Security for Data Scientists

Pascal Lafourcade

Mars 2018

Notation

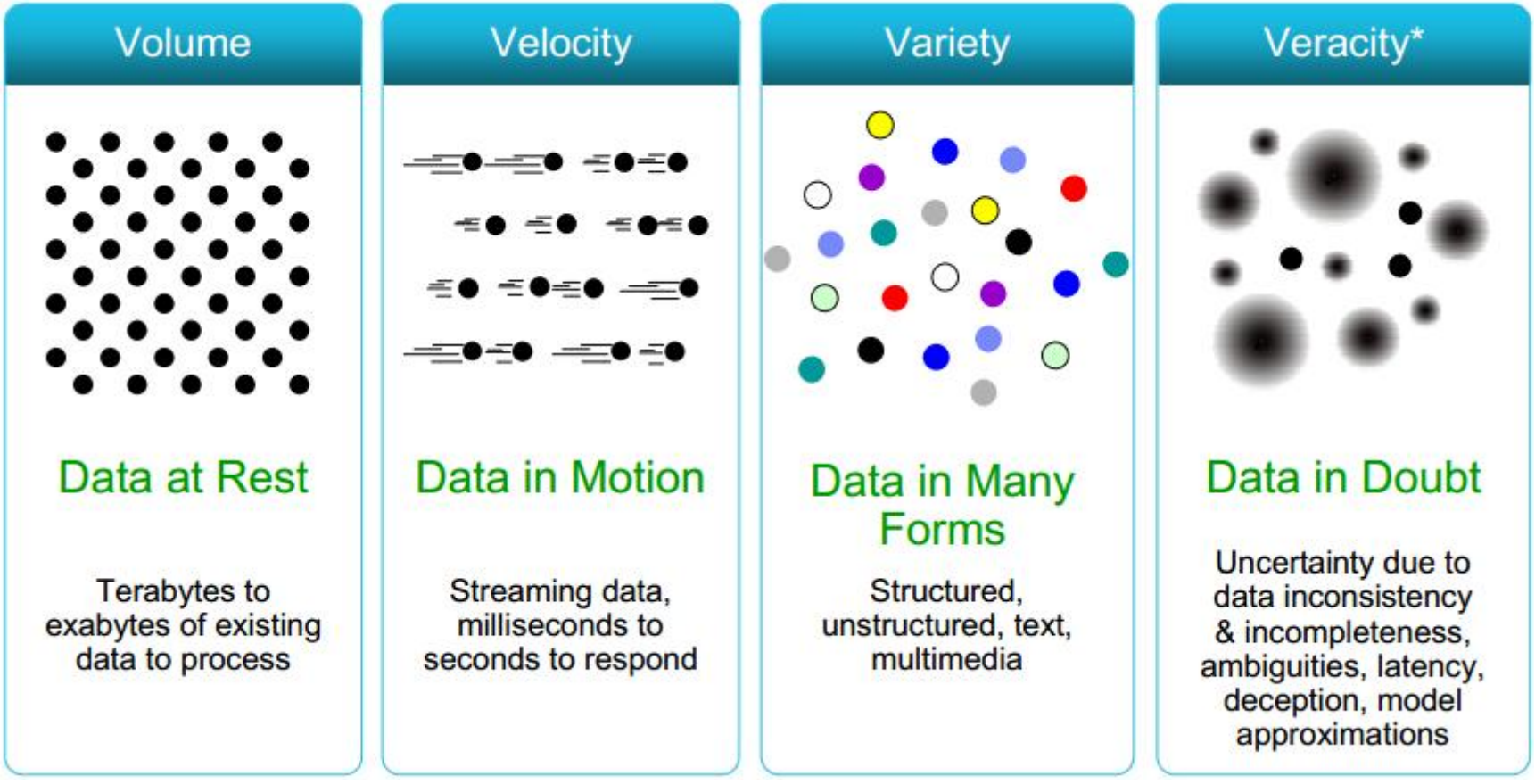
$2 \times$ 2h00 + 2h00 + 2h00 TP

$$\text{Note} = 70\% \text{ Projet PIA} + 30\% \text{ TP}$$

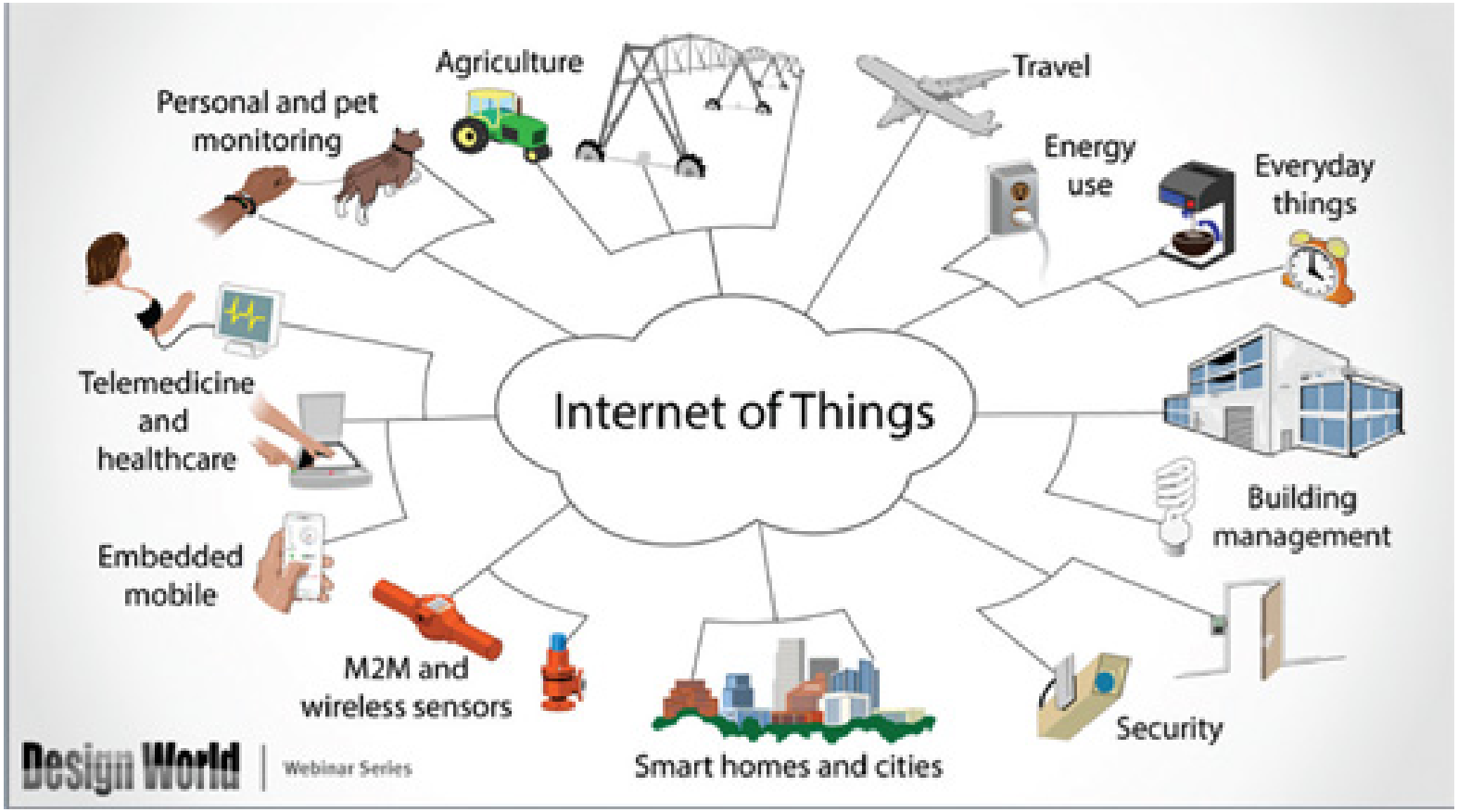
TP : python + REDIS + SSE

Outline

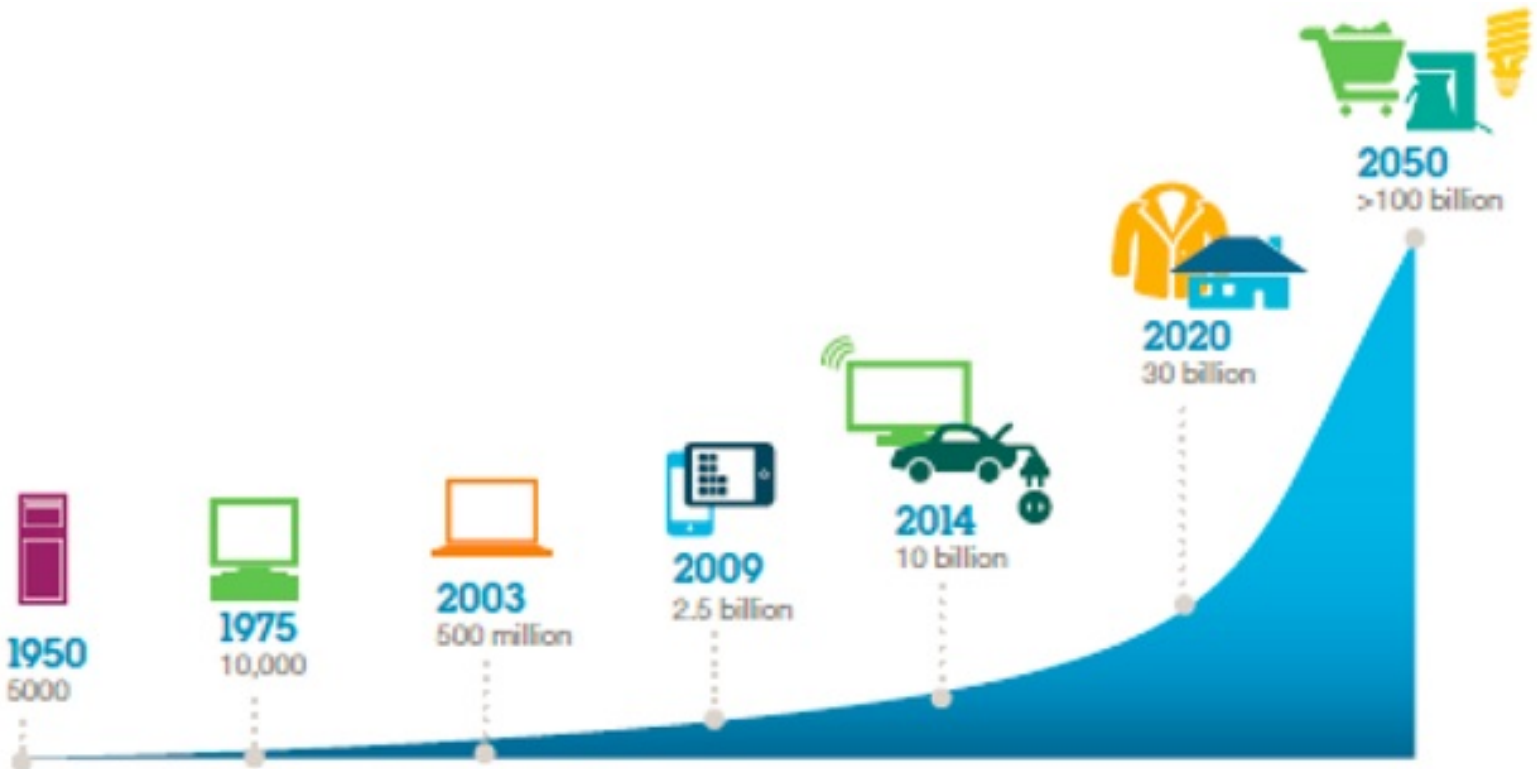
Big Data


Volume
Data at Rest
Terabytes to exabytes of existing data to process
Velocity
Data in Motion
Streaming data, milliseconds to seconds to respond
Variety
Data in Many Forms
Structured, unstructured, text, multimedia
Veracity*
Data in Doubt
Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

IoT

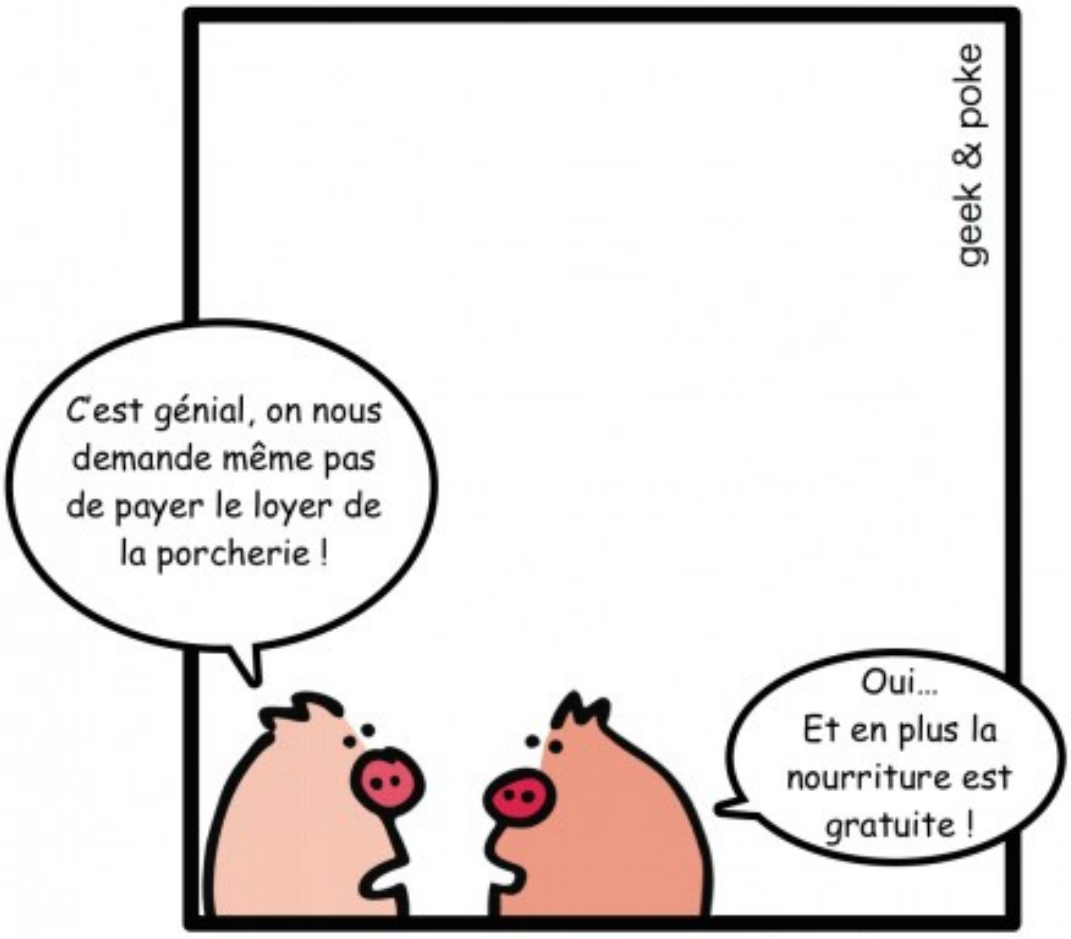

Agriculture
Travel
Personal and pet monitoring
Energy use
Everyday things
Telemedicine and healthcare
Internet of Things
Building management
Embedded mobile
M2M and wireless sensors
Smart homes and cities
Security
Design World
Webinar Series

IoT


1950
5000
1975
10,000
2003
500 million
2009
2.5 billion
2014
10 billion
2020
30 billion
2050
>100 billion

Big Data and Security

Free ?



Deux cochons discutant du modèle « gratuit »

Free ?

If it is free then you are the product

Data Privacy ?

Outline

CNIL créé en 1978

Commission nationale de l'informatique et des libertés

BUT

Protéger les données personnelles, accompagner l'innovation, préserver les libertés individuelles

ANSSI créée le 7 juillet 2009.

STAD

Système de Traitement Automatisé de Données

"Tout ensemble composé d'une ou plusieurs unités de traitement, de mémoire, de logiciel, de données, d'organes d'entrées-sorties et de liaisons, qui concourent à un résultat déterminé, cet ensemble étant protégé par des dispositifs de sécurité".

Aucune définition précise dans la loi

Dans les faits c'est presque tout :





3 acteurs

Utilisateur

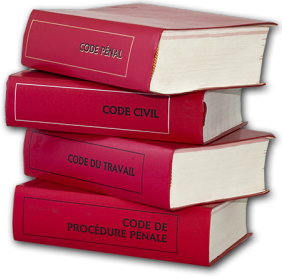
Responsable

Pirate

L'utilisateur

Droits

- D'accès : demander directement au responsable d'un fichier s'il détient l'intégralité de ces données
- De rectification
- D'opposition dêtre dans un fichier
- Déréférencement sur le web par rapport au nom et prénom



Le responsable

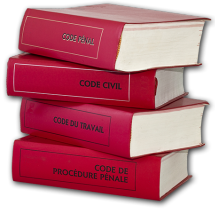
Et le sous-traitant via le contrat.

Devoirs

- Déclarer les traitements de données personnelles
  5 ans & 300 000
- Prendre toutes précautions pour la sécurité des données selon
  - la nature des données
  - les risques présentés par le traitement

  5 ans & 300 000

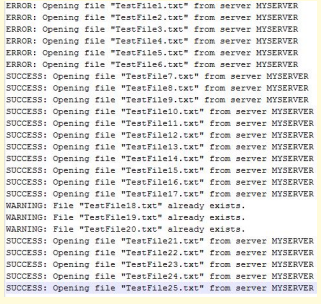
Lois informatique et libertés : Article 22 et Article 34.
Guide de la CNIL : La sécurité des données personnelles



Conservation des logs

LCEN 2004

- 1 an pour les logs (jurisprudence de la BNP Paribas)
- Décret 2011-219 du 25 février 2011 relatif à la conservation et à la communication des données permettant d'identifier toute personne ayant contribué à la création d'un contenu mis en ligne:
  - ip, url, protocole, date heure, nature de l'opération
  - éventuellement les données utilisateurs
  - éventuellement données bancaires
  - accédées dans le cadre d'une réquisition
  - conservées un an
  - données utilisateurs pendant un an après la clôture



Article 226-20 : les logs ont une date de péremption

Le pirate

Risques (STAD (Article 323-1))

- accès frauduleux ou maintien frauduleux de l'accès **2 ans & 60 000**
- suppression ou modification des données **3 ans & 100 000**
- si données à caractère personnel **5 ans & 150 000**
- altération du fonctionnement **5 ans et de 75 000**
- si données à caractère personnel **7 ans & 100 000**

Risques encourus

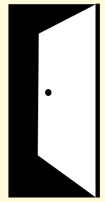
En pratique

- Atteintes aux intérêts fondamentaux de la nation (Sécurité nationale) Article 410-1 à 411-6
- Secret des communication pour l'autorité publique et FAI **3 ans et 45 000** Article 432-9
- Usurpation d'identité **5 ans et de 75 000** Article 434-23
- Importer, détenir, offrir ou mettre à disposition un moyen de commettre une infraction est puni

Sauf si

Pas de condamnation si

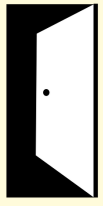
- aucune protection
- aucune mention de confidentialité
- accessible via les outils de navigation grand public
- même en cas de données nominatives



Sauf si

Pas de condamnation si

- aucune protection
- aucune mention de confidentialité
- accessible via les outils de navigation grand public
- même en cas de données nominatives



Il est donc important de protéger ces données

Outline

Réglement européen : 25 mai 2018

Règlement Général sur la Protection des Données RGPD

Histoire

Invalidation du “safe harbor” par la Cour de Justice de l’Union européenne : une décision clé pour la protection des données, 07 octobre 2015
Quel niveau de protection des données personnelles transférées aux Etats-Unis ?

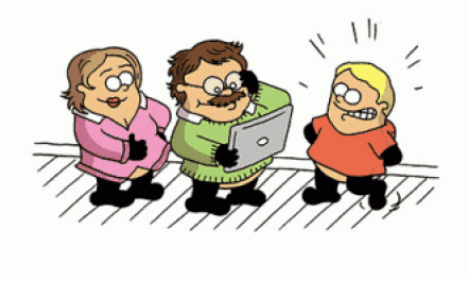
Plus de droits pour vos données !

Sanction

Plus de transparence

Droit à l'oubli

Guichet unique



Protection des mineurs

Portabilité

Objectifs?

Renforcer la transparence:

- Quelles données sont collectées?
- Dans quels buts?
- Pour combien de temps?

