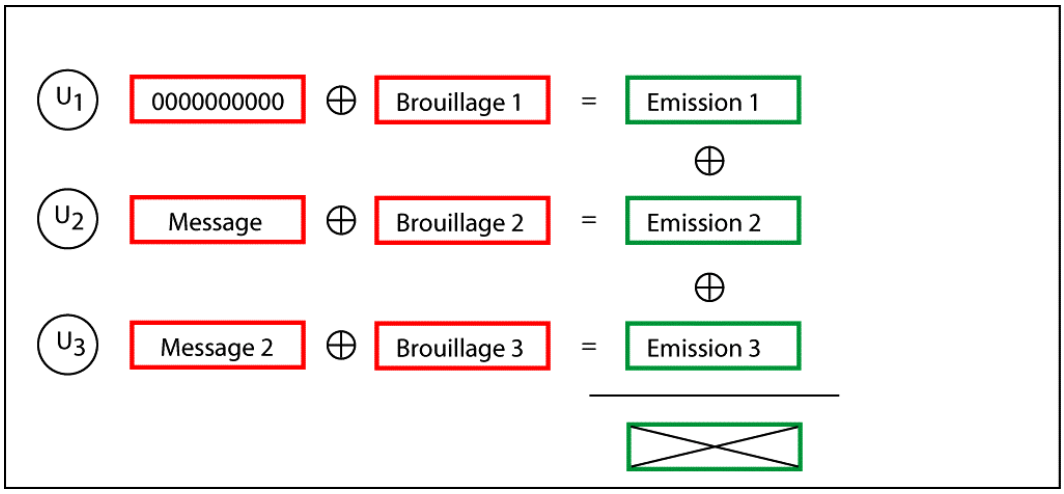
Protocole par tour : à chaque tour :

- Chacun diffuse un message ou du bourrage
- Chacun fait le XOR de tout ce qu'il a reçu
- Les bourrages sont générés de façon à s'annuler par XOR

  --> résultat = XOR(messages)

  - Si pas de message : résultat = 0
  - Si un seul message : tous les participants reçoivent le message (en clair)
  - Si plusieurs messages : collision --> résolution "aloha"

Envoi superposé



Bourrage s'auto-annulant

- ∀ i, j ∈ {cryptographes}, i et j partagent une chaîne secrète de bits aléatoires de longueur infinie : $S_{i,j} = S_{j,i}$

- A chaque tour k :
  - Si i ne veut pas émettre de message, il diffuse $B_i = XOR_{i \neq j}$ (k-ième tranche $(S_{i,j})$)
  - Si i veut émettre le message M, il diffuse M XOR B_i
  - $XOR_{i=1..n}(B_i) = 0 \Rightarrow$ résultat = M (si un seul message)

Débit Max DCnet

- Nombre de XOR par round proportionnel à **n**
  → Débit % **~1/n**
- Video conf seulement pour petits groupes
  (débit-latence)
- Videostreaming (?) ou transferts de gros fichiers
  limités à 8 users max (débit)
- Audio possible pour des centaines d'utilisateurs
  ex: VoIP (débit-latence)

Private Information Retrieval (PIR)

- Exemple : PIR "parfaitement" sûr
  - Base de données répliquée
  - Composée de N éléments de taille fixe
  - 2 Requêtes :
    - 1 chaîne S de N bits aléatoires -> serveur 1
    - même chaîne sauf le k-ième bit inversé -> serveur 2
  - Réponse de chaque serveur = XOR de tous les éléments i tels que $S_i = 1$
  - Réponse = XOR des deux réponses
- Avec des méthodes cryptographiques (chiffrements homomorphiques {a + b} = {a}+{b}, résidus quadratiques et non-quadratiques, …), on peut réaliser des PIR "computationnellement" sûrs sans réplication

Émission/réception non observables

- Thèse de Carlos Aguilar (LAAS, 2006)

