



Introduction à la sécurité
dans le domaine des Big Data.

**Protection des données personnelles
et
Introduction à la cryptographie**

Thierry Baritaud

Plan de la présentation

- Introduction à la sécurité des SI
 - impact du Big Data sur la sécurité,
 - nouveaux challenges et menaces
- Identification, authentification
- Protection des données personnelles, privacy
- Contexte réglementaire et normatif de la sécurité
- Anonymisation ,pseudonymisation
- Introduction à la cryptographie pour la protection des données
 - Crypto à clé secrète et à clé publique, certificats, chiffrement homomorphe



Evolutions du secteur Télécoms : des ruptures technologiques historiques à fort impact de sécurité



1 • Broadband everywhere



- Images
- X-DSL
- GigaEthernet

2 • Mobility everywhere



- Wifi, Bluetooth, NFC
- Wimax, 3G, 4G, 5G
- IoT, LoRa SigFox

3 • IT platforms on open networks



- E-Commerce
- Instant Messaging
- Web and Intermediation services
- Cloud computing, Big data
- Customer intimacy, Trust and Privacy



Security

4 • Innovative multi-access terminals



■ Terminals

■ Home Gateways

■ Voice services

■ M2M, internet of things

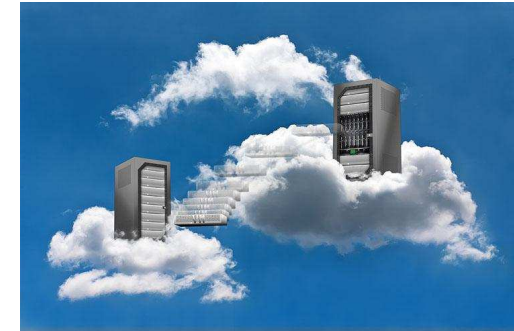


© Thierry BARITAUD

Quelle sécurité pour les nouvelles technologies ?



Cloud



Systèmes industriels SCADA



Intelligence artificielle

© Thierry BARITAUD

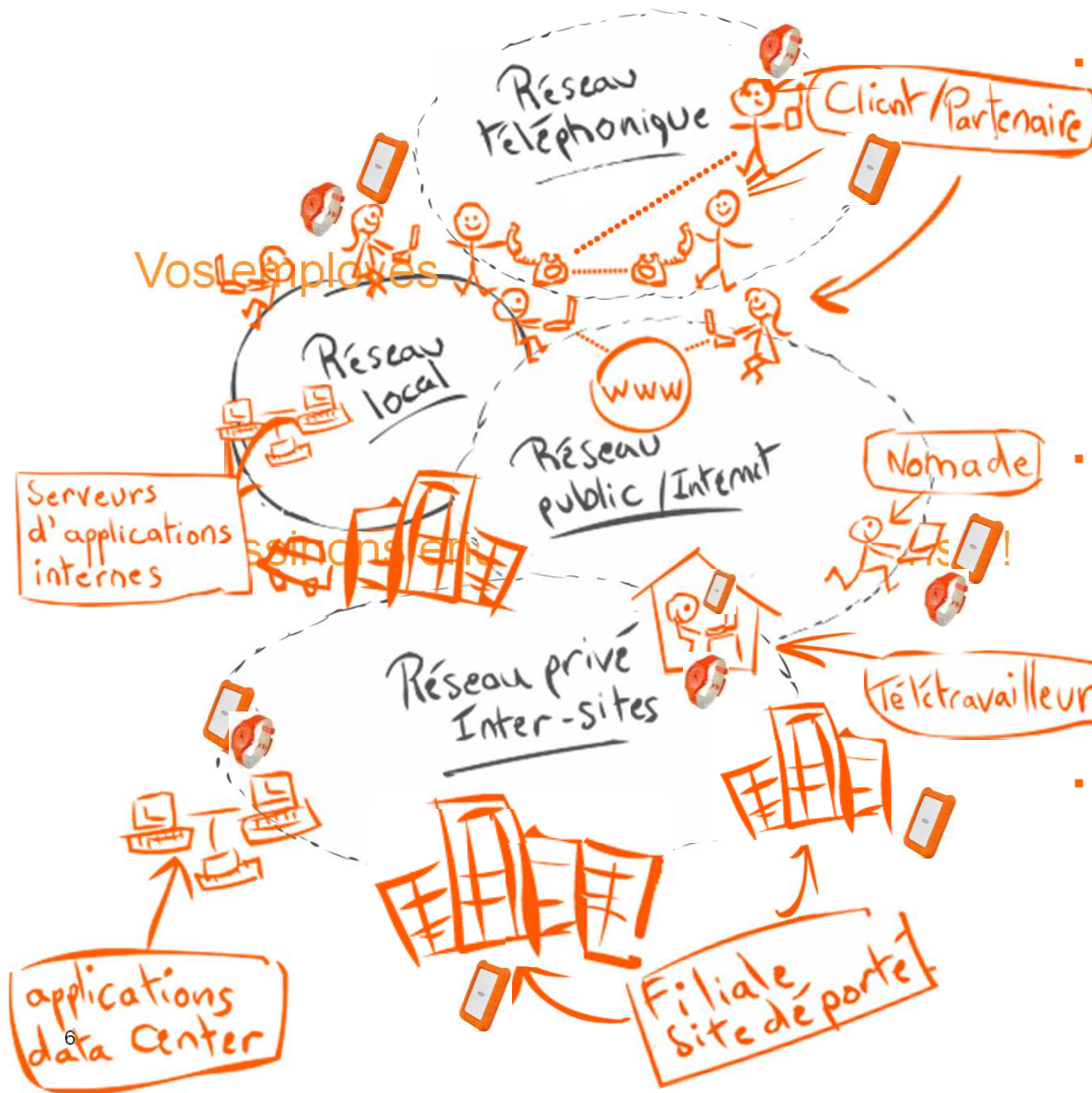
Nouveaux environnements : nouvelles menaces

- source : ENISA

| Top Threats | Current Trends | Top 10 Threat Trends in Emerging Areas | | | | | | |
|------------------------------------|----------------|--|------------------|-------------------|-----------------|-----------------|----------|--------------------|
| | | Critical Infrastr. | Mobile Computing | Social Networking | Cloud Computing | Trust Infrastr. | Big Data | Internet of Things |
| 1. Drive-by Downloads | ↑ | ↑ | ↑ | ↑ | | ↑ | ↑ | |
| 2. Worms/Trojans | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ |
| 3. Code Injection | ↑ | ↑ | ↑ | ↔ | ↑ | ↑ | ↑ | |
| 4. Exploit Kits | ↑ | ↔ | ↑ | ↑ | ↑ | ↑ | ↑ | |
| 5. Botnets | ↔ | ↑ | ↑ | ↑ | ↑ | | | |
| 6. Physical Damage/Theft/Loss | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ |
| 7. Identity Theft/Fraud | ↑ | | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ |
| 8. Denial of Service | ↑ | ↑ | | | ↑ | | | ↑ |
| 9. Phishing | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ |
| 10. Spam | ↔ | | | ↑ | | | | ↑ |
| 11. Rogueware/Ransomware/Scareware | ↑ | | | | | | | |
| 12. Data Breaches | ↑ | | ↑ | | ↑ | ↑ | ↑ | ↑ |
| 13. Information Leakage | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ |
| 14. Targeted Attacks | ↑ | ↑ | | | | ↔ | ↑ | ↑ |
| 15. Watering Hole | ↑ | | | | | | | |

Legend: Trends: ↓ Declining, ↔ Stable, ↑ Increasing

L'entreprise se transforme et les risques augmentent: *maintenir le niveau de sécurité d'origine devient difficile*



▪ L'ouverture des entreprises engendre:

- De nouveaux accès géographiques
- De nouveaux réseaux d'accès
- Des accès à de nouvelles applications
- Une multiplicité des moyens et technologies d'accès

▪ Cette multiplicité des accès offre des opportunités pour :

- Les utilisateurs légitimes
- Les personnes malveillantes souhaitant s'introduire sur les SI / réseaux

▪ Objectifs:

- Simplifier les procédures d'accès
- Gérer les risques en maintenant un niveau de sécurité et de confiance fort

Vers une cybercriminalité facilitée...?



- Convergence vers le Tout IP : **fin de la confiance ?**
 - Fin des réseaux étanches, dépendances de technos dominées par d'autres pays/acteurs
- Nouveaux usages : **multiplication des accès légitimes ou malveillants**
 - Nomadisme, accès aux mêmes ressources via plusieurs médias
 - Accès permanent : ADSL, sans contact, bluetooth, Wifi, Wimax, UMTS, 4G...
- Echanges dématérialisés : **argent, biens de valeurs sur le réseau**
 - Services à valeur ajoutée: E-commerce/banking, E-monnaie, musique, videos, images
 - Stockage réseau/distant de données sensibles ou stratégiques, cloud, big data
- Connaissances des techniques d'attaques : **fraudeurs en position de force**
 - Explosion du nb de « hackers » maîtrisant les attaques sophistiquées
 - Multiplication des outils d'attaques sur le web, et sources d'informations sur les failles



Et une sécurisation difficile...

- Utilisateurs « humains » non sensibilisés donc vulnérables...
 - Démunis face à la complexité technique, phishing, virus, vol de données.
- Sentiment d'anonymat et d'impunité pour les attaquants
 - Législations difficilement applicables, entr'aide internationale variable
 - Volumétrie des infos à traiter, nombres d'infractions à sanctionner



- Une double hétérogénéité rend la sécurisation technique complexe
 - Hétérogénéité des réseaux/services : niveaux de sécurité internes variables
 - Hétérogénéité des solutions complémentaires de sécurisation du marché

Login : toto
Password : X&t*\$K7u



- En entreprise: la sécurité souvent vue comme un coût ou frein à l'innovation
 - Solutions souvent locales, dépendant des métiers, contextes, réglementations
 - Rarement une approche globale de gestion des risques de sécurité
 - Comment concilier sécurité, ergonomie, et développement des usages ?

Sécurité des SI et des réseaux : nécessité d'une approche globale par la gestion des risques



Système d'information :

Sécuriser les informations et les supports d'information



- les données et applications
 - annuaires d'identité, serveurs de fichiers, bases de données, big data..
 - en local, lors de leur transfert, hébergées à distance hors de l'entreprise, cloud..
 - applications logicielles, web et intranet, portails, applications et processus métiers...
- les infrastructures
 - les réseaux locaux, internes, inter-sites et les passerelles d'interconnexions
 - les postes de travail et supports amovibles
 - les serveurs applicatifs, et les systèmes industriels
 - les équipements bureautiques (ex: imprimantes, mopieurs...)
- les interfaces et réseaux d'accès sans fils
 - réseaux, terminaux et applications mobiles : 2G, 3G, 4G, 5G
 - Wifi, sans contact, capteurs et objets connectés...
 - accès distants, solutions VPN, solutions de mobilité
- les outils de communications
 - téléphonie, messagerie,
 - outils collaboratif audio, vidéo, partage de documents,
- les acteurs
 - employés, fournisseurs, partenaires, clients...

De nombreux défauts sont à l'origine des failles

- Défauts dans la conception des protocoles et des applications
 - Aspects sécurité non traités ou partiellement
- Défauts dans l'implémentation logicielle de protocoles et applications
 - Aspects sécurité non traités ou partiellement
 - Erreurs de programmation: Estimation entre 5 à 15 erreurs sur 1000 lignes de code.
 - Appli de jeu mobile : 100.000 lignes de codes, World of WarCraft : 6 M
 - Facebook : 60 M ; Windows Vista: 50 M ; Mac OS X Tiger : + 80 M
- Défauts dans la configuration des systèmes et des réseaux
 - L'autorisation de certaines actions peut permettre des actions illicites
 - La combinaison

Difficultés de prise en compte de la sécurité: *Compréhension insuffisante des enjeux...*



Un problème d'éducation sur la valeur de l'information

- L'information a une valeur importante pour l'entreprise, pour les concurrents, pour les États « guerre de l'information ».

Un problème de formation

- Des dirigeants qui n'ont pas tous une culture sécurité ;
- Des mobilités vers le poste de RSSI, sans formation complémentaire adéquate :
 - personnel issu de la technique : administrateur réseau, système...
 - personnel issu de la qualité : responsable qualité...

Un coût lié à la sécurité qui rebute en période difficile :

- Authentification forte : achats de jetons/carte à puce
- Plan de secours : acheter en double certains équipements ;
- Personnel : former aux bonnes pratiques en sécurité...

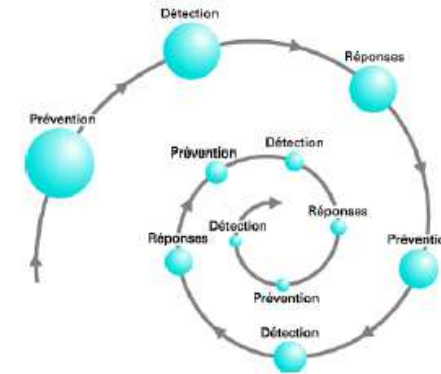
Difficultés de prise en compte de la sécurité:

Implication nécessaire de la direction...



- Rien ne peut se faire sans l'aval de l'exécutif.
 - Le chef d'entreprise doit être conscient des enjeux de sécurité. La PSSI est une réflexion stratégique : elle permet de prévoir l'avenir de l'organisation ;
 - Être **proactif** plutôt que réactif. Anticiper les problèmes et incidents
 - **Prendre le temps** de comprendre, ne pas être absorbé que par ses marchés, ses clients, ses concurrents, son relationnel...
- La sécurité va au-delà de la technique.
 - La sécurité ne doit pas rester un domaine d'experts
 - L'humain joue un rôle central. La sécurité est l'affaire de tous.
 - Ce n'est pas qu'une contrainte coûteuse, c'est aussi un investissement.
- La dynamique sécurité vient de la direction : montrer l'exemple !
 - Définir la politique de sécurité de l'organisation pour répondre aux enjeux
 - Responsabiliser : en désignant un responsable de la coordination, qui distribuera les tâches au sein des équipes ;
 - Réagir en cas d'attaque avérée : mettre des ressources à disposition, permettre l'expertise juridique et porter plainte ;
 - Impliquer le personnel, le sensibiliser, le former.

Méthode de protection classique (1/2)



Prévention...

- Définir une architecture de sécurité
 - Référencement des besoins, analyse du risque
 - Implanter des équipements de sécurité : Firewalls, IDS.
 - Sécuriser les équipements (plate-formes, serveurs, postes de travail) et les applications.
 - Évaluation des nouveaux produits (logiciels et matériels) introduits dans le système
 - N'autoriser que ce qui est strictement nécessaire.

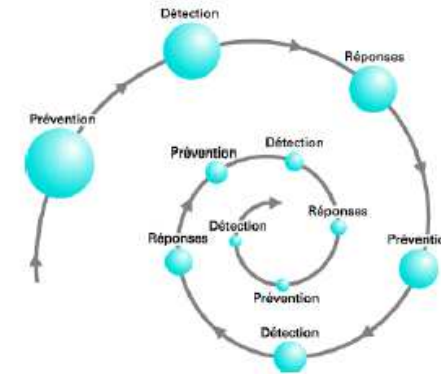
- Faire vivre la sécurité : politique de sécurité
 - Administrer les équipements de manière régulière
 - Ce ne sont pas des boîtes noires délaissées une fois qu'elles fonctionnent.
 - Avoir du personnel compétent et ayant du temps dédié à la sécurité !
 - Appliquer les mises à jour de sécurité sur les systèmes et les applications
 - Veille sur les attaques, failles de produits, vulnérabilités...

- Audit de sécurité
 - Évaluer régulièrement le niveau de sécurité d'un système et son adéquation avec les risques
 - Pour y déceler les failles et ajuster les configurations

Méthode de protection classique (2/2)

Détection...

- Se rendre compte que le bien à été endommagé
- Utilisation de systèmes de détection d'intrusion (IDS)
- Analyse des logs



Réaction...

- Avoir une politique de réaction sur incident (piratage, sauvegarde...) et l'appliquer.
 - Journaliser les attaques réalisées en cas de poursuites.
- Prendre en compte les avis des CERT : Computer Emergency Response Team
 - Cert-IST, Cert-Renater, Cert-A
 - centralisation des demandes d'assistance suite aux incidents de sécurité
 - traitement des alertes et réaction aux attaques informatiques
 - établissement et maintenance d'une base de donnée des vulnérabilités
 - prévention par diffusion d'informations sur les précautions à prendre
- Réponse aux attaques
 - Changer l'architecture de sécurité et mettre à jour la politique de sécurité
 - Réparer les dommages

Impact du Big Data sur la sécurité

Pourquoi le Big Data bouscule la sécurité... ?

