Technologies de protection de la vie privée sur Internet

Yves Deswarte
deswarte@laas.fr

LAAS-CNRS, Toulouse

Sécurité & protection de la vie privée

- *"Privacy"* ≈ confidentialité de données (et méta-données) personnelles
  PII : Personally Identifiable Information

- = sous-ensemble de "sécurité" (*CIA*)

- Mais…

… "*the devil is in the details*"

- Garder les justificatifs, en cas de litige
- Traçabilité des actions
- Authentification forte
- … danger pour la vie privée !!!

Sommaire

- *"Privacy"* : Définition, Règlementation
- Principes de base
- PETs : *Privacy Enhancing Technologies*
  - Gestion d'identités multiples
  - Communications et accès anonymes
  - Autorisation respectant la vie privée
  - Gestion des données personnelles
- Projet Prime

"Privacy" : définitions

- Intimité, respect/protection de la vie privée, …
- Critères Communs (ISO 15408) :
  une classe fonctionnelle, 4 propriétés :
  - Anonymat : impossibilité (pour d'autres utilisateurs) de déterminer le véritable nom de l'utilisateur associé à un sujet, une opération, un objet
  - "Pseudonymat" : idem, sauf que l'utilisateur peut être tenu responsable de ses actes
  - Non-"chaînabilité" : impossibilité (pour d'autres utilisateurs) d'établir un lien entre différentes opérations faites par un même utilisateur
  - Non-observabilité : impossibilité (pour d'autres utilisateurs) de déterminer si une opération est en cours

Pseudonymat < anonymat < non-chaînabilité < non-observabilité

Réglementation (1)

- Déclaration universelle des droits de l'homme : Art. XII: "Nul ne sera l'objet d'immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance…" *Ass. Géné. ONU, décembre 1948*

- Internationale : Guides pour l'utilisation de données personnelles informatisées et leurs transmissions internationales : *OCDE en septembre 1980, Assemblée Générale de l'ONU, en décembre 1990.*

- Européenne : Protection des données à caractère personnel : Convention 108 du Conseil de l'Europe (26/01/81), directives 95/46/EC (libre mouvement) et 2002/58/CE (communications électroniques, remplaçant la directive 97/66/CE) + directive conservation données 2006-24-EC

- Française : 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* »

Réglementation (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.

Principes de base

1er Principe pour protéger la vie privée :

- **Minimisation des données personnelles**
  ne transmettre une information qu'à ceux qui en ont besoin pour réaliser la tâche qu'on leur confie
  -> "Besoin d'en connaître" ("*need-to-know*")
  puis destruction/oubli

- … sur Internet comme dans le monde réel

- …avec des limites : certaines informations personnelles doivent pouvoir être fournies aux autorités judiciaires en cas de litige ou d'enquête (lutte contre le blanchiment d'argent sale, par exemple) : "pseudonymat" plutôt qu'anonymat total

- Liens : minimisation <--> proportionnalité et finalités légitimes

Exemple : commerce électronique (1)

- Parties impliquées :
  un client, un marchand, un service de livraison, des banques, un émetteur de carte de crédit, un fournisseur d'accès Internet, …

- Le marchand n'a pas besoin (en général) de l'identité du client, mais doit être sûr de la validité du moyen de paiement.

- La société de livraison n'a pas besoin de connaître l'identité de l'acheteur, ni ce qui a été acheté (sauf les caractéristiques physiques), mais doit connaître l'identité et l'adresse du destinataire.

Exemple : commerce électronique (2)

- La banque du client ne doit pas connaître le marchand ni ce qui est acheté, seulement la référence du compte à créditer, le montant …
- La banque du marchand ne doit pas connaître le client…
- Le f.a.i. ne doit rien connaître de la transaction, sinon les caractéristiques techniques de la connexion …

2ème Principe pour protéger la vie privée :

- "Souveraineté" : garder le contrôle sur ses [méta-] données personnelles

  -> stockage sur un dispositif personnel (carte à puce, PDA, PC…)

  -> si ces données sont divulguées à un tiers, imposer des **obligations** sur leur usage

  - Date de péremption
  - Notification en cas de transfert ou d'usage non prévu
  - etc…

Technologies de protection de la vie privée

PETs : *Privacy-Enhancing Technologies*

- Gestion d'identités multiples
- Communications et accès anonymes
- Autorisation respectant la vie privée
- Gestion des données personnelles

1° PET : gestion d'identités multiples

- Identité = représentation d'une personne physique
- Réduire/contrôler les liens entre une personne et les données (et méta-données) la concernant (contrôler la *chaînabilité*)
  - on présuppose la non-chaînabilité des communications et des accès
- Mais : accès personnalisés / privilégiés : pseudonymes
  - Préférences (ex: météo) -> « *cookies* »
  - "Rôles" différents -> pseudonymes différents
    - Ex: contribuable et électeur
  - Authentification adaptée au risque d'usurpation d'identité (et à la responsabilité)
  - Durée de vie liée aux besoins de chaînabilité -> pseudonymes "jetables"
- Identités virtuelles multiples vs. "*single-sign-on*"
  Liberty Alliance <http://www.projectliberty.org>
  vs. Microsoft Passport

Adresse IP= donnée identifiante

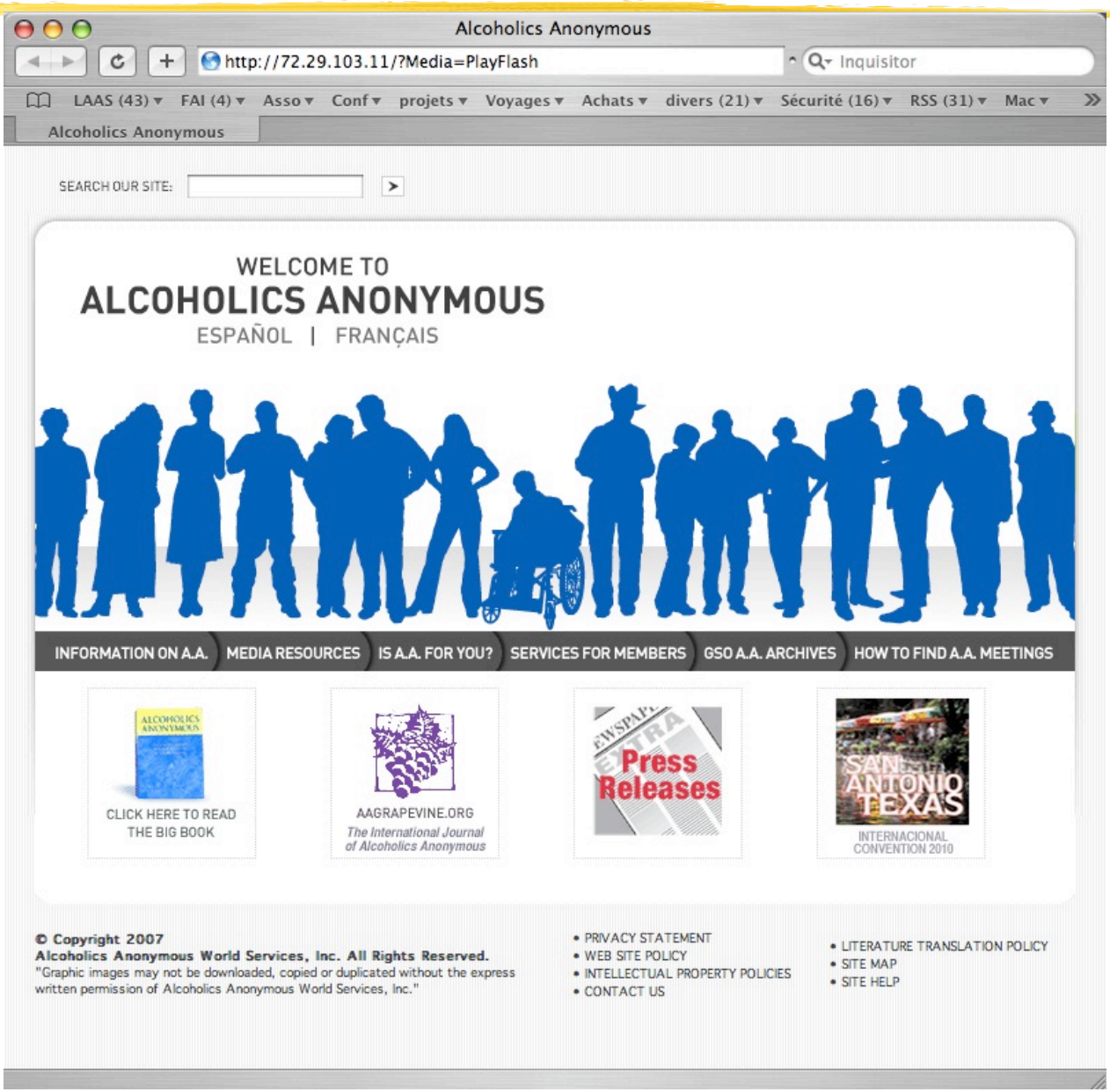
Exemple :

```
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
```