Faciliter l'exercice des droits

- droit à la rectification
- droit à la portabilité : récupération et communication à un autre traitement
- droit à l'oubli : suppression de données personnelles
  - dès qu'elles ne sont plus nécessaires au traitement
  - dès que le consentement de l'utilisateur a été retiré
  - dès que la personne s'y oppose

Règles d'or de la CNIL

1. Licéité du traitement
2. Finalité du traitement
3. Pertinence et proportionnalité des données; principe de minimisation
4. Conservation limitée des données
5. Exactitude, intégrité et confidentialité des données : principe de sécurité
6. Renforcement de la transaparence et exercice des doits facilité

Nouveautés

RESPONSABILITATION de TOUS les acteurs

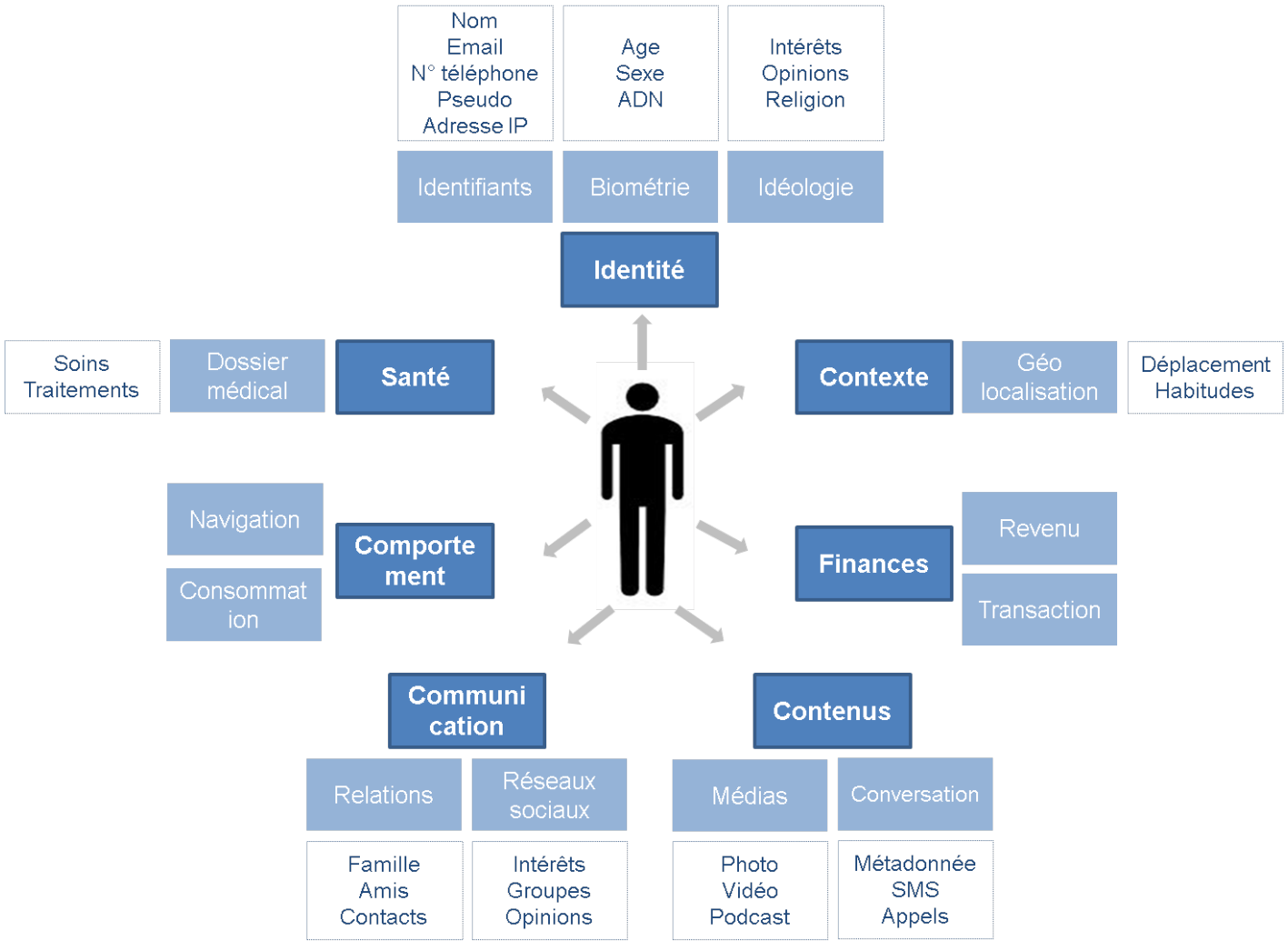
Principes

Tous responsables et tous auditables
Privacy by design
Security by default
DPO (Data Protection Officer)

- conformité au RGPD
- Point de contact avec les autorités

Analyse d'impact (PIA: Privacy Impact Assessment)

Qu'est-ce qu'une donnée personnelle ?



Qu'est-ce qu'une donnée personnelle ?

Information qui permet d'identifier une personne physique, directement ou indirectement.

- un nom,
- une photographie,
- une adresse IP,
- un numéro de téléphone,
- un identifiant de connexion informatique,
- une adresse postale,
- une empreinte,
- un enregistrement vocal,
- un numéro de sécurité sociale,
- un mail, etc.

Qu'est-ce qu'une donnée personnelle **sensible**?

Données liés à de la discrimination ou des préjugés :

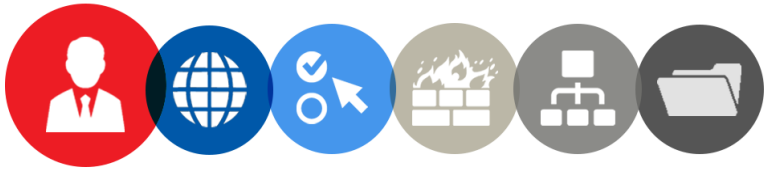
- Une opinion politique,
- une sensibilité religieuse,
- un engagement syndical,
- une appartenance ethnique,
- une orientation sexuelle,
- une situation médicale ou des idées philosophiques sont des données sensibles.

Toute collecte sans consentement préalable écrit, clair et explicite est interdite !

RPGD : en 6 étapes @CNIL

1. Désigner un pilote
2. Cartographier
3. Prioriser
4. Gérer les risques
5. Organiser
6. Documenter

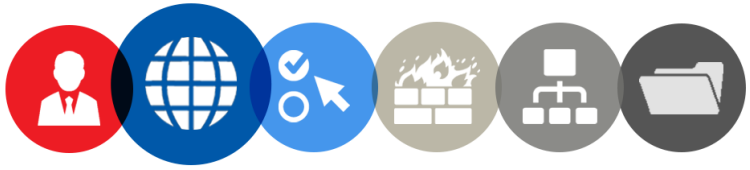
Étape 1 : Désigner un pilote



Délégué à la protection des données

Mission d'information, de conseil et de contrôle en interne.
Conformité au RGPD.

Étape 2 : Cartographier

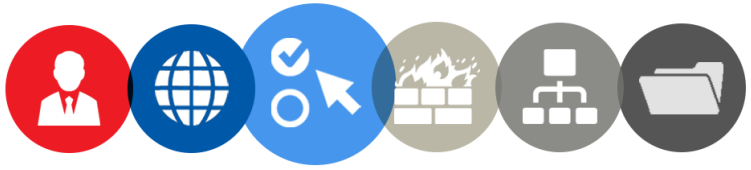


Tenir une documentation interne complète sur leurs traitements de données personnelles

- Catégories les données traitées
- Recenser précisement vos traitements de données personnelles (**Registre des traitements**)
- Lister les objectifs
- Identifier les acteurs
- Identifier les flux des données

But : Assurer que ces traitements respectent bien le règlement.

Étape 3 : Prioriser



1. Collecter et traiter **que les données nécessaires**.
2. **Base juridique du traitement** : consentement de la personne, contrat, obligation légale ...
3. Révisez vos **mentions d'information** : articles 12, 13 et 14: droits de la personne concernée : Transparence, Information et Transitivité
4. Vérifier vos **sous-traitants** et clause des contrats
5. Prévoyez les **modalités d'exercice des droits** des personnes concernées : droit d'accès, de rectification, droit à la portabilité, retrait du consentement...
6. Vérifiez les **mesures de sécurité** mises en place.

Étape 3 : VIGILANCE, des **types** de données

- origine prétendument **raciale ou ethnique**, les opinions politiques, philosophiques ou religieuses, l'appartenance syndicale,
- la **santé** ou l'orientation sexuelle,
- génétiques ou **biométriques**,
- infraction ou de condamnation **pénale**,
- sur les **mineurs**.

Étape 3 : VIGILANCE, votre traitement

- la surveillance **systématique** à grande échelle d'une zone accessible au public
- l'évaluation **systématique** et approfondie d'aspects personnels, y compris le profilage, sur la base de laquelle vous prenez des décisions produisant des effets juridiques à l'égard d'une personne physique ou l'affectant de manière significative.

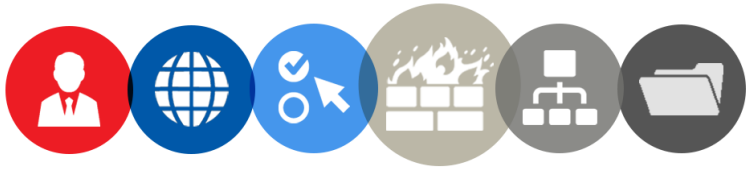
Étape 3 : VIGILANCE **transfert** des données hors UE

- Vérifiez que le pays vers lequel vous transférez les données est reconnu comme adéquat par la Commission européenne ;
- Dans le cas contraire, encadrez vos transferts.

Étape 4 : Gérer les risques

Privacy Impact Assessment (PIA)

Data protection impact assessment



- Principes et droits fondamentaux, **non négociables**, de la loi
- Gestion des **risques sur la vie privée** des personnes concernées, pour déterminer les mesures techniques et d'organisation pour protéger les données personnelles.

Un PIA contient :

- Une **description** du traitement étudié et de ses **finalités**.
- Une **évaluation de la nécessité et de la proportionnalité** des opérations de traitement au regard des finalités
- Une **évaluation des risques** pour les droits et libertés des personnes, les mesures envisagées pour faire face aux risques.

Étape 4 : Qui participe au PIA?

- **Le responsable de traitement** : valide et applique le PIA.
- **Le délégué à la protection des données** : élabore le plan d'action et se charge de vérifier son exécution ;
- **Le(s) sous-traitant(s)** : fournit les informations nécessaires à l'élaboration du PIA ;
- **Les métiers (RSSI, maîtrise d'ouvrage, maîtrise d'œuvre)** : aident à la réalisation du PIA en fournissant les éléments adéquats ;
- **Les personnes concernées** : donnent leurs avis sur le traitement.

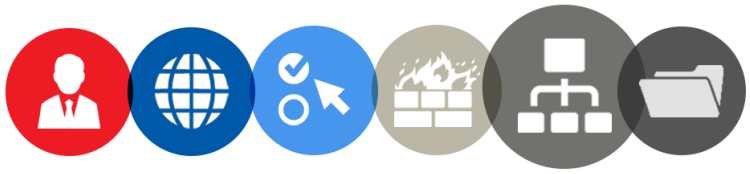
Étape 4 : **PIA obligatoire** Art. 35

Pour tout traitement susceptible d'engendrer des **risques élevés** pour les droits et libertés des personnes concernées.

1. Evaluation ou notation;
2. Décision automatisée avec effet juridique significatif;
3. Surveillance systématique ;
4. Données sensibles ou données à caractère hautement
5. Données personnelles traitées à grande échelle ;
6. Croisement d'ensembles de données ;
7. Données concernant des personnes vulnérables ;
8. Usage innovant ou application de nouvelles solutions technologiques ou organisationnelles ;
9. Exclusion du bénéfice d'un droit, d'un service ou contrat.

Si **au moins 2 de ces critères**, alors faire un PIA.

Étape 5 : Organiser



- Protection des données personnelles **dès la conception**
- **Sensibiliser et d'organiser la remontée d'information**
- Traiter les **réclamations et les demandes** des personnes concernées quand à l'exercice de leurs droits
- **Anticiper les violations de données**, dans les 72 heures aux autorités et personnes concernées

Étape 6 : Documenter

Prouver la conformité = Avoir la documentation nécessaire

- Traitements
- Information des personnes
- Contrat pour les acteurs

Étape 6 : Documenter les traitements

- Le **registre des traitements** (pour les responsables de traitements) ou des **catégories d'activités de traitements** (pour les sous-traitants)
- **PIA** pour les traitements à risque
- L'**encadrement des transferts** de données hors de l'Union européenne.

Étape 6 : Documenter l'information

- Les **mentions d'information**
- Les modèles de **recueil du consentement** des personnes concernées,
- Les procédures **mises en place** pour l'exercice des droits

Étape 6 : Documenter les contrats

- Les **contrats avec les sous-traitants**
- Les **procédures internes** en cas de violations de données
- Les **preuves** que les personnes concernées ont donné leur consentement lorsque le traitement de leurs données repose sur cette base.

Objectif

Responsabilisation de tous les acteurs impliqués dans le traitement des données personnlles dès lors qu'elle concernent des résidents européens.

- Obligation de transparence et traçabilité
  - Contrtat écrit entre les acteurs
  - Autorisation écrite des traitement
  - Démontrer le respect de vos obligations
  - Ternir un **registre des traitements**
- Protection by design et by default (paramètres, accès, purge)
- Obligation de garantir la sécurité des données traitées
- Obligation d'assistance, d'alerte et de conseil (immédiate)

Registre des catégories d'activités de traitement

- nom et les coordonnées de chaque client
- le nom et les coordonnées de chaque sous-traitant
- le nom et les coordonnées du délégué à la protection des données
- les catégories de traitements effectués
- les transferts de données hors UE
- une description générale des mesures de sécurité techniques et organisationnelles que vous mettez en place

Qui est touché ?

TOUT LE MONDE !

- les prestataires de services informatiques
- les agences de marketing ou de communication
- tout organisme offrant un service ou une prestation
- Un organisme public ou une association

qui traite les données personnelles.

Sanctions

Jusqu'à 10 ou 20 millions d'euros, ou 2% ou 4% du chiffre d'affaires annuel mondial de l'exercice précédent.
En France la CNIL devient autorité de contrôle

Sanctions pour le sous-traitant

- si vous agissez en dehors des instructions licites de votre client ou contrairement à ces instructions ;
- si vous n'aidez pas votre client à respecter ses obligations
- si vous ne mettez pas à la disposition de votre client les informations permettant de démontrer le respect des obligations ou pour permettre la réalisation d'audits
- si vous n'informez pas votre client qu'une instruction constituerait une violation du règlement européen
- si vous sous-traitez sans autorisation préalable de votre client
- si vous fait es appel à un sous-traitant qui ne présente pas de garanties suffisantes
- si vous ne désignez pas un délégué à la protection des données
- si vous ne tenez pas de registre des catégories d'activités de traitement

Bilan après 4 mois

- 24500 organismes ont 1 DPO: 13000 DPO contre 5000 CIL
- 600 notifications de violations, environ 7 par jour
- 3 millions de visites sur le site de la CNIL
- 150 000 téléchargements du registre simplifié de la CNIL
- 3767 plaintes soit une augmentation de 64%
- plus de 200 plaintes transfrontalières

24 Juillet 2018 : Sanction de 50 000 € de la CNIL à l'encontre de la société DAILYMOTION, mot de passe stocké en clair temporairement.
15 Aout 2018 : Sanction de 30 000 € par la CNIL à l'encontre de l'Office Public de l'Habitat de Rennes Métropole

4 Croyances sur le RGPD par Florence BONNET

I. La probabilité de faire l'objet d'un contrôle de la CNIL est faible

- obligation de notifier et de communiquer les violations de données personnelles à l'autorité et aux personnes concernées le cas échéant.
- toute personne a le droit de réclamer auprès d'une autorité de contrôle et d'exercer son droit d'obtenir réparation.

4 Croyances sur le RGPD par Florence BONNET

II. En cas de contrôle, il suffira de collaborer avec la CNIL et de faire preuve de réactivité pour éviter une sanction

Absence de mesures élémentaires de sécurité = non-conformité.

- Mise en ligne d'un site sans test
- Exposition aux données sans authentification (mot de passe suffisamment robustes)
- Ne doivent pas être conservés ou transmis en clair mais de manière sécurisée
- Les connexions et flux de données doivent être sécurisés
- Les connexions à une plateforme des paiements doivent être tracés
- Le dispositif de communication bluetooth doit être sécurisé
- La connexion à distance doit être sécurisée (VPN, IP)
- Les données les sensibles doivent être conservées et sécurisées
- Le chiffrement doit être à l'état de l'art ! Pas de MD5 !

4 Croyances sur le RGPD par Florence BONNET

II.

- Protection du secret : le sel doit être conservé dans un espace distinct de celui où sont stockés les mots de passe ;
- Les numéros de carte bancaire ne doivent pas être conservés en clair avec les cryptogrammes .
- Les accès aux données doivent être strictement limités aux seules personnes ayant besoin d'en connaitre
- il appartient au responsable de traitement, d'adapter les conditions d'usage de ce logiciel à sa propre population

4 Croyances sur le RGPD par Florence BONNET

III. Se croire à l'abri parce qu'il existe forcément une politique de sécurité dans l'entreprise

- il est vain pour la société de chercher à se dégager de sa responsabilité en invoquant de supposées procédures préventives en la matière (procédure sécurité conforme ISO 27001, exigences du Règlement CRBF 97-02)

4 Croyances sur le RGPD par Florence BONNET

IV. c'est le sous-traitant qui sera responsable

- Il convient de tracer et de documenter les échanges avec le prestataire
- L'intervention d'un prestataire crée une responsabilité supplémentaire de contrôle effectif des agissements du prestataire et des solutions utilisées par ce dernier.
- La prestation de service doit obligatoirement faire l'objet d'un contrat encadrant les obligations du sous-traitant en matière de sécurité et de confidentialité des données à caractère personnel

Note: 70%

Faire un PIA pour votre entreprise ou une partie de votre entreprise.

Outline

Démarche

Publiée en octobre 2005 et révisée en 2013.

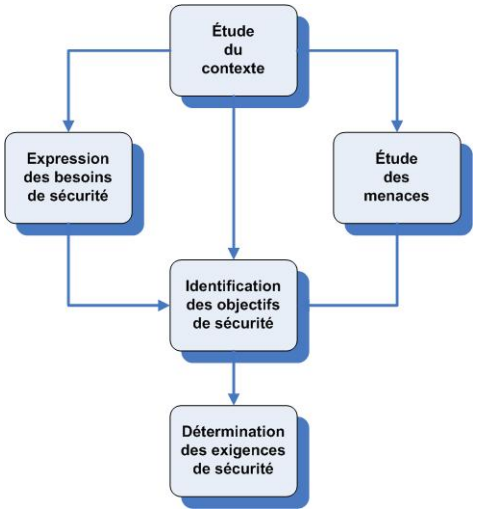
1. Phase d'établissement
2. Phase d'implémentation
3. Phase de maintien
4. Phase d'amélioration

SMSI : Système de management de la sécurité de l'information

Phase d'établissement (PLAN)

1. Définir la politique et le périmètre du SMSI
2. Identifier et évaluer les risques liés à la sécurité et élaborer la politique de sécurité (EBIOS)
3. Traiter le risque et identifier le risque résiduel par un plan de gestion (Évitement, réduction, transfert, acceptation)
4. Choisir les mesures de sécurité à mettre en place

EBIOS



1. Identifier les actifs ;
2. Identifier les personnes responsables ;
3. Identifier les vulnérabilités ;
4. Identifier les menaces ;
5. Identifier leurs impacts sur les actifs à défendre ;
6. Èvaluer la vraisemblance ou potentialité du risque ;
7. Estimer les niveaux de risque, fonction de leur potentialité et de leur impact.

Phase d'implémentation (DO)

1. Établir un plan de traitement des risques
2. Déployer les mesures de sécurité
3. Générer des indicateurs:
   - De performance pour savoir si les mesures de sécurité sont efficaces
   - De conformité qui permettent de savoir si le SMSI est conforme à ses spécifications
4. Former et sensibiliser le personnel

Phase de maintien (Check)

Gérer le SMSI au quotidien et à détecter les incidents

- Le contrôle interne (s'assurer en permanence que les processus fonctionnent normalement)
- Les audits internes (vérifier la conformité et l'efficacité du système de management.
- Les revues (ou réexamens) qui garantissent périodiquement l'adéquation du SMSI avec son environnement.

Phase d'amélioration (Act)

Actions correctives, préventives ou d'amélioration pour les incidents et écarts constatés lors de la phase Check.

Outline

Ethique

éthique = (Larousse) Ensemble des principes moraux qui sont à la base de la conduite de quelqu'un.
Science de la morale / un art du comportement.

2 principes s'imposent (Jérôme Béranger)

1. L'information est agrégée en connaissance, mais cette connaissance est une connaissance pratique, finalisée dans l'action. C'est moins un savoir qu'un savoir-utiliser.
2. A une description de processus on préfèrera une description d'état. L'enjeu de l'éthique, est le passage d'un état de savoirs complexes, désorganisés et flous vers un état de savoirs simples, structurés et orientés vers une fin.

Les données médicales

4 principes par Tom Beauchamp et James Childress, Principles of Biomedical Ethics (2001).

1. bienfaisance avec deux règles précises :
   - elle doit être bénéfique,
   - et elle doit être utile (avoir un rapport coût-bénéfice positif)
2. l'autonomie : le fait qu'une personne se donne à elle-même sa règle de conduite. Ce principe vise à la participation du patient au processus de décision
3. la "non-malfaisance" : éviter le mal à celui dont on a la responsabilité, lui épargner préjudices ou souffrances qui n'auraient pas de sens pour lui
4. la justice : notion d'égalité et d'équité.

Problème

Le système impose des règles d'attribution et d'accès à l'information qui diffèrent en fonction du statut. La dissymétrie de connaissances est discriminante et remet en cause la transparence de l'information.
Quelle est leur utilisation et diffusion ?
$\Rightarrow$ La simplification des données transmises entraîne un usage et un accès plus efficace, avec une meilleure saisie et une plus grande sécurité. Elle aboutit en revanche à une moins bonne intégrité des données.
De ce fait, la hiérarchisation des données simplifie le travail des divers utilisateurs, mais induit une plus grande complexité technique pour le concepteur du système d'information.

Se questionner

- quels sont les objectifs, les buts, les enjeux et le sens de cette étape ?
- Quelles données vais-je utiliser ?
- Des données partielles ou totales ?
- Comment vais-je les utiliser ?
- À quel endroit ? Auprès de quels utilisateurs ?
- Plus globalement, comment exploiter l'ensemble hétérogène des données accumulées et stockées dans un système d'information ?
- Quelle sera sa pertinence par rapport à ma situation ?
- Cela ne va-t-il pas dénaturer la valeur informative initiale ?
- L'intégrité du message final sera-t-elle conservée ?

Outline

Définition

La maîtrise et la protection de l'information stratégique utile pour tout acteur économique.

3 piliers

- Maîtrise de l'information, management des connaissances
- Protection du patrimoine informationnel
- Stratégie d'influence et lobbying

La compétitivité est la finalité de l'IE
(Intelligence = renseignement)

Maîtriser l'Information

- Identifier les sources
- Collecter l'information (veille, reseaux sociaux ...)
- Exploitation : analyse et aide à la décision
- Diffusion :

Protection de l'Information

"Seuls les paranoïaques survivent", Andy GROVE, Cofondateur d'Intel en 1968

1. Classification de l'information
2. Diagnostic
3. Protection des accès
4. Sensibilisation
5. Surveillance, détection

Stratégies d'Influence

- Presse, média
- Blog, réseaux sociaux
- Communication en cas de crise infomration/désinformation

Outline

La sécurité numérique est déjà là

Mais prendre de bonnes habitudes ça prend du temps ...

même quand c'est important

Devenir acteur de sa sécurité numérique

Devenir acteur de sa sécurité numérique

car la sécurité c'est pas automatique.