- Le *big data* bouscule la sécurité en raison de son ambivalence :
 - Il se révèle à la fois profitable, pour les opérateurs économiques
 - connaître les besoins des consommateurs,
 - aider les acteurs de la santé, prédire des épidémies...
 - aider les gouvernements dans la lutte contre le terrorisme...
 - et potentiellement néfaste, principalement pour les individus.
 - risques de détournement de fichiers et d'interconnexion des données ayant trait à la vie privée des individus.
 - surveillance des personnes...
- Le Big Data bouscule les responsabilités de chaque acteur:
 - besoin de sécurisation du stockage d'une telle masse de données, dans un data center privé, dans le cloud, ou hébergé par un prestataire,...
 - prévoir dans les contrats des clauses de sécurité et des clauses précisant les modalités de la protection des données personnelles...

Une variété de situations critiques

- Les problèmes varient en fonction de :
 - l'origine des données (publiques, privées ou mixtes),
 - la loyauté de leur recueil,
 - la présence ou non, directe ou indirecte, de données personnelles
 - l'objectif poursuivi (recherche scientifique, avantage concurrentiel...)
 - la transparence ou l'opacité des buts poursuivis,
 - les infrastructures (publiques, privées ou mixtes) de stockage et de calculs mises en œuvre
 - et le caractère ouvert ou fermé des traitements algorithmiques.

- Les attaques sont donc multiples:
 - attaques informatiques classiques, atteintes aux infrastructures
 - usages détournés des puissances de calculs
 - falsification, clonage ou manipulation des données et de l'information
 - atteinte à la dignité ou vie privée des individus

Sécurité en environnement Big Data :

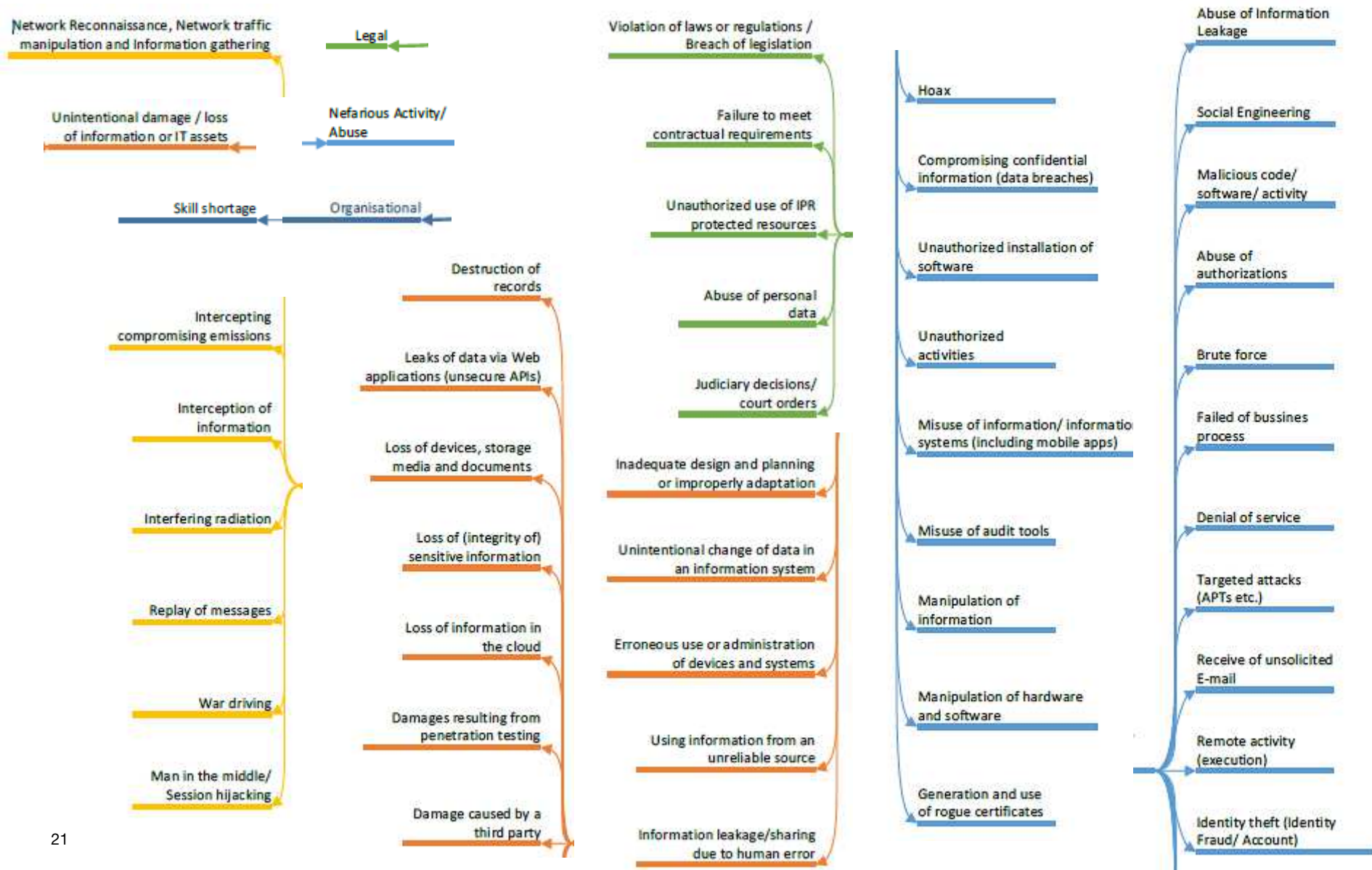
Quelques questions qui prennent encore plus de sens...

- Quelle confiance porter aux bases de données ?
- Comment protéger les sources, processus et décisions contre le vol et la corruption ?
- Comment est assurée la confidentialité des informations, quelles politiques et processus ont été mis en place vis à vis des employés ?
- Comment assurer le stockage sécurisé d'une telle masse de données ?
- Quelles sont vos actes qui peuvent être exploités par nos adversaires ?
- A quels types appartiennent les informations qui sont collectées, et quels sont les défis juridiques et réglementaires ?
- Quelles sont les responsabilités des acteurs lors de l'hébergement de données dans un data center privé, dans le Cloud, ou hébergé par un prestataire externe ?

Quelques challenges de sécurité pour les nouveaux environnements: Big Data, cloud...

- Insuffisance des protections périmétriques:
 - virtualisation et ubiquité, constitutives des architectures massives, augmentent les surfaces d'attaques et les délocalisent.
- Origine des données
- Contrôle d'accès légitime aux données
- Communication et échanges sécurisés entre acteurs
- Stockage sécurisé étanche entre bases de données
- Chiffrement cryptographique adapté
- Supervision des activités et des connexions
- Sécurité des supports d'information
- Politiques de sécurité globale et gestion des droits
- Privacy
- Transferts de responsabilité de sécurisation des données

Exemples de menaces en environnement Big Data



Contexte juridique

notamment liés à la protection des données à caractère personnel dans les Systèmes d'Information et les réseaux

une évolution/révolution en cours



Rappel sur les données personnelles numériques

Données à caractère personnel :

- *toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres.*
 - *Ex : nom, n° d'immatriculation, n° de téléphone, photographie, date de naissance, commune de résidence, empreinte digitale...*
- Données d'identification
 - Etat-civil, Coordonnées, Relations (familles, collègues, anciens camarades)
- Données comportementales
 - Centres d'intérêt : du général (cinéma, musique) au particulier (film, chanteur)
 - Mode de vie : habitudes (agenda), consommation...
 - Vie sociale : tendance politique, opinions, vie sentimentale...
 - Enregistrements visuels de vidéosurveillance permettant d'identifier directement ou indirectement une personne
- Données professionnelles
 - Secteur d'activité de l'entreprise
 - Fonction, nature du travail en cours
 - Orientation interne et politique industrielle de l'entreprise

Caractère identifiant des données

| Niveau de gravité | Description du caractère identifiant |
|-------------------|---|
| 1 – négligeable | Il semble quasiment impossible d'identifier les personnes à l'aide des données les concernant (ex. : identifier quelqu'un au sein d'une population en ne connaissant que son prénom) |
| 2 - limité | Il semble difficile d'identifier les personnes à l'aide des données les concernant, bien que |
| | cela soit possible dans certains cas (ex. : identifier quelqu'un au sein d'une population en connaissant son nom et son prénom) |
| 3 - important | Il semble relativement facile d'identifier les personnes à l'aide des données les concernant (ex. : identifier quelqu'un au sein d'une population en connaissant son nom, son prénom et sa date de naissance) |
| 4 - maximal | Il semble extrêmement facile d'identifier les personnes à l'aide des données les concernant (ex. : identifier quelqu'un au sein d'une population en connaissant son nom, son prénom, sa date de naissance et son adresse postale) |

Le droit des technologies de l'information et de la communication en France

- Un droit non codifié : des dizaines de codes en vigueur
- ... et difficile d'accès
 - Au carrefour des autres droits
 - En évolution constante et rapide
 - Issu de textes de toute nature /niveaux
 - Caractérisé par une forte construction jurisprudentielle*
- nécessitant un effort de veille juridique.



Quelques réglementations à prendre en compte

- RGPD (Règlement Général sur la Protection des Données) Date d'entrée en vigueur : 25 mai 2018
 1. • Ce nouveau règlement européen oblige les organisations à s'assurer du consentement explicite des individus quant à l'utilisation qui sera faite de leurs données. La transparence, la mise en œuvre d'alerte en cas de constatation d'une fuite de données, mais également la mise en place d'une structure interne
 2. dédiée à la protection des données sont notamment exigées.
- LPM (Loi de Programmation Militaire) Entrée en vigueur des mesures de cyber-sécurité: Juillet 2016
 1. • Cette réglementation concerne les entreprises classées « Opérateurs d'Importance Vitale » (OIV) qui sont tenues de renforcer leur niveau de sécurité (contrôles réguliers, détection des événements, alerte suite à un incident) sous peine de dispositions pénales.
- La directive européenne NIS (juillet 2016), déclinée en droit français en février 2018), pour les Opérateur de services essentiels.
- DSP 2 (Directive sur les services de paiement 2) Date d'entrée en vigueur : 13 janvier 2018

• Cette directive européenne définit les règles concernant les nouveaux acteurs sur le marché des paiements (FinTechs). Les services d'agrégation d'information (permettant une vision consolidée des comptes bancaires) ou les services d'initiation de paiement sont dorénavant encadrés et des mesures de sécurité exigées (sécurisation des API, authentification forte, etc.).

Passer des formalités préalables à une logique de responsabilité

- Loi Informatique et Libertés :
 - les obligations des organismes reposent en grande partie sur les formalités préalables (déclaration, autorisation),
- Règlement européen sur la protection des données
 - repose sur une logique de responsabilisation et de transparence.

Cette notion de responsabilité (accountability) se traduit notamment par :

- la prise en compte de la protection des données dès la conception d'un service ou d'un produit (Security by Design)
- la mise en place d'une organisation, de mesures et d'outils internes garantissant une protection optimale des personnes dont les données sont traitées.

Règlement Général sur la Protection des Données (GDPR)

- RÈGLEMENT (UE) 2016/679 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE
- Applicable depuis Mai 2018

- Table des matières : 99 articles
- *Préambule*: 31 pages
- *Chapitre I: Dispositions générales* (4 articles)
- *Chapitre II: Principes* (7 articles)
- *Chapitre III: Droits de la personne concernée* (12 articles)
- *Chapitre IV: Responsable du traitement et sous-traitant* (20 articles)
- *Chapitre V: Transfert vers des pays tiers* (7 articles)
- *Chapitre VI: Autorités de contrôle indépendantes* (9 articles)
- *Chapitre VII: Coopération et cohérence* (16 articles)
- *Chapitre VIII: Voies de recours, responsabilité et sanctions* (8 articles)
- *Chapitre IX: Dispositions relatives à des situations particulières* (7 articles)
- *Chapitre X: Actes délégués et actes d'exécution* (2 articles)
- *Chapitre XI: Dispositions finales* (7 articles)



RGPD : ce qui a changé en 2018...

| Droits des individus | | Obligations des entreprises | |
|---|---|---|---|
|  | Transparence et information lors de la collecte |  | Privacy by design default/security by design (voir page 12) |
|  | Accès aux données |  | Respect des principes de traitement des données personnelles |
|  | Rectification des données |  | Relations avec les sous-traitants |
|  | Effacement/ droit à l'oubli |  | Registre des traitements |
|  | Limitation du traitement |  | Sécurité du traitement |
|  | Portabilité des données |  | Détection des incidents liés à la sécurité |
|  | Opposition au traitement |  | Notification à la CNIL et communication à la personne concernée |
|  | Testament numérique |  | Analyse d'impact |
| | |  | Transfert des données à l'étranger |

Sous traitance

- **Sous-traitant :** « *Toute personne traitant des données à caractère personnel pour le compte du responsable du traitement* »
 - Agit sous l'autorité et sur instruction du responsable du traitement
 - « *Doit présenter des garanties suffisantes pour assurer la mise en œuvre des mesures de sécurité et de confidentialité* »
 - Le responsable du traitement n'est pas déchargé de « *son obligation de veiller au respect des mesures* »
 - Le contrat écrit doit prévoir l'indication des obligations du sous-traitant en matière de sécurité et de confidentialité
 - Le sous-traitant agit sous instruction du responsable

- **Cas particulier d'un prestataire de service de certification électronique**
 - peut traiter des données pour les seuls besoins de la délivrance et conservation des certificats liés aux signatures électroniques
 - mais uniquement si les données sont directement collectées auprès de la personne concernée aux fins en vue desquelles elles ont été recueillies

Cas du cloud et externalisation de données (1/2)

- **Responsable du traitement :**

- personne physique ou morale, autorité publique, service ou autre organisme qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement;

- **Sous-traitant :**

- personne physique ou morale, autorité publique, service ou autre organisme qui traite des données à caractère personnel pour le compte du responsable du traitement;

- **Profilage :**

- toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique;

- **Pseudonymisation :**

- traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable;

Confidentialité des données en cas d'externalisation (Cloud): *recommandations pratiques pour définir le partage des responsabilités*

- Avant tout recours à une prestation d'externalisation (cloud), il est nécessaire de:
 - Identifier clairement les données et les traitements qui passeront dans le cloud
 - Définir ses propres exigences de sécurité techniques et juridiques
 - Conduire une analyse de risques afin d'identifier les mesures de sécurité nécessaires
 - Identifier le type de cloud pertinent pour le traitement envisagé
 - Choisir un prestataire présentant des garanties suffisantes
 - Revoir la politique de sécurité interne
 - Surveiller les évolutions dans le temps.
- Ces 7 étapes préalables permettent de :
 - Déterminer la qualification juridique du prestataire:
 - *le prestataire est-il simple sous-traitant ou responsable conjoint de traitement ?*
 - Evaluer le niveau de protection assuré par le prestataire aux données traitées :
 - *le niveau de sécurité offert par le prestataire est-il supérieur ou égal à celui du client ?*
 - Rédiger des clauses pour définir clairement les responsabilités dans un contrat
- Recommandations CNIL:
 - Obligations de sécurité et périmètre de responsabilités, doivent être clairement établis.
 - La distribution des responsabilités doit être actée en amont, par ex dans un contrat
 - La responsabilité conjointe est retenue lorsque le client ne peut pas donner d'instructions à son prestataire ni contrôler l'effectivité des garanties de sécurité du prestataire.

Confidentialité des données en cas d'externalisation (Cloud): *définition contractuelle des responsabilités du client et du prestataire*

- Information transparente par le prestataire sur les modalités de traitement
 - Décrire les moyens de traitement à mettre en œuvre, et les mesures de sécurité appliquées
 - Informer le client, du recours à des sous-contractants et obtenir son accord préalable
 - Proposer des procédures simples de respect des droits (ex: accès des personnes à leurs données)
 - Mettre à disposition un système de remontées des plaintes et des failles de sécurité

- Information claire sur les lieux de stockage des données et sur les transferts indiquant:
 - Les pays hébergeant les serveurs du prestataire et de ses sous-contractants
 - L'existence d'une protection suffisante des données (si hébergées en-dehors de l'UE)
 - La possibilité de limiter le transfert de données vers des pays au niveau de protection suffisant
 - Toute requête provenant d'une autorité administrative ou judiciaire étrangère

- Garanties mises en œuvre par le prestataire :
 - Respect des durées de conservation des données, destruction ou restitution des données
 - Devoir de coopération avec les autorités
 - Possibilité pour le client de diligenter des audits

- Mesures de sécurité sur les données hébergées
 - Indication des obligations du prestataire en matière de sécurité des données
 - Politique de sécurité et mesures de sécurité retenues,
 - Procédure permettant l'audit chez le prestataire/sous-traitant
 - Réversibilité/portabilité des données, traçabilité, continuité de service, sauvegardes...