Adresse IP= contenu sensible

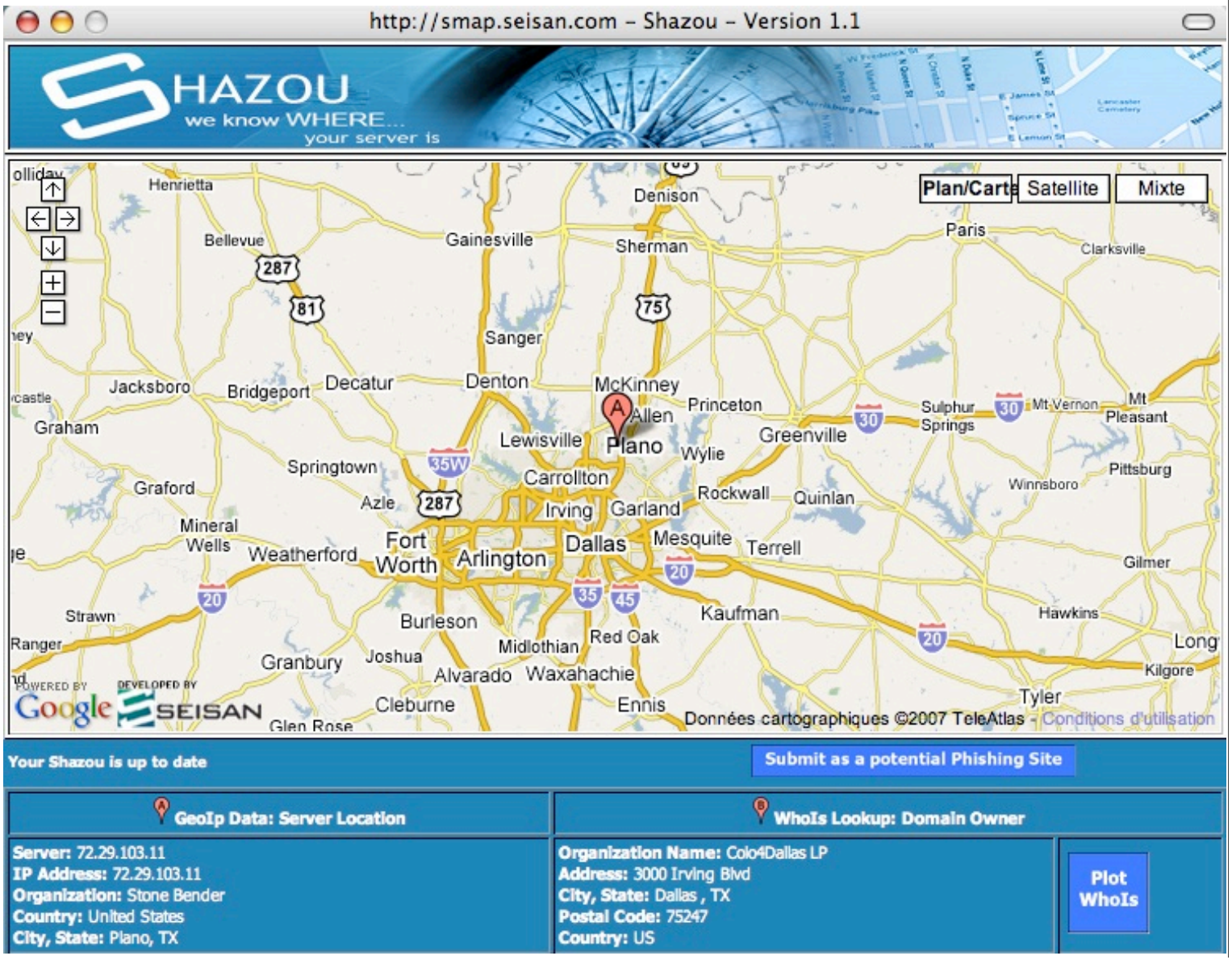
Exemple :

http://72.29.103.11/



Adresse IP= localisation

Exemple :

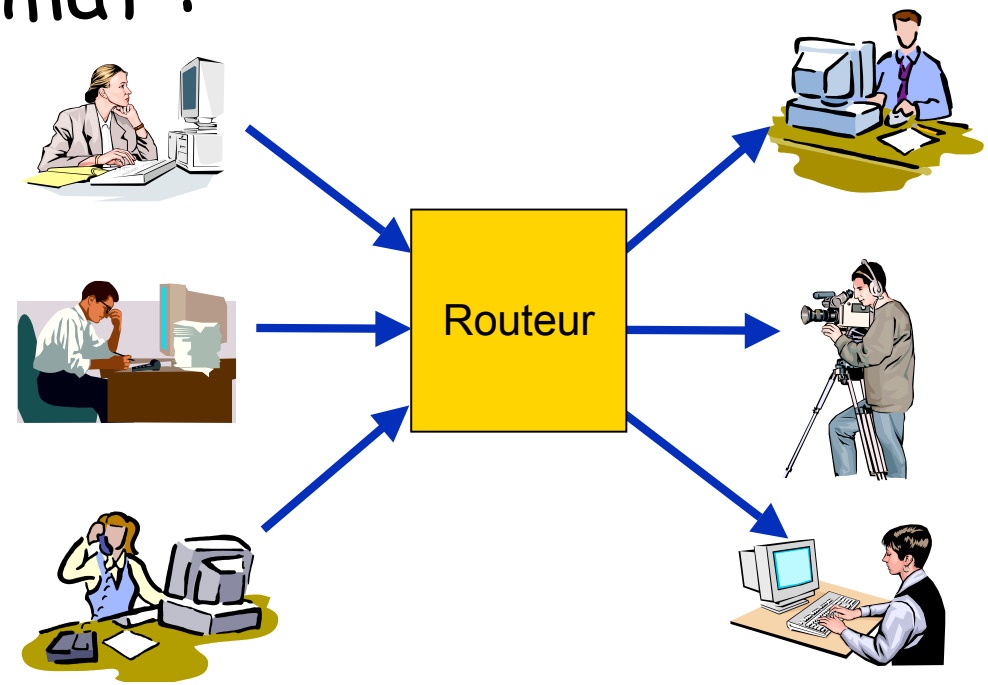


IP V6, réseaux ad hoc, …

- Demain : IP partout (*pervasive/ubiquitous computing, intelligence ambiante, sensor networks, RFID, convergence 4G …*)
- chaque "machin" aura une adresse IP implicite *unique et permanente* (basée sur un numéro de fabrication)
- chaque personne aura plusieurs machins …
- … qui se connecteront aux machins proches (réseaux ad hoc)
- … qui s'identifieront, routeront leurs communications, fourniront des infos contextuelles, etc.