Sécurité de mes mots de passe

Sécurité de mes mots de passe

Top 25 en 2014

1. 123456
2. password
3. 12345
4. 12345678
5. qwerty
6. 123456789
7. 1234
8. baseball
9. dragon
10. football
11. 1234567
12. monkey
13. letmein
14. abc123
15. 111111
16. mustang
17. access
18. shadow
19. master
20. michael
21. superman
22. 696969
23. 123123
24. batman
25. trustno1

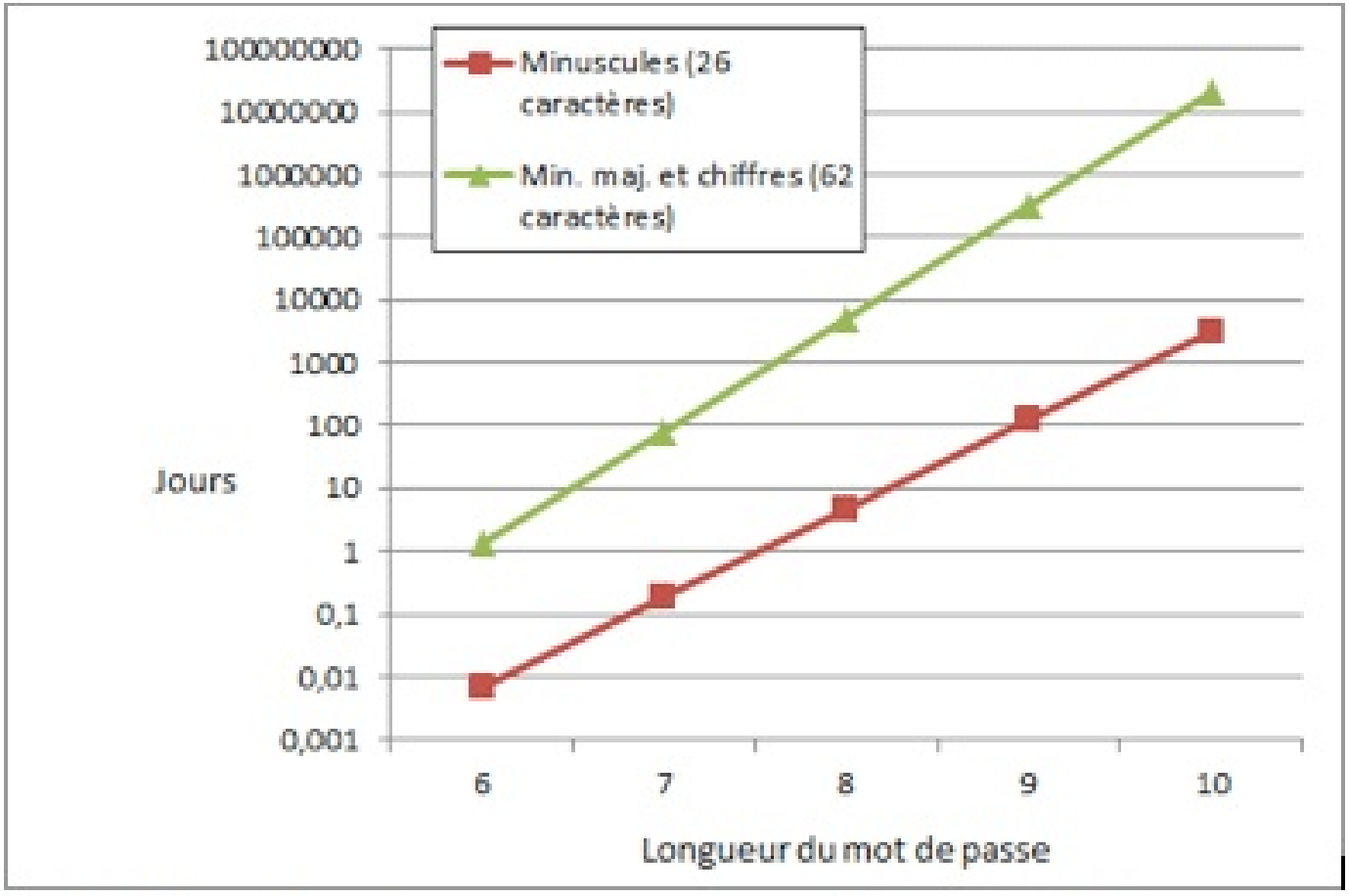
Top 25 en 2015

1. 123456 (Unchanged)
2. password (Unchanged)
3. 12345678 (Up 1)
4. qwerty (Up 1)
5. 12345 (Down 2)
6. 123456789 (Unchanged)
7. football (Up 3)
8. 1234 (Down 1)
9. 1234567 (Up 2)
10. baseball (Down 2)
11. **welcome**
12. **1234567890**
13. abc123 (Up 1)
13. **1qaz2wsx**
14. dragon (Down 7)
15. master (Up 2)
16. monkey (Down 6)
17. letmein (Down 6)
18. **login**
19. **princess**
20. **qwertyuiop**
21. **solo**
22. **passw0rd**
23. **starwars**

Top 25 en 2016

1. 123456 (Unchanged)
2. 123456789 (Up 5)
3. qwerty (Up 1)
4. 12345678 (Down 1)
5. 111111 (Up 9)
6. **1234567890**
7. 1234567 (Up 1)
8. password (Down 6)
9. **123123**
10. **987654321**
11. **qwertyuiop**
12. **mynoob**
13. **123321**
14. **666666**
15. **18atcskd2w**
16. **7777777**
17. **1q2w3e4r**
18. **654321**
19. **555555**
20. **3rjs1la7qe**
21. **google**
22. **1q2w3e4r5t**
23. **123qwe**
24. **zxcvbnm**
25. **1q2w3e**

Passwords: Brute force



Quelques chiffres

number of Characters	Numbers only	Upper or lower case letters	upper or lower case letters mixed	numbers, upper and lower case letters	numbers, upper and lower case letters, symbols
3	Instantly	Instantly	Instantly	Instantly	Instantly
4	Instantly	Instantly	Instantly	Instantly	Instantly
5	Instantly	Instantly	Instantly	3 secs	10 secs
6	Instantly	Instantly	8 secs	3 mins	13 mins
7	Instantly	Instantly	5 mins	3 hours	17 hours
8	Instantly	13 mins	3 hours	10 days	57 days
9	4 secs	6 hours	4 days	1 year	12 years
10	40 secs	6 days	169 days	106 years	928 years
11	6 mins	169 days	16 years	6k years	71k years
12	1 hour	12 years	600 years	108k years	5m years
13	11 hours	314 years	21k years	25m years	423m years
14	4 days	8k years	778k years	1bn years	5bn years
15	46 days	212k years	28m years	97bn years	2tn years
16	1 year	512m years	1bn years	6tn years	193tn years
17	12 years	143m years	36bn years	374tn years	14qd years
18	126 years	3bn years	1tn years	23qd years	1qt years

Key:

k – Thousand (1,000 or 10^{-3})
m – Million (1,000,000 or 10^{-6})
bn – Billion (1,000,000,000 or 10^{-9})
tn – Trillion (1,000,000,000,000 or 10^{-12})
qd – Quadrillion (1,000,000,000,000,000 or 10^{-15})
qt – Quintillion (1,000,000,000,000,000,000 or 10^{-18})

Calculer la « force » d'un mot de passe

Type de mot de passe	Taille de clé équivalente	Force	Commentaire
Mot de passe de 8 caractères dans un alphabet de 70 symboles	49	Très faible	Taille usuelle
Mot de passe de 10 caractères dans un alphabet de 90 symboles	65	Faible	
Mot de passe de 12 caractères dans un alphabet de 90 symboles	78	Faible	Taille minimale recommandée par l'ANSSI pour de passe ergonomiques ou utilisés de façon locale.
Mot de passe de 16 caractères dans un alphabet de 36 symboles	82	Moyen	Taille recommandée par l'ANSSI pour des mots plus sûrs.
Mot de passe de 16 caractères dans un alphabet de 90 symboles	104	Fort	
Mot de passe de 20 caractères dans	130	Fort	Force équivalente à la plus petite taille de clé de l'

Suite aux fuites ...

Suite aux fuites ...

En réalité

En réalité


©Crown Copyright, 2012
www.dukeandduchessofcambridge.org


MilFlip Logon Details
RMS
ATTENTION ALL OHMS STUDENTS

Quelques conseils

Un mot de passe

1. ne se prête pas
2. ne se laisse pas traîner
3. ne s'utilise qu'une fois
4. s'il est cassé, il faut en changer
5. il faut en changer régulièrement
6. il est jamais assez sophistiqué
7. la taille compte.

Quelques conseils

Un mot de passe

1. ne se prête pas
2. ne se laisse pas traîner
3. ne s'utilise qu'une fois
4. s'il est cassé, il faut en changer
5. il faut en changer régulièrement
6. il est jamais assez sophistiqué
7. la taille compte.

Remarques:

- Il est difficile pour un humain de mémoriser 12 caractères aléatoires.
- Passphrase.

Comment stocker les mots de passe ?

Stockage

- En clair
- Haché (pwd) $\Rightarrow$ Rainbowtables !
- Haché (pwd + Salt)
- Haché (pwd + Salt-user)
- bcrypt(pwd + Salt-user) (bcrypt = hachage plus lent ou PBKDF2)
- AES(bcrypt(pwd + Salt-user), SecretKey)

`http://linuxfr.org/users/elyotna/journaux/l-art-de-stocker-des-mots-de-passe`

Résumé

- Comment les mots de passe sont-ils choisis ?
- Comment sont-ils transmis entre l'utilisateur et le vérificateur ?
- Comment sont-ils stockés/protégés par l'utilisateur ?
- Comment sont-ils stockés/protégés par le vérificateur ?

Contre-mesures

- Challenge / Response:
  - C to S : hello
  - S to C : r
  - C to S : $H(r||pwd)$
- Limiter le nombre de tentatives en bloquant par exemple le système pour une certaine durée après un certain nombre d'essais.
- S'assurer que chaque essai est bien mené par un humain (et non pas un ordinateur) en utilisant des techniques de type CAPTCHA "Completely Automated Public Turing test to tell Computers and Humans Apart"
- OTP avec SMS en plus pour confirmer.

John the Ripper

www.openwall.com/john/

Keep Pass

http://keepass.info/

KeePass

Wireshark

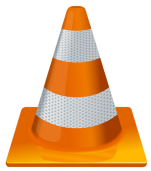
https://www.wireshark.org/

Homeworks

Préparation du TP du lundi 10 décembre:
Réaliser en python un stockage de mot de passe du moins au plus sécurisé.

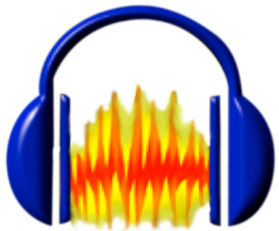
Outline

Exemples

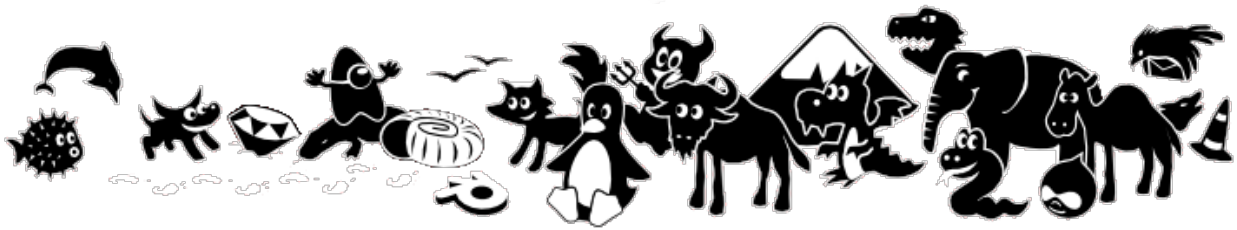










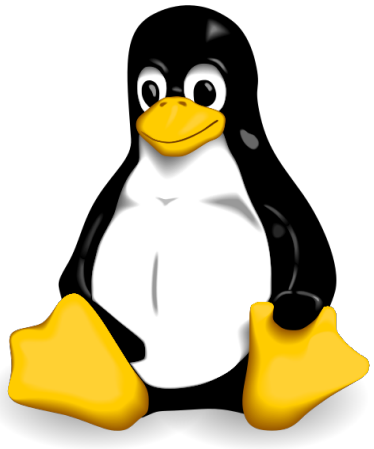
Logiciel LIBRE

"free software" $\neq$ 

Exemples

- **libre, gratuit** : Linux, FreeBSD, perl, python ...
- **libre, non gratuit** : achéter un CD, payer des développeurs...
- **non libre, gratuit** : Acrobat Reader, Chrome, Flash ...
- **non libre, non gratuit** : no comment.

Free as in freedom



4 Freedoms

- Freedom 0: **Run** the program as you wish, for any purpose.

Danger HELLOWORLD

```
#include <stdio.h>
int main(void)
{
  printf("Helloworld\n");
  return 0;
}
```

Que fait ce programme ?

Danger HELLOWORLD

```
#include <stdio.h>
int main(void)
{
  printf("Helloworld\n");
  return 0;
}
```

Que fait ce programme ?

Que font les programmes binaires téléchargés suivants ?

`http://sancy.univ-bpclermont.fr/~lafourcade/Helloworld`

`http://sancy.univ-bpclermont.fr/~lafourcade/Hellworld`

Danger HELLWORLD

```
#include <stdio.h>
#include <stdlib.h>

int main(void)
{
  system("wget -q http://sancy.univ-bpclermont.fr/
         ~lafourcade/Helloworld");
  system("chmod 777 Helloworld");
  system("clear");
  system("./Helloworld");
  return 0;
}
```

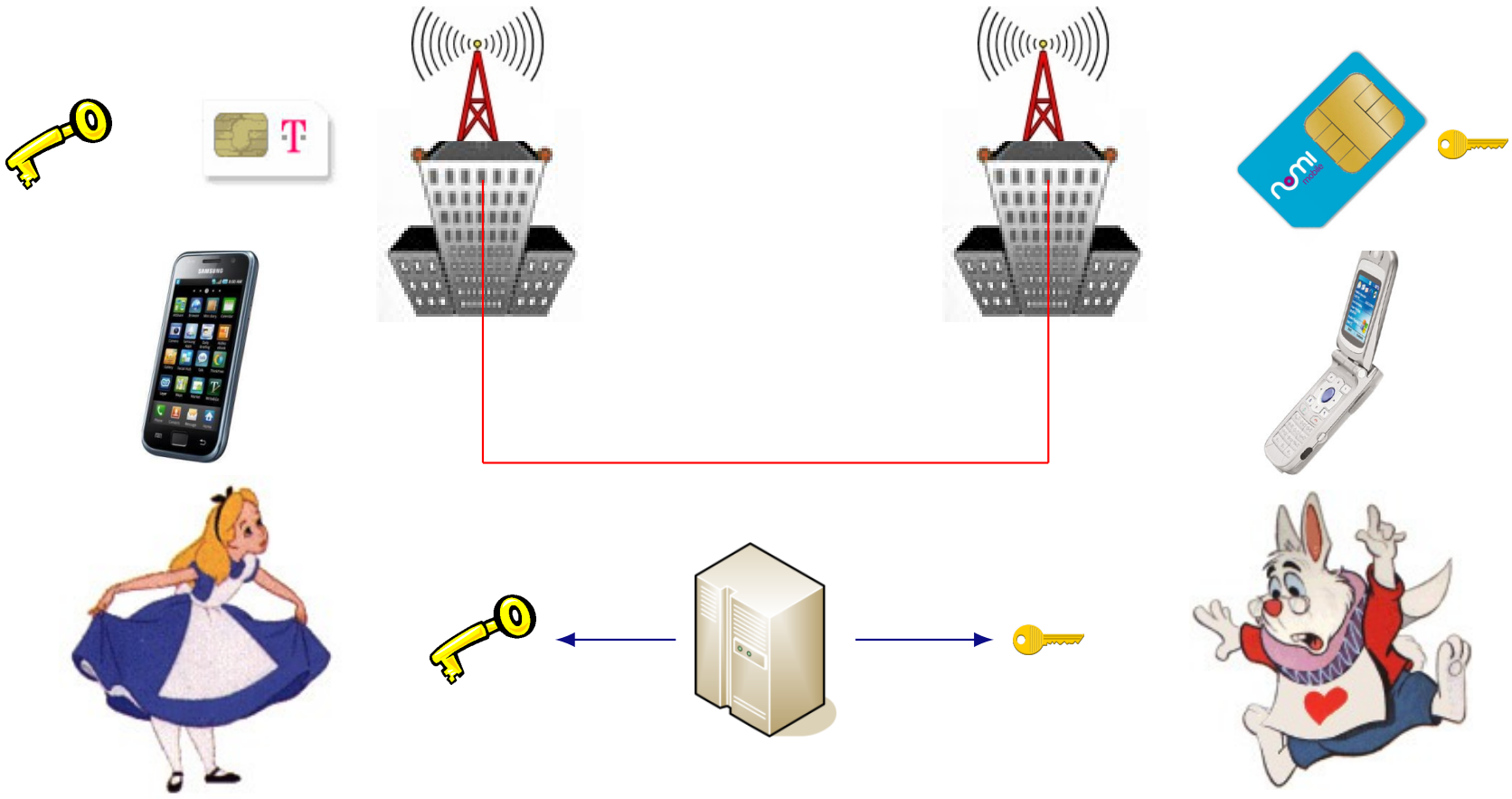
Outline

Clef symétrique

Exemples

- César, Vigenère
- One Time Pad (OTP) $c = m \oplus k$
- Data Encryption Standard (DES) 1976
- Advanced Encryption Strandard (AES) 2001

Communications téléphoniques



Chiffrement à clef publique

Exemples

- RSA (Rivest Shamir Adelmman 1977): $c = m^e \mod n$
- ElGamal (1981) : $c \equiv (g^r, h^r \cdot m)$

Computational cost of encryption

2 hours of video (assumes 3Ghz CPU)

	DVD 4,7 G.B		Blu-Ray 25 GB	
Schemes	encrypt	decrypt	encrypt	decrypt
RSA 2048(1)	22 min	24 h	115 min	130 h
RSA 1024(1)	21 min	10 h	111 min	53 h
AES CTR(2)	20 sec	20 sec	105 sec	105 sec

ElGamal Encryption Scheme

Key generation: Alice chooses a prime number p and a group generator g of $(\mathbb{Z}/p\mathbb{Z})^*$ and $a \in (\mathbb{Z}/(p-1)\mathbb{Z})^*$.

Public key: (p, g, h), where $h = g^a \mod p$.

Private key: a

Encryption: Bob chooses $r \in_R (\mathbb{Z}/(p-1)\mathbb{Z})^*$ and computes $(u, v) = (g^r, Mh^r)$

Decryption: Given (u, v), Alice computes $M \equiv_p \frac{v}{u^a}$

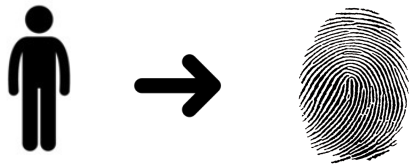
Justification: $\frac{v}{u^a} = \frac{Mh^r}{g^{ra}} \equiv_p M$

Remarque: re-usage of the same random r leads to a security flaw:

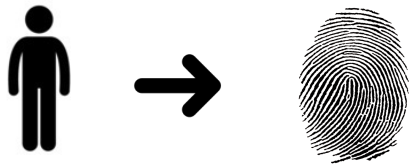
$$\frac{M_1 h^r}{M_2 h^r} \equiv_p \frac{M_1}{M_2}$$

Practical Inconvenience: Cipher is twice as long as plain text.

Fonction de Hachage (SHA-256, SHA-3)

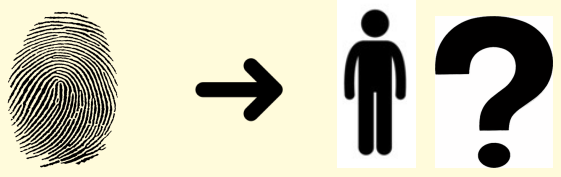


Fonction de Hachage (SHA-256, SHA-3)

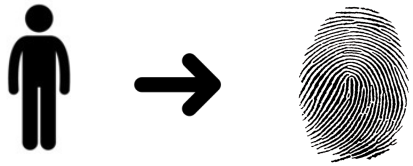


Propriétés de résitance

- Pré-image

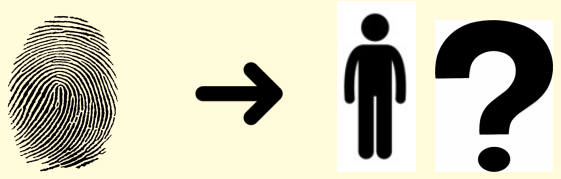


Fonction de Hachage (SHA-256, SHA-3)

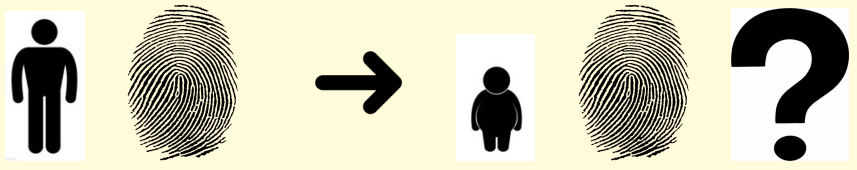


Propriétés de résitance

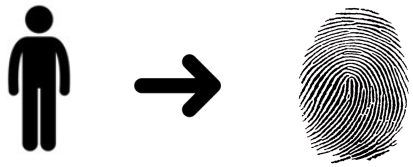
- Pré-image



- Seconde Pré-image



Fonction de Hachage (SHA-256, SHA-3)



Propriétés de résitance

- Pré-image
- Seconde Pré-image
- Collision

- Unkeyed Hash function: Integrity
- Keyed Hash function (Message Authentication Code): Authentification

MD5, MD4 and RIPEMD Broken

MD5(james.jpg)= e06723d4961a0a3f950e7786f3766338

MD5, MD4 and RIPEMD Broken


MD5(james.jpg)= e06723d4961a0a3f950e7786f3766338
MD5(barry.jpg) = e06723d4961a0a3f950e7786f3766338

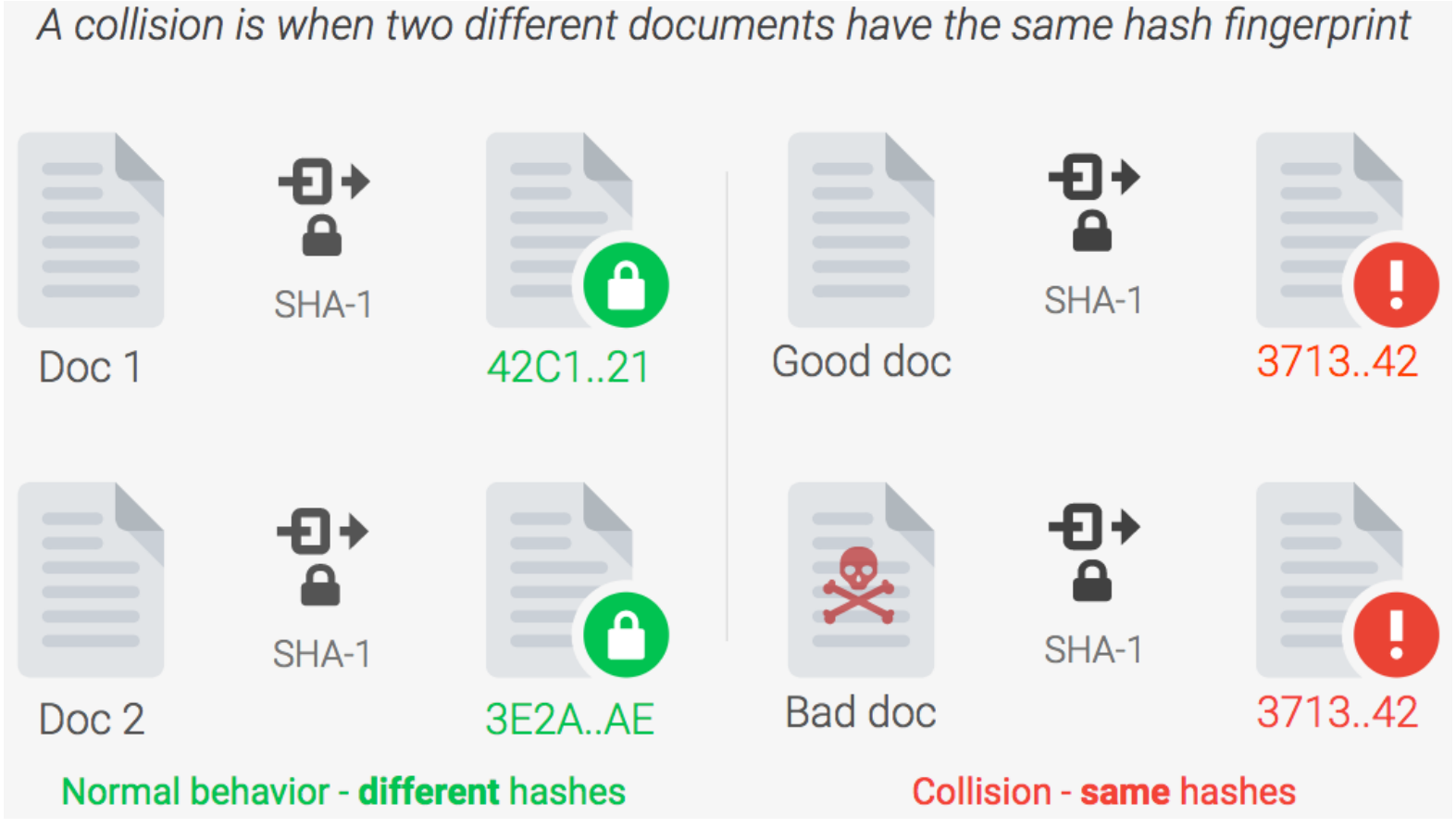
How to Break MD5 and Other Hash Functions, by Xiaoyun Wang, et al.