- Formalités à effectuer auprès de la CNIL par le client ou prestataire (si resp conjoint)

la Réglementation Européenne en résumé



□ 3 points principaux

- ✓ Droits de la personne
- ✓ Responsabilisation de l'entreprise
- ✓ Accountability (Obligation de rendre compte)

□ 2 types de données

- ✓ du Personnel du contractant : (Co-)Responsable de Traitement du Personnel
- ✓ des Clients finaux : (Co-)Responsable de Traitement ou Sous-traitant

□ les domaines d'actions

- ✓ Gestion du consentement
- ✓ Contrats (obligation de conseil du sous-traitant)
- ✓ Gestion des incidents de sécurité (remontée d'incidents)
- ✓ Droits de la personne
- ✓ Registre des traitements
- ✓ Sécurisation des données personnelles (organisationnelle et technique)

Anonymisation - pseudonymisation

L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et ce de manière irréversible. (CNIL)

Lorsque l'anonymisation est effective, le RGPD ne s'applique plus aux données ainsi anonymisées, celles-ci n'étant dès lors plus à caractère personnel.

Anonymisation

- Le processus d'anonymisation vise à éliminer toute possibilité de ré-identification : **il implique donc une nécessaire perte de qualité des données**. Leur exploitation future est ainsi limitée à certains types d'utilisation.
- Ces contraintes sont à prendre en compte dès le début du projet (privacy by design).
- Pour construire un processus d'anonymisation pertinent, il est conseillé de :
 - supprimer les éléments d'identification directe ainsi que les valeurs rares qui pourraient permettre une réidentification aisée des personnes (par exemple, la connaissance précise de l'âge des individus présents dans un jeu de données peut permettre dans certains cas de réidentifier très facilement les personnes centenaires) ;
 - distinguer les informations importantes des informations secondaires ou inutiles (c'est-à-dire supprimables) ;
 - définir la finesse idéale et acceptable pour chaque information conservée ;
 - définir les priorités (par exemple, est-il plus important de conserver une grande finesse sur telle information ou de conserver telle autre information ?)

Anonymisation traditionnelle : des précautions de base...

- Ne collecter les données qu'au niveau de finesse strictement nécessaire
- Répartir les données, dont le croisement risque de lever l'anonymat, dans des fichiers ou des systèmes informatiques distincts
- Dans la procédure de collecte ou de saisie des données, cloisonner la collecte et la saisie des données en les répartissant auprès de personnels ou organismes différents
- Ne pas fournir systématiquement un logiciel d'interrogation généraliste permettant de croiser n'importe quels critères
- Interdire certains croisements
- Dans les requêtes d'interrogation, ne pas fournir de résultat si le nombre est trop faible

Anonymisation et pseudonymisation

- **L'anonymisation** : technique appliquée aux données à caractère personnel afin d'empêcher leur identification de façon **irréversible**.

- **La pseudonymisation** : article 4 du RGPD
 - « traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable. »

- En l'absence d'irréversibilité, les techniques de pseudonymisation, ne permettent pas de se soustraire à la réglementation RGPD relative aux données personnelles.
 - Elles réduisent simplement la corrélation d'un ensemble de données avec l'identité originale d'une personne concernée,

Différences entre anonymisation et pseudonymisation

- La pseudonymisation est un traitement de données personnelles réalisé de manière à ce qu'on ne puisse plus attribuer les données relatives à une personne physique sans avoir recours à des informations supplémentaires.
 - En pratique la pseudonymisation consiste à remplacer les données directement identifiantes (nom, prénom, etc.) d'un jeu de données par des données indirectement identifiantes (alias, numéro dans un classement, etc.).
- La pseudonymisation permet ainsi de traiter les données d'individus sans pouvoir identifier ceux-ci de façon directe.
 - En pratique, il est toutefois bien souvent possible de retrouver l'identité de ceux-ci grâce à des données tierces.
 - C'est pourquoi des données pseudonymisées demeurent des données personnelles.
- La pseudonymisation est réversible, contrairement à l'anonymisation.

Techniques cryptographiques de pseudonymisation

Il s'agit de processus réversibles qui consistent à remplacer un attribut par un autre au sein d'un enregistrement.

- Système cryptographique à clé secrète : le détenteur de la clé peut ré-identifier chaque personne concernée en déchiffrant l'ensemble des données
- Fonction de hachage : la fonction de hachage renvoie un résultat de taille fixe, quelle que soit la taille de l'entrée.
- Fonction de hachage avec clé secrète : il s'agit d'une fonction de hachage qui utilise une clé secrète comme entrée supplémentaire

Intérêt de la pseudonymisation des données personnelles

- La pseudonymisation, assimilable à une anonymisation réversible et temporaire, remplace un attribut par une autre valeur.
 - Elle peut nécessiter une **clé (secret)** pour établir le lien entre les données pseudonymisées et l'identité des personnes auxquelles elles se rapportent
 - Cette clé doit être protégée contre toute utilisation illicite (confidentialité, accès...)
- La pseudonymisation d'une partie des données peut être utile lors de la **création de bases de test** pour les développeurs ou les responsables du support d'une entreprise.
 - Les équipes peuvent ainsi utiliser les données en conditions réelles, afin d'étudier des corrélations, sans pour autant avoir accès à des informations personnelles.
 - Le système utilisé va par exemple remplacer le nom d'un individu par un alias ou un numéro d'identification.

En résumé



- Techniques d'anonymisation
 - Plusieurs techniques, mais aucune n'est infaillible
 - Très difficile d'aboutir à des données véritablement anonymes
 - Applicabilité de la réglementation à toute donnée non anonyme

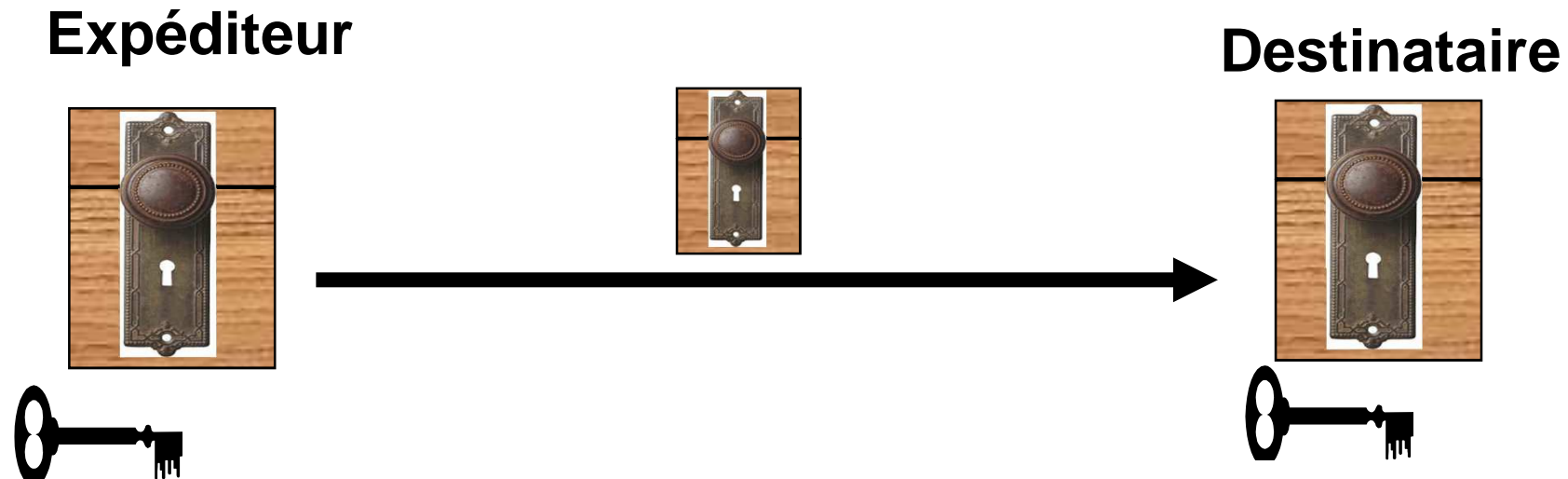
- La pseudonymisation,
 - Encouragée par le RGPD
 - Plus simple à mettre en œuvre que l'anonymisation
 - Permet de se conformer aux nouveaux principes de protection dès la conception (privacy by design) et de protection par défaut (privacy by default).

Introduction à la cryptographie

Où trouve-t-on de la cryptographie ?



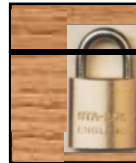
Serrurerie symétrique... (à clé secrète)



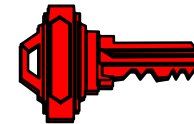
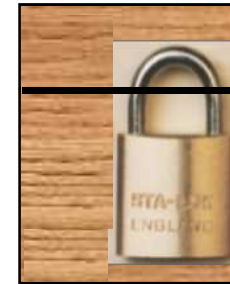
- La même clé est nécessaire pour fermer la boîte et pour l'ouvrir
- Seuls l'expéditeur et le destinataire peuvent ouvrir ou fermer la boîte
- Expéditeur et destinataire doivent maintenir leur clé secrète

Serrurerie asymétrique... (à clé publique)

Expéditeur

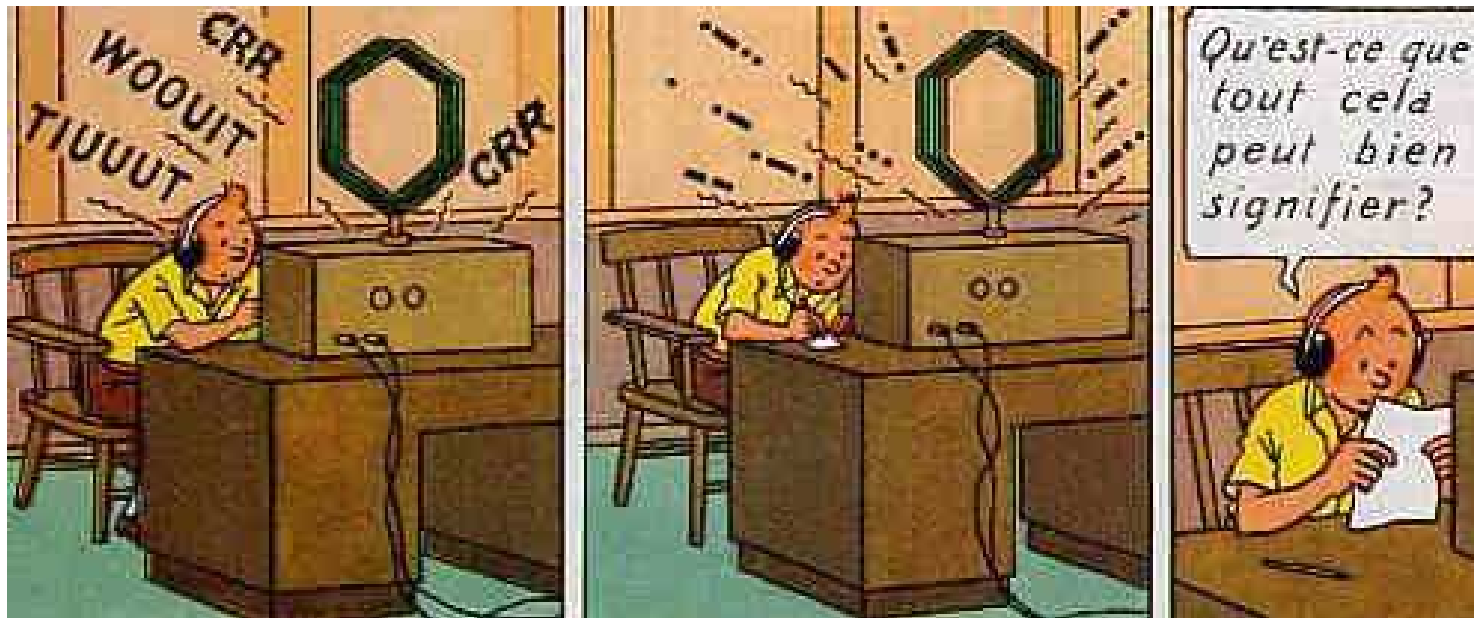


Destinataire



- Une fois que l'expéditeur a récupéré le cadenas (ouvert) appartenant au destinataire, il peut fermer le cadenas et lui envoyer une boîte fermée
- Aucune clé n'est nécessaire pour fermer la boîte.
- Seul le destinataire peut ouvrir la boîte
- Seul le destinataire doit maintenir sa clé privée

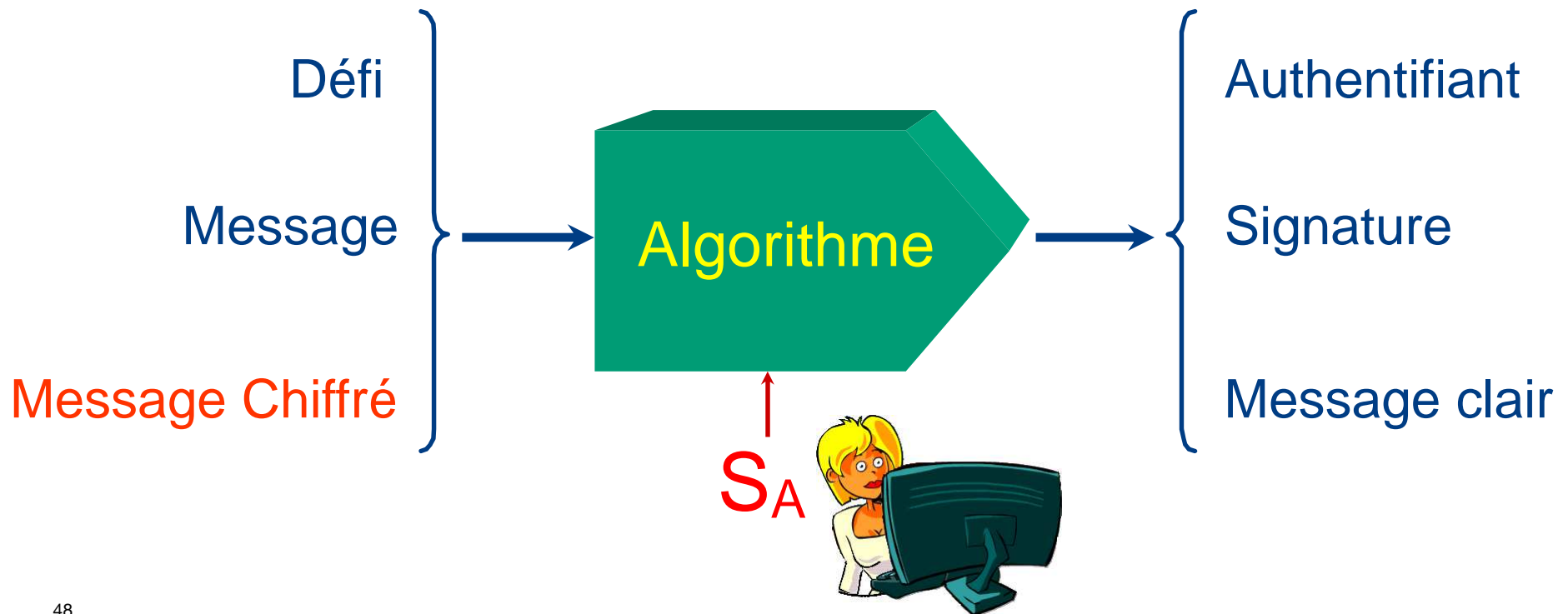
Cryptologie et cryptographie



- Cryptologie, étymologiquement la *science du secret* (*KRYPTOS* = *caché*), englobe :
 - la cryptographie — l'écriture secrète —
 - et la cryptanalyse — l'analyse de cette dernière.
 - s'attachant à protéger des messages
 - assurant confidentialité, authenticité et intégrité en s'aidant souvent de *secrets* ou clés.
- La cryptographie se scinde en deux parties nettement différenciées :
 - la cryptographie à clef secrète, encore appelée *symétrique* ou bien *classique* ;
 - la cryptographie à clef publique, dite également *asymétrique* ou *moderne*.

Algorithmes cryptographiques

Un algorithme cryptographique va permettre à Alice de s'authentifier, de signer ou de déchiffrer un message, en utilisant son secret S_A , mais *sans le révéler*.