Réception / Émission	Diffusion	PIR
Bourrage chiffré	EBBS	pMIX
Envoi superposé	Serveur DC-Net	pDC-Net

Connexion IP nomade anonymisée

Roaming : PC portable, PDA, téléphone …

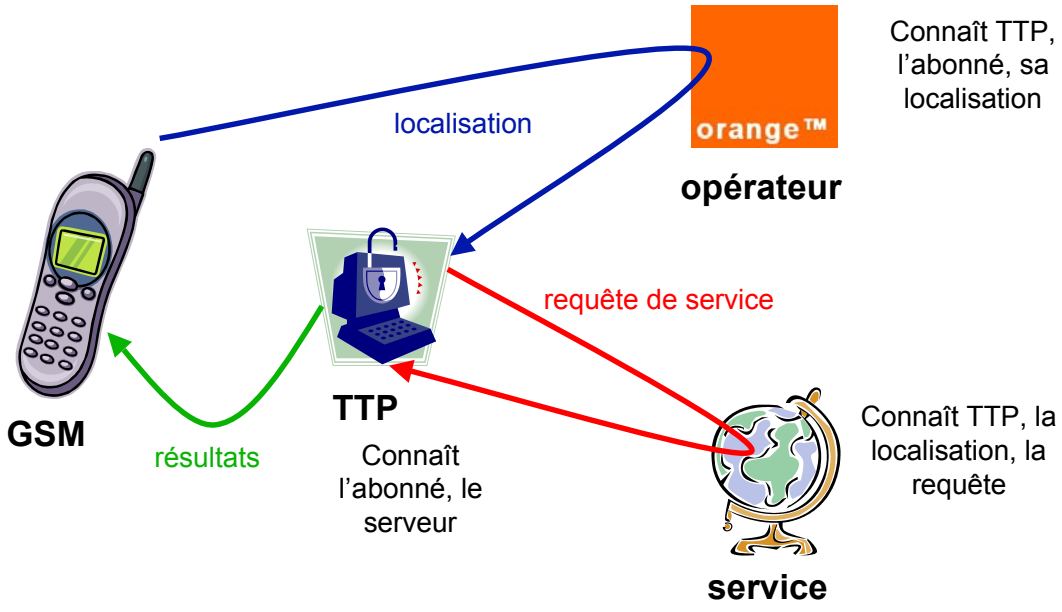
1. Génération d'1 @MAC aléatoire
2. Obtention d'1 @IP temporaire
3. Tunnel vers un TTP de roaming
4. Génération d'une autre @IP
5. Authentification sur FAI

2°bis PET: Accès anonyme à des services

- Relais d'anonymat (*anonymity proxy*) : unidirectionnels (ou bidirectionnels?)
  - e-mail, news (Usenet)
    - anon.penet.fi (700 000 utilisateurs en 1996 !)
    - Cypherpunks
  - ftp
  - Web : ex: proxify.com
  - …
- Serveur de pseudonymes :
  - e-mail
  - Identités multiples fournies par des f.a.i. (adresses mél)