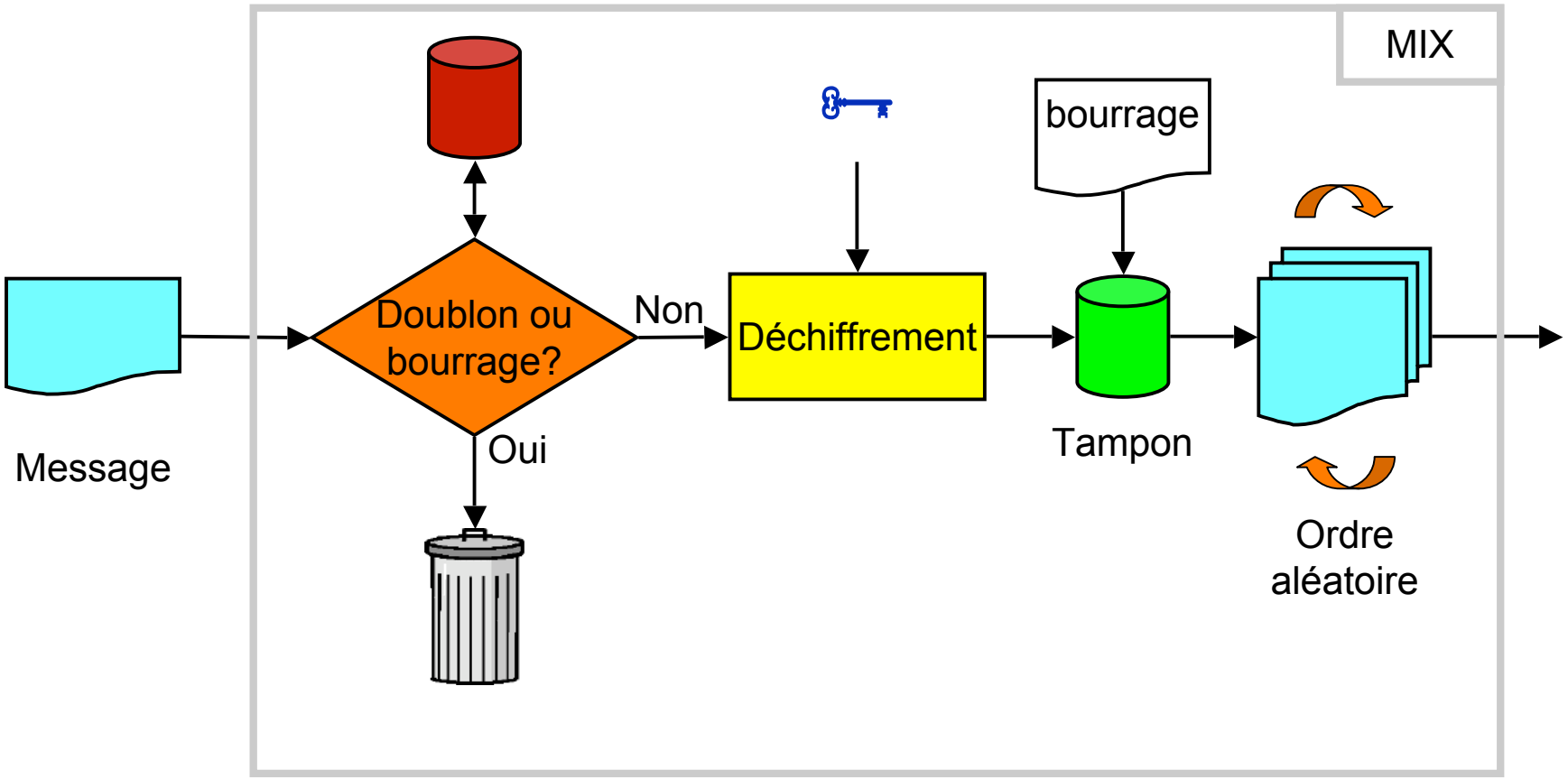
2° PET : Communications et accès anonymes

- Protéger les adresses IP :
  affectation dynamique des adresses IP
  (DHCP, PPP, NAT, ...)
- Routeurs d'anonymat :
  - MIX
  - Onion Routing
  - Crowds

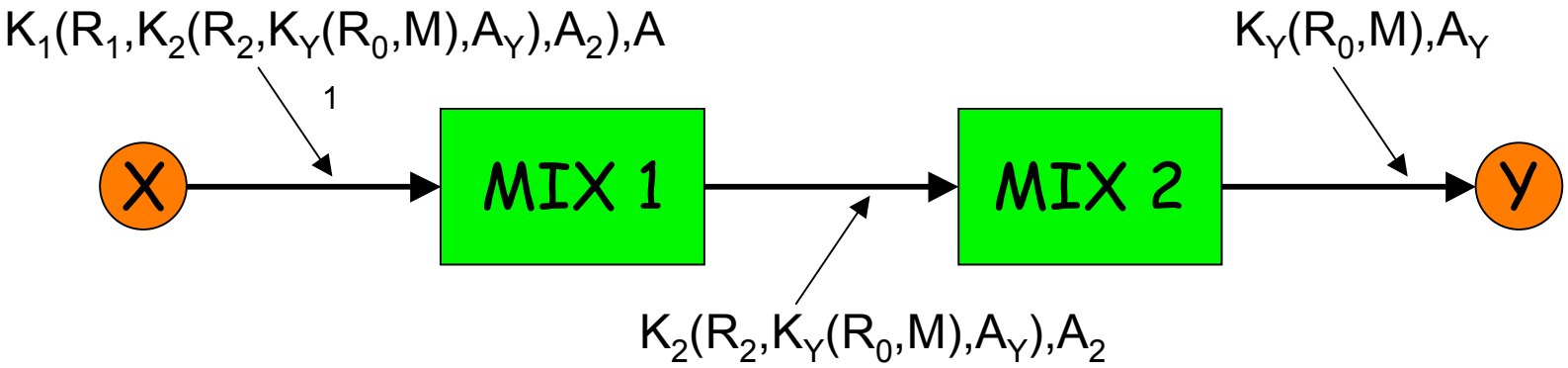


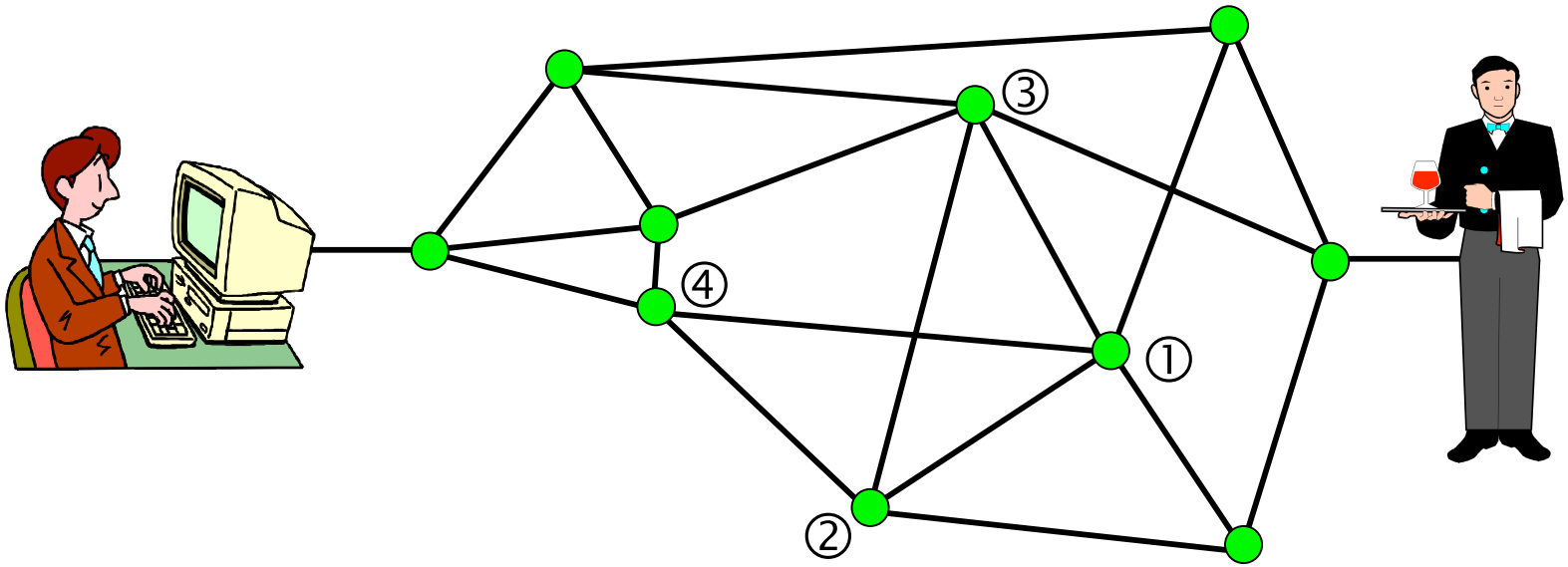
MIX : comment ça marche ?