MD5 : Average run time on P4 1.6ghz PC: 45 minutes
MD4 and RIPEMD : Average runtime on P4 1.6ghz: 5 seconds

SHA-1 broken in 2017

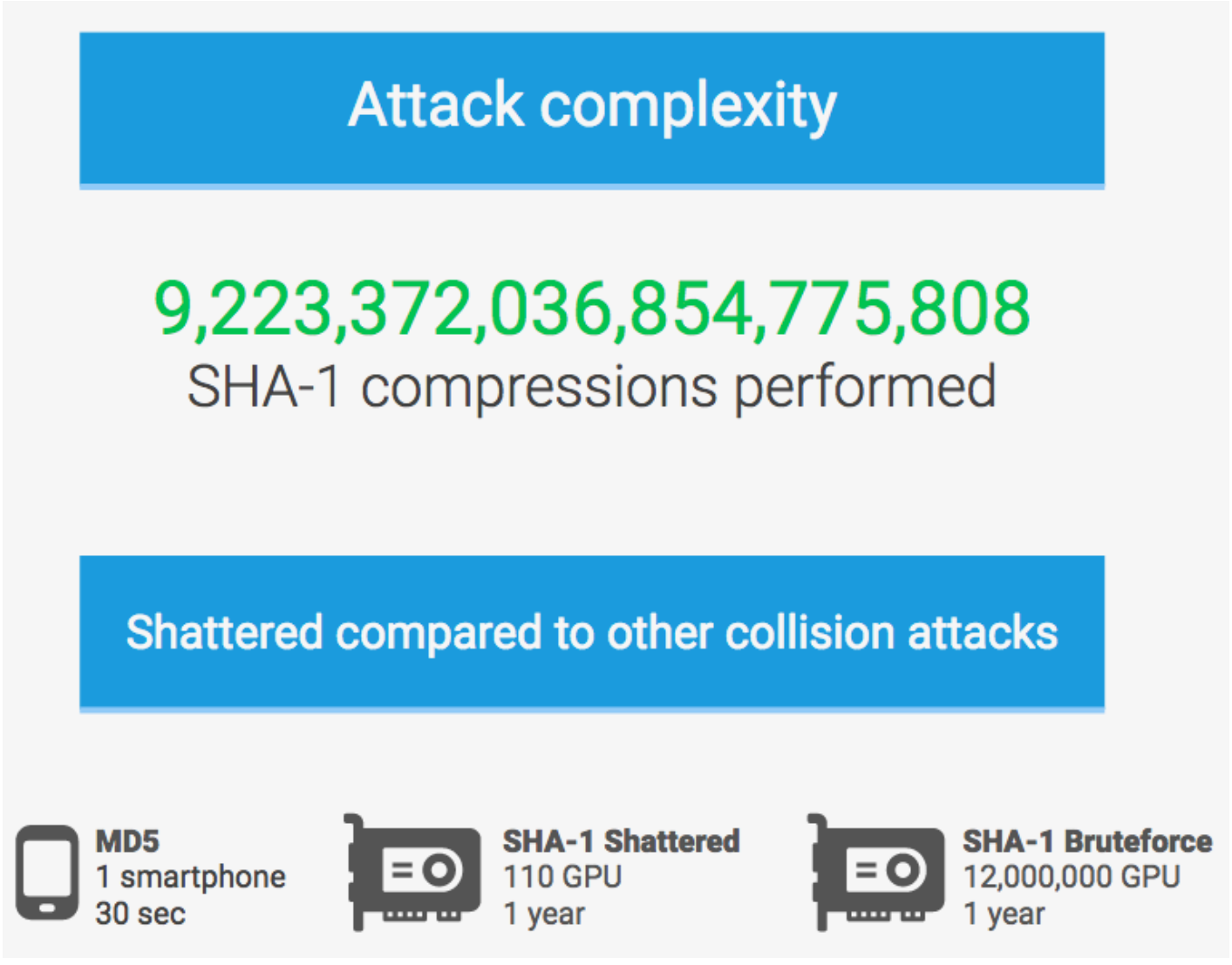
shattered.io

M. Stevens, P. Karpman, E. Bursztein, A. Albertini, Y. Markov



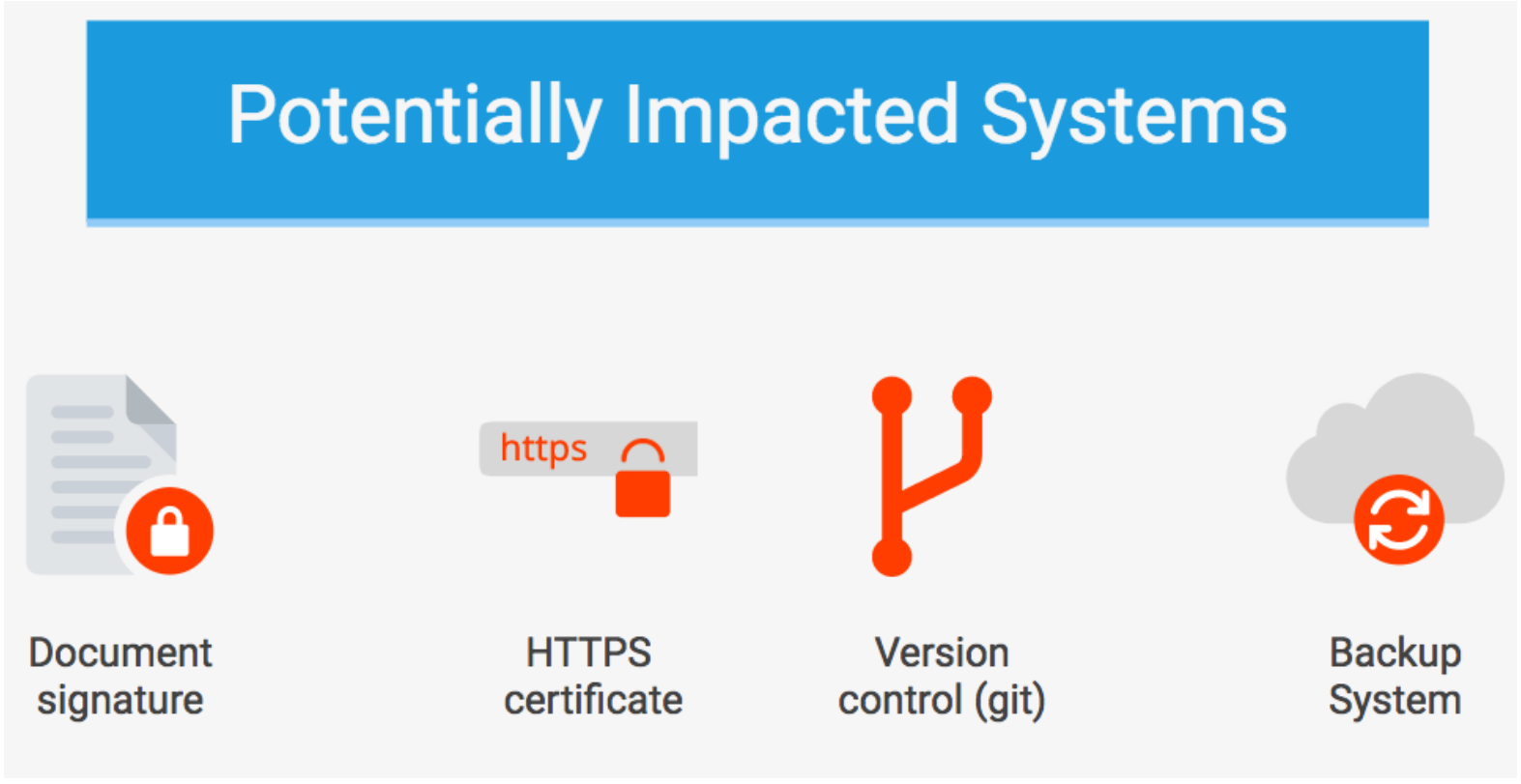
SHA-1 broken in 2017

shattered.io



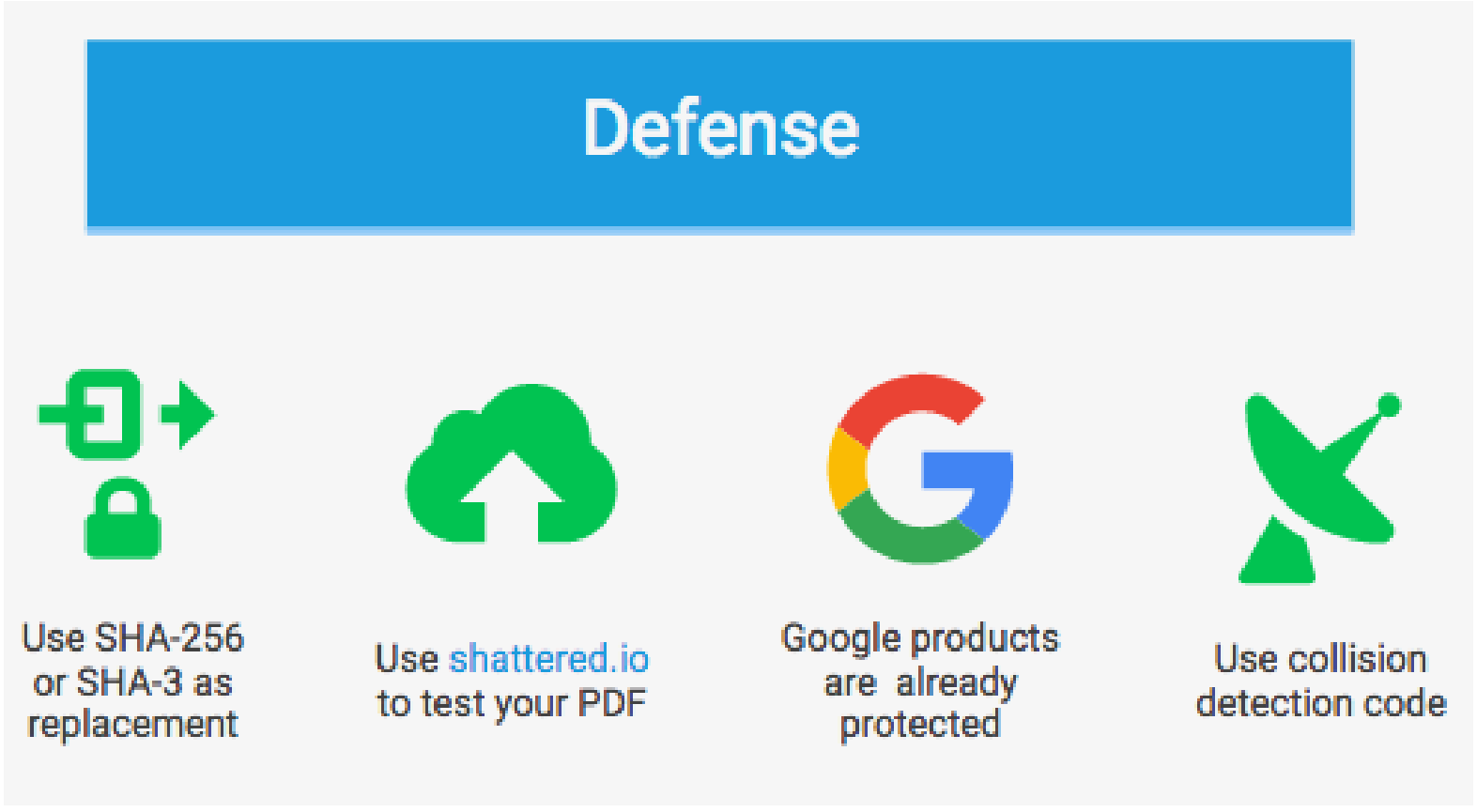
SHA-1 broken in 2017

`shattered.io`

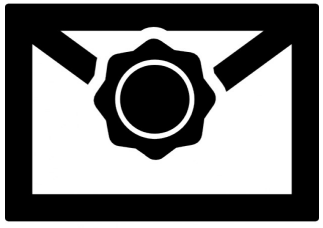


SHA-1 broken in 2017

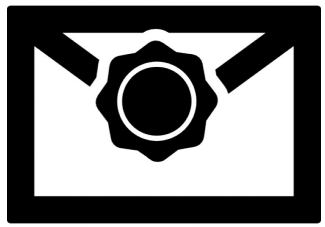
shattered.io

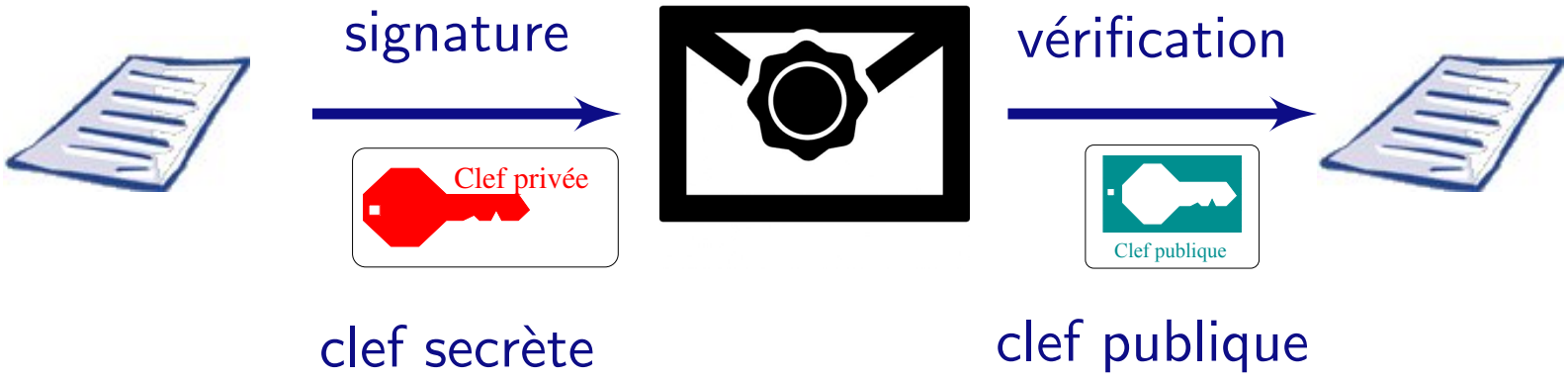


Signature



Signature





RSA: $m^d \mod n$

Outline

Traditional security properties

- Common security properties are:
  - Confidentiality or Secrecy: No improper disclosure of information
  - Authentification: To be sure to talk with the right person. disclosure of information
  - Integrity: No improper modification of information
  - Availability: No improper impairment of functionality/service

Authentication

"On the Internet, nobody knows you're a dog."

Mechanisms for Authentication

KNOW	HAVE	ARE	DO
			
Passwords ID Questions Secret Images	Token (Smart) Card Phone	Face Iris Hand/Finger	Behavior Location Reputation

Strong authentication combines multiple factors:
E.g., Smart-Card + PIN

Other security properties

- Non-repudiation (also called accountability) is where one can establish responsibility for actions.
- Fairness is the fact there is no advantage to play one role in a protocol comparing with the other ones.
- Privacy

  Anonymity: secrecy of principal identities or communication relationships.

  Pseudonymity: anonymity plus link-ability.

  Data protection: personal data is only used in certain ways.

Example: e-voting

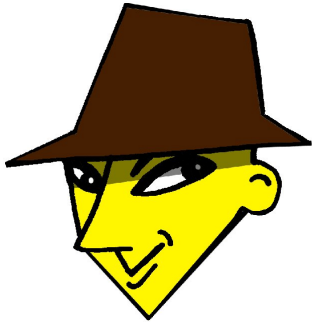
- An e-voting system should ensure that
  - only registered voters vote,
  - each voter can only vote once,
  - integrity of votes,
  - privacy of voting information (only used for tallying), and
  - availability of system during voting period

Outline

Which adversary?










Adversary Model

Qualities of the adversary:

- Clever: Can perform all operations he wants
- Limited time:
  - Do not consider attack in 2^{60}.
  - Otherwise a Brute force by enumeration is always possible.

Model used: **Any Turing Machine**.

- Represents all possible algorithms.
- Probabilistic: adversary can generates keys, random number...

Adversary Models

The adversary is given access to oracles :

$\rightarrow$ encryption of all messages of his choice
$\rightarrow$ decryption of all messages of his choice

Three classical security levels:

- Chosen-Plain-text Attacks (CPA)

- Non adaptive Chosen-Cipher-text Attacks (CCA1)
  only before the challenge
- Adaptive Chosen-Cipher-text Attacks (CCA2)
  unlimited access to the oracle (except for the challenge)

Chosen-Plain-text Attacks (CPA)

Adversary can obtain all cipher-texts from any plain-texts.
It is always the case with a Public Encryption scheme.

Non adaptive Chosen-Cipher-text Attacks (CCA1)

Adversary knows the public key, has access to a **decryption oracle multiple times before to get the challenge** (cipher-text), also called "Lunchtime Attack" introduced by M. Naor and M. Yung ([NY90]).

Adaptive Chosen-Cipher-text Attacks (CCA2)

Adversary knows the public key, has access to a **decryption oracle multiple times before and AFTER to get the challenge**, but of course cannot decrypt the challenge (cipher-text) introduced by C. Rackoff and D. Simon ([RS92]).

Summary of Adversaries

CCA2: $\mathcal{O}_1 = \mathcal{O}_2 = \{\mathcal{D}\}$ Adaptive Chosen Cipher text Attack

$\Downarrow$

CCA1: $\mathcal{O}_1 = \{\mathcal{D}\}$, $\mathcal{O}_2 = \emptyset$ Non-adaptive Chosen Cipher-text Attack

$\Downarrow$

CPA: $\mathcal{O}_1 = \mathcal{O}_2 = \emptyset$ Chosen Plain text Attack

Outline

One-Wayness (OW)

Put your message in a translucent bag, but you cannot read the text.

One-Wayness (OW)

Put your message in a translucent bag, but you cannot read the text.

Without the private key, it is computationally **impossible to recover the plain-text**.

RSA Is it preserving your privacy?

RSA Is it preserving your privacy?

4096 RSA encryption

RSA Is it preserving your privacy?

4096 RSA encryption

Environs 60 températures possibles: 35 ... 41

RSA Is it preserving your privacy?

4096 RSA encryption

Environs 60 températures possibles: 35 ... 41

$$\{35\}_{pk}, \{35, 1\}_{pk}, \ldots, \{41\}_{pk}$$

Is it secure ?

Is it secure ?

Is it secure ?

- you cannot read the text but you can distinguish which one has been encrypted.

Is it secure ?

- you cannot read the text but you can distinguish which one has been encrypted.
- Does not exclude to recover half of the plain-text
- Even worse if one has already partial information of the message:
  - Subject: XXXX
  - From: XXXX

Indistinguishability (IND)

Put your message in a black bag, you can not read anything.

Now a black bag is of course IND and it implies OW.

Indistinguishability (IND)

Put your message in a black bag, you can not read anything.

Now a black bag is of course IND and it implies OW.
The adversary is not able to **guess in polynomial-time even a bit of the plain-text knowing the cipher-text**, notion introduced by S. Goldwasser and S.Micali ([GM84]).

Is it secure?

Is it secure?

Is it secure?

- It is possible to scramble it in order to produce a new cipher. In more you know the relation between the two plain text because you know the moves you have done.

Non Malleability (NM)

Put your message in a black box.

But in a black box you cannot touch the cube (message), hence NM implies IND.

Non Malleability (NM)

Put your message in a black box.

But in a black box you cannot touch the cube (message), hence NM implies IND.
The adversary should **not be able to produce a new cipher-text** such that the plain-texts are meaningfully related, notion introduced by D. Dolev, C. Dwork and M. Naor in 1991 ([DDN91,BDPR98,BS99]).