Service basé sur la localisation

- Ex: PRIME : pharmacie la + proche



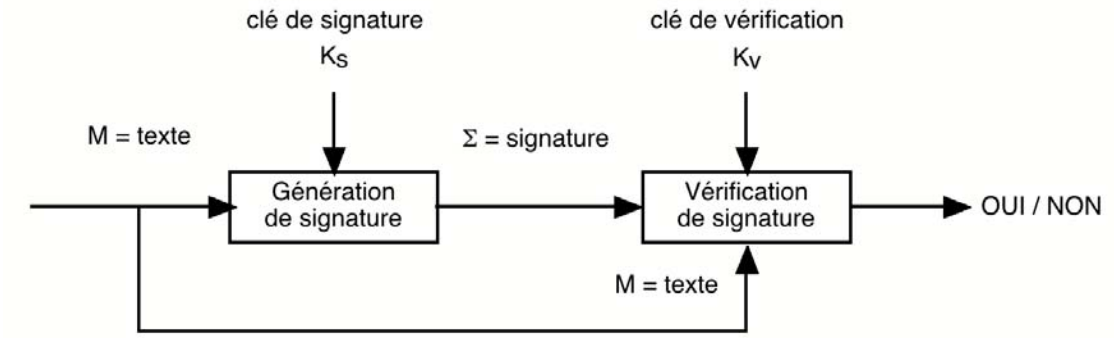
3° PET: Autorisation sur Internet

- Aujourd'hui : client-serveur
  le serveur accorde ou refuse des privilèges au client en fonction de son identité déclarée (éventuellement vérifiée par des mécanismes d'authentification)
- Le serveur doit enregistrer des données personnelles : preuves en cas de litige
- Ces données peuvent être utilisées à d'autres fins (profilage des clients, marketing direct, revente de fichiers clients, chantage…)
- Action P3P (W3C) : *Platform for Privacy Preferences Project*
  vérification automatique de politiques de sécurité/privacy "déclarées"

Ce schéma est dépassé

- Les transactions sur Internet mettent en jeu généralement plus de deux parties
  (ex : commerce électronique)
- Ces parties ont des intérêts différents (voire opposés) : suspicion mutuelle
- Nocif pour la vie privée :
  opposé au "besoin d'en connaître"

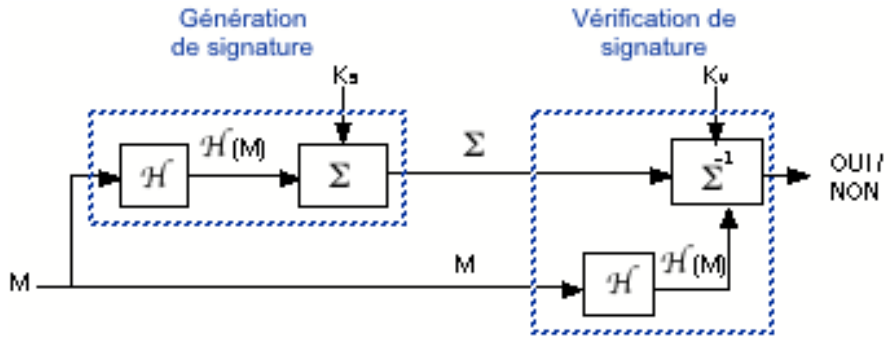
Rappel : signature numérique



- Ks = clé de signature Kv = clé de vérification
- Intégrité :
  - Sans connaître Ks, "impossible" de générer une signature valide
  - Il est "impossible" de trouver Ks, connaissant M et Σ (clair connu)
  - Il est "impossible" de trouver Ks, en choisissant M (clair choisi)