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



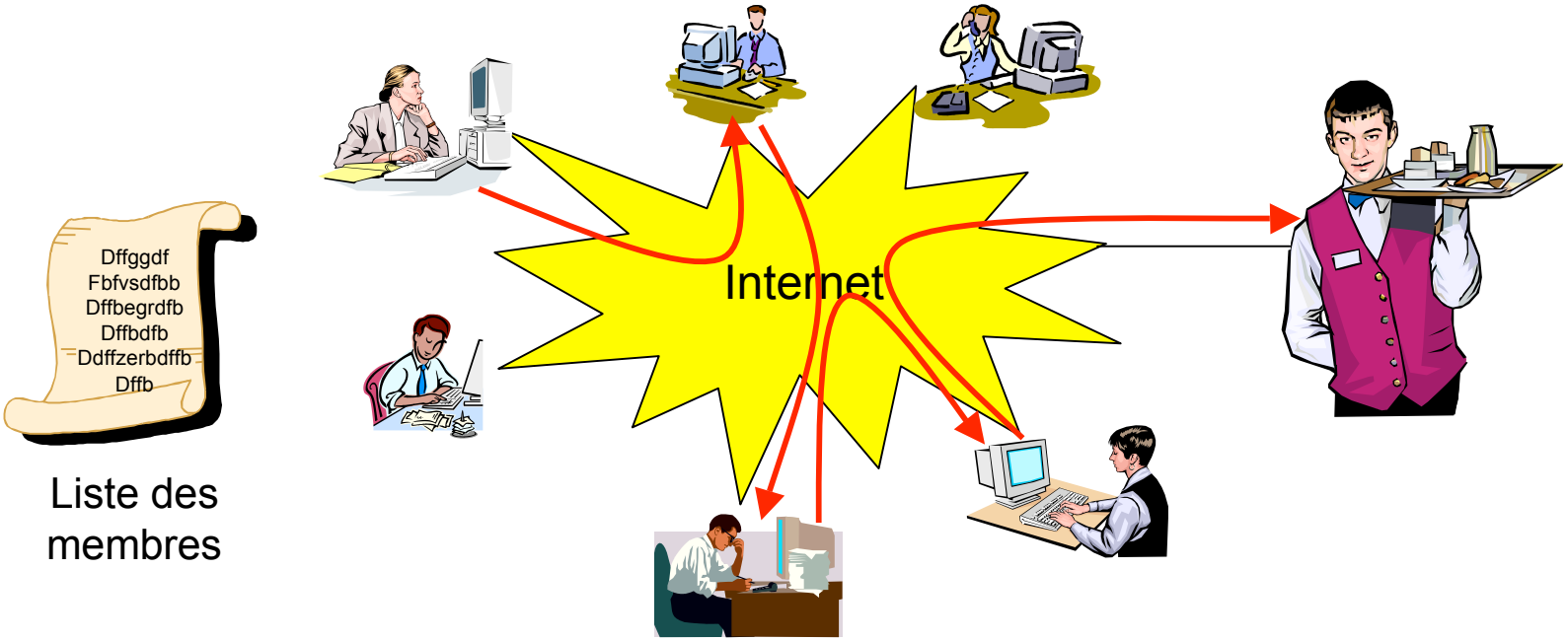
MIX / Onion Routing / Crowds





Crowds/Hords : pair-à-pair

- Chaque membre est un MIX pour les autres
- Probabilité *p* d'envoi au destinataire
  (1-*p*) d'envoi à un autre membre au hasard



Inconvénients des MIX

- Coût (# de messages, chiffrement, latence, …)
- OK pour mail, Web, … mais pas VoIP, …
- Vulnérables à la collusion entre les MIX
  --> indépendance entre les MIX ?
- Vulnérables à un observateur global
  (analyses statistiques)
  --> distribution sur Internet ?
- Interactivité : canal retour + anonymat de relation
- Mal adapté aux réseaux locaux…

Le dîner des cryptographes

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

DC-network

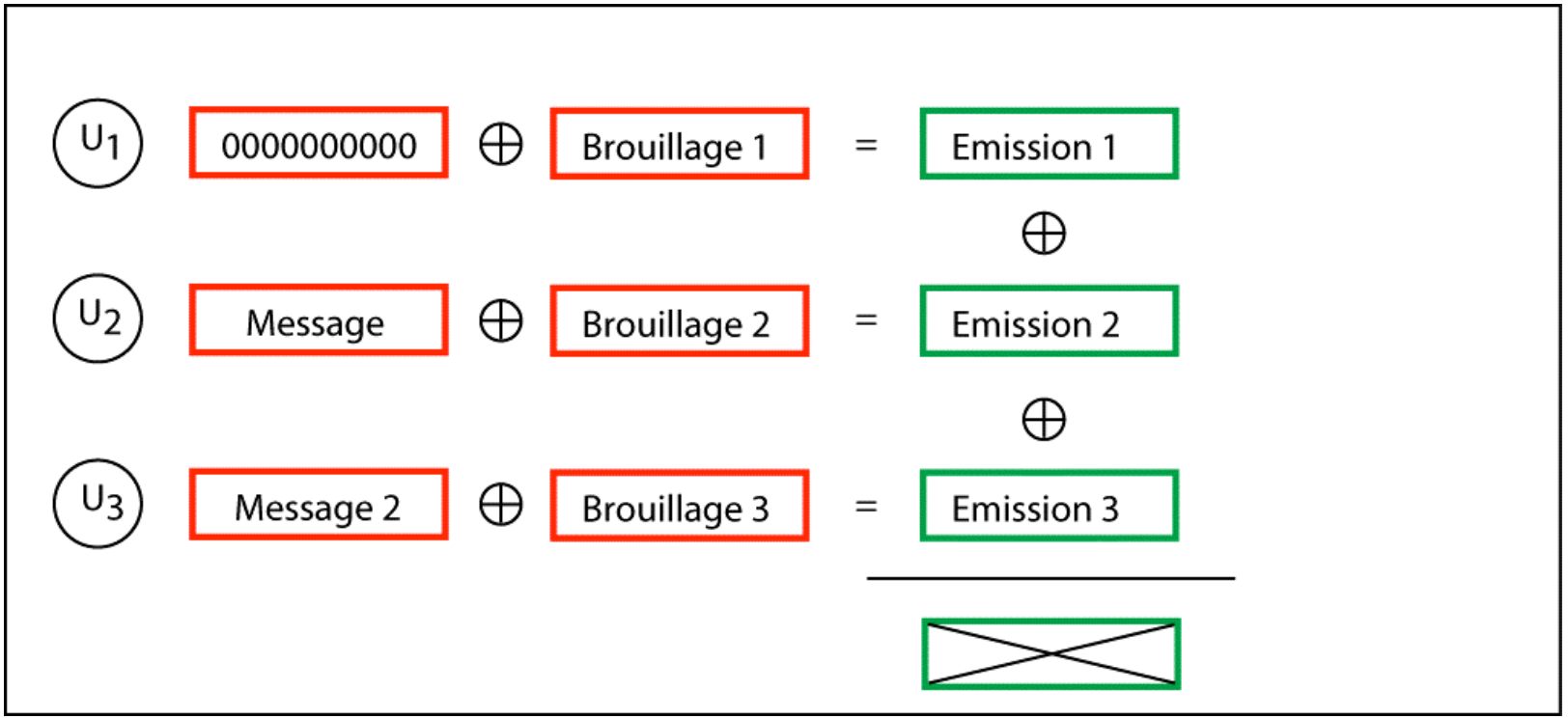
Protocole par tour : à chaque tour :