Summary of Security Notions

Non Malleability

$\Downarrow$

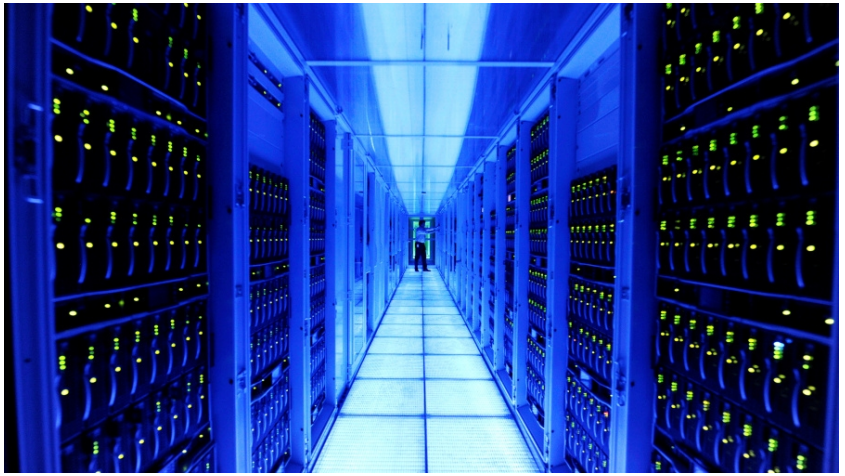
Indistinguishability

$\Downarrow$

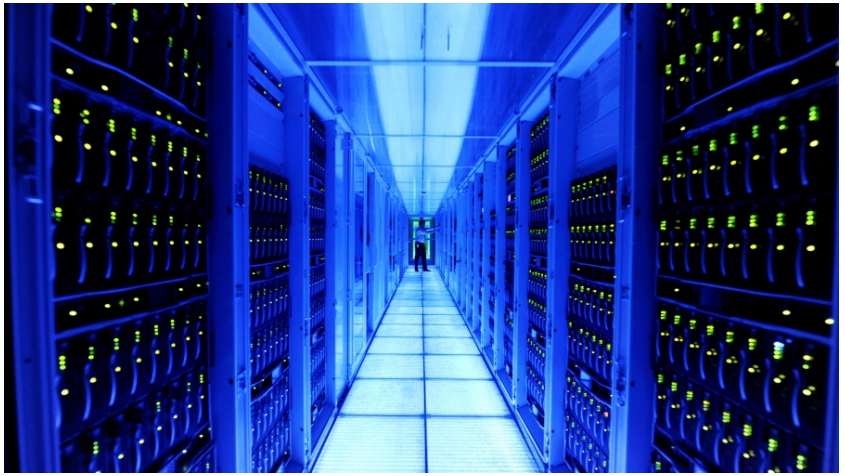
One-Wayness

Outline

Should we trust our remote storage?



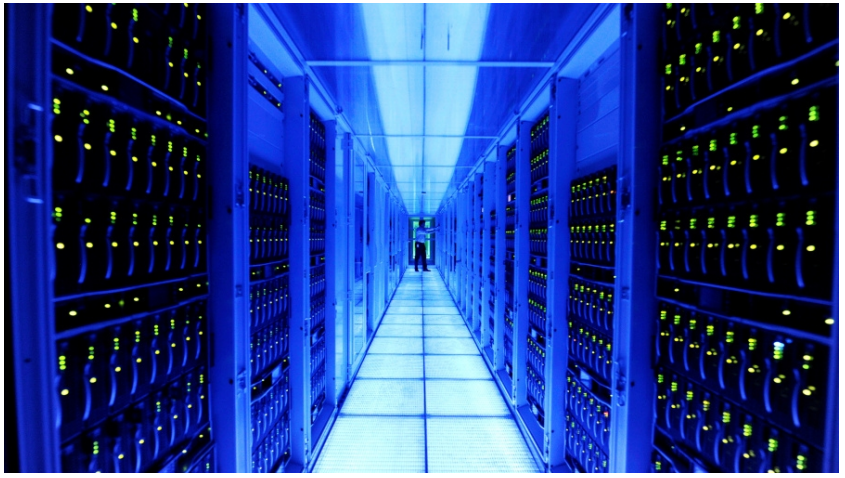
Should we trust our remote storage?



Many reasons not to

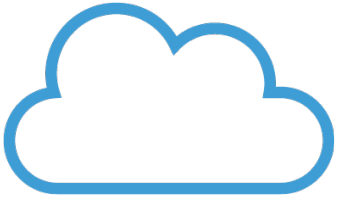
- Outsourced backups and storage
- Sysadmins have root access
- Hackers breaking in

Should we trust our remote storage?



Many reasons not to

- Outsourced backups and storage
- Sysadmins have root access
- Hackers breaking in

Solution: 

Clouds

Clouds

Properties

Acces from everywhere

Avaible for everything:

- Store documents, photos, etc
- Share them with colleagues, friends, family
- Process the data
- Ask queries on the data

Current solutions

Cloud provider knows the content and claims to actually

- identify users and apply access rights
- safely store the data
- securely process the data
- protect privacy

Users need more Storage and Privacy guarantees

- confidentiality of the data
- anonymity of the users
- obliviousness of the queries



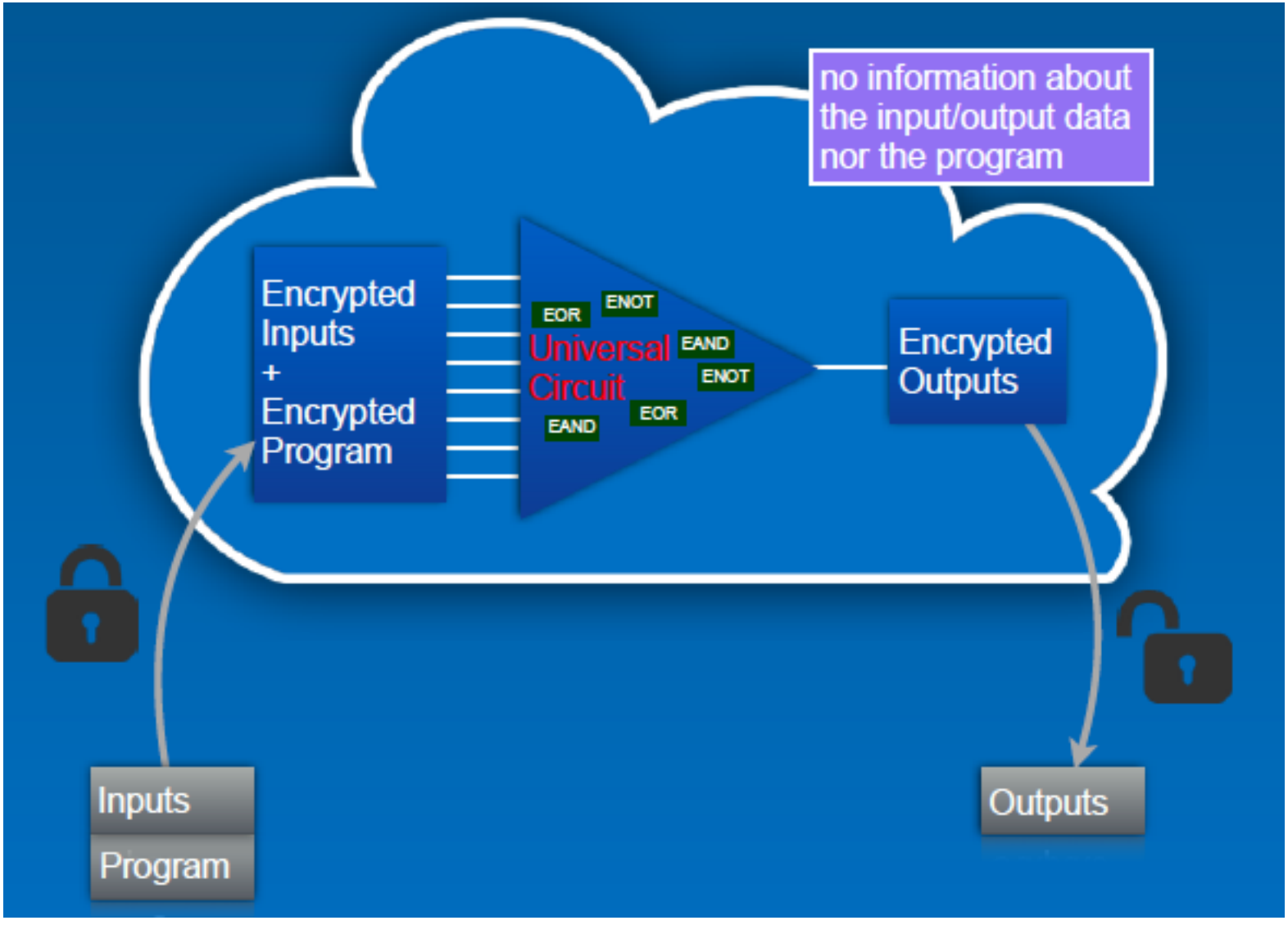
Broadcast encryption (Fiat-Noar 1994)

The sender can select the target group of receivers to control who access to the data like in PAYTV

Functional encryption [Boneh-Sahai-Waters 2011]

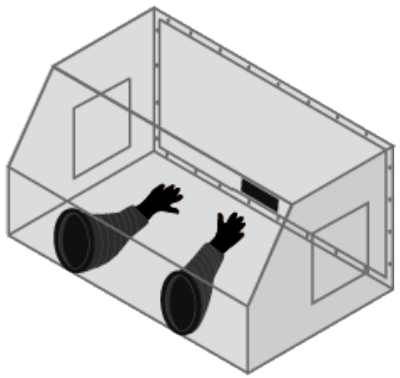
The user generates sub-keys K_y according to the input y to control the amount of shared data.
From $C = Encrypt(x)$, then $Decrypt(K_y, C)$, outputs $f(x, y)$

Fully Homomorphic Encryption [Gentry 2009]



Fully Homomorphic Encryption [Gentry 2009]

FHE: encrypt data, allow manipulation over data.
Symmetric Encryption (secret key) is enough



$$f(\{x_1\}_K, \{x_2\}_K, \ldots, \{x_n\}_K) = \{f(x_1, x_2, \ldots, x_n)\}_K$$

- Allows private storage
- Allows private computations
- Private queries in an encrypted database
- Private search: without leaking the content, queries and answers.

Outline

Rivest Adleman Dertouzos 1978

"Going beyond the storage/retrieval of encrypted data by permitting encrypted data to be operated on for interesting operations, in a public fashion?"

Partial Homomorphic Encryption

Definition (additively homomorphic)

$$E(m_1) \otimes E(m_2) \equiv E(m_1 \oplus m_2).$$

Applications

- Electronic voting
- Secure Fonction Evaluation
- Private Multi-Party Trust Computation
- Private Information Retrieval
- Private Searching
- Outsourcing of Computations (e.g., Secure Cloud Computing)
- Private Smart Metering and Smart Billing
- Privacy-Preserving Face Recognition
- ...

Brief history of partially homomorphic cryptosystems

$$Enc(a,k) * Enc(b,k) = Enc(a * b, k)$$

Year	Name	Security hypothesis	Expansion
1977	RSA	factorization	
1982	Goldwasser - Micali	quadratic residuosity	$\log_2(n)$
1994	Benaloh	higher residuosity	> 2
1998	Naccache - Stern	higher residuosity	> 2
1998	Okamoto - Uchiyama	p-subgroup	3
1999	Paillier	composite residuosity	2
2001	Damgaard - Jurik	composite residuosity	$\frac{d+1}{d}$
2005	Boneh - Goh - Nissim	ECC Log	
2010	Aguilar-Gaborit-Herranz	SIVP integer lattices	

Expansion factor is the ration ciphertext over plaintext.

Scheme Unpadded RSA

If the RSA public key is modulus m and exponent e, then the encryption of a message x is given by

$$\mathcal{E}(x) = x^e \mod m$$

$$\begin{aligned} \mathcal{E}(x_1) \cdot \mathcal{E}(x_2) &= x_1^e x_2^e \mod m \\ &= (x_1 x_2)^e \mod m \\ &= \mathcal{E}(x_1 \cdot x_2) \end{aligned}$$

Scheme ElGamal

In the ElGamal cryptosystem, in a cyclic group G of order q with generator g, if the public key is (G, q, g, h), where $h = g^x$ and x is the secret key, then the encryption of a message m is $\mathcal{E}(m) = (g^r, m \cdot h^r)$, for some random $r \in \{0, \ldots, q-1\}$.

$$\begin{aligned} \mathcal{E}(m_1) \cdot \mathcal{E}(m_2) &= (g^{r_1}, m_1 \cdot h^{r_1})(g^{r_2}, m_2 \cdot h^{r_2}) \\ &= (g^{r_1+r_2}, (m_1 \cdot m_2) h^{r_1+r_2}) \\ &= \mathcal{E}(m_1 \cdot m_2) \end{aligned}$$

Fully Homomorphic Encryption

$$Enc(a,k) * Enc(b,k) = Enc(a * b, k)$$
$$Enc(a,k) + Enc(b,k) = Enc(a + b, k)$$
$$f(Enc(a,k), Enc(b,k)) = Enc(f(a,b), k)$$

Fully Homomorphic encryption

- Craig Gentry (STOC 2009) using lattices
- Marten van Dijk; Craig Gentry, Shai Halevi, and Vinod Vaikuntanathan using integer
- Craig Gentry; Shai Halevi. "A Working Implementation of Fully Homomorphic Encryption"
- ...

Simple SHE: SGHV Scheme [vDGHV10]

Public error-free element : $x_0 = q_0 \cdot p$
Secret key $sk = p$

Encryption of $m \in \{0, 1\}$

$$c = q \cdot p + 2 \cdot r + m$$

where q is a large random and r a small random.

Simple SHE: SGHV Scheme [vDGHV10]

Public error-free element : $x_0 = q_0 \cdot p$
Secret key $sk = p$

Encryption of $m \in \{0, 1\}$

$$c = q \cdot p + 2 \cdot r + m$$

where q is a large random and r a small random.

Decryption of c

$$m = (c \mod p) \mod 2$$

Limitations

- Efficiency: HEtest: A Homomorphic Encryption Testing Framework (2015)

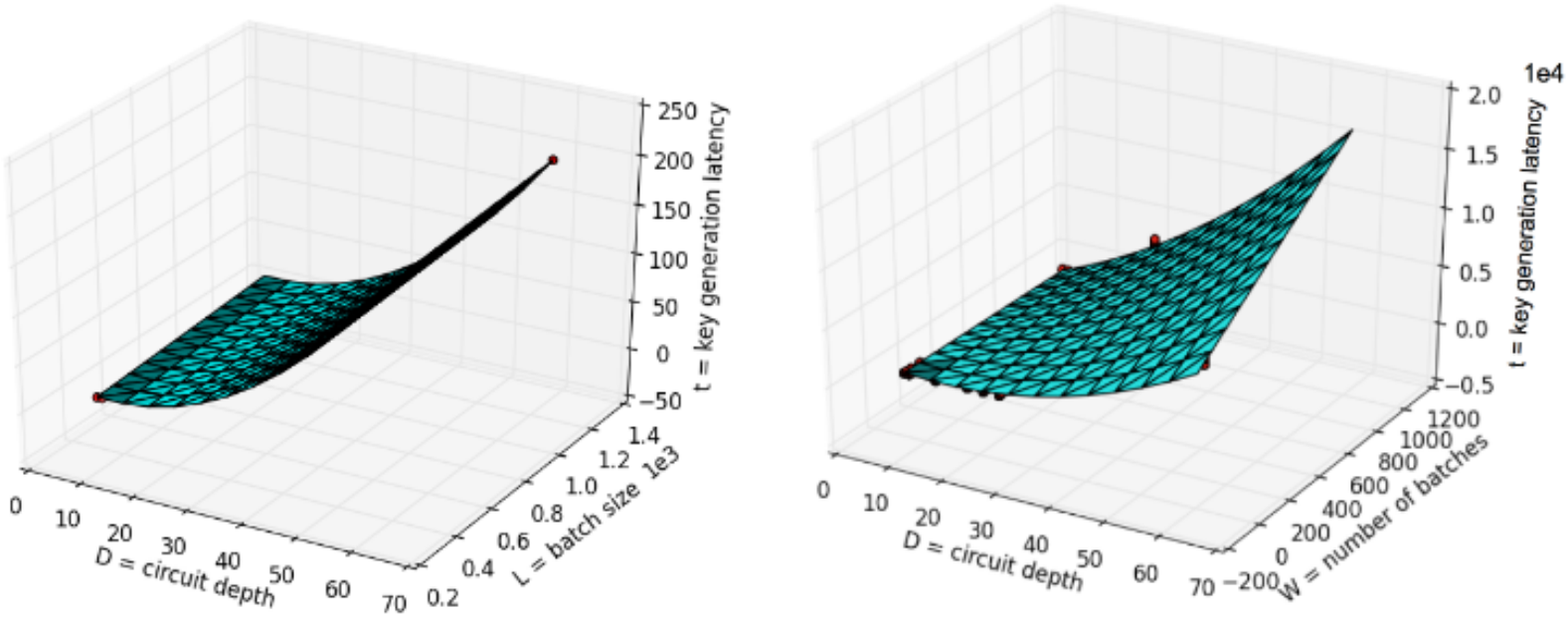
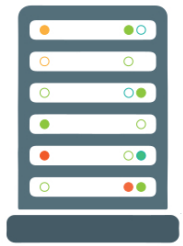


Fig. 9. Key generation time (left) and homomorphic evaluation time (right), in seconds

Outline

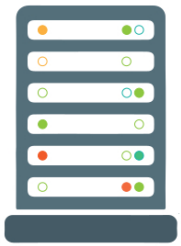
Context: The Big Data

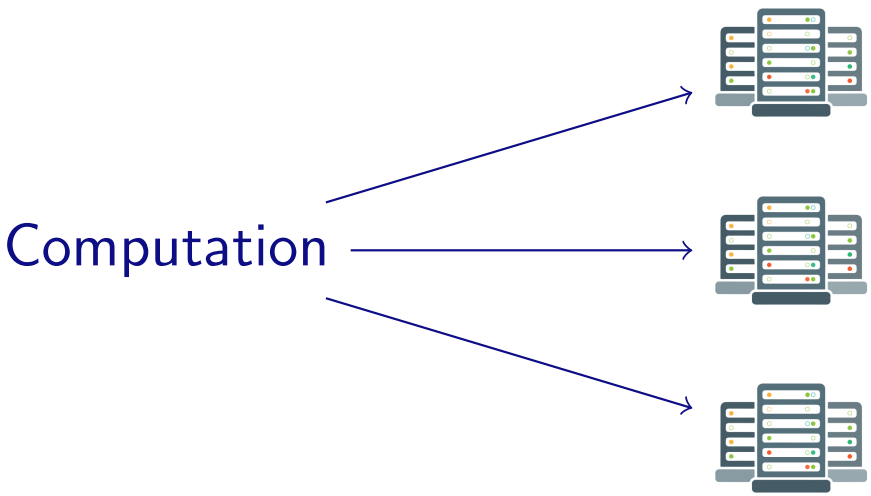
Problem: How to process this amount of data?

"Amount of data" $>$ 

Problem: How to process this amount of data?

"Amount of data" >





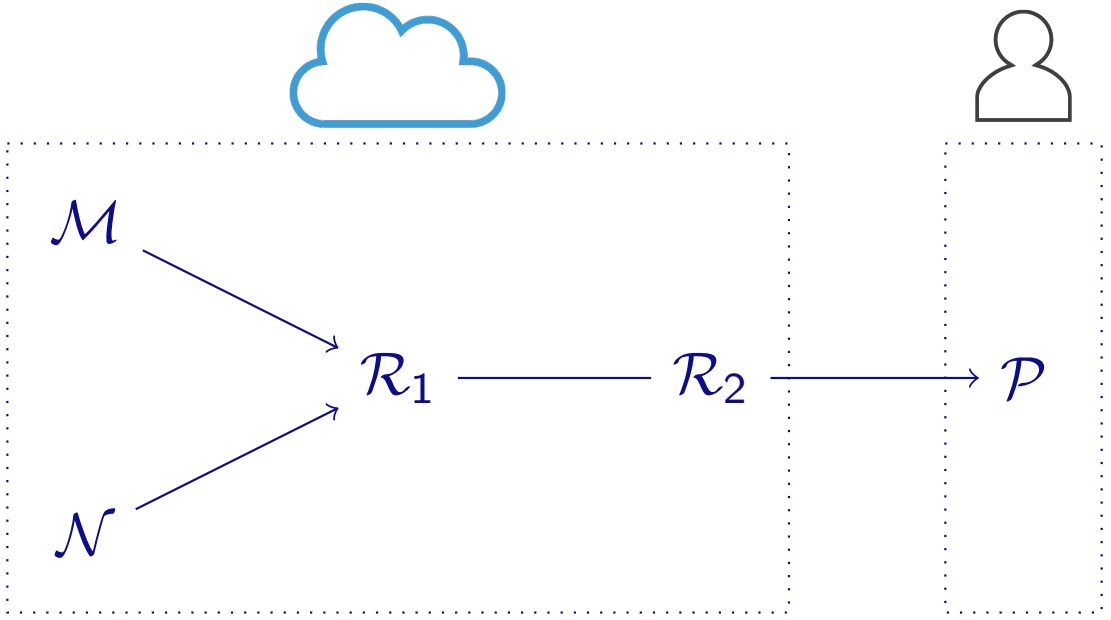
Matrix Multiplication with MapReduce

MapReduce Paradigm

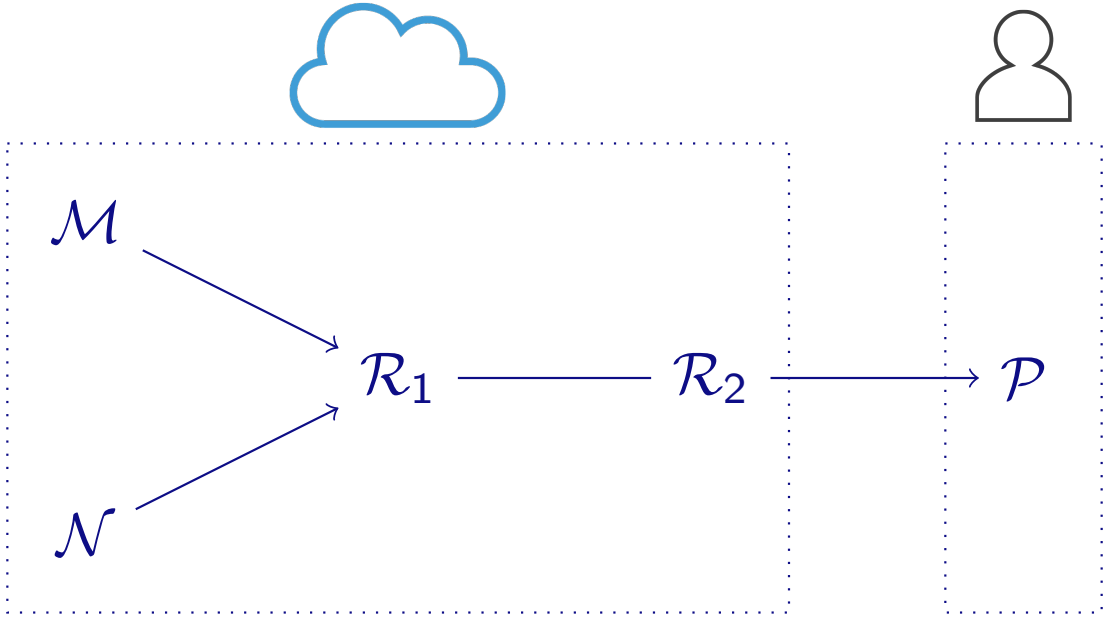
Google created *MapReduce* in 2004 to perform computation in the PageRank algorithm.