Signatures à clé publique : Ks ≠ Kv

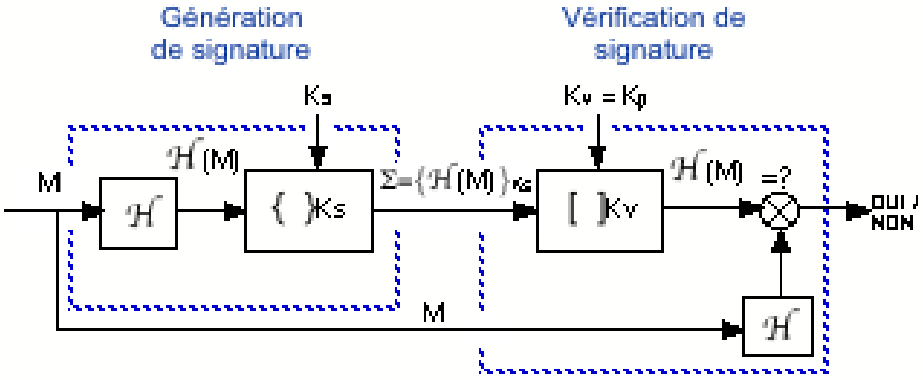
- Exemple : DSA



- Fonction de hachage : SHA-1
- Signature/vérification : el Gamal

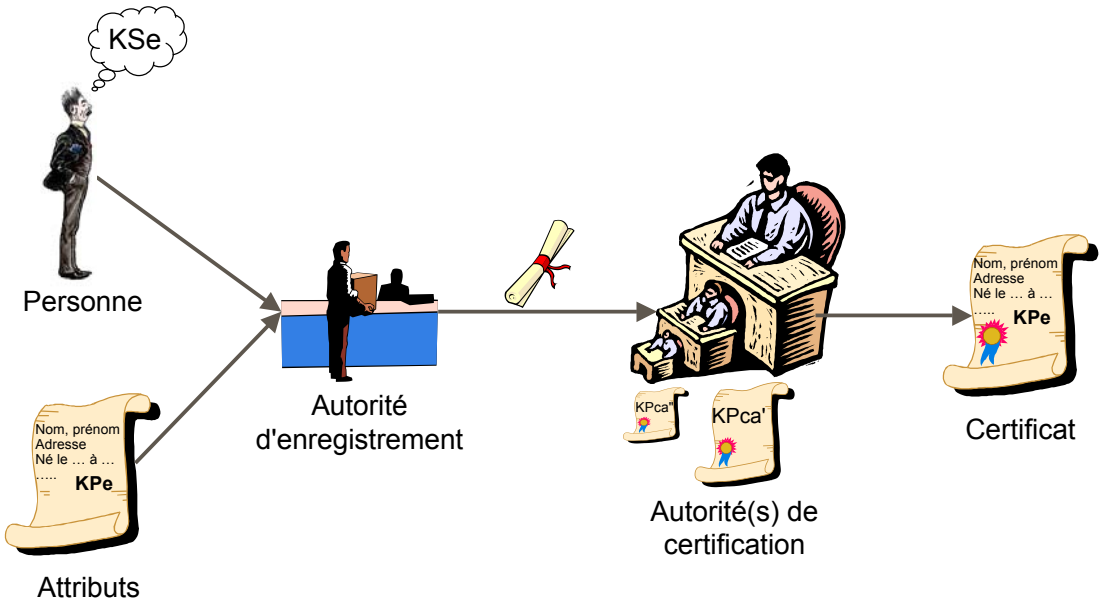
Signature par chiffres à clé publique

- Exemple : RSA



- Ks = clé de signature = clé de chiffrement Kc privée
- Kv = clé de vérification = clé de déchiffrement Kd publique

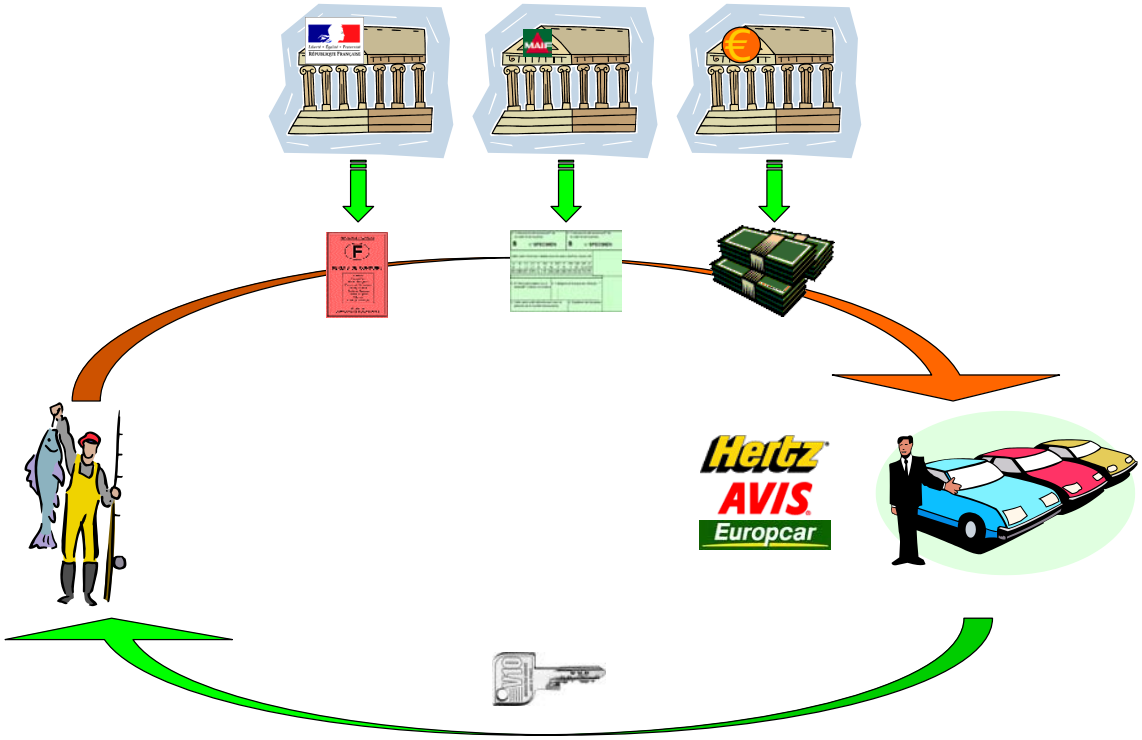
Certificats - IGC (PKI)