Cryptographie symétrique (à clé secrète)

- Repose sur des techniques simples (XOR, permutations de bits, additions...)
- Nécessité d'un partage/échange préalable du secret entre chaque utilisateur
- Une clé S par couple ou groupe d'utilisateurs

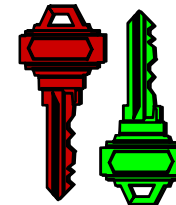


- La clé secrète S doit être maintenue secrète :
 - Chez chaque utilisateur
 - Lors de sa création et transfert/échange
- L'algorithme de déchiffrement est l'inverse de l'algorithme de chiffrement

49

Cryptographie asymétrique (à clé publique)

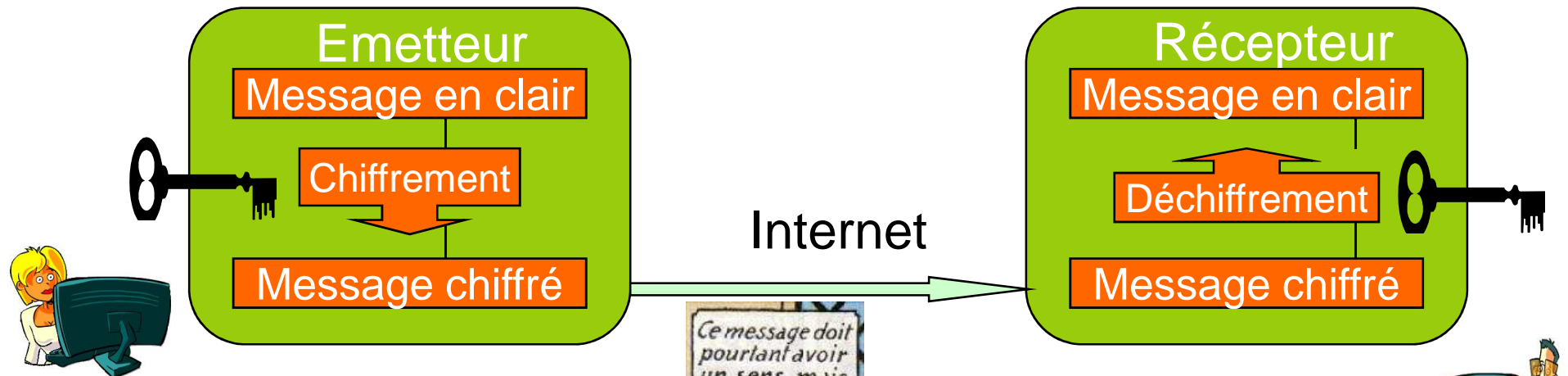
- Repose sur des technos mathématiques complexes
- Pas de partage au préalable d'un secret pour chaque couple d'utilisateurs
- Un couple de clés personnelles (S , P) pour chaque utilisateur



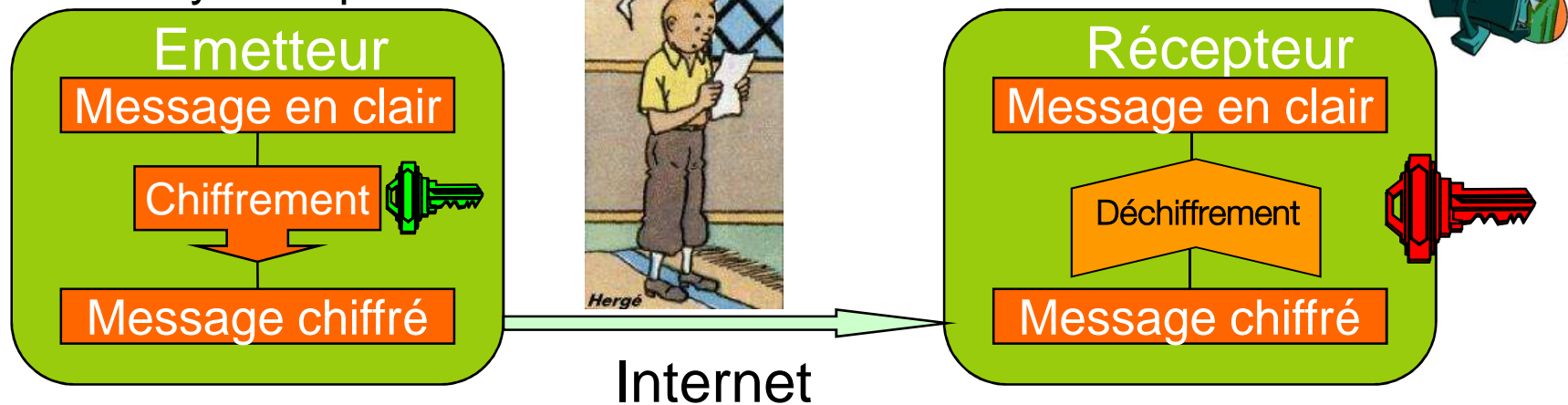
- P est dérivée de S par une fonction à sens unique
 - P est rendue publique
 - P définit la fonction de chiffrement utilisée par toute personne désirant établir une communication sécurisée avec celui qui l'a publiée
- S est gardée secrète/privée par son propriétaire
 - S définit la fonction déchiffrement de messages reçus

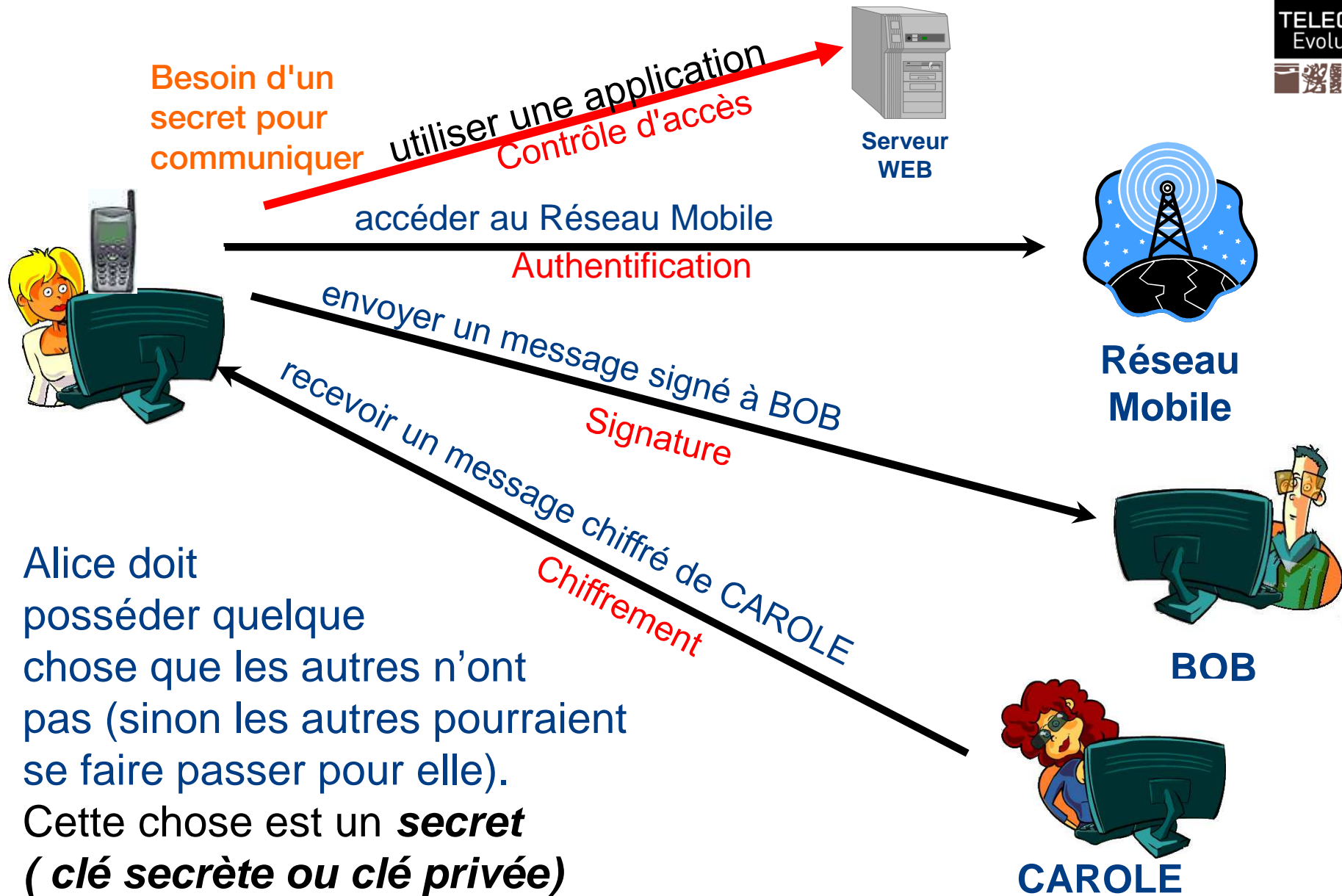
Exemple dans le cas du chiffrement : *confidentialité des données*

Chiffrement symétrique



Chiffrement asymétrique





Crypto symétrique

Alice partage son secret

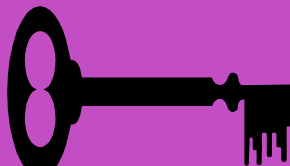


Le réseau Mobile
authentifie Alice



Réseau
Mobile

avec un
algorithme symétrique



Bob vérifie la
signature d'Alice



BOB

Carole chiffre un
message pour Alice



CAROLE

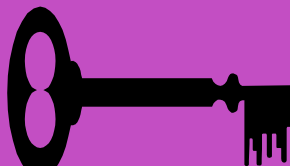
Crypto symétrique



Le réseau Mobile
authentifie Alice



avec un
algorithme symétrique



Bob vérifie la
signature d'Alice



BOB

Carole chiffre un
message pour Alice



CAROLE

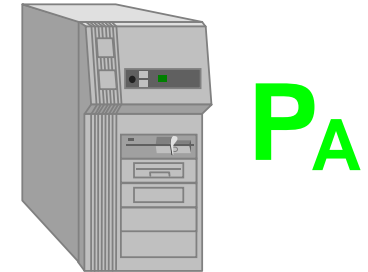
N'est pas adaptée aux systèmes « ouverts »

Crypto asymétrique

Alice donne sa clé publique

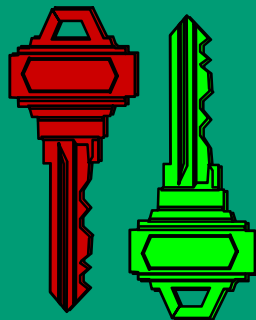


le serveur WEB
authentifie Alice



Serveur
WEB

avec un
algorithme asymétrique



Bob vérifie la
signature d'Alice



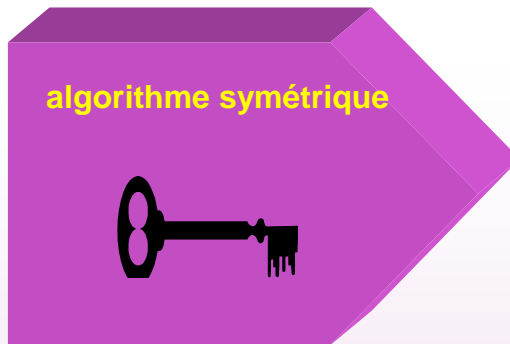
BOB

Carole chiffre un
message pour Alice



CAROLE

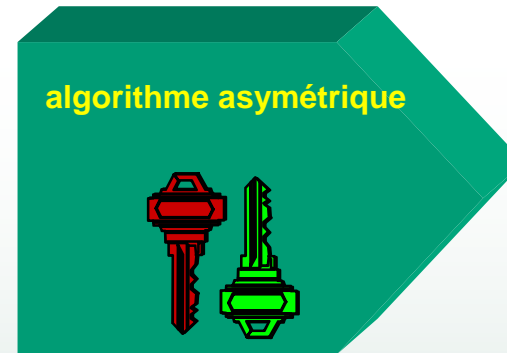
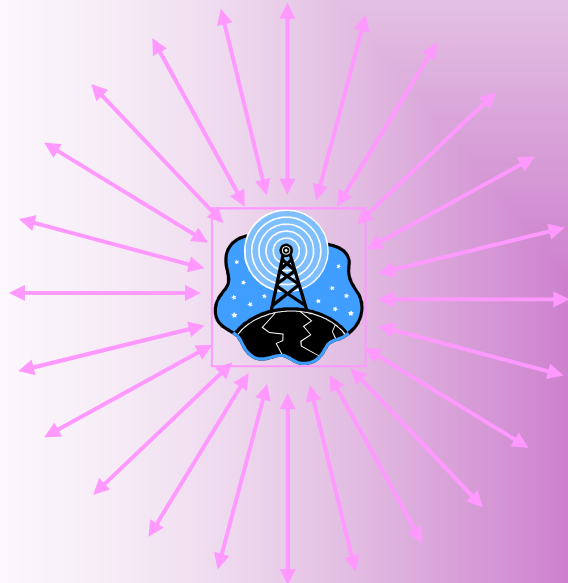
Algorithmes : symétrique - asymétrique



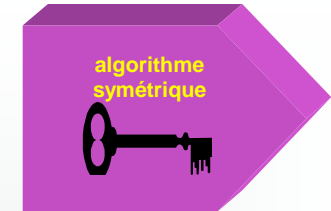
Systèmes « fermés »

Relation en « étoile »

Exemple : les réseaux mobiles



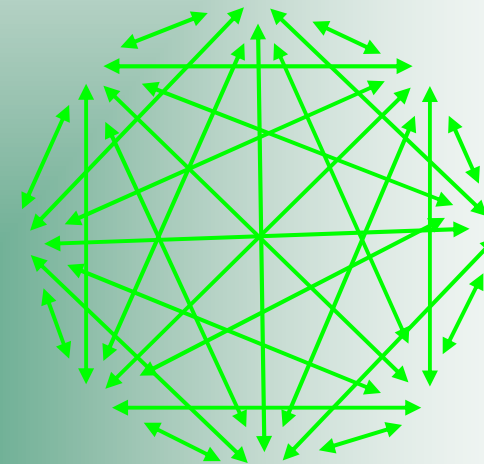
+



Systèmes « ouvert »

Relation « peer to peer »

Exemple : Internet



Cryptographie à clés secrètes (symétrique)

Algorithmes Symétriques ou à clé secrète

- Il ya bien longtemps !

- Chiffre de César
- Vigenère

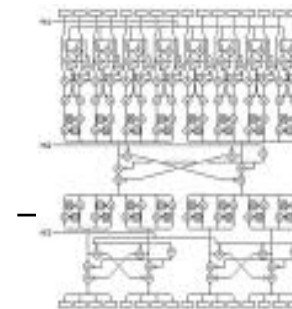
A→D
B→E
C→F
...
Z→C

- Électromécanique : enigma

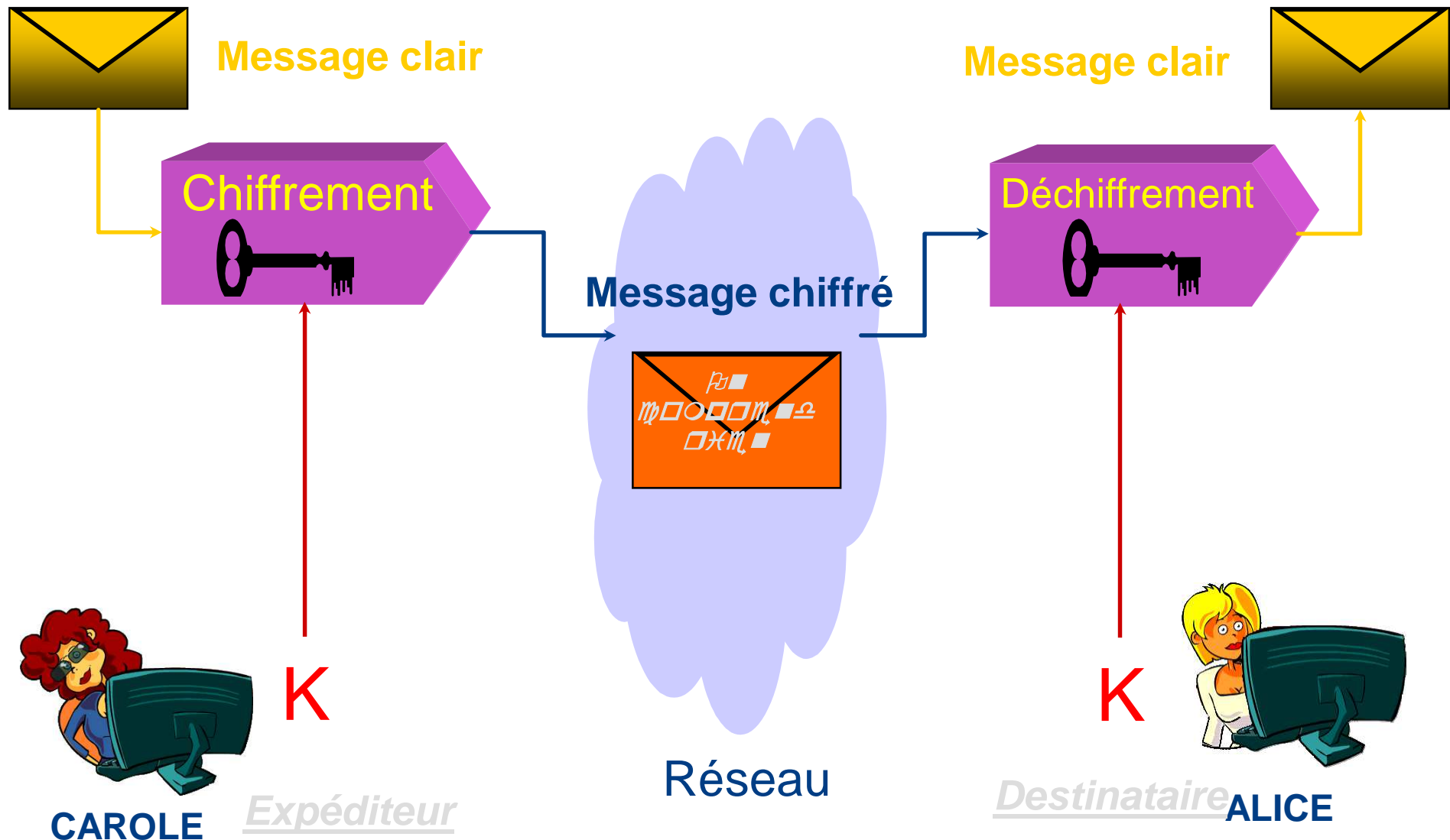


- Avec l'électronique et l'informatique :