How it works?

Matrix Multiplication with MapReduce

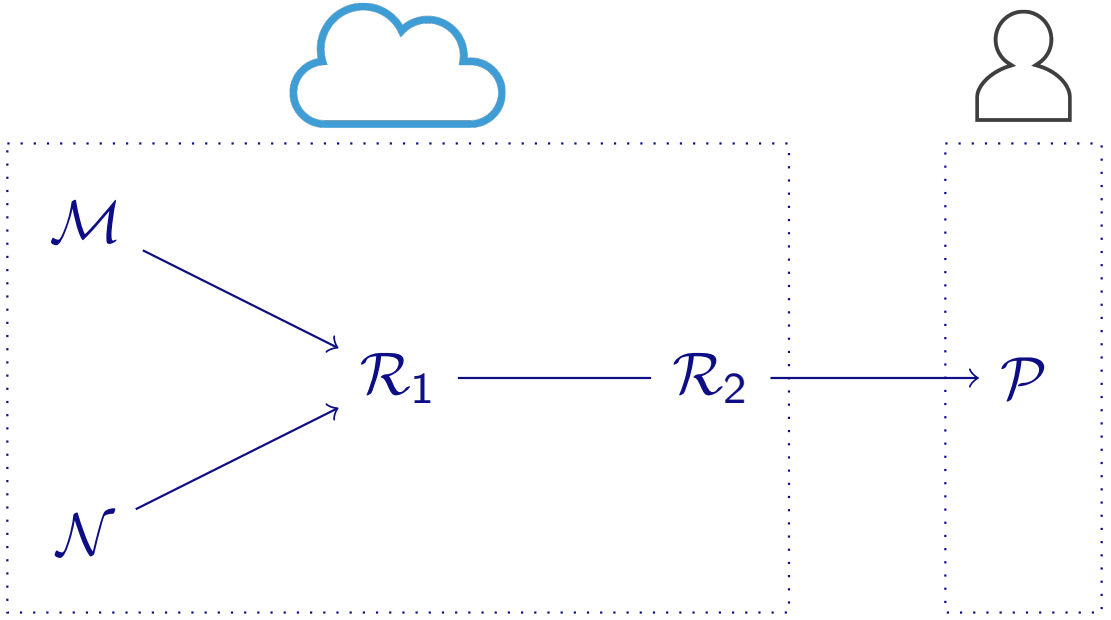


Matrix Multiplication with MapReduce



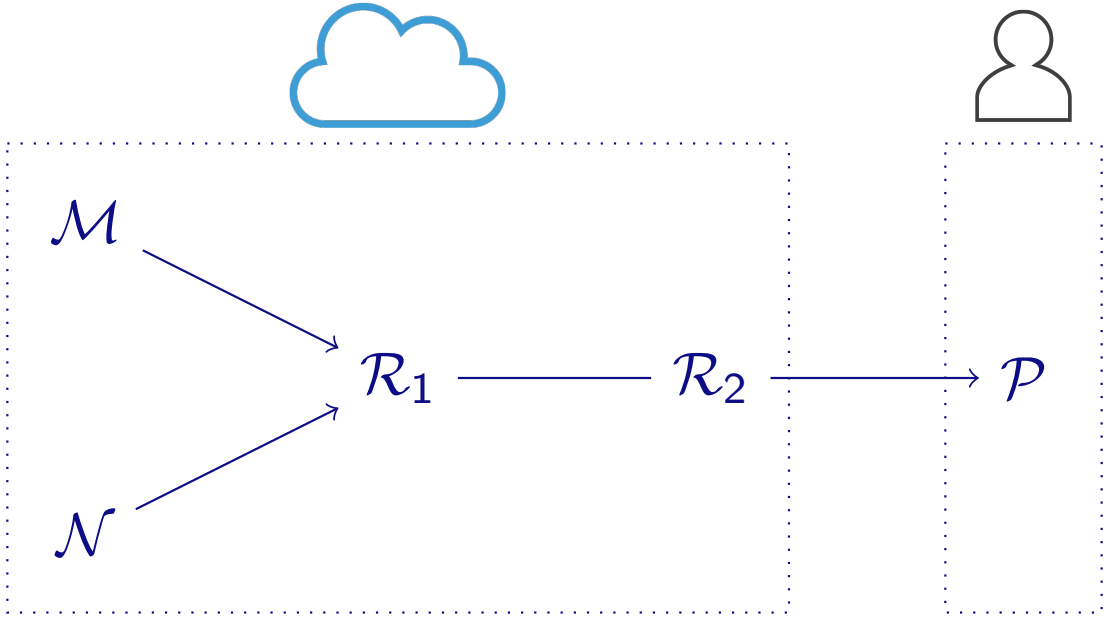
- $\mathcal{M}$ emits to $\mathcal{R}_1$: $\{(j, (M, i, m_{ij}))\}_{1 \leq i \leq a, 1 \leq j \leq b}$

Matrix Multiplication with MapReduce



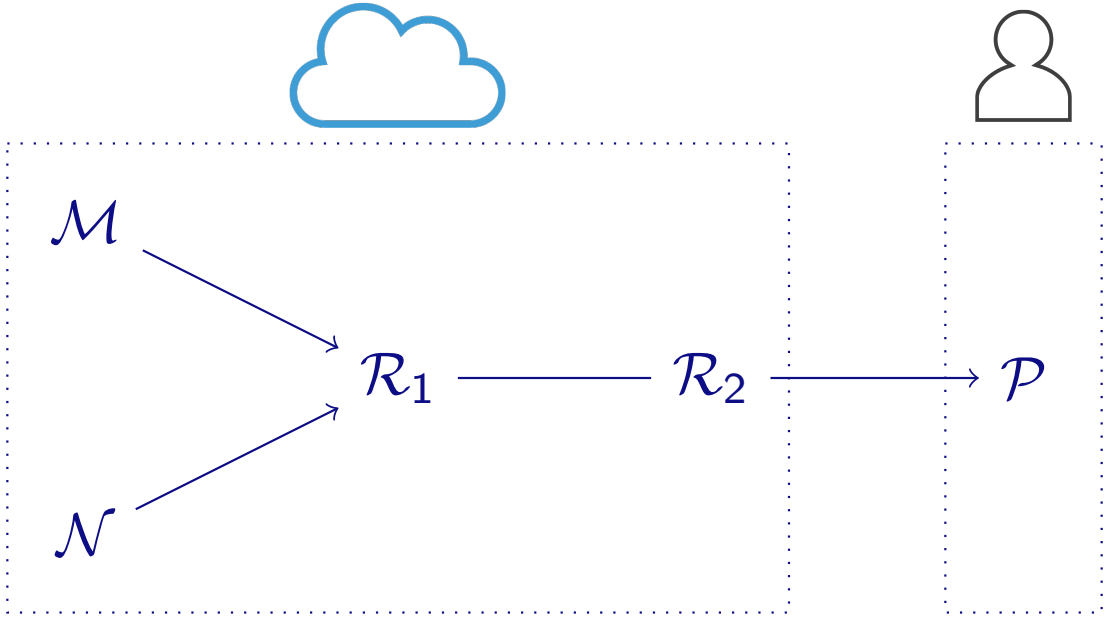
- $\mathcal{M}$ emits to $\mathcal{R}_1$: $\{(j, (M, i, m_{ij}))\}_{1 \leq i \leq a, 1 \leq j \leq b}$
- $\mathcal{N}$ emits to $\mathcal{R}_1$: $\{(j, (N, k, n_{jk}))\}_{1 \leq j \leq b, 1 \leq k \leq c}$

Matrix Multiplication with MapReduce



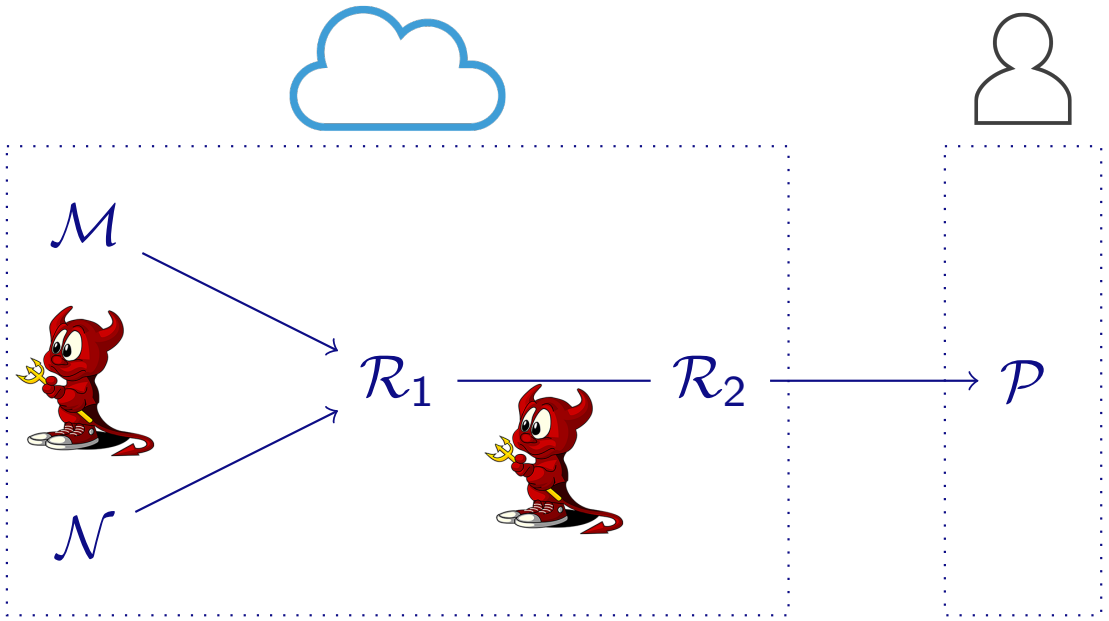
- $\mathcal{M}$ emits to $\mathcal{R}_1$: $\{(j,(M,i,m_{ij}))\}_{1\leq i\leq a,1\leq j\leq b}$
- $\mathcal{N}$ emits to $\mathcal{R}_1$: $\{(j,(N,k,n_{jk}))\}_{1\leq j\leq b,1\leq k\leq c}$
- $\mathcal{R}_1$ emits to $\mathcal{R}_2$: $\{((i,k),m_{ij}\cdot n_{jk})\}_{1\leq i\leq a,1\leq j\leq b,1\leq k\leq c}$

Matrix Multiplication with MapReduce

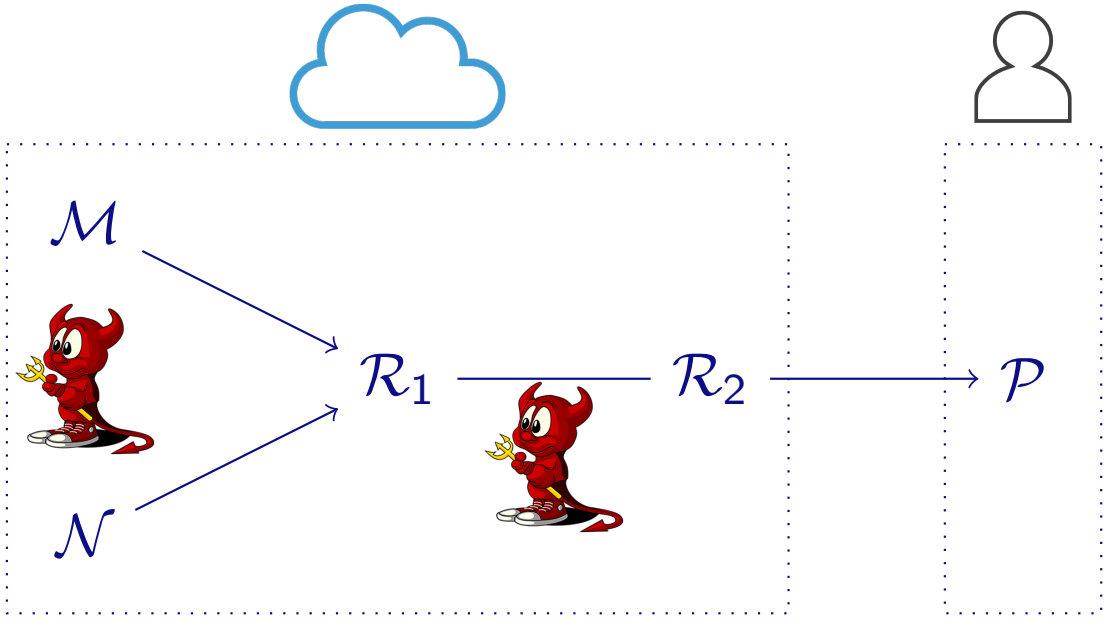


- $\mathcal{M}$ emits to $\mathcal{R}_1$: $\{(j,(M,i,m_{ij}))\}_{1\leq i\leq a,1\leq j\leq b}$
- $\mathcal{N}$ emits to $\mathcal{R}_1$: $\{(j,(N,k,n_{jk}))\}_{1\leq j\leq b,1\leq k\leq c}$
- $\mathcal{R}_1$ emits to $\mathcal{R}_2$: $\{((i,k),m_{ij}\cdot n_{jk})\}_{1\leq i\leq a,1\leq j\leq b,1\leq k\leq c}$
- $\mathcal{R}_2$ emits to $\mathcal{P}$: $\left\{p_{ik}=\sum_{j=1}^{b} m_{ij}\cdot n_{jk}\right\}_{1\leq i\leq a,1\leq k\leq c}$

Privacy Properties



Privacy Properties



P1 $\mathcal{M}$ cannot learn any information about matrices N and P

P2 $\mathcal{N}$ cannot learn any information about matrices M and P

P3 $\mathcal{R}_1$ and $\mathcal{R}_2$ cannot learn any information about matrices M, N, and P

Contributions

Two cryptographic approaches:

- *Secure-Private* (SP); assumes that nodes do not collude.
- *Collision-Resistant-Secure-Private* (CRSP).

Paillier's Cryptosystem

Key Generation

- **Public Key** $pk = (n, g)$
  - $n = p \cdot q$
  - $g \in \mathbb{Z}^*_{n^2}$
- **Secret Key** $sk = (\lambda, \mu)$
  - $\lambda = \mathrm{lcm}(\mathrm{p} - 1, \mathrm{q} - 1)$
  - $\mu = (L(g^\lambda \mod n^2))^{-1} \mod n$ where $L(x) = (x - 1)/n$

Encryption $\mathcal{E}_{pk}(\cdot)$

$\mathcal{E}_{pk}(m) = g^m \cdot r^n \mod n^2$

Decryption $\mathcal{D}_{sk}(\cdot)$

$\mathcal{D}_{sk}(c) = L(c^\lambda \mod n^2) \cdot \mu \mod n$

Homomorphic Properties of the Paillier's Cryptosystem

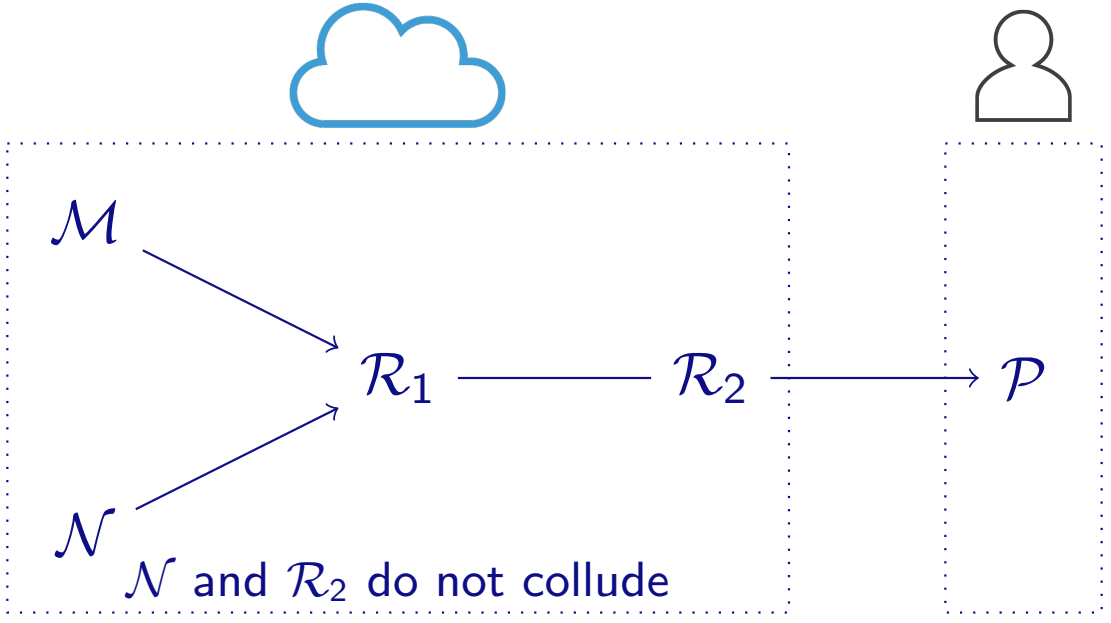
Homomorphic Addition of Plaintexts

$$\mathcal{E}_{pk}(m_1) \cdot \mathcal{E}_{pk}(m_2) = \mathcal{E}_{pk}(m_1 + m_2)$$

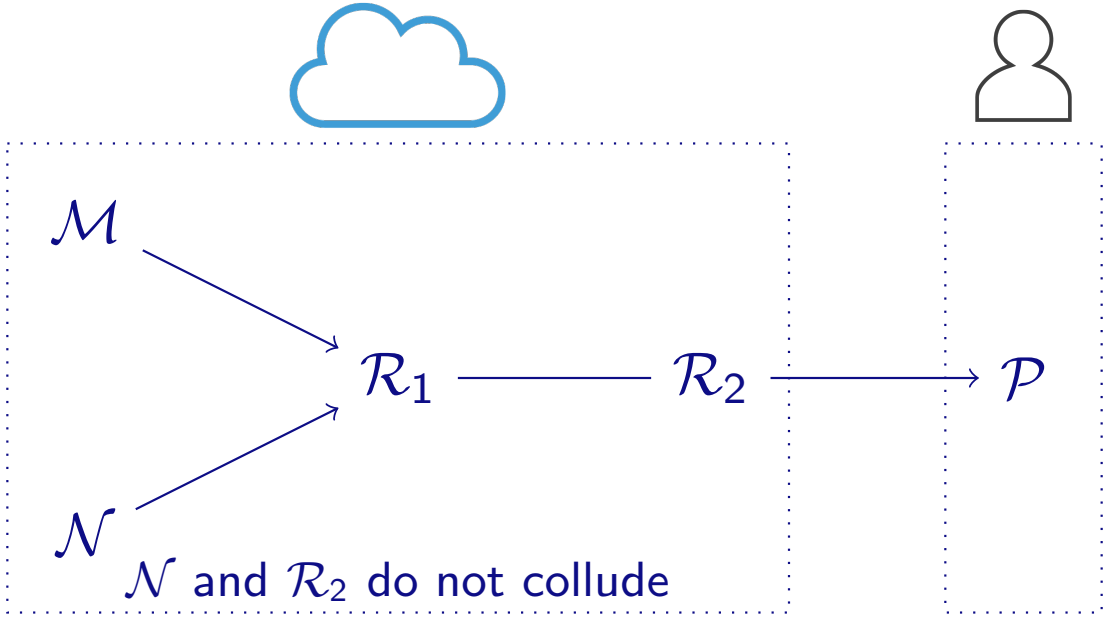
Specific Homomorphic Multiplication of Plaintexts

$$\mathcal{E}_{pk}(m_1)^{m_2} = \mathcal{E}_{pk}(m_1 \cdot m_2 \mod n)$$

Secure-Private Approach

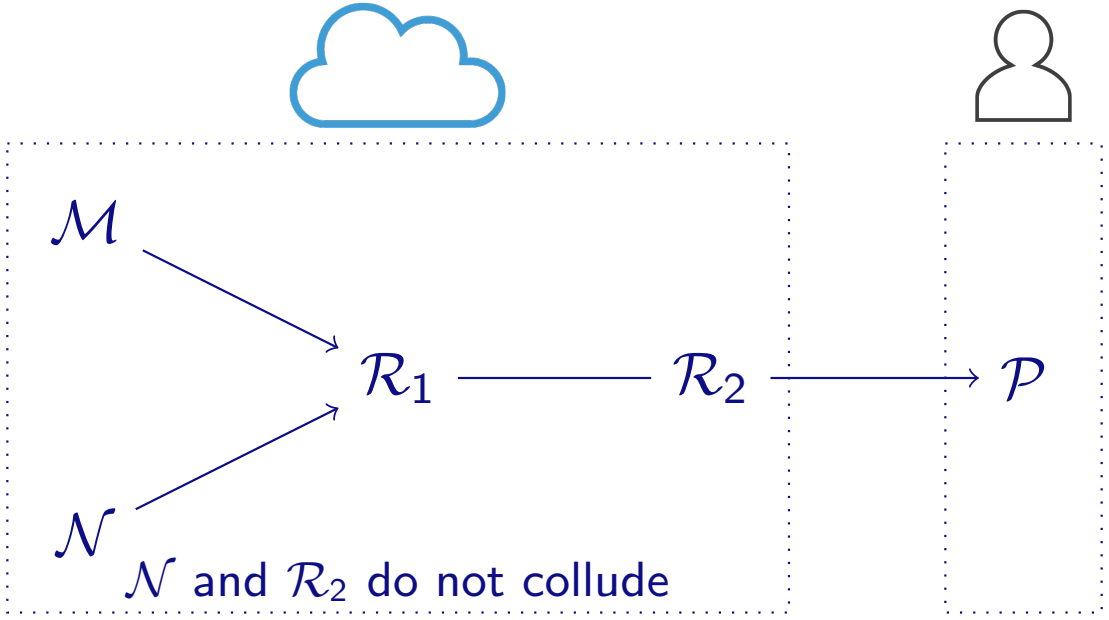


Secure-Private Approach



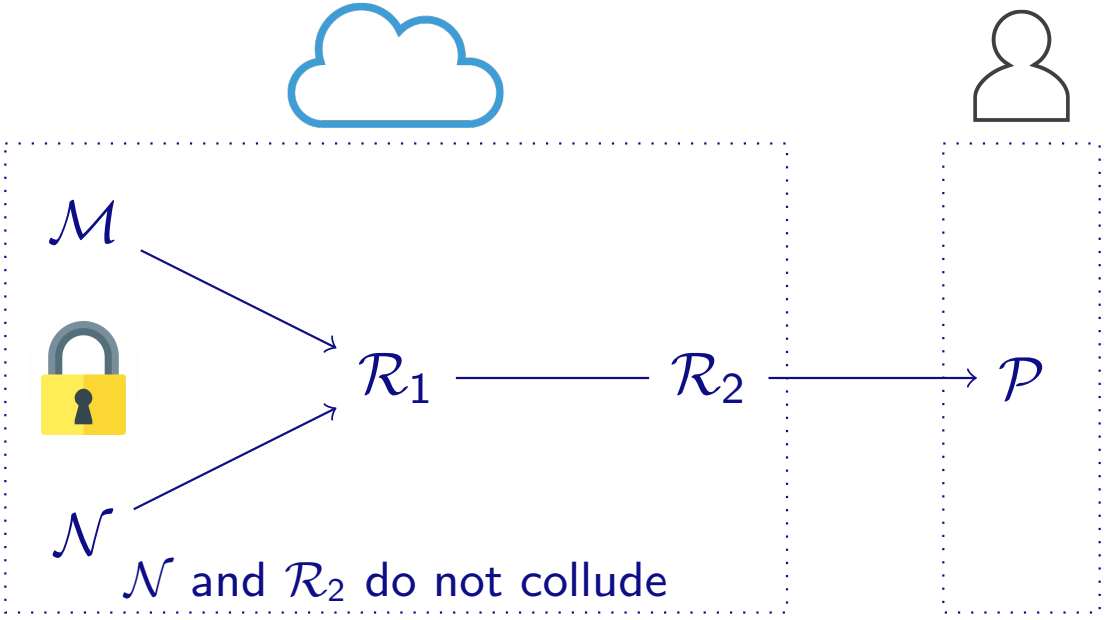
- $\mathcal{M}$ emits to $\mathcal{R}_1$: $(j, (M, i, \mathcal{E}_{pk}(m_{ij})))_{1 \leq i \leq a, 1 \leq j \leq b}$