Preuves d'autorisation: credentials

- *Credential* = garantie, accréditation
- Certificats multiples :
  ex: SPKI : certificats d'attributs/d'autorisation
  - cartes d'abonnement, de membre d'association, …
  - permis de conduire, carte d'électeur…
- Problèmes: "chaînabilité" (confiance dans l'AC ?,une seule clé publique pour plusieurs certificats ?), gestion des certificats/clés, authentification, préservation des preuves, révocation, …
- Certificats restreints :
  - "Partial Revelation of Certified Identity"
    Fabrice Boudot, CARDIS 2000

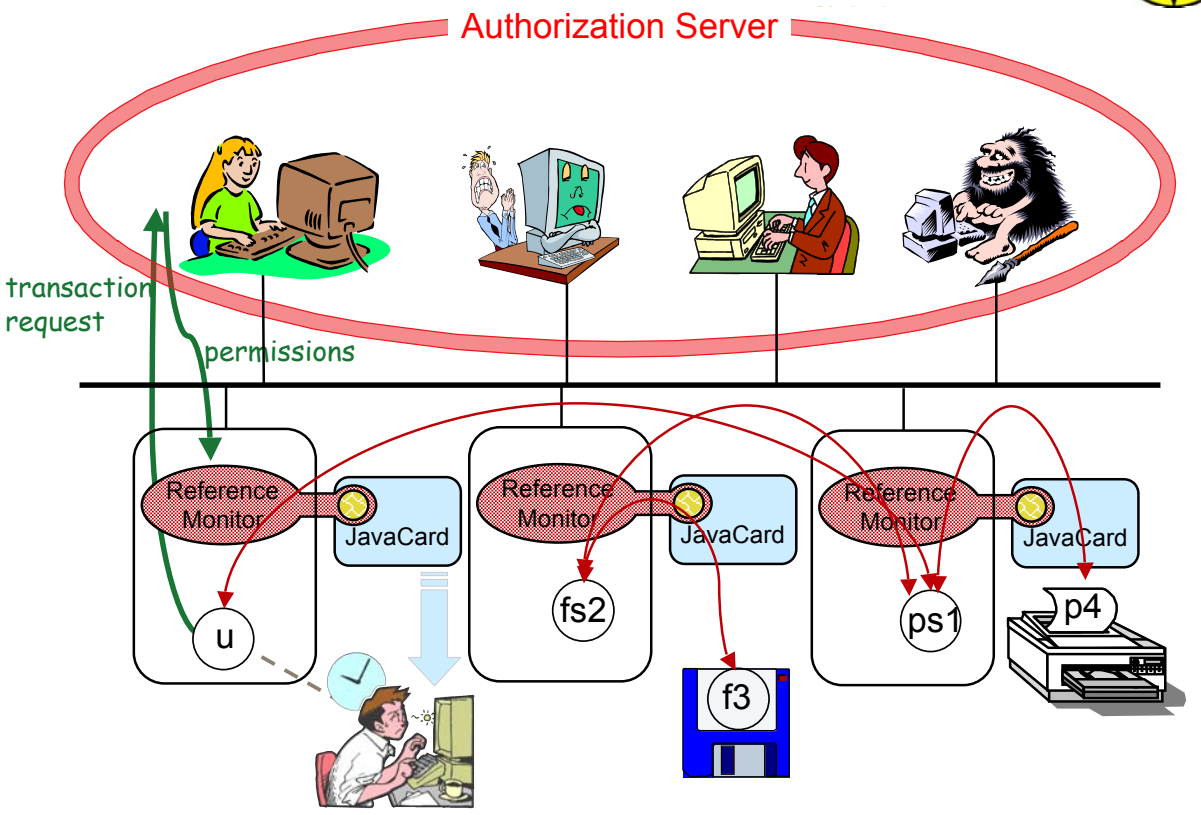
"Anonymous Credentials" (Idemix)