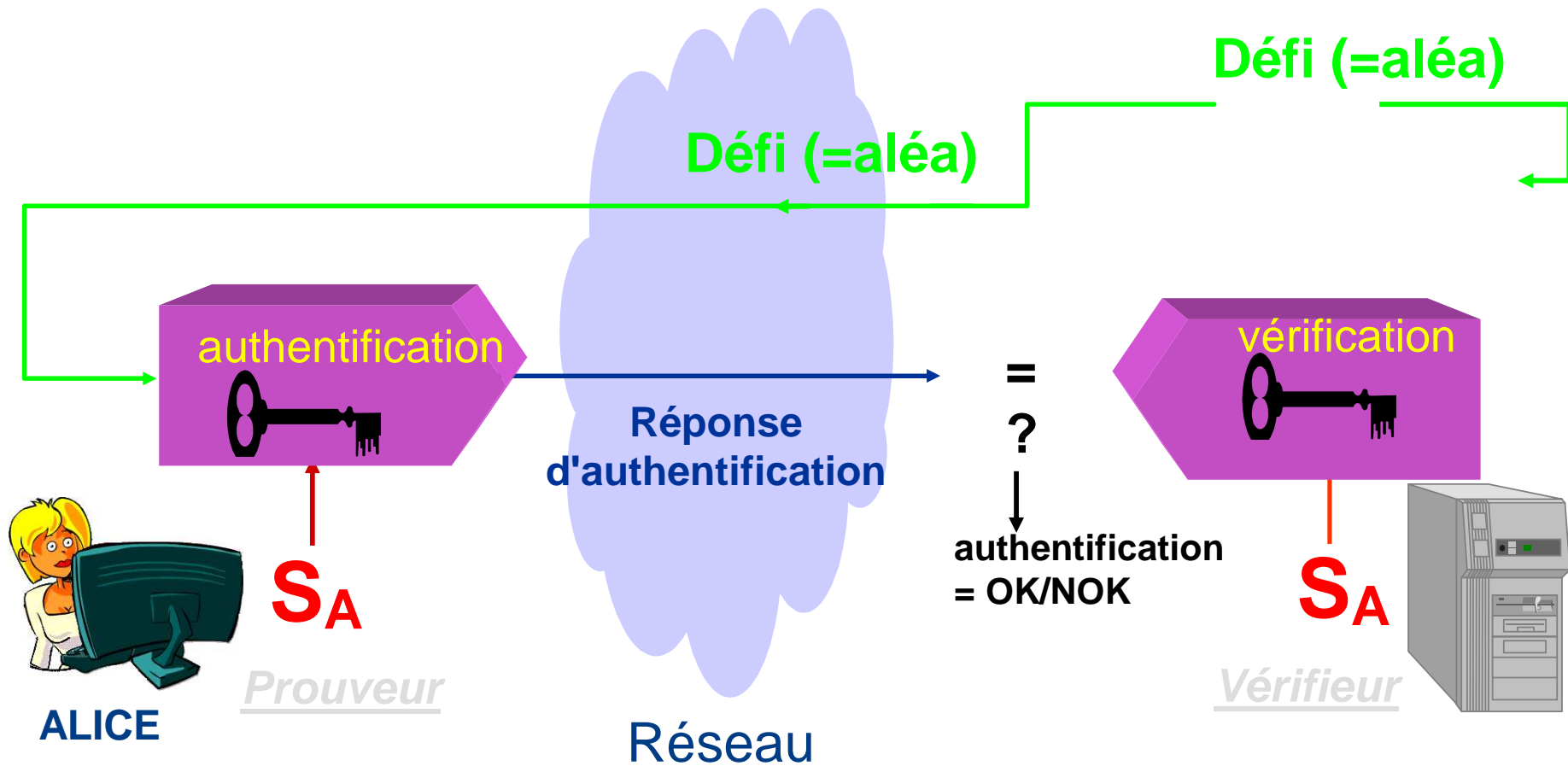
- Data Encryption Standart (DES) 1970
- IDEA, RC2, RC4 ... 1980 - 1990
- Advanced Encryption Standart : (AES – Rijndael) Oct 2000



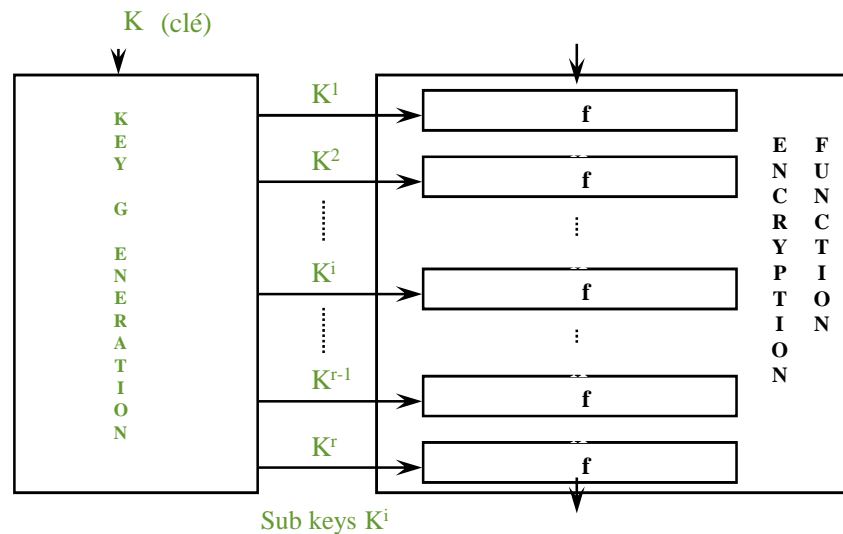
Chiffrement symétrique (à clé secrète)



L'authentification (forte) à clé secrète



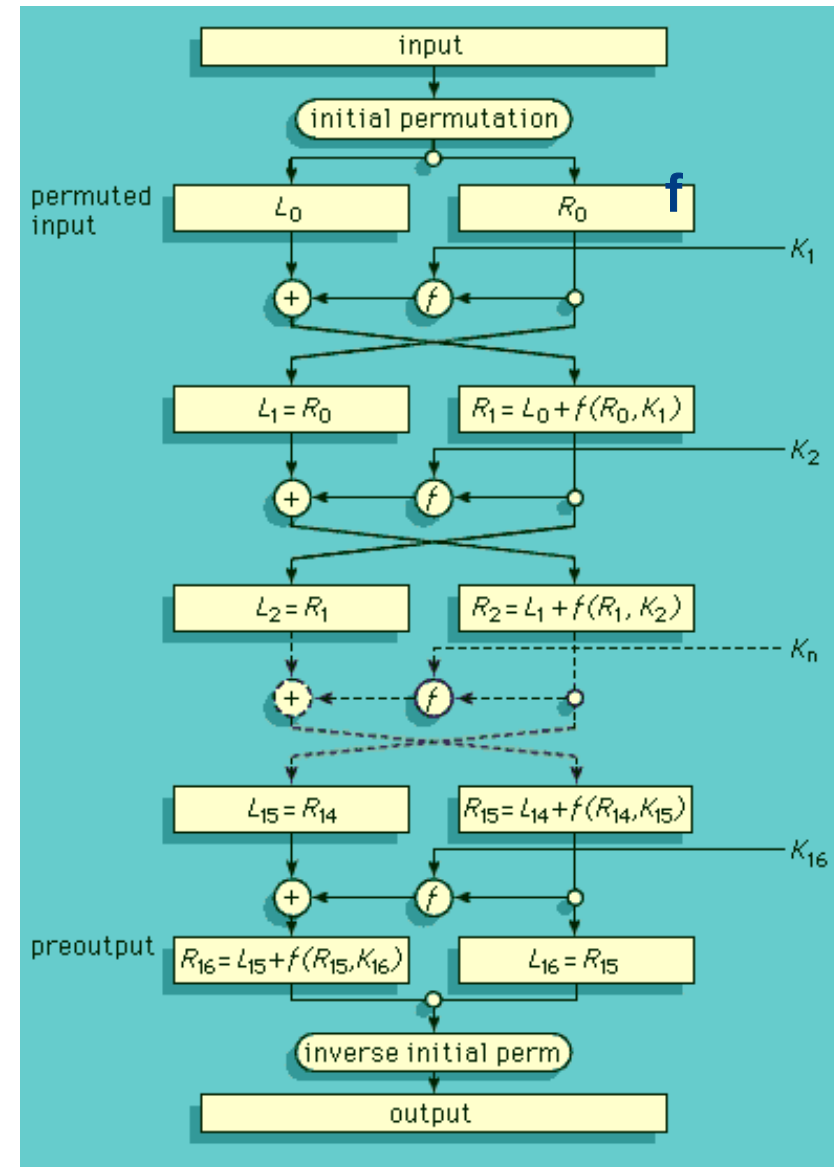
DES (Data Encryption Standard)



Horst Feistel



Don Coppersmith



Les attaques cryptographiques sur algos à clés secrètes (blocs ou flux)

Plus la clé est longue, plus le système est sûr :

- « 64 bits » signifie: 18446744073709551616 clés possibles de longueur 64 bits
- « 256 bits » signifie: 15792089237316195423570985008687907853269984665640564039457584007913129639936 clés

Toujours considérer que les algos sont connus de tous. Le seul élément secret doit être la clé secrète

- Recherche exhaustive : force brute (Balayage de l'ensemble des clefs)
 - Sans faiblesse cryptanalytique connue, une attaque nécessite en moyenne 2^{L-1} exécutions (L = longueur de clé)
 - La complexité dépend peu de l'algorithme considéré (facteur <10 en pratique). Ordres de grandeur :
 - loi de Moore : Puissance de calcul * 2 tous les ans (= gain de 1 bit/an pour le balayage des clés)
- Attaques par cryptanalyse : spécifiques aux algos, compliquées, compétence mathématique pointue
 - Étude des moyens de casser un algorithme crypto, typiquement déchiffrer un message, sans connaître la clef
 - Utilisation d'effets de bord des algo. pour décoder l'information sans connaître la clef
 - Chiffré connu : on ne connaît que des messages chiffrés.
 - Clair connu : on connaît des messages clairs et leurs chiffrés
 - Clair choisi : on peut choisir des messages clairs et obtenir leurs chiffrés
- Faiblesses liées aux implémentations
 - Principalement dans les générateurs de nombres aléatoires
 - Simplifications ou erreurs dans la programmation

Tailles de clés recommandées

Plus la clé est longue, plus le système est sûr? ...

- « 64 bits » : 18446744073709551616 clés possibles de 64 bits de longueur
- « 256 bits » : 15792089237316195423570985008687907853269984665640564039457584007913129639936 clés

| Key Size | Possible combinations |
|---------------|-----------------------|
| 1-bit | 2 |
| 2-bit | 4 |
| 4-bit | 16 |
| 8-bit | 256 |
| 16-bit | 65536 |
| 32-bit | 4.2×10^9 |
| 56-bit (DES) | 7.2×10^{16} |
| 64-bit | 1.8×10^{19} |
| 128-bit (AES) | 3.4×10^{38} |
| 192-bit (AES) | 6.2×10^{57} |
| 256-bit (AES) | 1.1×10^{77} |

| Key Size (bits) | Number of Alternative Keys | Time required at 10^6 Decryption/ μ s |
|-----------------|--------------------------------|---|
| 32 | $2^{32} = 4.3 \times 10^9$ | 2.15 milliseconds |
| 56 | $2^{56} = 7.2 \times 10^{16}$ | 10 hours |
| 128 | $2^{128} = 3.4 \times 10^{38}$ | 5.4×10^{18} years |
| 168 | $2^{168} = 3.7 \times 10^{50}$ | 5.9×10^{30} years |

- Taille de clés minimale recommandée en crypto symétrique : 128 bits
- DES cracker (EFF deep crack) en 1998 : 5 jours pour trouver une clé
- (90 milliards de clés par seconde, horloge = 2 M Hz



- Imaginons aujourd'hui... ☺

Conseils pour un chiffrement efficace de données

- Choisir un algorithme prouvé solide et adapté au besoin
 - Utiliser les algorithmes qui ont fait leurs preuves et les clés les plus longues possibles.
 - Un chiffrement matériel apporte un plus en accélérant la vitesse de chiffrement.
 - Les données stockées chiffrées restent logiquement isolées des autres.
- Protéger ses clés de chiffrement
 - Sécuriser le cycle de vie des clés de chiffrement: génération, transfert, stockage, renouvellement, effacement
 - Ne pas stocker les clés de chiffrement auprès des données (PIN sur carte bancaire)
- Ne jamais communiquer les clés
 - Chiffrer les données lorsqu'elles sortent de l'entreprise et les déchiffrer lorsqu'elles reviennent.
 - exemple pour mise de données dans le Cloud: VPN + stockage chiffré
- Attention à la manipulation directe des données chiffrées
 - Comment assurer le traitement des données chiffrées et leur sécurité ?
- S'assurer de la sécurité des utilisateurs et de leur environnement de travail
 - employés nomades, utilisation de réseau publics, télétravail
 - coexistence de données perso et pro sur le même équipement,
 - Cloud: qui peut avoir accès aux données , en local au stockage ou à distance ?

Limites des techniques à clé secrète

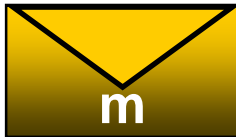
- Partage préalable des clés : Chaque couple doit préalablement s'échanger une clé commune de façon sûre, ou sur un canal autre que le canal à protéger.
- Une clé par couple d'utilisateur : Problème dans un réseau maillé lorsque N utilisateurs veulent dialoguer avec les N-1 autres (messagerie sécurisée) => $N.(N-1)/2$ clés différentes.
- Problème du maintien de la confidentialité des clés secrètes
- Pas de véritable signature vérifiable par un tiers : pas de non répudiation de signatures => pas d'arbitrage possible en cas de «disputes» entre deux interlocuteurs.

Fonction de condensation / hachage

- Une fonction de hachage « H » (on trouve aussi « hash function » ou « fonction de condensation ») transforme une entrée de données de taille variable « m » et en une sortie de taille inférieure et fixe : h
 - $h = H(m)$
- Une fonction de condensation doit remplir les conditions suivantes :
 - L'entrée peut être de dimension variable.
 - La sortie doit être de longueur fixe.
 - $H(m)$ doit être relativement facile à calculer.
 - $H(m)$ doit être une fonction à sens unique (non inversible).
 - $H(m)$ doit être sans collision (difficile de trouver deux messages ayant même haché).
- Une fonction de condensation permet de détecter toutes modification de m, fortuite ou malfaisante
 - un code détecteur d'erreur a des propriétés semblable mais ne permet pas de détecter des modification malfaisantes

Fonction de condensation

Ensemble des
messages

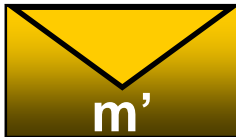
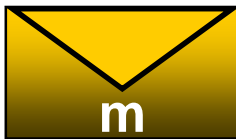


H

Ensemble des mots de p
bits (ex. : $p = 128$)

$H(m)$

Peut on signer $H(m)$ à la place de m ?