- 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

- $\forall$ i, j $\in$ {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 …

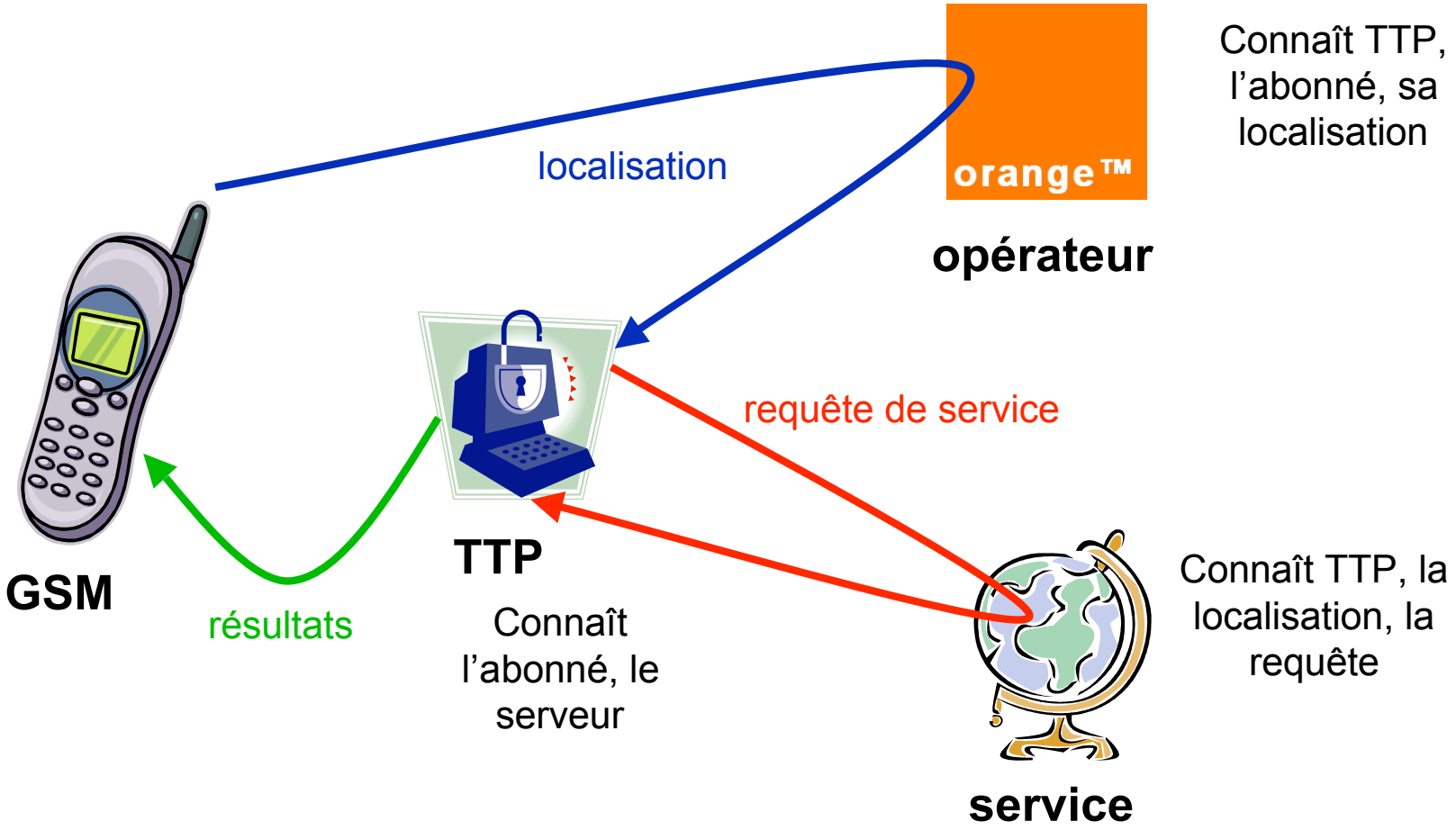
1. Génération d'1 @MAC aléatoire
2. Obtention d'1 @IP temporaire
3. Tunnel vers un TTP de roaming
4. Génération d'une autre @IP
5. Authentification sur FAI

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"

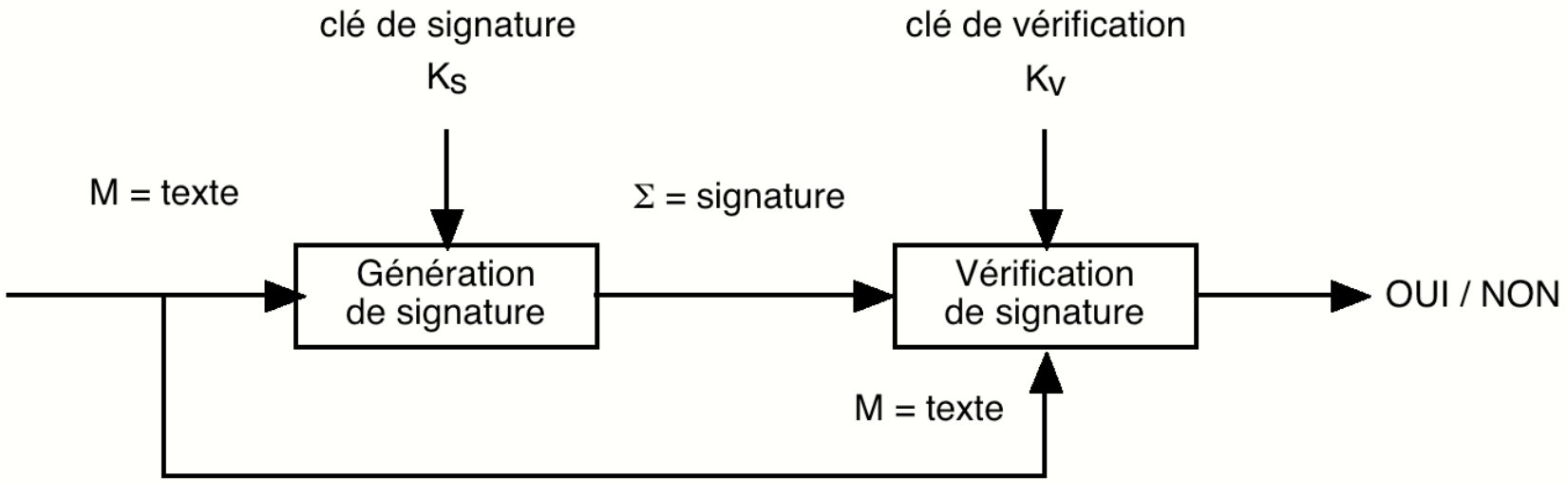
Ce schéma est dépassé

- 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"