Signature de groupe

- Une clé publique de vérification de signature, *n* clefs privées de génération de signature.
- Le responsable de groupe distribue une clef privée à chaque membre du groupe.
- Pour prouver qu'on est membre du groupe (= possède une garantie anonyme), on chiffre un message aléatoire, vérifiable, signé par le groupe.
- La vérification de la signature est une preuve d'appartenance, donc de garantie.
- Seul le responsable de groupe peut vérifier quel membre a signé.

Autorisation dans MAFTIA



e-Cash (1)

- Propriétés souhaitées :
  - Anonymat : un billet n'identifie pas la personne pour laquelle il a été émis
  - Impossibilité de fabriquer des faux
  - Impossibilité de dépenser deux fois
  - Transmissibilité : un billet peut être échangé entre personnes
  - Liquidité : un billet peut être divisé en petites coupures, ou agrégé en coupures supérieures

e-Cash (2) : signature aveugle (*blind sign.*)

- Alice génère un nombre aléatoire R, le multiplie par un facteur secret S, et l'envoie signé à sa banque: A->B: $[R*S, \text{valeur}]_A$
- La banque débite le compte d'Alice de la valeur, et renvoie le billet signé à Alice : B->A: $[R*S, \text{valeur}]_B$
- Alice "désaveugle" le billet $[R, \text{valeur}]_B$, et le dépense chez un marchand
- Le marchand transmet le billet à la banque : M->B: $[R, \text{valeur}]_B$
- La banque vérifie la signature, enregistre le billet comme dépensé, et crédite le compte du marchand de la valeur, et notifie le marchand, qui donne un reçu à Alice
- Si Alice (ou le marchand) essaye de redépenser le billet, la banque trouvera le billet dans la liste des billets dépensés

4° PET : gestion des données personnelles

- Négociation entre l'individu et l'entreprise
  ex: coupons de réduction en échange d'une publicité ciblée
- Auto-détermination : celui qui fournit des informations sur lui-même doit pouvoir contraindre l'usage qui pourrait en être fait --> Obligations
  ex: à effacer dans 48 h.
- Minimisation des données personnelles
  -> répartition : séparation des pouvoirs, fragmentation des données
  -> anonymisation + appauvrissement
  ex: remplacer le code postal par l'identifiant de la région

  -> Private Information Retrieval (PIR)

4°-bis PET : Accès aux données

- Principe du moindre privilège : un individu ne doit avoir que les droits minimaux nécessaires à sa tâche
- Politique de sécurité et mécanismes de protection : le détenteur d'une information en est responsable (art 34 de la loi « informatique et libertés »)
- Ces données peuvent être très critiques :
  ex: dossiers médicaux
  - Disponibilité : temps de réponse (urgence), pérennité
  - Intégrité : nécessaire à la confiance, éléments de preuve
  - Confidentialité : vie privée <-> intérêts économiques
- Privacy = contrôle d'accès + obligations