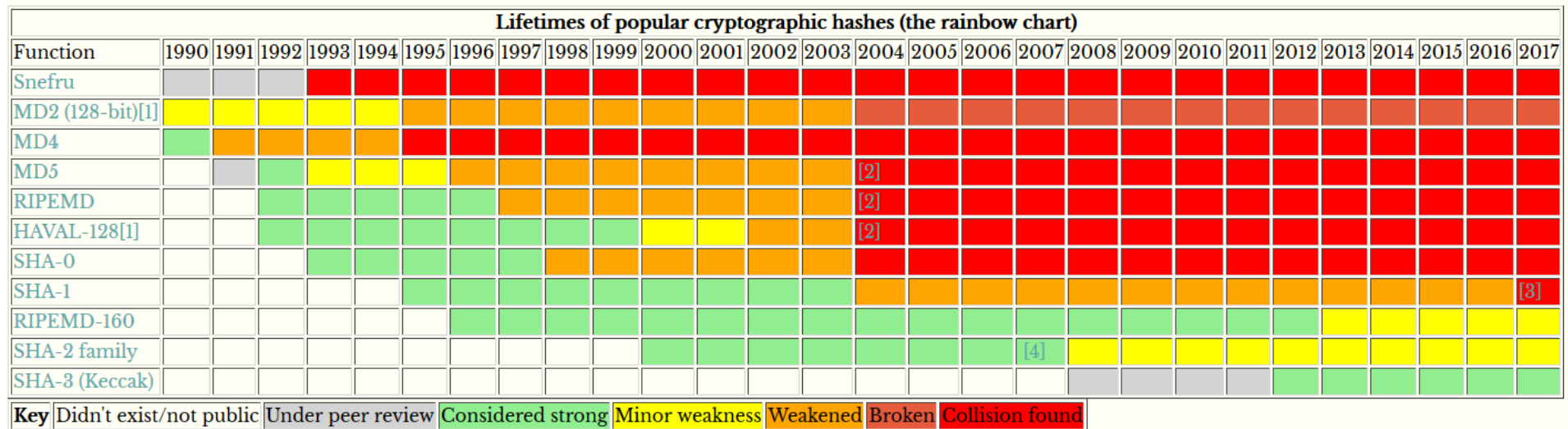


Collision !

$H(m) = H(m')$

**H est une fonction de condensation s'il est
« informatiquement » impossible de construire une collision**

Fonctions de hachage: comparaisons



Source : <http://valerieaurora.org/hash.html>

Source :
<https://www.enisa.europa.eu/publications/algorithms-key-size-and-parameters-report-2014>

| Primitive | Output Length | Classification | |
|------------|---------------|------------------|--------|
| | | Legacy | Future |
| SHA-2 | 256, 384, 512 | ✓ | ✓ |
| SHA3 | 256,384,512 | ✓ | ✓ |
| Whirlpool | 512 | ✓ | ✓ |
| SHA3 | 224 | ✓ | ✗ |
| SHA-2 | 224 | ✓ | ✗ |
| RIPEMD-160 | 160 | ✓ | ✗ |
| SHA-1 | 160 | ✓ ^[1] | ✗ |
| MD-5 | 128 | ✗ | ✗ |
| RIPEMD-128 | 128 | ✗ | ✗ |

Table 3.3: Hash Function Summary

Anonymisation / pseudonymisation des données par fonction de hachage

- calculer une valeur numérique à partir des données directement ou indirectement nominatives « textuelles » d'un individu
- cette valeur est ensuite substituée aux données à partir desquelles elle a été calculée.
- Le caractère irréversible de l'anonymisation
 - taux très faible des collisions
 - bonnes performances informatiques des algorithmes de hachage

Cryptographie asymétrique (à clé publique)

Algorithmes Asymétriques ou à clé publique

- 1976 : Diffie – Hellman

- Échange de clés






- 1977 : RSA (Rivest Shamir Adelman)

- Signature électronique
 - Chiffrement (lent et lourd)



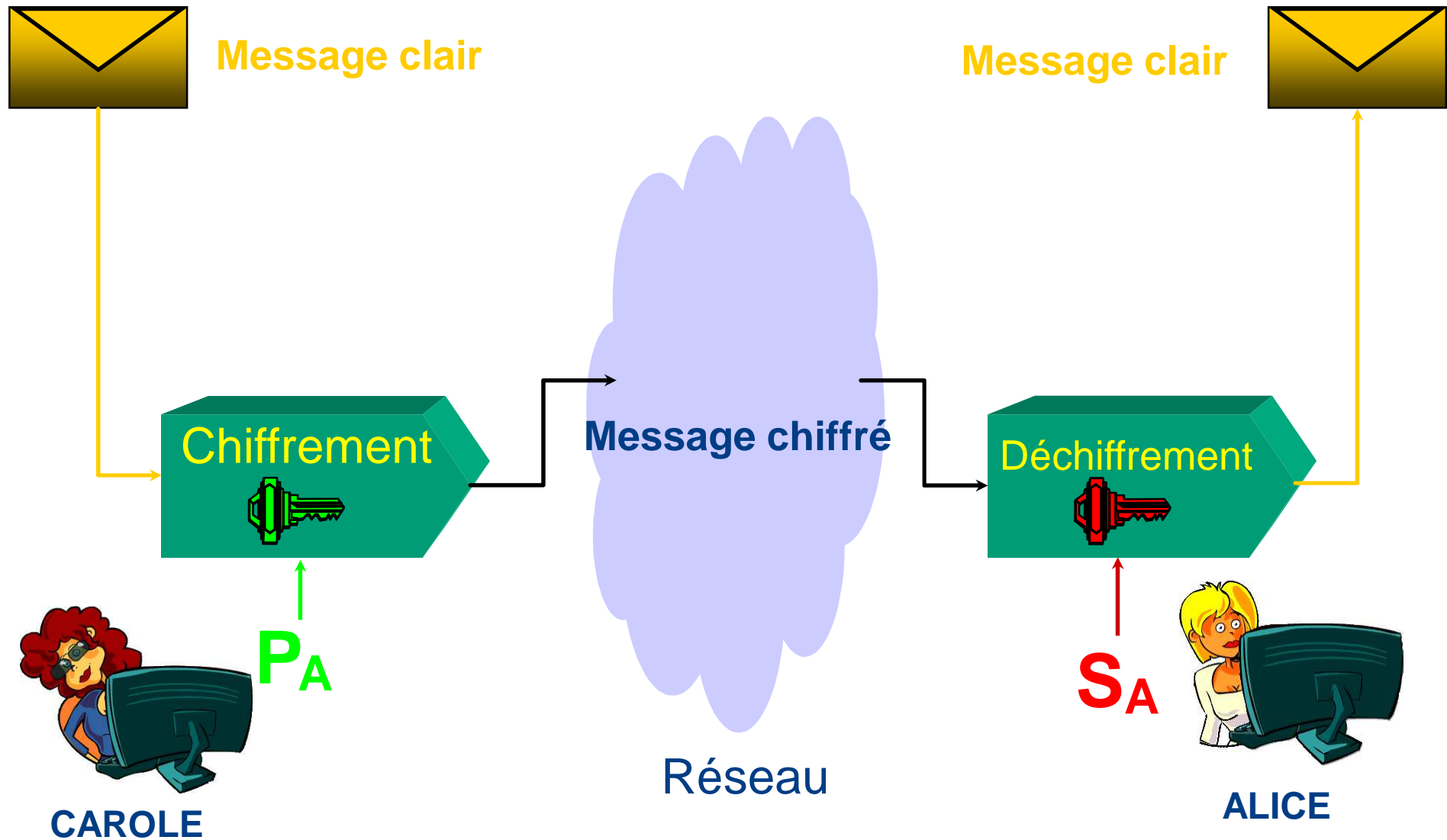
- Protocoles zéro-knowledge
- Certificat numérique et infrastructure à clé publique
- Aujourd'hui :
 - Signature aveugle, signature de groupe...
 - Anonymat révocable
 - Vote, enchère, concours anonyme, jeux en ligne

Cryptographie à clé publique

- Repose sur des techniques mathématiques complexes
- Pas de partage au préalable d'un secret pour chaque couple d'utilisateurs
- Un couple de clés personnelles (S , P) pour chaque utilisateur 
- P est dérivée de S par une fonction à sens unique
- P est rendue publique et définit la fonction de chiffrement utilisée par toute personne désirant établir une communication sécurisée avec celui qui l'a publiée 
- S est gardée secrète/privée par son propriétaire et définit la fonction de déchiffrement des messages reçus 

Dès qu'un utilisateur a choisi S et P, et publié P, toute autre personne peut lui envoyer des messages confidentiels...

Chiffrement (asymétrique)



RSA (1977)

- Pour générer ses clés, Alice
 - Tire au sort p et q premiers [> 512 bits chacun]
 - Choisit e premier avec (p-1)(q-1)
 - Calcule $n = pq$ [> 1024 bits]
 - Calcule d tel que $ed = 1 \mod ppcm((p-1), (q-1))$

Clé publique d'Alice : $P_A = (n, e)$

Clé privée d'Alice : $S_A = d$
- Pour chiffrer un message pour Alice ou vérifier une signature d'Alice :

Bob utilise la fonction publique P_A d'Alice

$X \rightarrow Y = P_A(X) = X^e \mod n$
- Pour déchiffrer un message de Bob ou signer un message :

Alice utilise sa fonction privée S_A

$Y \rightarrow S_A(Y) = Y^d \mod n = (X^e)^d \mod n = X^{ed} \mod n = X^1 \mod n = X$

On a la propriété Fondamentale :

$$\mathcal{P} \circ \mathcal{S} = \mathcal{S} \circ \mathcal{P} = Id$$

Car :

$$(X^e)^d = (X^d)^e = X^{ed} = X$$

Grâce au théorème de Fermat (ou d'Euler) :

$$X^{(p-1)(q-1)} = 1 \mod pq$$

Exemple numérique

$$p = 3, q = 11$$

$$n = pq = 3 \times 11 = 33$$

$$(p-1)(q-1) = 2 \times 10 = 20$$

$$e = 3; d = 7; ed \equiv 1 \pmod{(p-1)(q-1)}$$

$$\text{Car : } 3 \times 7 \equiv 1 \pmod{20}$$

On a alors : $\forall X \in \{0, 1, \dots, 32\}$

Si $Y = X^3 \pmod{33}$ alors $Y^7 = X \pmod{33}$ et

Si $Y = X^7 \pmod{33}$ alors $Y^3 = X \pmod{33}$

Exemple : $X = 29$

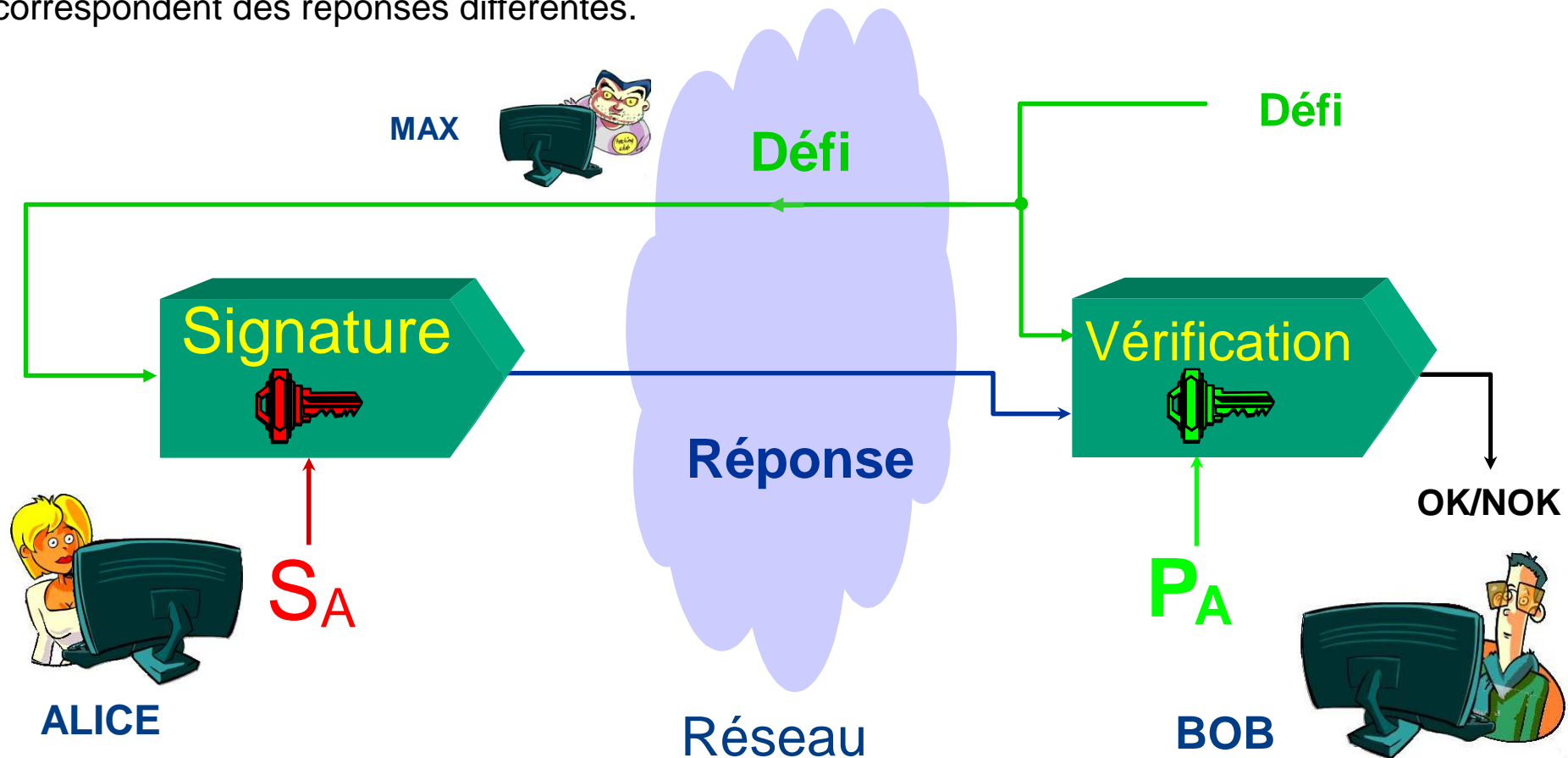
$$Y = X^e = 29^3 = 24389 = 2 + 24387 = 2 + (739 \times 33) \equiv 2 \pmod{33}$$

$$Y^d = 2^7 = 128 = 29 + 99 \equiv 29 \pmod{33} = X$$

L'authentification (forte) à clé publique

Seul le possesseur du secret S_A (Alice) peut construire la réponse au défi de Bob.

En observant les transactions, Max ne peut répondre à aucun défi car à des défis différents correspondent des réponses différentes.



Exemple avec RSA:

Alice génère Réponse = $(\text{Défi})^d \bmod n$

Exemple avec RSA:

Bob vérifie $(\text{Réponse})^e \bmod n \stackrel{?}{=} \text{Défi}$

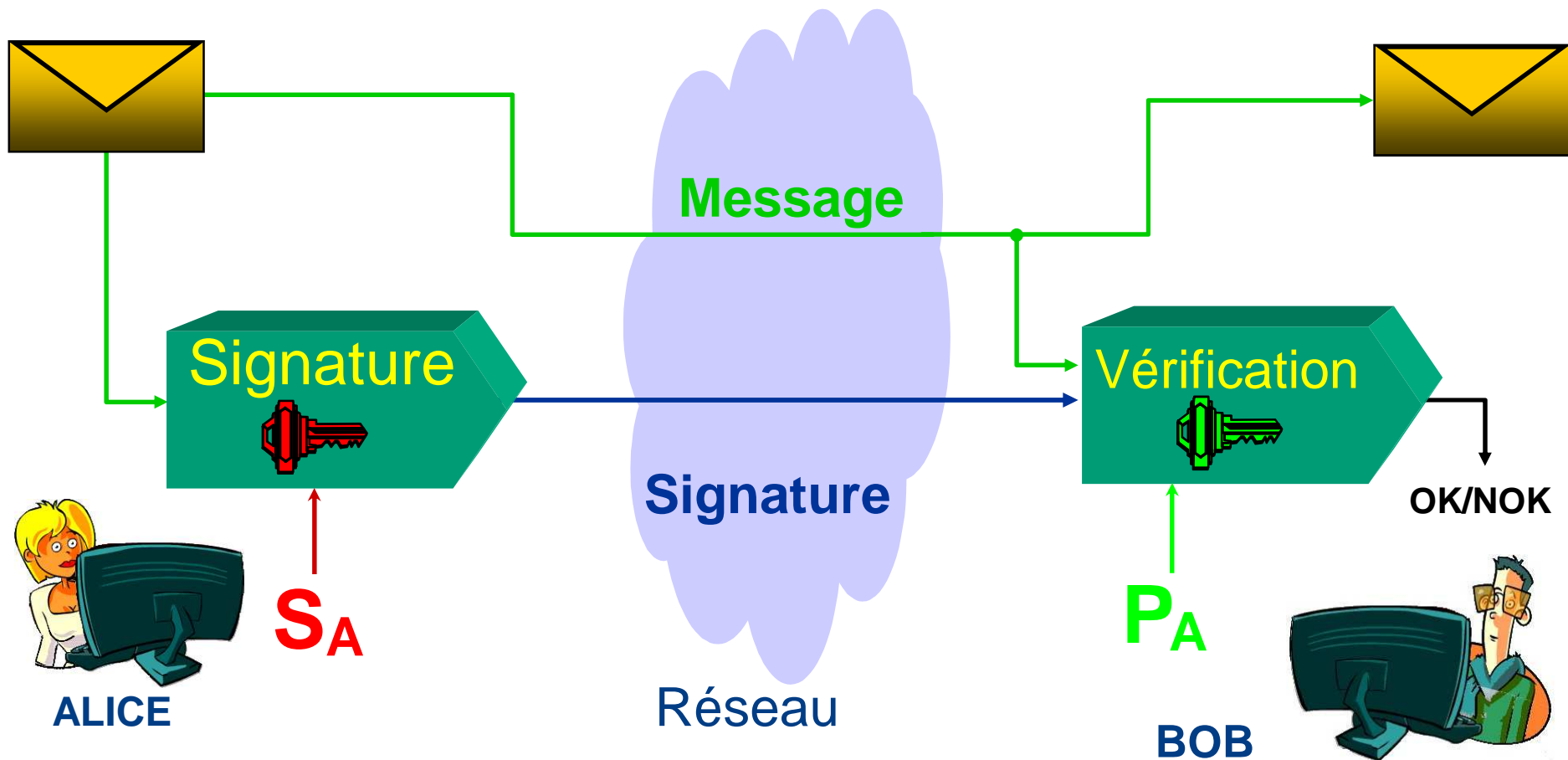
Signature électronique

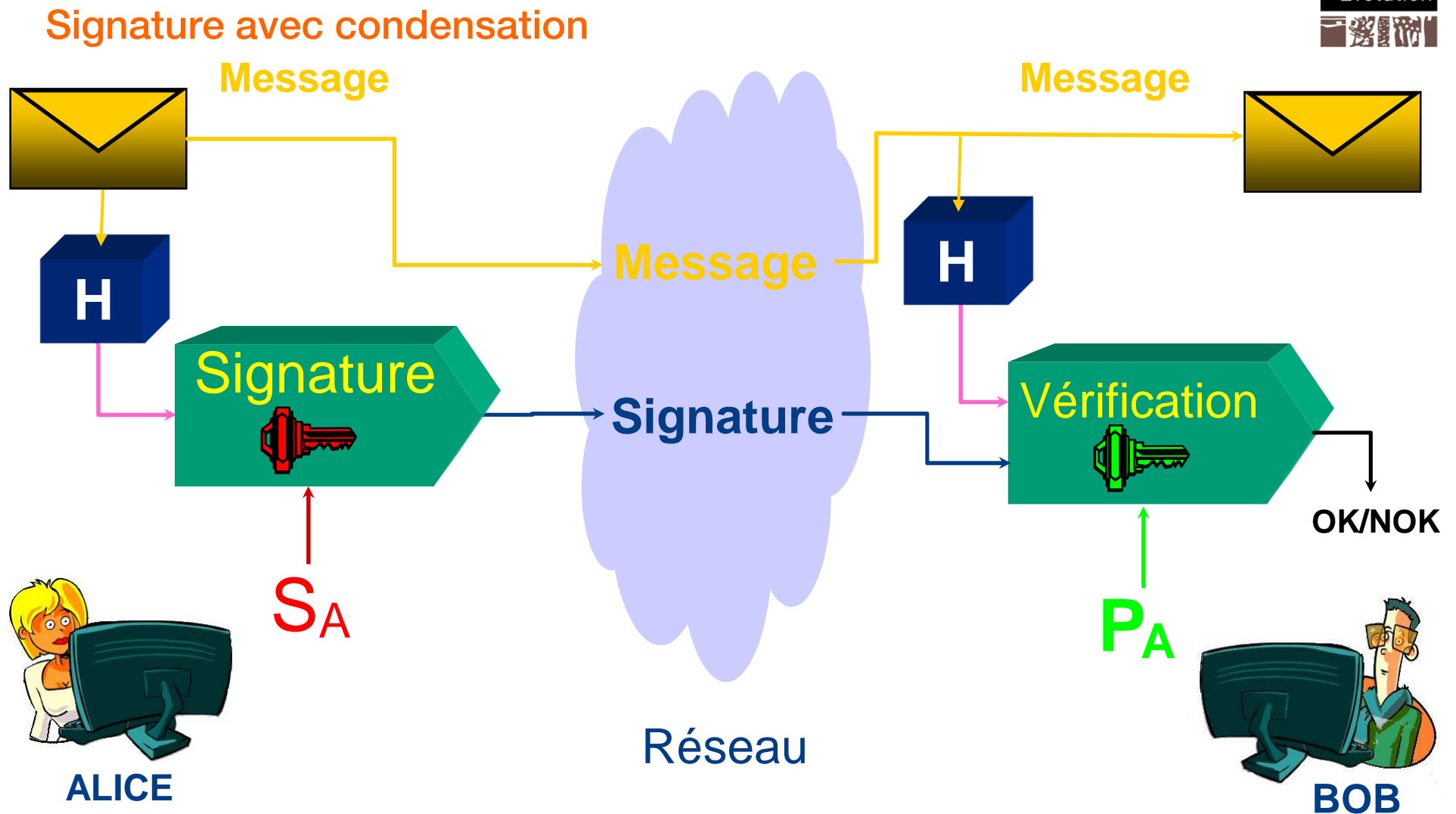
- La signature électronique remplit deux rôles majeurs, qui tendent à établir les conditions de la confiance dans les échanges numériques et donc à rendre possible la dématérialisation :
 - la signature électronique d'un document (un contrat par exemple) confère à celui-ci une **valeur juridique équivalente à celle d'un document papier signé de manière manuscrite**, en marquant l'engagement de la personne qui a apposé la signature ;
- des fonctions connexes à la signature électronique (cachet, horodatage...) servent à offrir des conditions de **sécurité technique en garantissant sa** provenance, son intégrité, ou encore la date de sa réalisation.
- une signature électronique apporte :
 - la garantie de l'intégrité du document ;
 - un lien certain avec l'identité du signataire.

Signature

Seul le possesseur du secret S_A (Alice) peut construire la signature du message.

Tout le monde peut vérifier : il suffit de posséder la clé publique P_A





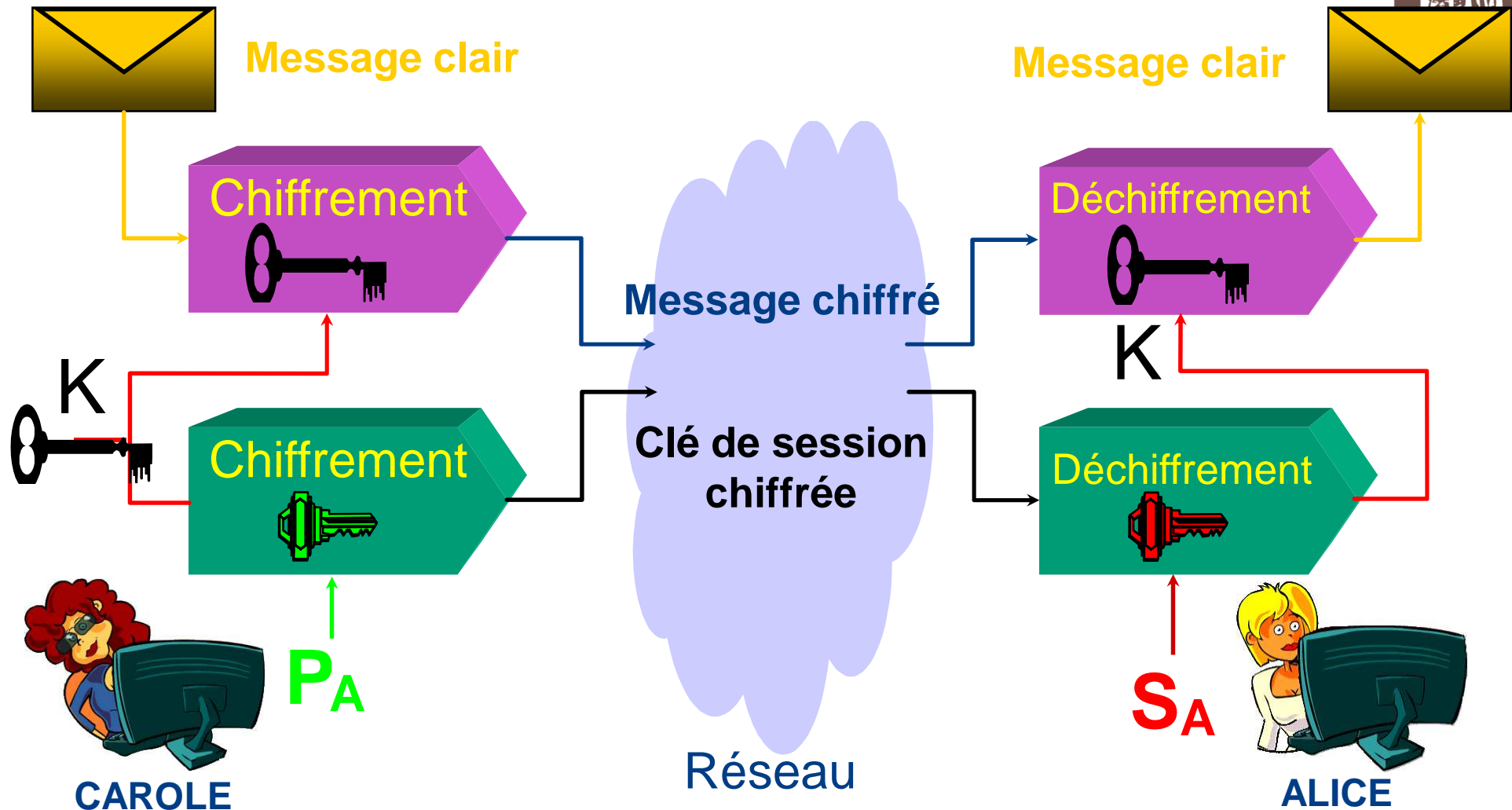
Exemple avec RSA:

Alice génère **Signature** = $(H(\text{Message}))^d \bmod n$

Exemple avec RSA:

Bob vérifie **(Signature)^e mod n** = $? = H(\text{Message})$

Chiffrement avec échange de clé de session



Exemple avec RSA:
Clé de session chiffrée = $K^e \bmod n$

M. chiffré = $C_K(\text{M. clair})$

Exemple avec RSA:
K = $(\text{Clé de session chiffrée})^d \bmod n$

M. clair = $D_K(\text{M. chiffré})$

À retenir

- Chaque individu a :
 - sa propre clé privée qu'il garde secrète
 - sa propre clé publique dérivée qu'il diffuse à ses interlocuteurs
 - ce couple de clés est différent d'un utilisateur à l'autre
- Je chiffre avec la clé publique de mon interlocuteur
 - Afin que mon interlocuteur soit le seul à pouvoir déchiffrer mon message avec sa propre clé privée qu'il est le seul à connaître
- Je signe avec ma propre clé privée, que je suis le seul à connaître
 - Afin que tout les interlocuteurs puissent vérifier ma signature à l'aide de ma clé publique que je leur ai donnée

Avantages/inconvénients des systèmes à clés publiques



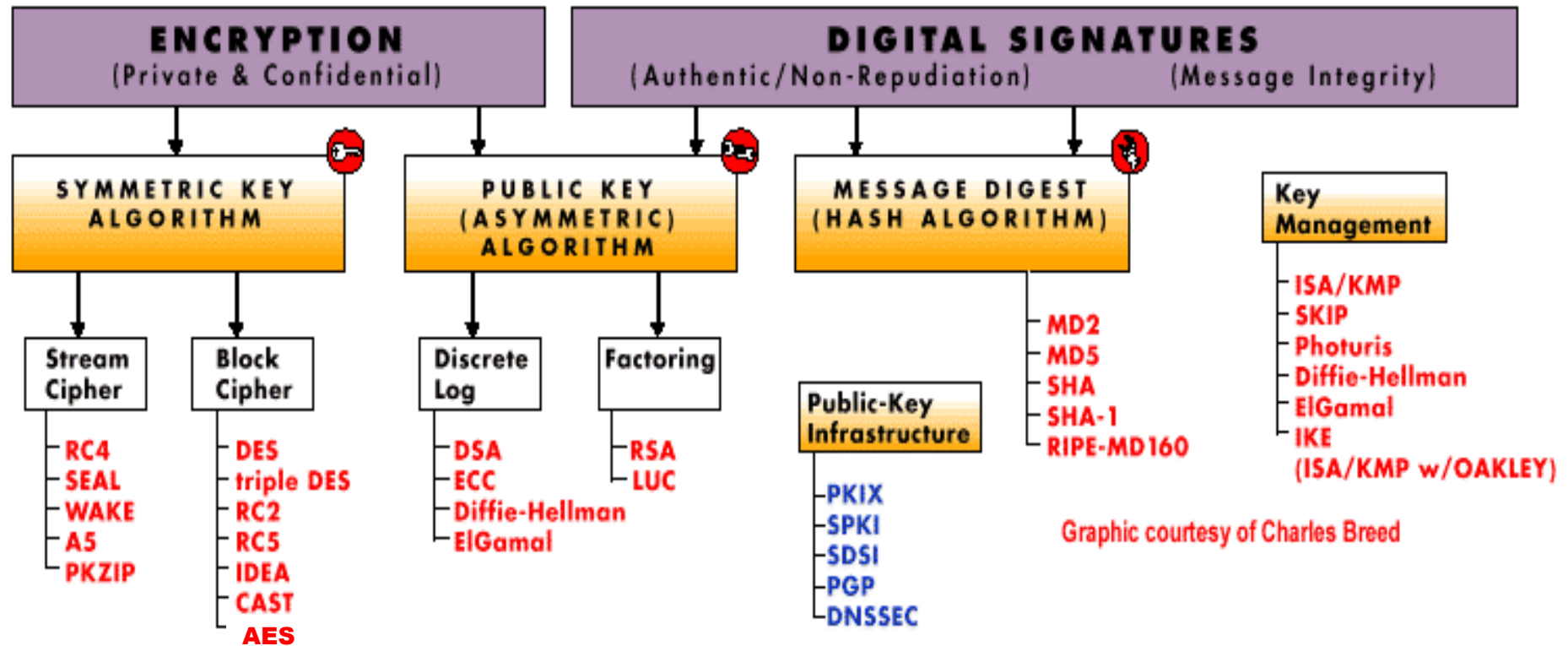
- Avantages

- Absence de distribution préalable de secret commun
- Simplification (???) de la gestion des clés dans les applications décentralisées
- Pas de conservation de la gestion des clés utilisateurs dans des centres d'authentification ou de distribution sécurisés
- La vérification d'une signature n'exige pas de secret
- Quiconque peut envoyer un message chiffré (confidentiel) à un autre utilisateur
- Non répudiation de signature = signature vérifiable par un tiers (ex : un juge)

- Inconvénients

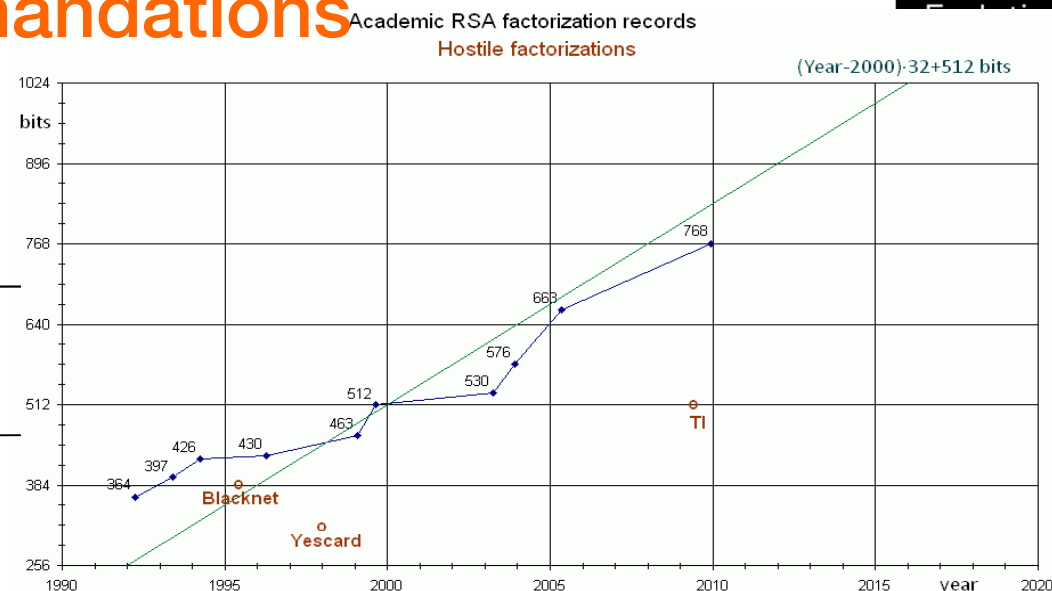
- Peu d'algorithmes disponibles et maths complexes
- Coût, lenteur, puissance de calcul nécessaire >> ceux des algos à clés secrètes
- Pas de systèmes efficaces de chiffrement de données
- Longueur des clés (milliers de bits)
- Besoin d'un annuaire et/ou certification de clés publiques (PKI)