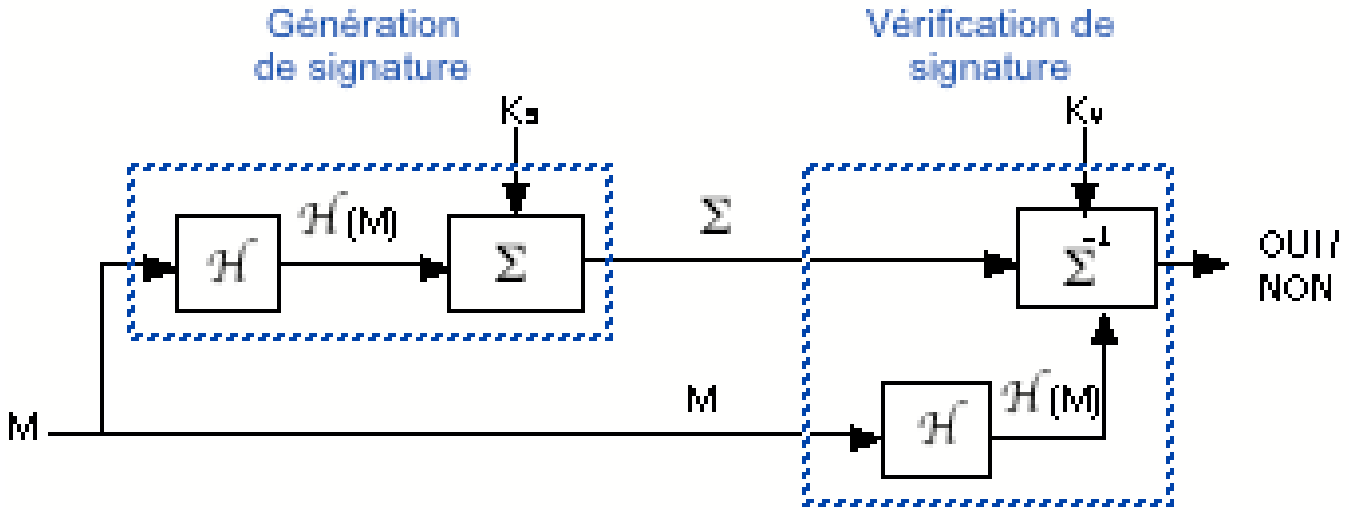
Rappel : signature numérique



- 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

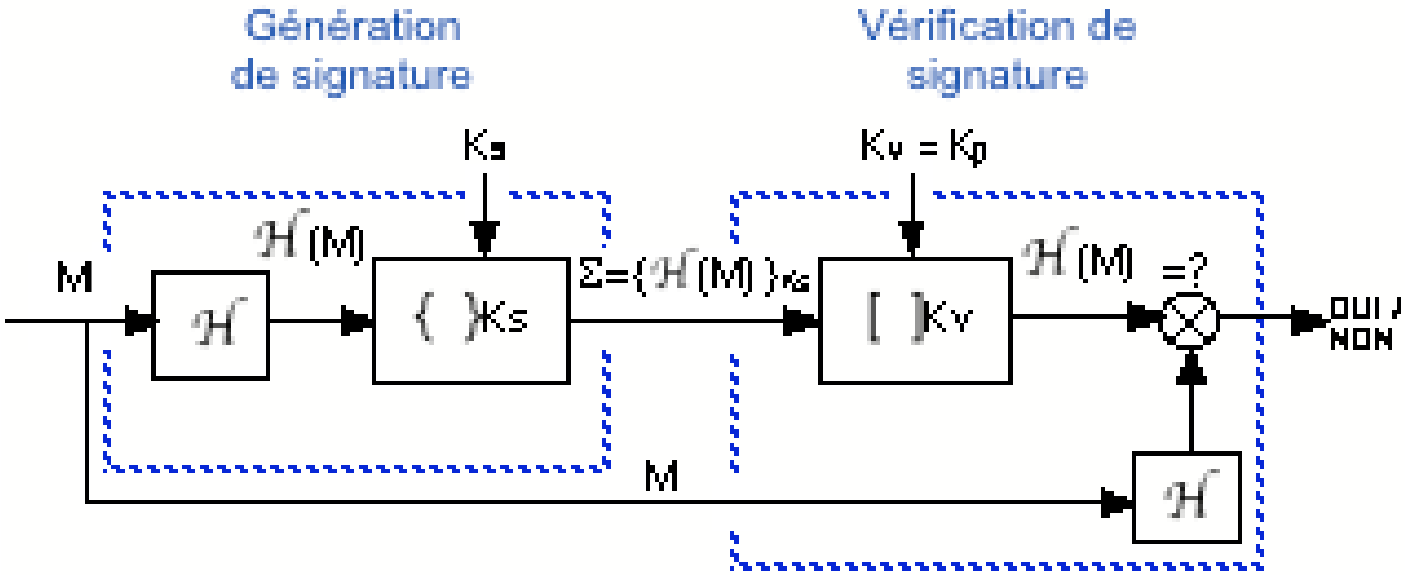
- Exemple : DSA



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

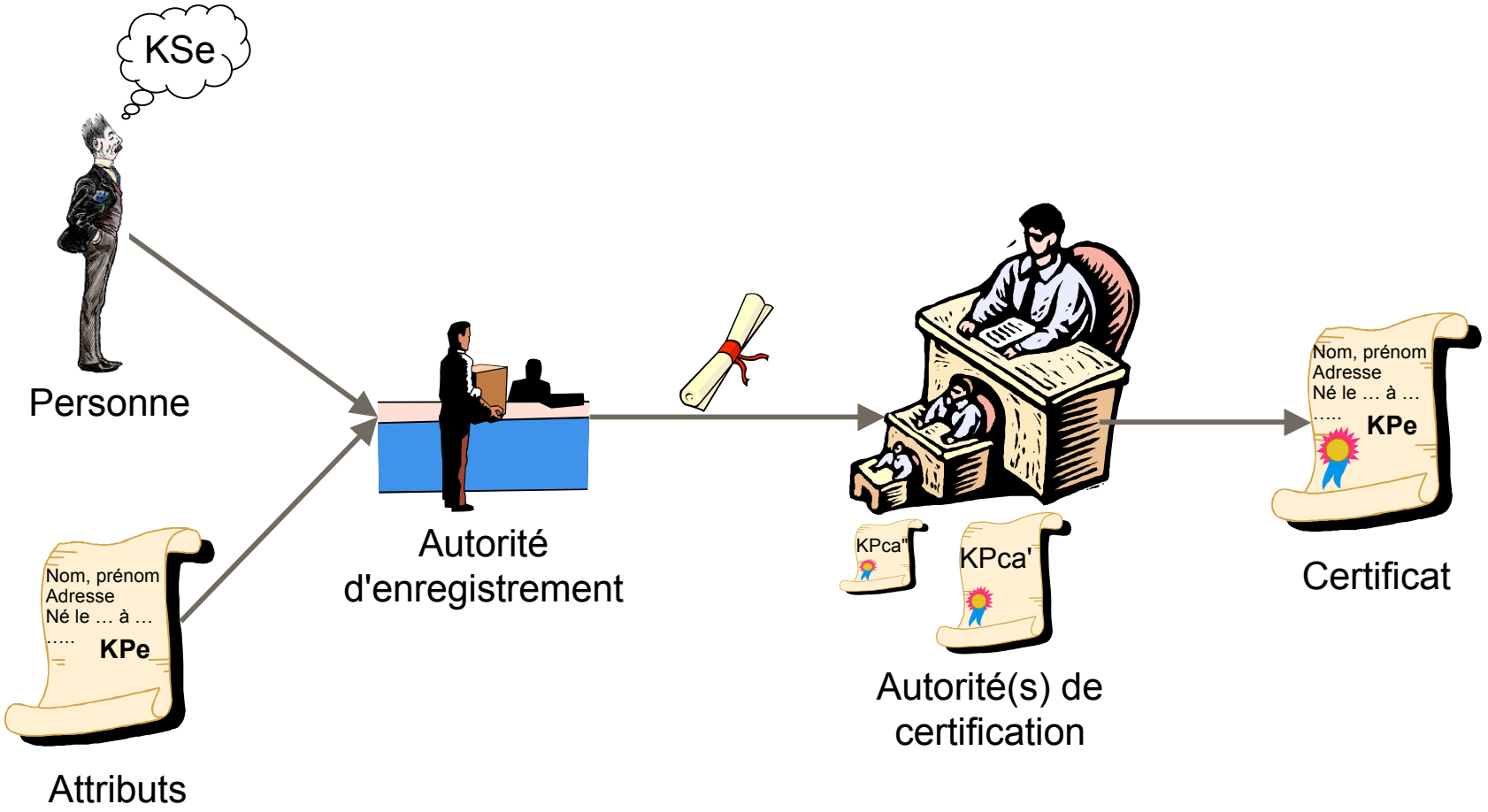