Contrôle d'accès aux données

- Séparation entre décision de contrôle d'accès et mise en œuvre
  - Décision : à un niveau élevé (ex. transaction)
    - Cohérence de l'ensemble des opérations
    - Décision sur la « sémantique » de la transaction
    - Moindre privilège : le privilège d'exécuter la transaction est inférieur à celui d'exécuter les opérations élémentaires

    Si OK --> génération de preuves d'autorisation
  - Mise en œuvre : à chaque opération élémentaire : fournir ou bloquer l'accès en fonction de l'opération et de ses paramètres vs. les preuves d'autorisation

Exemple : virement bancaire

- Transaction : virer 2000 € du compte 184-948449 au compte 946448-658
  - Lire le solde du compte 184-948449
  - Tester si le solde est supérieur à 2000 €
  - Si oui :
    - solde := solde - 2000; écrire solde 184-948449
    - Lire le solde du compte 946448-658
    - solde := solde + 2000; écrire solde 946448-658
  - Si non : retourner « solde insuffisant ».

Donner confiance aux utilisateurs…

… que leur vie privée est protégée?

- Certification & labellisation
- Approche Trusted Computing Group (TCG)
  - Support matériel : TPM
  - Bootstrap sûr
  - Vérification sceau S/W avant chargement
  - Vérifiable à distance, sans dévoiler d'identité (DAA)

PRIME (03/2004 - 02/2008)