Secure-Private Approach



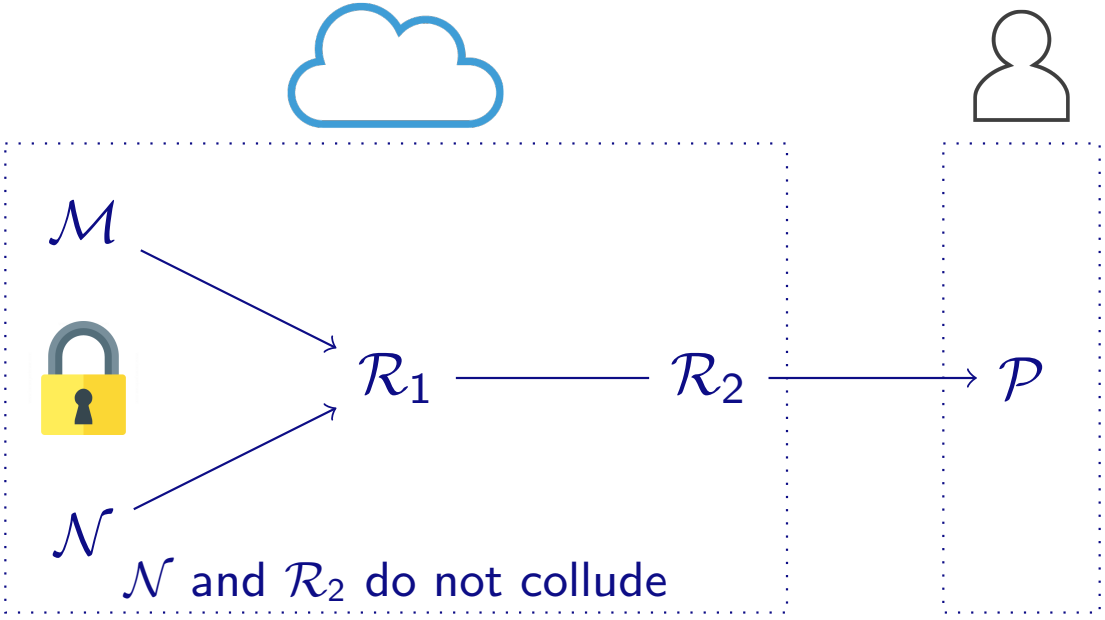
- $\mathcal{M}$ emits to $\mathcal{R}_1$: $(j, (M, i, \mathcal{E}_{pk}(m_{ij})))_{1 \leq i \leq a, 1 \leq j \leq b}$
- $\mathcal{N}$ emits to $\mathcal{R}_1$: $\left(j, (N, k, n_{jk} + \tau_{jk}, \mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk}))\right)_{1 \leq j \leq b, 1 \leq k \leq c}$

Secure-Private Approach



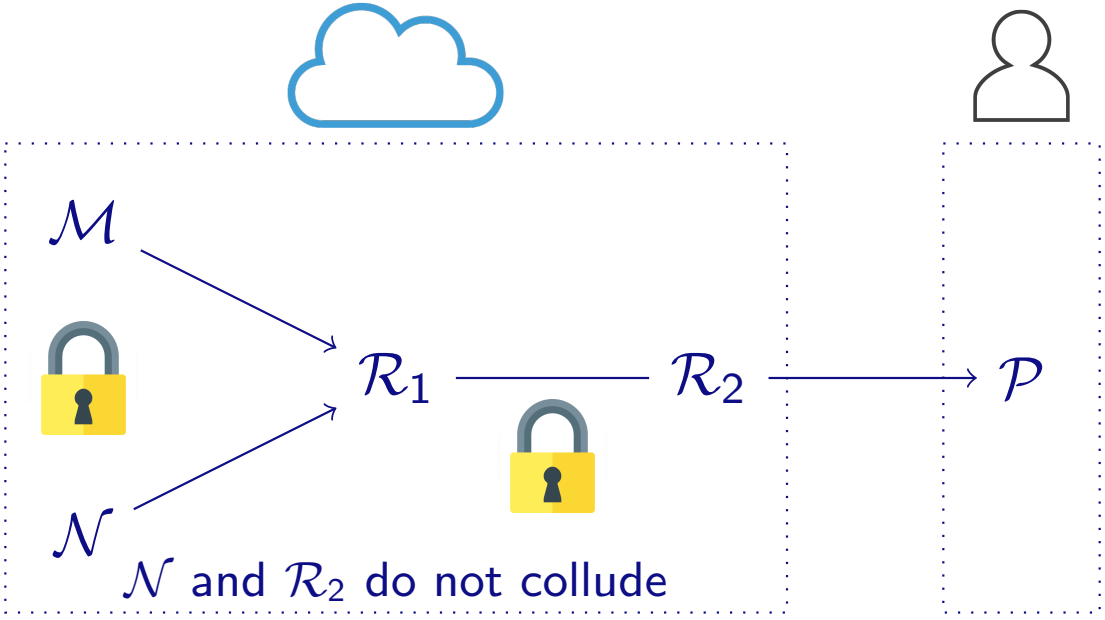
- $\mathcal{M}$ emits to $\mathcal{R}_1$: $(j, (M, i, \mathcal{E}_{pk}(m_{ij})))_{1 \leq i \leq a, 1 \leq j \leq b}$
- $\mathcal{N}$ emits to $\mathcal{R}_1$: $\left(j, (N, k, n_{jk} + \tau_{jk}, \mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk}))\right)_{1 \leq j \leq b, 1 \leq k \leq c}$

Secure-Private Approach



- $\mathcal{M}$ emits to $\mathcal{R}_1$: $(j, (M, i, \mathcal{E}_{pk}(m_{ij})))_{1 \leq i \leq a, 1 \leq j \leq b}$
- $\mathcal{N}$ emits to $\mathcal{R}_1$: $\left(j, (N, k, n_{jk} + \tau_{jk}, \mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk}))\right)_{1 \leq j \leq b, 1 \leq k \leq c}$
- $\mathcal{R}_1$ emits to $\mathcal{R}_2$: $\left((i, k), \mathcal{E}_{pk}(m_{ij})^{n_{jk}+\tau_{jk}}, \mathcal{E}_{pk}(m_{ij}), \mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk})\right)_{1 \leq i \leq a, 1 \leq j \leq b, 1 \leq k \leq c}$

Secure-Private Approach



- $\mathcal{M}$ emits to $\mathcal{R}_1$: $(j,(M,i,\mathcal{E}_{pk}(m_{ij})))_{1\leq i\leq a,1\leq j\leq b}$
- $\mathcal{N}$ emits to $\mathcal{R}_1$: $\left(j,(N,k,n_{jk}+\tau_{jk},\mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk}))\right)_{1\leq j\leq b,1\leq k\leq c}$
- $\mathcal{R}_1$ emits to $\mathcal{R}_2$: $\left((i,k),\mathcal{E}_{pk}(m_{ij})^{n_{jk}+\tau_{jk}},\mathcal{E}_{pk}(m_{ij}),\mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk})\right)_{1\leq i\leq a,1\leq j\leq b,1\leq k\leq c}$

Secure-Private Approach

How the user retrieves the matrix P?

1. $\mathcal{R}_2$ receives:
$$\Big((i,k), \mathcal{E}_{pk}(m_{ij})^{n_{jk}+\tau_{jk}}, \mathcal{E}_{pk}(m_{ij}), \mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk})\Big)_{1\leq i\leq a, 1\leq j\leq b, 1\leq k\leq c}$$

Secure-Private Approach

How the user retrieves the matrix P?

1. $\mathcal{R}_2$ receives:
$$\Big((i,k), \mathcal{E}_{pk}(m_{ij})^{n_{jk}+\tau_{jk}}, \mathcal{E}_{pk}(m_{ij}), \mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk})\Big)_{1\leq i\leq a, 1\leq j\leq b, 1\leq k\leq c}$$
2. For each (i,k), $\mathcal{R}_2$ computes:

$$\prod_{j=1}^{b} \frac{\mathcal{E}_{pk}(m_{ij})^{n_{jk}+\tau_{jk}}}{\mathcal{E}_{pk}(m_{ij})^{\mathcal{D}_{sk_{\mathcal{R}_2}}(\mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk}))}} = \prod_{j=1}^{b} \mathcal{E}_{pk}(m_{ij} \cdot n_{jk})$$
$$= \mathcal{E}_{pk}(p_{ik})$$

Secure-Private Approach

How the user retrieves the matrix P?

1. $\mathcal{R}_2$ receives:
$$\Big((i,k), \mathcal{E}_{pk}(m_{ij})^{n_{jk}+\tau_{jk}}, \mathcal{E}_{pk}(m_{ij}), \mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk})\Big)_{1\leq i\leq a, 1\leq j\leq b, 1\leq k\leq c}$$
2. For each (i,k), $\mathcal{R}_2$ computes:
$$\prod_{j=1}^{b} \frac{\mathcal{E}_{pk}(m_{ij})^{n_{jk}+\tau_{jk}}}{\mathcal{E}_{pk}(m_{ij})^{\mathcal{D}_{sk_{\mathcal{R}_2}}(\mathcal{E}_{pk_{\mathcal{R}_2}}(\tau_{jk}))}} = \prod_{j=1}^{b} \mathcal{E}_{pk}(m_{ij} \cdot n_{jk}) = \mathcal{E}_{pk}(p_{ik})$$
3. Then, $\mathcal{R}_2$ sends to $\mathcal{P}$ all $\mathcal{E}_{pk}(p_{ik})$.

Interactive Homomorphic Multiplication of Ciphertexts [Cramer et al.]

How to compute $\mathcal{E}_{pk}(m_1 \cdot m_2)$?

Interactive Homomorphic Multiplication of Ciphertexts [Cramer et al.]

How to compute $\mathcal{E}_{pk}(m_1 \cdot m_2)$?

(sk, pk)

$c_1 = \mathcal{E}_{pk}(m_1)$ and $c_2 = \mathcal{E}_{pk}(m_2)$

Interactive Homomorphic Multiplication of Ciphertexts [Cramer et al.]

How to compute $\mathcal{E}_{pk}(m_1 \cdot m_2)$?

(sk, pk)

$c_1 = \mathcal{E}_{pk}(m_1)$ and $c_2 = \mathcal{E}_{pk}(m_2)$

Picks two randoms $\delta_1, \delta_2 \in \mathbb{Z}_n$

Interactive Homomorphic Multiplication of Ciphertexts [Cramer et al.]

How to compute $\mathcal{E}_{pk}(m_1 \cdot m_2)$?

(sk, pk)

$c_1 = \mathcal{E}_{pk}(m_1)$ and $c_2 = \mathcal{E}_{pk}(m_2)$

Picks two randoms $\delta_1, \delta_2 \in \mathbb{Z}_n$

$\alpha_1 = c_1 \cdot \mathcal{E}_{pk}(\delta_1)$

$\alpha_2 = c_2 \cdot \mathcal{E}_{pk}(\delta_2)$

Interactive Homomorphic Multiplication of Ciphertexts [Cramer et al.]

How to compute $\mathcal{E}_{pk}(m_1 \cdot m_2)$?

(sk, pk)

$c_1 = \mathcal{E}_{pk}(m_1)$ and $c_2 = \mathcal{E}_{pk}(m_2)$

Picks two randoms $\delta_1, \delta_2 \in \mathbb{Z}_n$

$\alpha_1 = c_1 \cdot \mathcal{E}_{pk}(\delta_1)$
$\alpha_2 = c_2 \cdot \mathcal{E}_{pk}(\delta_2)$

$\xleftarrow{\alpha_1, \alpha_2}$

$\mathcal{D}_{sk}(\alpha_1) = m_1 + \delta_1$
$\mathcal{D}_{sk}(\alpha_2) = m_2 + \delta_2$

Interactive Homomorphic Multiplication of Ciphertexts [Cramer et al.]

How to compute $\mathcal{E}_{pk}(m_1 \cdot m_2)$?

(sk, pk)

$c_1 = \mathcal{E}_{pk}(m_1)$ and $c_2 = \mathcal{E}_{pk}(m_2)$

Picks two randoms $\delta_1, \delta_2 \in \mathbb{Z}_n$

$\alpha_1 = c_1 \cdot \mathcal{E}_{pk}(\delta_1)$
$\alpha_2 = c_2 \cdot \mathcal{E}_{pk}(\delta_2)$

$\xleftarrow{\alpha_1, \alpha_2}$

$\mathcal{D}_{sk}(\alpha_1) = m_1 + \delta_1$
$\mathcal{D}_{sk}(\alpha_2) = m_2 + \delta_2$
$\beta = \mathcal{E}_{pk}((m_1 + \delta_1) \cdot (m_2 + \delta_2))$

Interactive Homomorphic Multiplication of Ciphertexts [Cramer et al.]

How to compute $\mathcal{E}_{pk}(m_1 \cdot m_2)$?

(sk, pk)

$c_1 = \mathcal{E}_{pk}(m_1)$ and $c_2 = \mathcal{E}_{pk}(m_2)$

Picks two randoms $\delta_1, \delta_2 \in \mathbb{Z}_n$

$\mathcal{D}_{sk}(\alpha_1) = m_1 + \delta_1$
$\mathcal{D}_{sk}(\alpha_2) = m_2 + \delta_2$
$\beta = \mathcal{E}_{pk}((m_1 + \delta_1) \cdot (m_2 + \delta_2))$

$\xleftarrow{\alpha_1, \alpha_2}$

$\xrightarrow{\beta}$

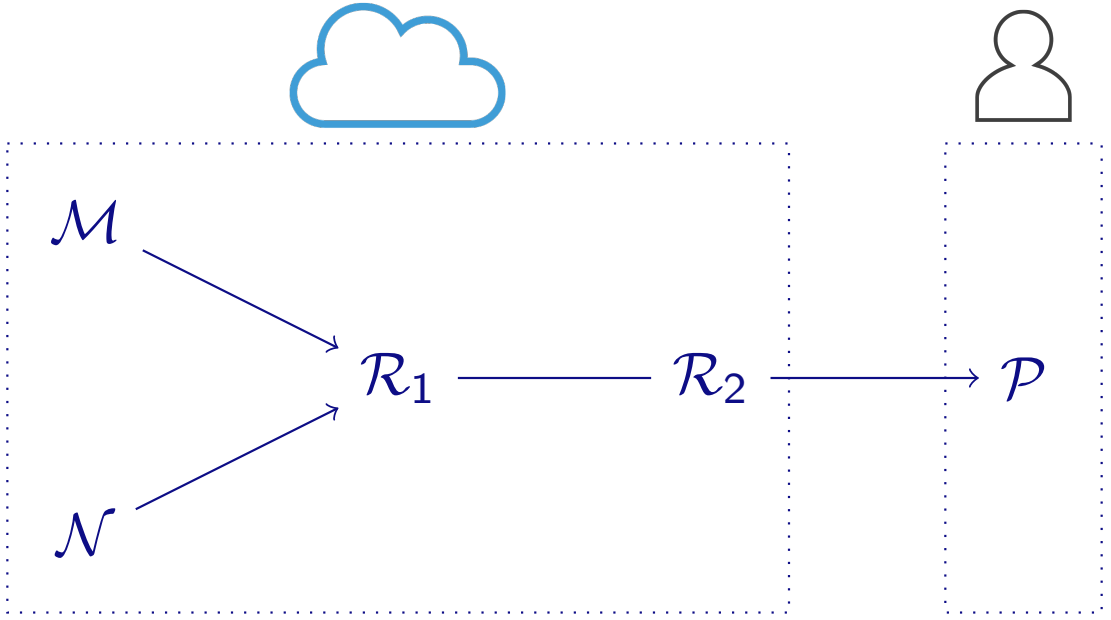
$\alpha_1 = c_1 \cdot \mathcal{E}_{pk}(\delta_1)$
$\alpha_2 = c_2 \cdot \mathcal{E}_{pk}(\delta_2)$

Computes $\frac{\beta}{\mathcal{E}_{pk}(\delta_1 \cdot \delta_2) \cdot c_1^{\delta_2} \cdot c_2^{\delta_1}}$

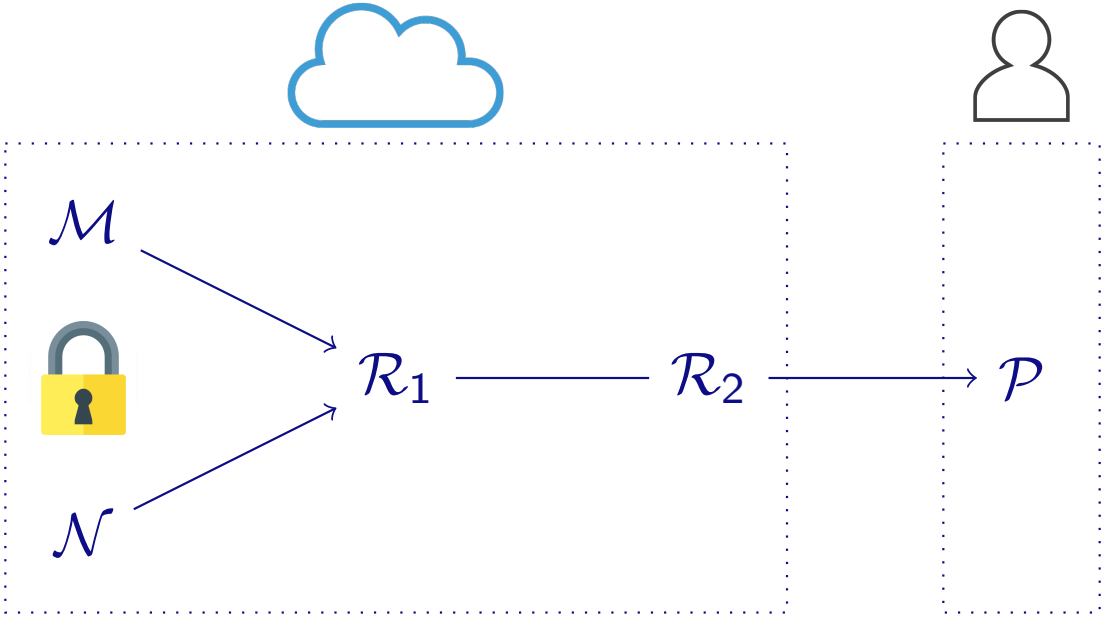
Verification of the Computation

$$
\begin{aligned}
\frac{\beta}{\mathcal{E}_{pk}(\delta_1 \cdot \delta_2) \cdot c_1^{\delta_1} \cdot c_2^{\delta_2}} &= \frac{\mathcal{E}_{pk}((m_1 + \delta_1) \cdot (m_2 + \delta_2))}{\mathcal{E}_{pk}(\delta_1 \cdot \delta_2) \cdot \mathcal{E}_{pk}(m_1)^{\delta_2} \cdot \mathcal{E}_{pk}(m_2)^{\delta_1}} \\
&= \frac{\mathcal{E}_{pk}(m_1 \cdot m_2 + m_1 \cdot \delta_2 + m_2 \cdot \delta_1 + \delta_1 \cdot \delta_2)}{\mathcal{E}_{pk}(\delta_1 \cdot \delta_2) \cdot \mathcal{E}_{pk}(m_1)^{\delta_2} \cdot \mathcal{E}_{pk}(m_2)^{\delta_1}} \\
&= \frac{\mathcal{E}_{pk}(m_1 \cdot m_2) \cdot \mathcal{E}_{pk}(m_1 \cdot \delta_2) \cdot \mathcal{E}_{pk}(m_2 \cdot \delta_1) \cdot \mathcal{E}_{pk}(\delta_1 \cdot \delta_2)}{\mathcal{E}_{pk}(\delta_1 \cdot \delta_2) \cdot \mathcal{E}_{pk}(m_1 \cdot \delta_2) \cdot \mathcal{E}_{pk}(m_2 \cdot \delta_1)} \\
&= \mathcal{E}_{pk}(m_1 \cdot m_2)
\end{aligned}
$$