Signature par chiffres à clé publique

- Exemple : RSA



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

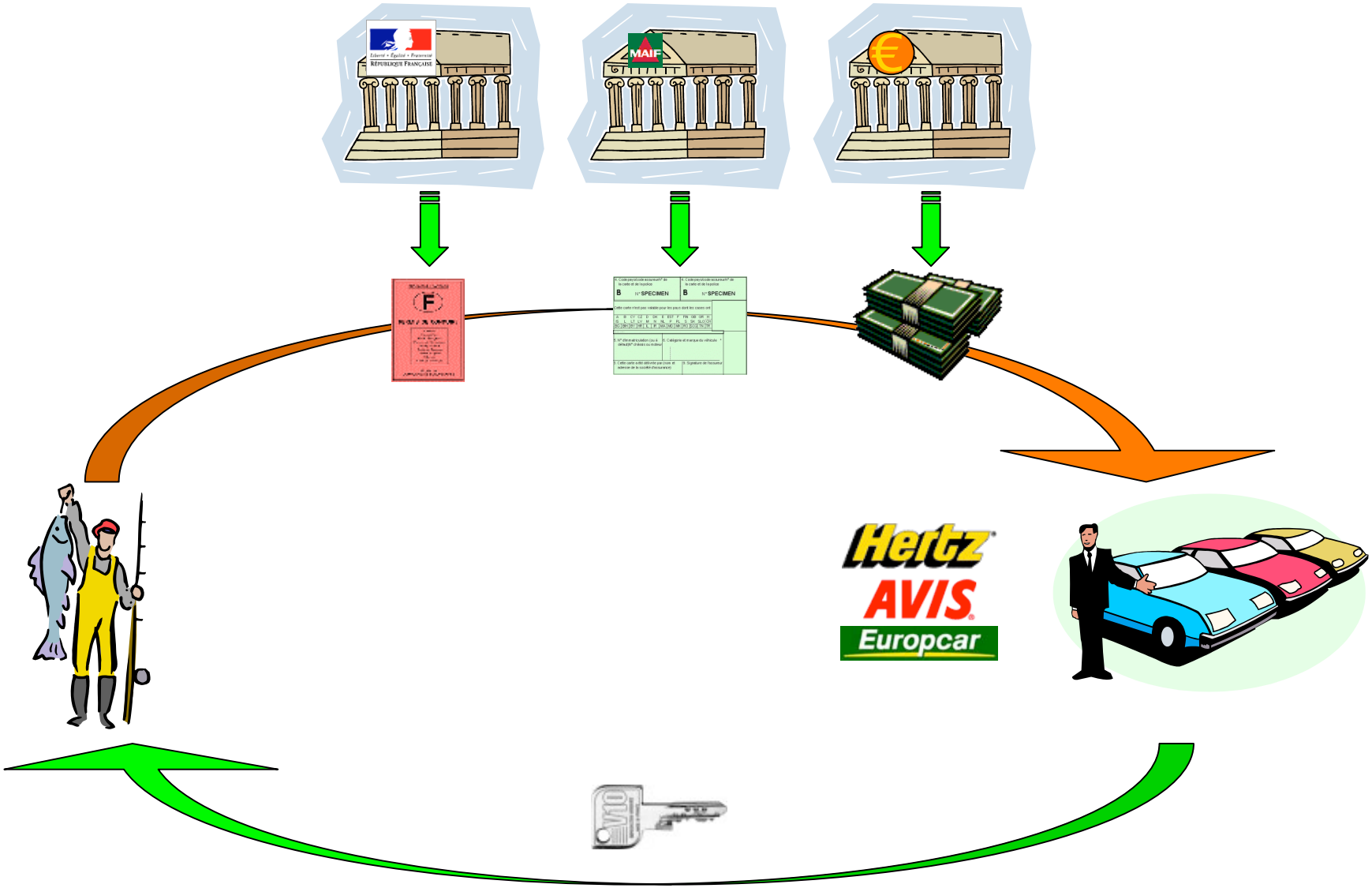
Certificats - IGC (PKI)



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

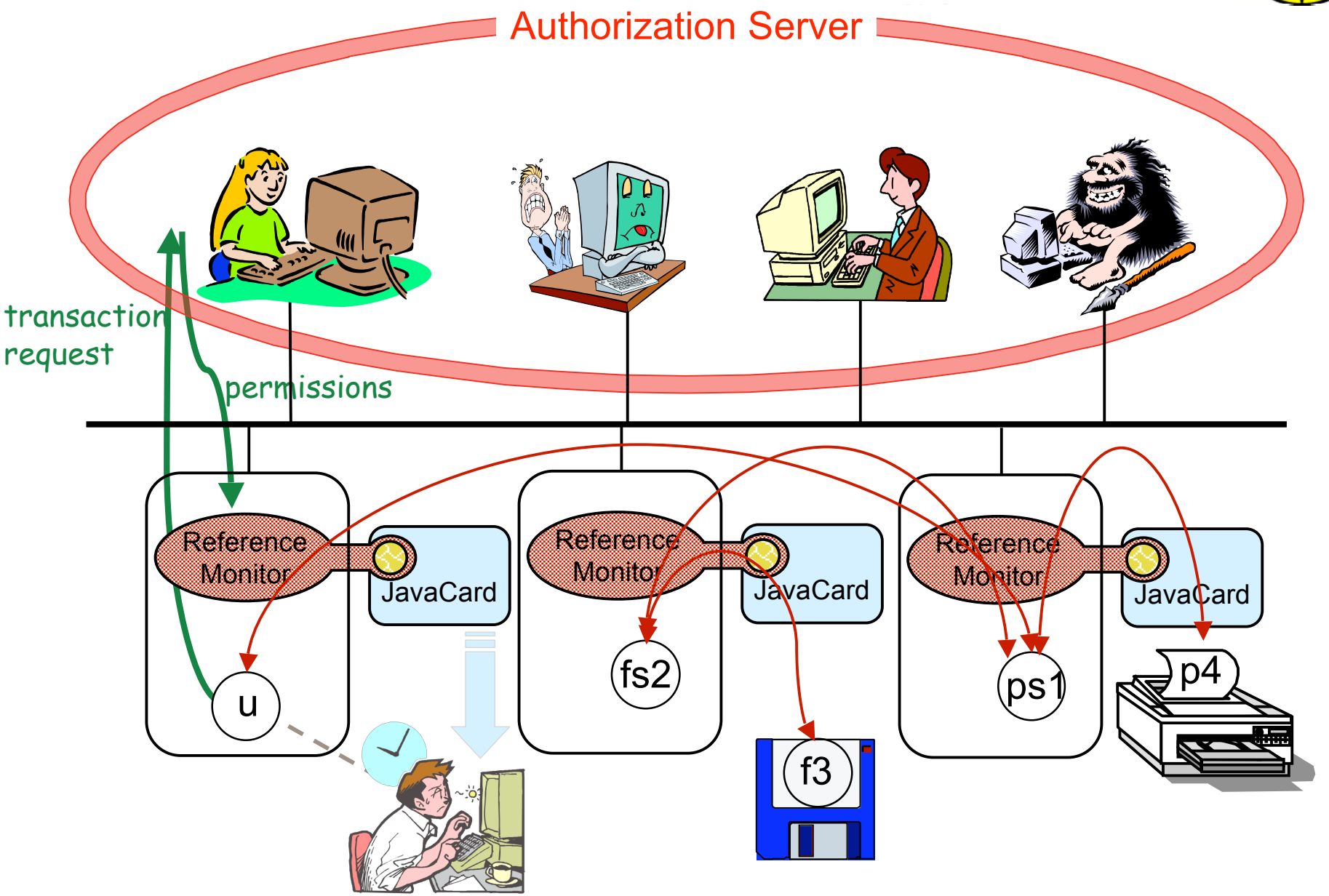
"Anonymous Credentials" (Idemix)