Exemples d'algorithmes cryptographiques



Tailles de clés: recommandations

| Symmetric | DH or RSA |
|-----------|-----------|
| 56 | 512 |
| 80 | 1024 |
| 112 | 2048 |
| 128 | 3072 |
| 192 | 7680 |
| 256 | 15360 |



| Method | Date | Symmetric | Asymmetric | Discrete Logarithm Key | Discrete Logarithm Group | Elliptic Curve | Hash |
|-------------------------|-------------|-----------|------------|------------------------|--------------------------|----------------|------|
| [1] Lenstra / Verheul ? | 2020 | 86 | 1881 1472 | 151 | 1881 | 161 | 171 |
| [2] Lenstra Updated ? | 2020 | 82 | 1387 1568 | 163 | 1387 | 163 | 163 |
| [3] ECRYPT II | 2016 - 2020 | 96 | 1776 | 192 | 1776 | 192 | 192 |
| [4] NIST | 2011 - 2030 | 112 | 2048 | 224 | 2048 | 224 | 224 |
| [5] ANSSI | 2010 - 2020 | 100 | 2048 | 200 | 2048 | 200 | 200 |

Utilisation des schémas à clés publiques



- Chiffrement de données de taille limitée (clés, secret, Identifiant,)
- Gestion (échange) de clés
- Authentification forte
- Signature électronique, non répudiation
- Signature (intégrité) de fichiers
- Idéal dans un environnement ouvert

- Exemples d'applications
 - Internet (Authentification forte, transfert de clés de chiffrement...)
 - Commerce électronique et messagerie
 - Téléprocédures..
 - Tous services nécessitant le recours à une signature non répudiable

Comment lier « identité » et « clé »

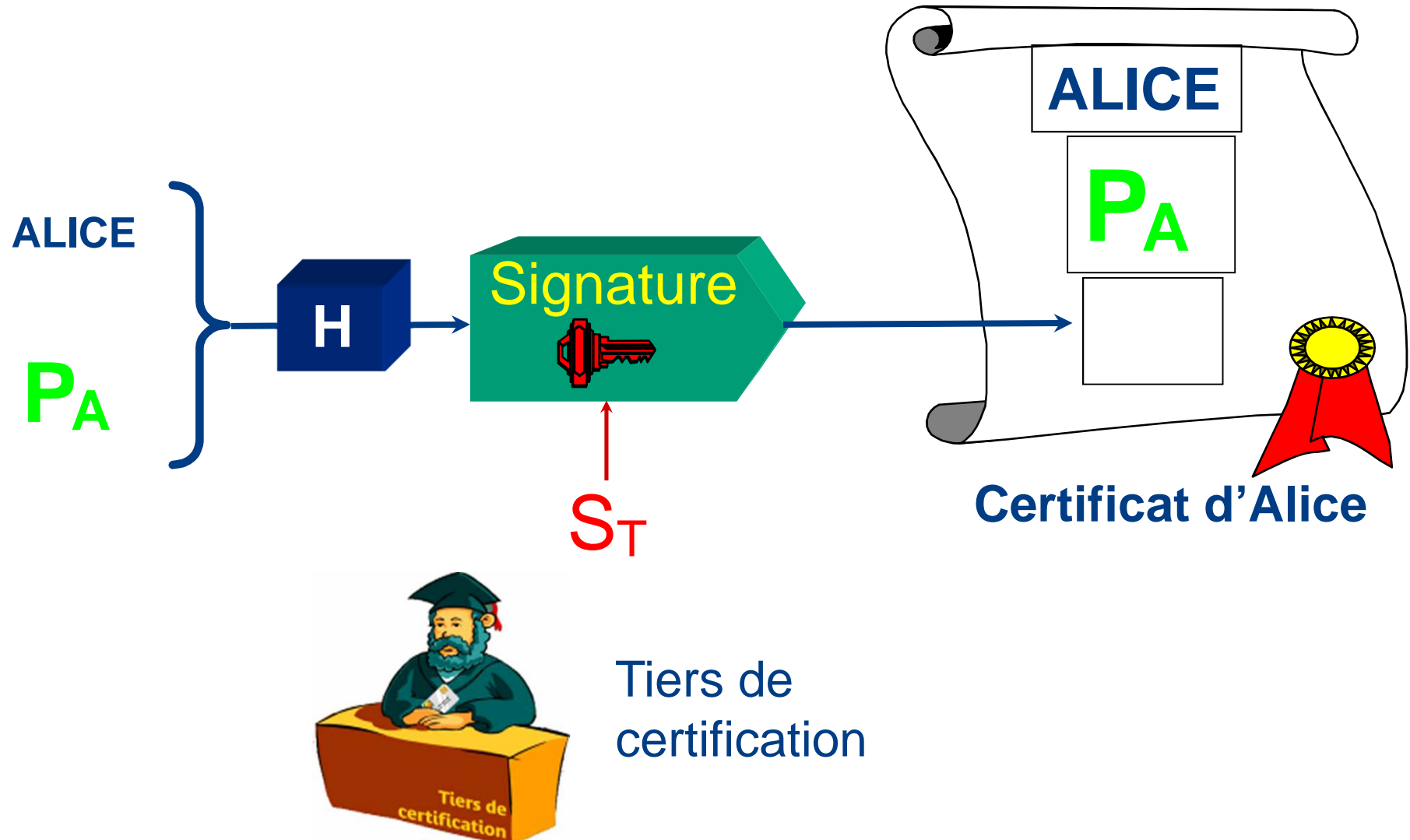


- L'identité est une chaîne de caractère qui représente concrètement l'individu ou l'objet qui réalise la transaction sécurisée.
- Une clé (secrète, privée ou publique) est une chaîne de bits qui n'a à priori aucun lien avec l'individu ou l'objet qui réalise la transaction. Pourtant c'est cette clé qui permet de sécuriser la transaction.
- **Il est donc indispensable de lier « identité » et « clé ».**
- Deux cas à distinguer :
 - Cas symétrique : on utilise la « ***diversification*** »
 - Cas asymétrique : on utilise la « ***certification*** ».

NB : dans le cas asymétrique il existe une technique particulière : les « ***schémas basés sur l'identité*** » où on a :
clé publique \equiv identité
et où la liaison clé - identité devient ainsi inutile.

Comment lier identité et clé : le cas asymétrique

« la certification »





Infrastructure à clés publiques (PKI)

- Définition d'une infrastructure à clés publiques:
 - L'ensemble des moyens matériels, logiciels et organisationnels permettant la gestion des clés et des certificats
- Le certificat est la pièce d'identité électronique du client, partenaire, fournisseur.
 - Il permet d'authentifier son porteur qui est la condition nécessaire pour accéder à un service.
 - Cette technologie apporte une sécurité fiable et à caractère légal.
 - Elle est disponible sous forme synchrone (SSTL/TLS) et asynchrone (S/MIME).
- Plus concrètement un système permettant de:
 - Délivrer, révoquer, publier renouveler, les certificats de leur porteurs
 - Générer ou séquestrer et remettre les clés numériques à leurs porteurs
 - Générer et publier la liste des certificats révoqués
- Mais:
 - Ce n'est qu'un moyen. Le but est le développement de services sur cette technologie.
 - Ce n'est pas que de la technique. Il est important qu'une concertation voire une coordination au niveau groupe soit mise en place

Les infrastructures à clé publique

- L'algorithmie asymétrique permet :
 - Simplification de la gestion des secrets
 - Procédés efficaces de distribution de clé
 - Mais surtout : une propriété d'ouverture...

- Les PKI : un système ouvert
 - Pour réaliser une transaction en toute confiance entre Alice et Bob, il faut et il suffit :
 - qu'Alice ait confiance dans l'autorité de certification de Bob et
 - que Bob ait confiance dans l'autorité de certification d'Alice.
 - Il n'est pas nécessaire :
 - qu'Alice connaisse Bob ou
 - qu'ils aient la même autorité de certification

- La confiance sur Internet :
 - Les PKI permettent d'établir une relation de confiance entre deux personnes qui ne se connaissent pas.
 - C'est l'outil indispensable au commerce électronique, aux téléprocédures, à toutes les transactions privées, commerciales ou professionnelles.

Fonctions liées à la certification

- *Enregistrement*

- Enregistrement des utilisateurs et arrivée d'un nouvel utilisateur
- Vérification des identités des demandeurs de certificats
- Génération de certificats et remise du certificat à l'utilisateur

- *Gestion des clés*

- Génération et distribution des clés
- Séquestre et recouvrement des clés

- *Gestion des certificats au quotidien*

- Renouvellement du certificat à fin date validité
- Gestion des mutations, mobilités, départs
- Gestion des oubli de code porteur, détérioration/perte/vol de cartes...
- Attribution de carte invité, délégation de signature

- *Révocation*

- Révocation de certificats
- Distribution ou émission périodique de la CRL

- *Publication*

- Publication des certificats de la communauté
- Publication de la dernière CRL à jour

➔ *Nécessité de définir de nouveaux acteurs, rôles et responsabilités*

➔ *Des autorités de confiance (AC, OC, AE) doivent intervenir dans chacune des fonctions de certification*

L'avenir :

Chiffrement homomorphe ???

Problématique: la manipulation des données chiffrées

- En environnement Cloud et Big Data, on stocke de gros volumes de données sur des serveurs distants, non maîtrisés.
 - Ces données peuvent être interceptées lors de leur transfert ou stockage
 - Les prestataires de services ont aussi accès à ces informations

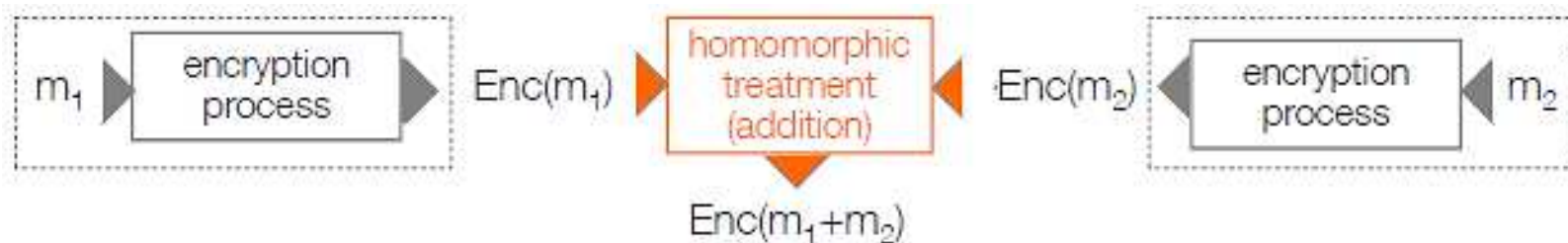
- **.Solution** : chiffrer les données avant leur transfert



- **Inconvénient**: on ne peut plus manipuler ces données à distance (retoucher ses photos, chercher des mots dans un texte, effectuer des calculs...).
- **Une nouvelle piste : la cryptographie homomorphe** : fournit la sécurité des données en permettant que les données chiffrées restent manipulables par les personnes autorisées.

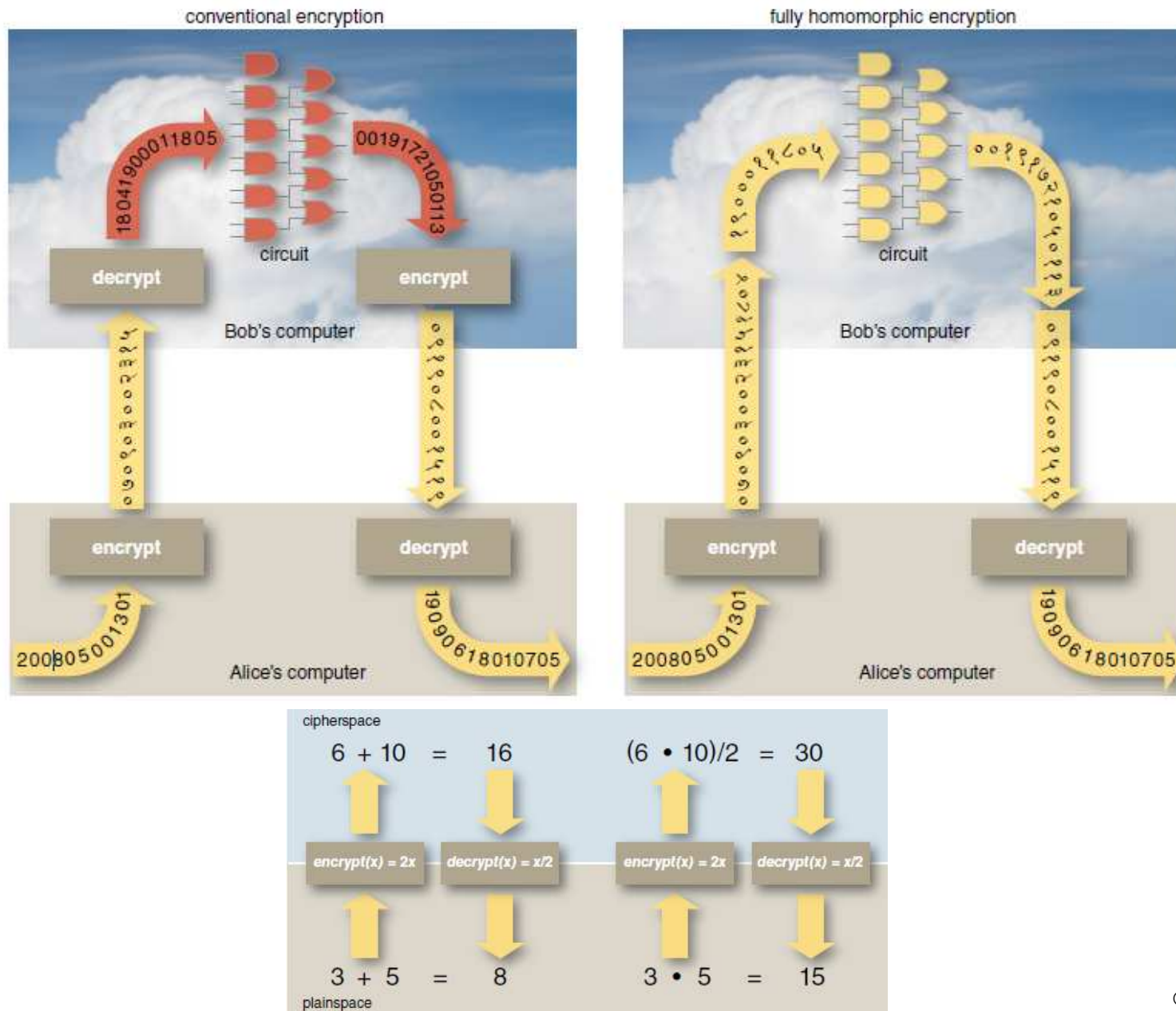
Chiffrement homomorphe : définition

- Un système de chiffrement homomorphe permet d'effectuer divers traitements sur un texte chiffré sans avoir à le déchiffrer.



- L'un des intérêts du chiffrement homomorphe est la délégation de calculs dans le cas où un utilisateur souhaite faire un calcul coûteux (et ne dispose pas nécessairement des ressources nécessaires pour l'exécuter) et aimerait faire appel à un service de cloud computing (auquel il ne fait pas nécessairement confiance) pour effectuer ses calculs.
 - Fondamental pour le traitement de données en environnement Cloud Computing, Big Data, Internet des Objets : un tiers peut faire des calculs sur les messages chiffrés sans les déchiffrer, et le résultat est utilisable, c'est-à-dire peut être déchiffré*
- Domaines d'applications :
 - Recherche par mot-clé, concaténation
 - Statistiques et datamining sur données chiffrées
 - Vérifier si deux fichiers chiffrés sont identiques
 - Détection de logiciel malveillant au sein de trafic chiffré
 - Fournir des publicités ciblées sans rien connaître du destinataire
 - Faire des recherches sur Internet et recevoir les réponses sans que le moteur de recherche ne sache quel était l'objet de notre requête

Chiffrement homomorphe



Une science encore immature

- Les algorithmes connus sont essentiellement partiellement homomorphes
- Les algorithmes génèrent une inflation des données
- Lenteur des calculs
- Adaptabilité à des cas concrets
- ...
- Mais beaucoup de progrès récents prometteurs...

Bibliographie

- Mooc sécurité des SI : <https://www.secnumacademie.gouv.fr/>
- Cryptographie appliquée B. Schneier (Vuibert)
- ANSSI : <http://www.ssi.gouv.fr/>
- CNIL : <http://www.cnil.fr/>
- AFNOR : Livre Blanc – Données massives/ Big Data
- ENISA : Big Data Threat Landscape and Good Practice Guide
- ENISA : State of the Art Analysis of Data Protection in Big Data Architectures (Privacy by Design in Big Data)