Secure Matrix Multiplication with MapReduce

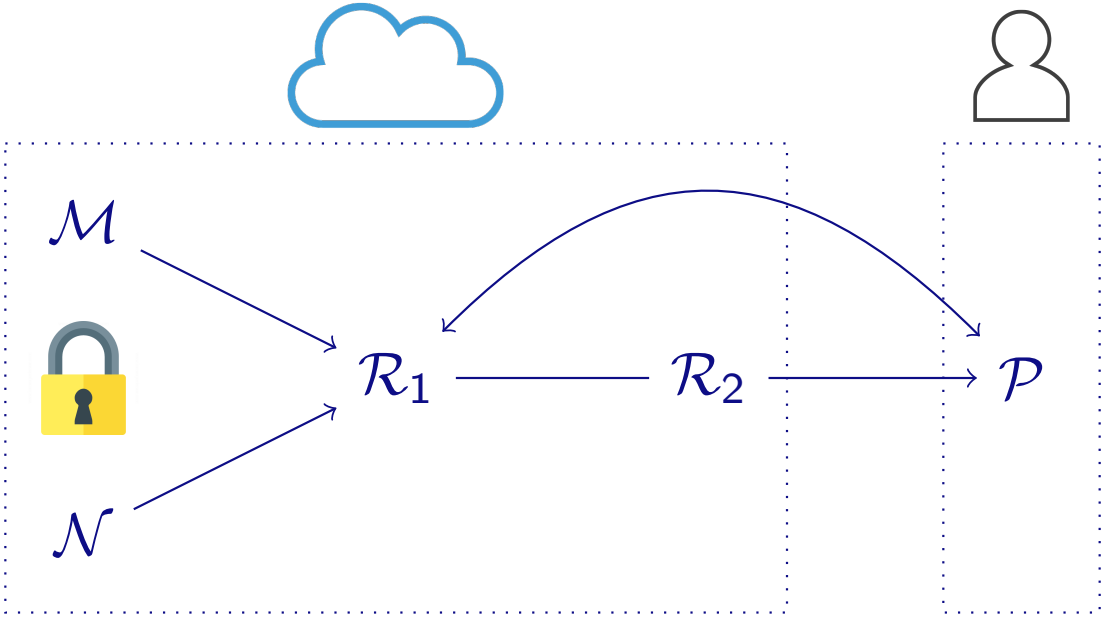


Secure Matrix Multiplication with MapReduce



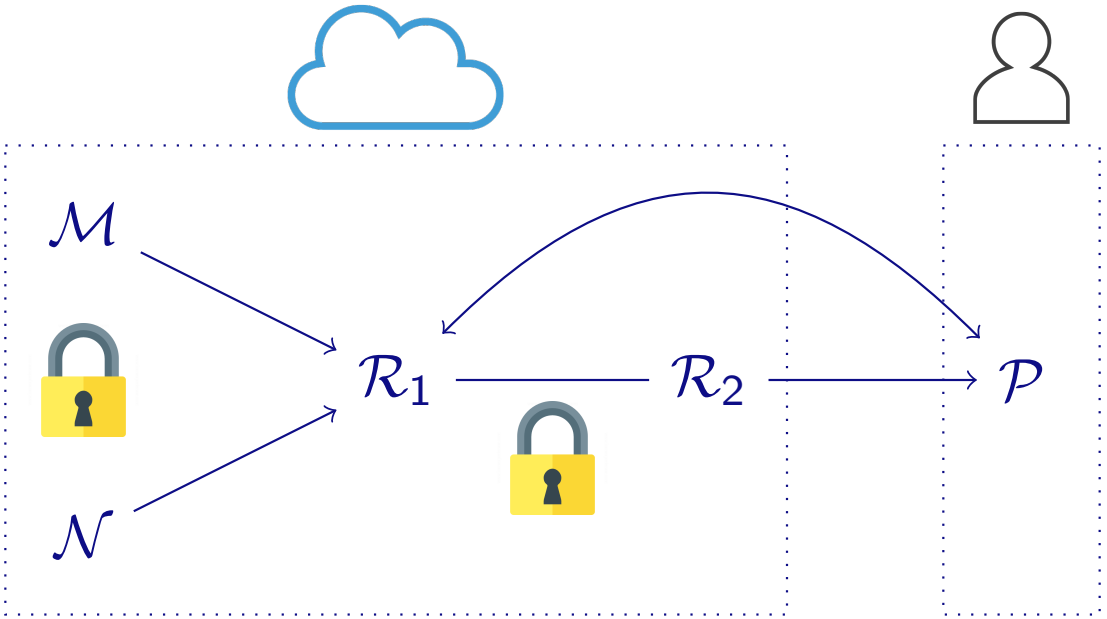
- $\mathcal{M}$ emits to $\mathcal{R}_1$: $(j, (M, i, \mathcal{E}_{pk}(m_{ij})))_{1 \leq i \leq a, 1 \leq j \leq b}$
- $\mathcal{N}$ emits to $\mathcal{R}_1$: $(j, (N, k, \mathcal{E}_{pk}(n_{jk})))_{1 \leq j \leq b, 1 \leq k \leq c}$

Secure Matrix Multiplication with MapReduce



- $\mathcal{M}$ emits to $\mathcal{R}_1$: $(j, (M, i, \mathcal{E}_{pk}(m_{ij})))_{1 \leq i \leq a, 1 \leq j \leq b}$
- $\mathcal{N}$ emits to $\mathcal{R}_1$: $(j, (N, k, \mathcal{E}_{pk}(n_{jk})))_{1 \leq j \leq b, 1 \leq k \leq c}$
- $\mathcal{R}_1$ emits to $\mathcal{R}_2$: $\mathcal{E}_{pk}(m_{ij} \cdot n_{jk})_{1 \leq i \leq a, 1 \leq j \leq b, 1 \leq k \leq c}$ (computed via *Interactive Homomorphic Multiplication*)

Secure Matrix Multiplication with MapReduce



- $\mathcal{M}$ emits to $\mathcal{R}_1$: $(j,(M,i,\mathcal{E}_{pk}(m_{ij})))_{1\leq i\leq a,1\leq j\leq b}$
- $\mathcal{N}$ emits to $\mathcal{R}_1$: $(j,(N,k,\mathcal{E}_{pk}(n_{jk})))_{1\leq j\leq b,1\leq k\leq c}$
- $\mathcal{R}_1$ emits to $\mathcal{R}_2$: $\mathcal{E}_{pk}(m_{ij}\cdot n_{jk})_{1\leq i\leq a,1\leq j\leq b,1\leq k\leq c}$ (computed via *Interactive Homomorphic Multiplication*)
- $\mathcal{R}_2$ emits to $\mathcal{P}$: $\left\{\mathcal{E}_{pk}(p_{ik})=\prod_{j=1}^{b}\mathcal{E}_{pk}(m_{ij}\cdot n_{jk})\right\}_{1\leq i\leq a,1\leq k\leq c}$

Computation/Communication Costs and Privacy

Algorithm	*Computation cost (big-O)*	*Comm. cost (big-O)*
Standard	$2n^2 + (C_\times + C_+)n^3$	$n^3 + 3n^2$
SP	$(C_+ + 2C_\mathcal{E})n^2 + (2C_\times + 2C_{\text{exp}} + C_\mathcal{D})n^3$	$3n^3 + 4n^2$
CRSP	$2C_\mathcal{E} n^2 + (4C_\mathcal{E} + 7C_\times + 2C_\mathcal{D} + 2C_{\text{exp}})n^3$	$4n^3 + 3n^2$

- $n = \max(a, b, c)$
- $C_+, C_\times, C_\mathcal{E}, C_\mathcal{D}, C_{\text{exp}}$ are the costs of the associated operations

Conclusion

Two secure approaches for matrix multiplication with MapReduce:

1. Secure-Private (SP): Assumes that nodes $\mathcal{N}$ and $\mathcal{R}_2$ do not collude.
2. Collision-Resistant-Secure-Private (CRSP): Resists to collisions but uses interaction between node $\mathcal{R}_1$ and $\mathcal{R}_2$.

Outline

Symmetric Searchable Encryption

Store data externally

- encrypted
- want to search data easily
- avoid downloading everything then decrypt
- allow others to search data without having access to plaintext

Context

Symmetric Searchable Encryption (SSE)

- Outsource a set of *encrypted data*.
- Basic functionnality: *single keyword query*.

(Client) 

$\longrightarrow$  (Server)

Symmetric Searchable Encryption

When searching, what must be protected?

- retrieved data
- search query
- search query outcome (was anything found?)

Scenario

- single query vs multiple queries
- non-adaptive: series of queries, each independent of the others
- adaptive: form next query based on previous results

Number of participants

- single user (owner of data) can query data
- multiple users can query the data, possibly with access rights defined by the owner

SSE by Song, Wagner, Perrig 2000

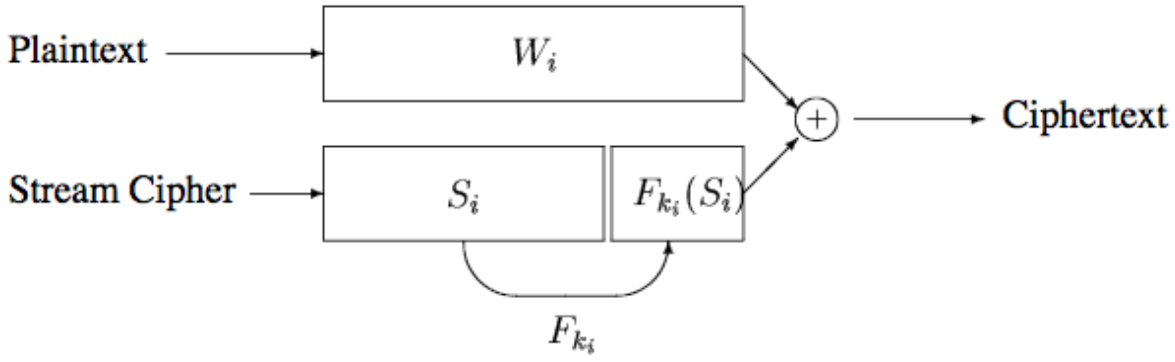


Figure 1. The Basic Scheme

Basic Scheme I

$$C_i = W_i \oplus < S_i, F_{k_i}(S_i) >$$

where S_i are randomly generated and $F_k(x)$ is a MAC with key k.

Basic Scheme

$$C_i = W_i \oplus < S_i, F_{k_i}(S_i) >$$

To search W :

- Alice reveals $\{k_i$, where W may occur$\}$
- Bob checks if $W \oplus C_i$ is of the form $< s, F_{k_i}(s) >$.

For unknown k_i, Bob knows nothing

Basic Scheme

$$C_i = W_i \oplus < S_i, F_{k_i}(S_i) >$$

To search W :

- Alice reveals $\{k_i$, where W may occur$\}$
- Bob checks if $W \oplus C_i$ is of the form $< s, F_{k_i}(s) >$.

For unknown k_i, Bob knows nothing

Problems for Alice !

- she reveals all k_i,
- or she has to know where W may occur !

Scheme II: Controlled Searching

Modifications

$$C_i = W_i \oplus < S_i, F_{k_i}(S_i) >$$

where S_i randoms, $F_k(x)$ is a MAC with key k; $k_i = f_{k'}(W_i)$

To search W :

- Alice only reveals $k = f_{k'}(W)$ and W.
- Bob checks if $W \oplus C_i$ is of the form $< s, F_k(s) >$

+ For unknown k_i, Bob knows nothing
+ Nothing is revealed about location of W.

Problem

- Still does not support hidden search (Alice reveals W)

Scheme III: Support for Hidden Searches

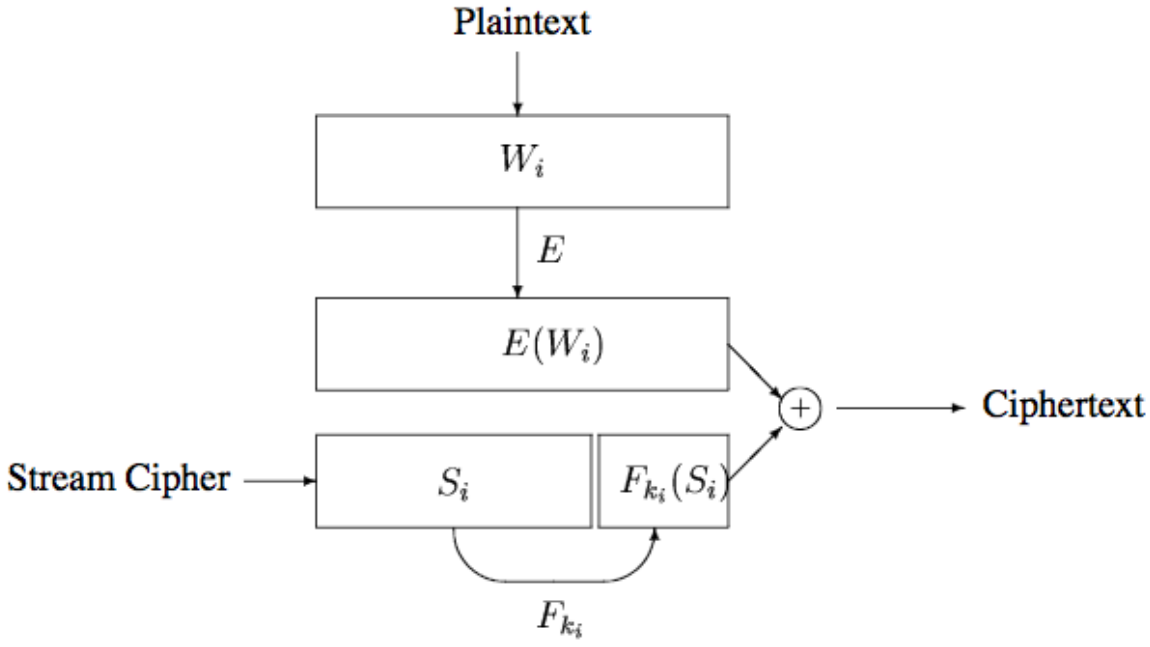


Figure 2. The Scheme for Hidden Search

Scheme III : Hidden Searches

$$C_i = E_{k''}(W_i) \oplus < S_i, F_{k_i}(S_i) >$$

S_i randoms and $F_k(x)$ is a MAC with k and $k_i = f_{k'}(E_{k''}(W_i))$

Scheme III: Support for Hidden Searches

$$C_i = E_{k''}(W_i) \oplus < S_i, F_{k_i}(S_i) >, where\ k_i = f_{k'}(E_{k''}(W_i))$$

To search W :

- Alice gives $X = E_{k''}(W)$ and $k = f_{k'}(X)$.
- Bob checks if $X \oplus C_i$ is of the form $< s, F_k(s) >$

Bob returns to Alice C_i

Scheme III: Support for Hidden Searches

$$C_i = E_{k''}(W_i) \oplus < S_i, F_{k_i}(S_i) >, where\ k_i = f_{k'}(E_{k''}(W_i))$$

To search W :

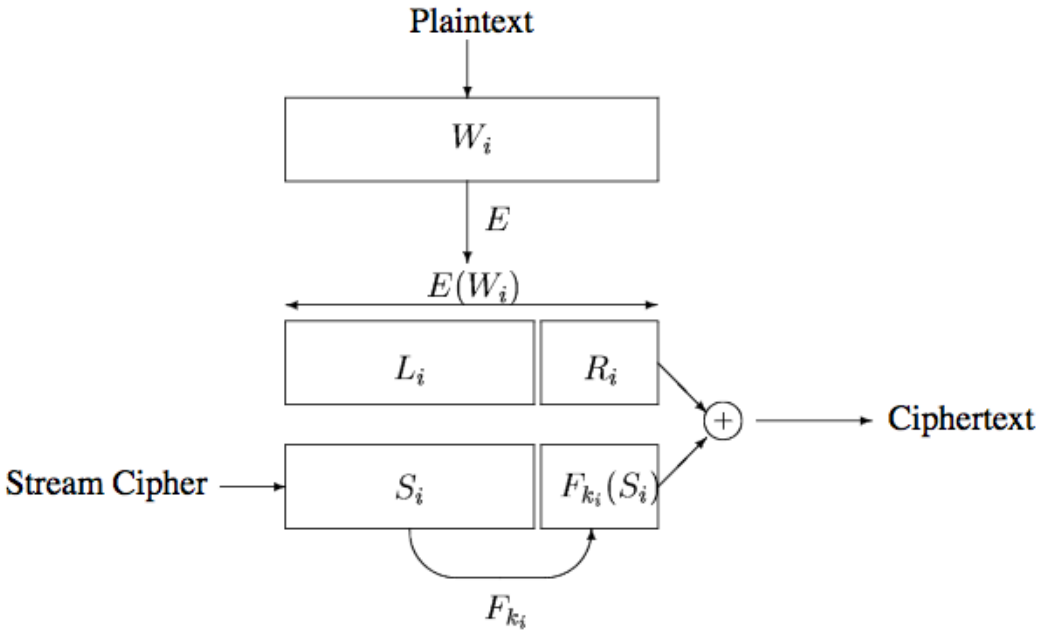
- Alice gives $X = E_{k''}(W)$ and $k = f_{k'}(X)$.
- Bob checks if $X \oplus C_i$ is of the form $< s, F_k(s) >$

Bob returns to Alice C_i

But Alice cannot recover the plaintext

She can recover S_i with X but not $F_{k_i}(S_i)$ because to compute $k_i = f_{k'}(E_{k''}(W_i))$ she needs to have $E_{k''}(Wi)$.
In this case, why do you need search ?

Final Scheme



Scheme IV : Final

$$C_i = X_i \oplus < S_i, F_{k_i}(S_i) >$$

where S_i randoms and $F_k(x)$ is a MAC with key k,
$X_i = E_{k''}(W_i) =< L_i, R_i >$ and $k_i = f_{k'}(L_i)$

Final Scheme (Ultimate TRICK !)

$$C_i = X_i \oplus < S_i, F_{k_i}(S_i) >$$

To search W :

- Alice gives $X = E_{k''}(W) =< L, R >$ and $k = f_{k'}(L)$
- Bob checks if $X \oplus C_i$ is of the form $< s, F_k(s) >$

Bob returns to Alice C_i

Alice recovers S_i and then $L_i = C_i \oplus S_i$. Then she computes $k_i = f_{k'}(L_i)$ and then $X = C_i \oplus < s, F_k(s) >$ and by decrypting with k'' to obtain W_i.

Alice only needs to remember k'' and k'.

Outline

Privacy vs. Confidentiality

Confidentiality

Prevent disclosure of information to unauthorized users

Privacy

- Prevent disclosure of personal information to unauthorized users
- Control of how personal information is collected and used

Data Privacy and Security Measures

Access control

Restrict access to the (subset or view of) data to authorized users

Inference control

Restrict inference from accessible data to additional data

Flow control

Prevent information flow from authorized use to unauthorized use

Encryption

Use cryptography to protect information from unauthorized disclosure while in transmit and in storage

2 kinds of data

- Personal data
- Anonymous data

CNIL:

"Dès lors qu'elles concernent des personnes physiques identifiées directement ou indirectement."

French Law:

"Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne."

How to evaluate the security?

Three criteria of robustness:

- is it still possible to single out an individual ?
  Singling out (Individualisation): the possibility to isolate some or all records which identify an individual in the dataset
- is it still possible to link records relating to an individual ?
  Linkability (Correlation): ability to link, at least, two records concerning the same data subject or a group of data subjects.
- can information be inferred concerning an individual?
  Inference (Deduction): deduce, with significant probability, the value of an attribute from the values of a set of other attributes

Example

ID	Age	CP	Sex	Pathology
Paul Sésame	75	75000	F	Cancer
Pierre Richard	55	78000	F	Cancer
Henri Poincarré	40	71000	M	Influe

Randomization

Alter veracity of the DB to remove the link

- **Noise addition:** modifying attributes in the dataset such that they are less accurate whilst retaining the overall distribution
- **Permutation:** shuffling the values of attributes in a table so that some of them are artificially linked to different data subjects,
- **Differential Privacy:** requires the outcome to be formally indistinguishable when run with and without any particular record in the data set.

Example

Q = select count() where Age = [20,30] and Diagnosis=B
Answer to Q on D1 and D2 should be indistinguishable, if Bob in D1 or Bob out D2.

Differential Privacy

C. Dwork : "Differential Privacy", International Colloquium on Automata, Languages and Programming , 2006.

Definition

Let ϵ be a positive real number and $\mathcal{A}$ be a randomized algorithm that takes a dataset as input (representing the actions of the trusted party holding the data). The algorithm $\mathcal{A}$ is ϵ-differentially private if for all datasets D_1 and D_2 that differ on a single element (i.e., the data of one person), and all subsets S of $\mathrm{im}\mathcal{A}$,

$$\Pr[\mathcal{A}(D_1) \in S] \leq e^{\epsilon} \times \Pr[\mathcal{A}(D_2) \in S]$$

where the probability is taken over the randomness used by the algorithm.

Pseudonymisation

ID	Age	CP	Sex	Pathology
1	75	75000	F	Cancer
2	55	78000	F	Cancer
3	40	71000	M	Influe

Replace identifier field by a new one called pseudonym.
Using Hash function
It does not ensure anonymity. Using several fields you can recover name like it has benn done by Sweeney in 2001.

Example

Sex + birthday date + Zip code are unique for 80 % of USA citizens. (record linkage attack)

k-Anonymity

- Identify the possible fields that can be used to recover data (generalisation).
- Modify them in order to have at least k different lines having the same identifiers.

It reduce the probabolity to guess something to $1/k$
Advantage: Analysis of data still give the same information that the orginal data base.

Example: k-Anonymity

Activity	Age	Pathology
M2	[22,23]	Cancer
M2	[22,23]	Blind
M2	[22,23]	VIH
PhD	[24,27]	Cancer
PhD	[24,27]	Allergies
PhD	[24,27]	Allergies
L	[20,21]	Cancer
L	[20,21]	Cancer
L	[20,21]	Cancer

3-Anonymity
Activity for student can be Master licence or PhD instead of name and activty, age can be ranged.

Disadvantages: k-Anonymity

- It leaks negative information. For instance you are not in all the other catergories.
- If all personn have the same value then the value is leaked.
- Main problem is to determine the right generalisation (it is difficult and expensive).

Minimum Cost 3-Anonymity is NP-Hard for $|\Sigma| = 2$ (Dondi et al. 2007)

l-diversity

Aims at avoiding that all person have the same values once they have been generalized.
/ values souhld be inside each field after generalisation. It allows to recover information by mixing information with some probability

Activity	Age	Pathology
M2	[22,23]	Cancer
M2	[22,23]	Allergies
M2	[22,23]	VIH
PhD	[24,27]	Cancer
PhD	[24,27]	VIH
PhD	[24,27]	Allergies
L	[20,21]	VIH
L	[20,21]	Allergies
L	[20,21]	Cancer

3-diversity, each category has 3 different values

t-closeness

Knowledge of global distribution of sensitive data of a class of equivalence.
It tries to reduce the weaknesses introduced by the l-diversity.
t is the factor that says how we are far from a global distribution.

- How to split data into partion to obtain all the same distribution.
- If all class of equivalence have the same number of data, what is the utility of any analysis of the data basis ?

Summary

Is Risky	Singling out	Linkability	Inference
Pseudonymisation	Yes	Yes	Yes
Noise addition	Yes	May not	May not
Substitution	Yes	Yes	May not
Aggregation or K-anonymity	No	Yes	Yes
L-diversity	No	Yes	May not
Differential privacy	May not	May not	May not

Outline

Things to bring home

- Date Security is cruciual
- Security should be done by experts!
- Security should be taken from the design and not after!



$$\textit{Protocol} + \textit{Properties} + \textit{Intruder} = \textit{Security}$$

Thank you for your attention.

Questions ?