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)

5°-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)

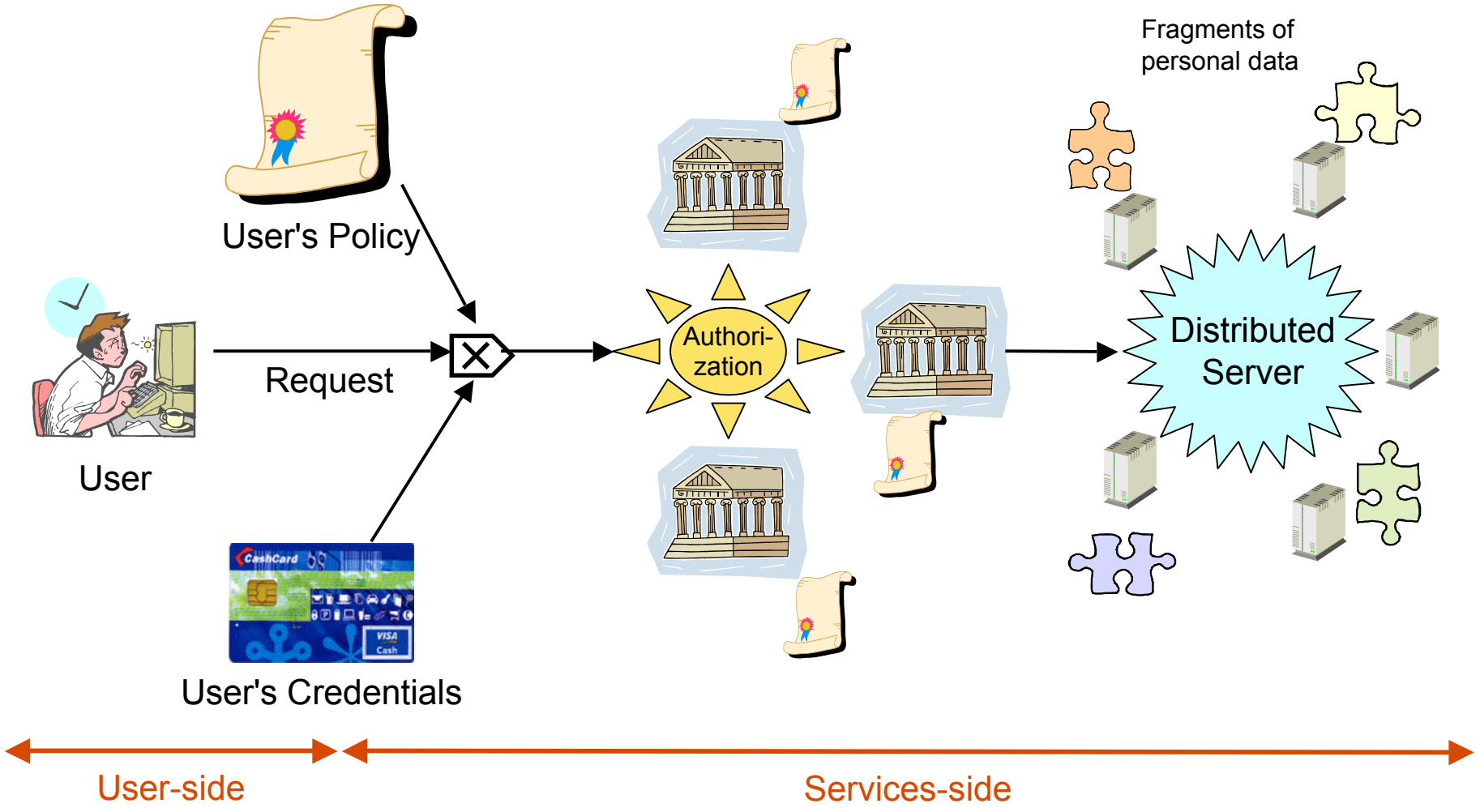
http://www.prime-project.eu/

- 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)

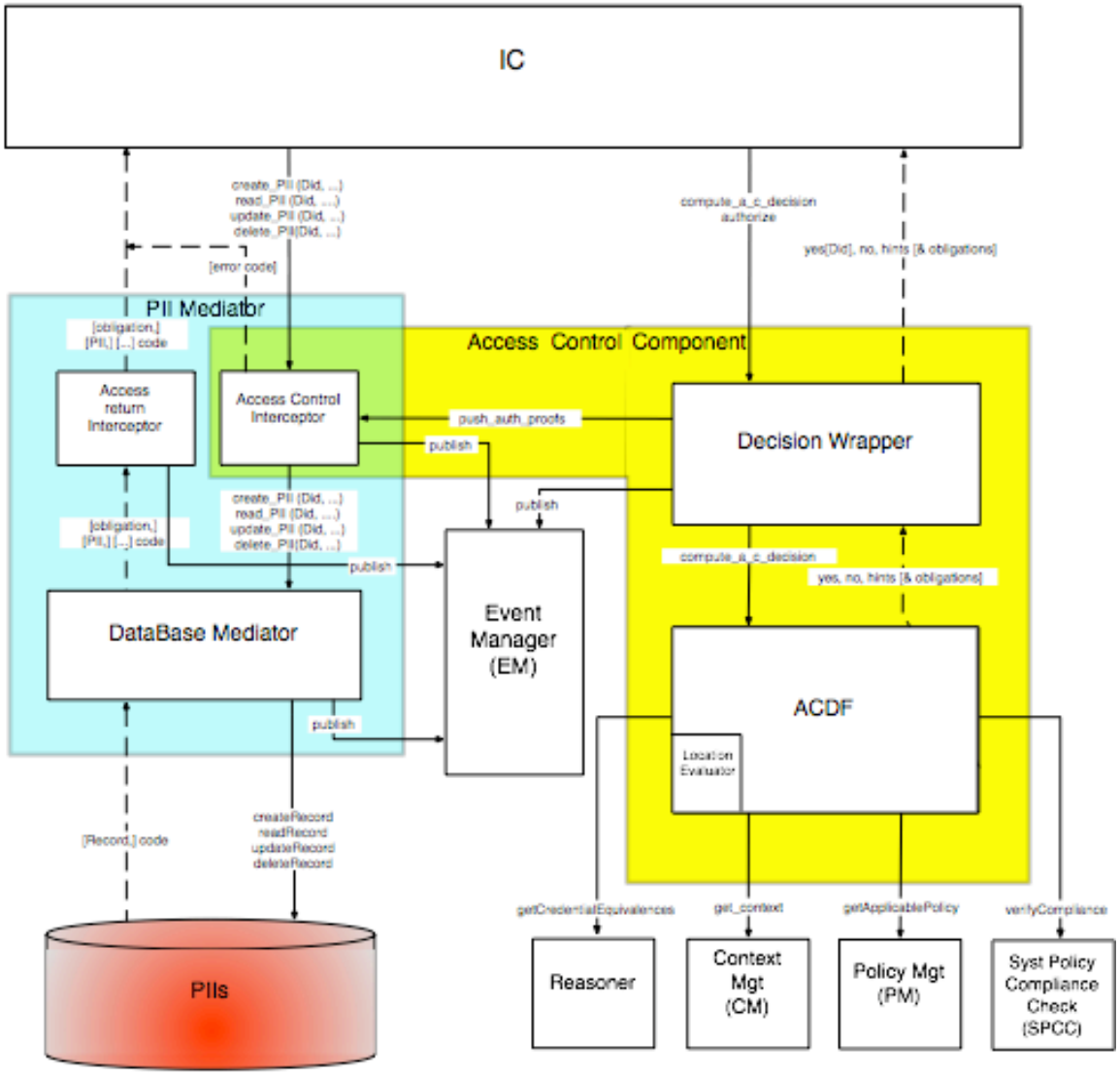
Principe :

- 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.