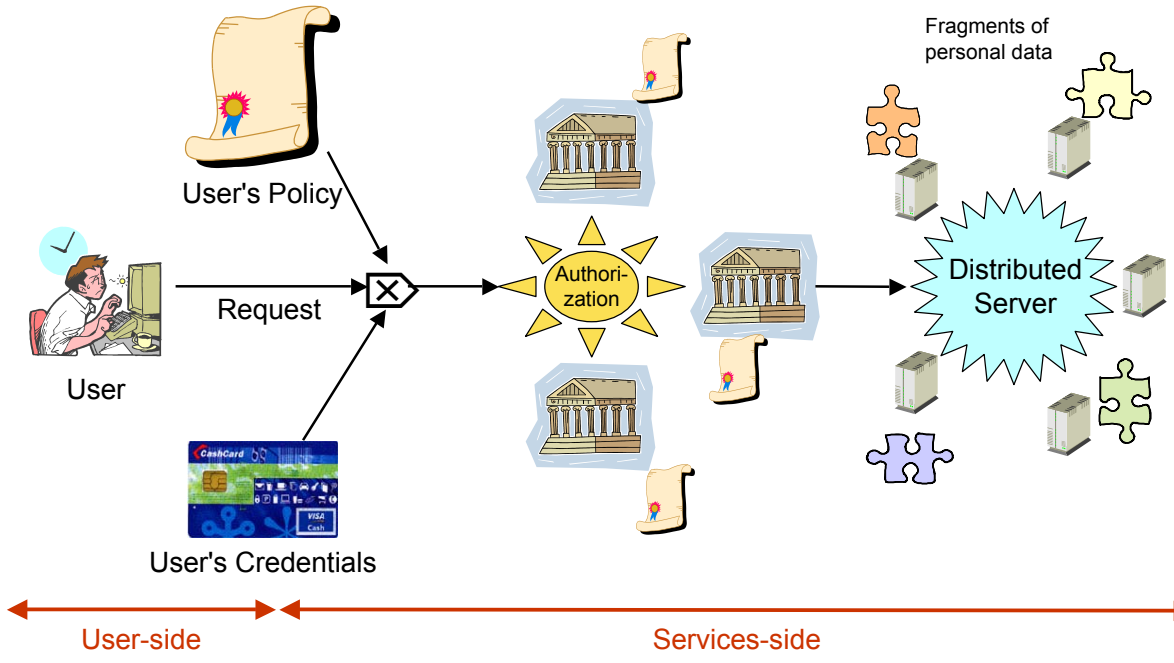
http://www.prime-project.eu/

- Privacy and Identity Management for Europe
  - Aspects juridico-socio-économiques
  - PET Côté utilisateur (dévelopt, utilisabilité)
  - PET Côté système, réseau, serveur
  - Applications réelles
- 20 Partenaires, 16 M€, subvention : ~10 M€
  - Fournisseurs (IBM, HP, …)
  - Labos (KUL, U. Dresde, U. Milan, Eurécom, LAAS…)
  - Utilisateurs (Lufthansa, T-Mobile, Swisscom, HSR)

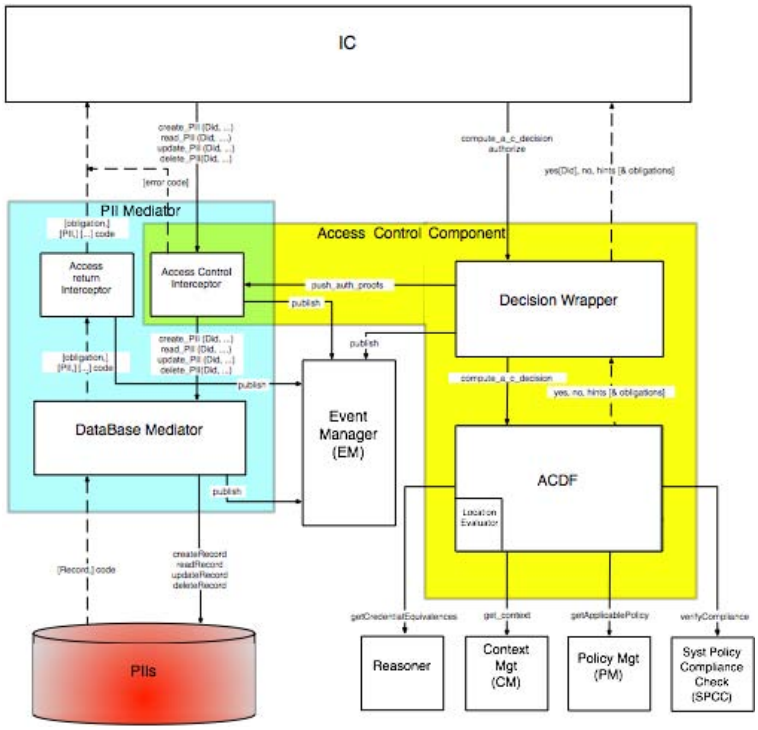
Principe :

- Identités différentes selon les besoins

Exemple d'architecture



Architecture du contrôle d'accès



Bibliographie

- *Sécurité des systèmes d'information V.2*, dir. Ludovic Mé & Yves Deswarte, Traité IC2, série Réseaux et télécommunications, Hermès, ISBN 2-7462-1259-5, 390 pp., juin 2006.
- Simone Fischer-Hübner, *IT-Security & Privacy*, LNCS 1958, Springer, 2001.
- Stefan A. Brands, *Rethinking Public Key Infrastructures and Digital Certificates*, MIT Press, 2000.
- Yves Deswarte, Carlos Aguilar-Melchor, "Current and Future Privacy Enhancing Technologies for the Internet", *Annales des Télécommunications*, vol.61, n°3-4, March/April 2006.
- Yves Deswarte, Carlos Aguilar-Melchor, Vincent Nicomette, Matthieu Roy, "Protection de la vie privée sur Internet", *Revue de l'Électricité et de l'Électronique* (*REE*), octobre 2006 (n°9), pp.65-74.
- Carlos Aguilar-Melchor, "Les communications anonymes à faible latence", Thèse de l'Institut National Polytechnique de Toulouse, 4 juillet 2006, LAAS n°06571.