Information System - Data Protection

Dr J Paul Gibson

Dept. INF

Office D311

paul.gibson@telecom-sudparis.eu

http://jpaulgibson.synology.me/~jpaulgibson/TSP/Teaching/CSC4104/CSC4104-InformationSystem-DataProtection.pdf


General
Data
Protection
Regulation

Bigger Responsibility, Bigger Repercussions

Fines of up to 4% of turnover
Organizations in breach of GDPR can be fined up to 4% of annual global turnover or €20 Million.

Breach notification within 72 hrs
Breaches must be reported within 72 hours of first having become aware of the breach.

Increased territorial scope
Applies to any company processing personal data of EU citizens, regardless of location.

Privacy by design
Data protection from the onset of the designing of systems, rather than a retrospective addition.

Consent matters
Explicit consent must be provided in an intelligible and easily accessible form.

Right to be forgotten
Entitles the data subject to have the data controller erase his/ her personal data (and potentially third parties, too).

Right to access and portability
Users can inquire whether and how their personal data is being processed.

Mandatory data protection officers
Appointed in certain cases, to facilitate the company's need to demonstrate GDPR compliance.

https://www.cnil.fr/fr

The 2016/679 European regulation of 27th April 2016 (known as "General Data Protection Regulation" or GDPR) specifies that protecting personal data requires taking "appropriate technical and organisational measures to ensure a level of security appropriate to the risk" (article 32).

GDPR Principles

https://www.itgovernance.eu/blog/en/the-gdpr-understanding-the-6-data-protection-principles

1. Lawfulness, fairness and transparency

The first principle is relatively self-evident: organisations need to ensure their data collection practices don’t break the law and that they aren’t hiding anything from data subjects.

To remain lawful, you need to have a thorough understanding of the GDPR and its rules for data collection. To remain transparent with data subjects, you should state in your privacy policy the type of data you collect and the reason you’re collecting it.

2. Purpose limitation

Organisations should only collect personal data for a specific purpose, clearly state what that purpose is, and only collect data for as long as necessary to complete that purpose.

Processing that’s done for archiving purposes in the public interest or scientific, historical or statistical purposes is given more freedom

3. Data minimisation

Organisations must only process the personal data that they need to achieve its processing purposes. Doing so has two major benefits.

First, in the event of a data breach, the unauthorised individual will only have access to a limited amount of data.

Second, data minimisation makes it easier to keep data accurate and up to date.

4. Accuracy

The accuracy of personal data is integral to data protection. The GDPR states that “every reasonable step must be taken” to erase or rectify data that is inaccurate or incomplete.

Individuals have the right to request that inaccurate or incomplete data be erased or rectified within 30 days.

5. Storage limitation

Similarly, organisations need to delete personal data when it's no longer necessary.

How do you know when information is no longer necessary? According to marketing company Epsilon Abacus, organisations might argue that they "should be allowed to store the data for as long as the individual can be considered a customer.

So the question really is: For how long after completing a purchase can the individual be considered a customer?"

The answer to this will vary between industries and the reasons that data is collected. Any organisation that is uncertain how long it should keep personal data should consult a legal professional.

6. Integrity and confidentiality

This is the only principle that deals explicitly with security. The GDPR states that personal data must be

> "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures".

The GDPR is deliberately vague about what measures organisations should take, because technological and organisational best practices are constantly changing.

Currently, organisations should encrypt and/or pseudonymise personal data wherever possible, but they should also consider whatever other options are suitable.

The seventh principle

The GDPR includes an additional principle, accountability, which acts as an overarching set of requirements related to the other six.

By achieving accountability, organisations demonstrate that they have the necessary documentation to prove that they are meeting their compliance requirements.

This is typically done through a combination of technical measures and documentation such as:

- Controller–processor contracts;
- Relevant policies and procedures;
- Privacy notices;
- Staff training records;
- Security monitoring and event logging records;
- Data breach records; and
- Data protection impact assessments.

Data Anonymization vs Pseudo Anonymization

Data masking: anonymization or pseudonymization?

https://gdpr.report/news/2017/09/28/data-masking-anonymization-pseudonymization/

What is pseudonymization?

Pseudonymization enhances privacy by replacing most identifying fields within a data record by one or more artificial identifiers, or pseudonyms. There can be a single pseudonym for a collection of replaced fields or a pseudonym per replaced field.

Specifically, the GDPR defines pseudonymization in Article 3, as “the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information.” To pseudonymize a data set, the “additional information” must be “kept separately and subject to technical and organizational measures to ensure non-attribution to an identified or identifiable person.”

Pseudonymization or Anonymization?

The legal distinction between anonymized and pseudonymized data is its categorization as personal data. Pseudonymous data still allows for some form of re-identification (even indirect and remote), while anonymous data cannot be re-identified.

Which data should be anonymized?

By definition, data anonymization techniques seek to conceal identity and thus identifiers of any nature. Identifiers can apply to any natural or legal person, living or dead, including their dependents, ascendants and descendants. Included are other related persons, direct or through interaction.

For example:

- Family names, patronyms, first names, maiden names, aliases.
- Postal addresses
- Telephones
- Postal codes + Cities
- IDs: social security number (e.g. Fiscal Code in Italy, National Insurance number in UK), bank account details (e.g. IBAN), credit card numbers, valid keys

Which techniques are available for anonymizing data?

- **Scrambling**
  Scrambling techniques involve a mixing or obfuscation of letters. The process can sometimes be reversible.For example: Annecy could become Yneanc
- **Masking**
  A masking technique allows a part of the data to be hidden with random characters or other data.For example: Pseudonymization with masking of identities or important identifiers. The advantage of masking is the ability to identify data without manipulating actual identities.
- **Personalized anonymization**
  This method allows the user to utilize his own anonymization technique. Custom anonymization can be carried out using scripts or an application.
- **Blurring**
  Data blurring uses an approximation of data values to render their meaning obsolete and/or render the identification of individuals impossible.

How to Mask Tables and Preserve Referential Integrity

CSC 4104

	EMPLOYEE_SID	LAST_NAME	DEPARTMENT_NAME	MANAGER_SID	MANAGER_NAME
1	301258594	Abel	Sales	821188916	Zlotkey
2	582268470	Ande	Sales	658874527	Errazuriz
3	388498749	Atkinson	Shipping	306289564	Fripp
4	497821851	Austin	IT	491928244	Hunold
5	831918075	Baer	Public Relations	526476627	Kochhar
6	441672618	Baida	Purchasing	223670049	Raphaely
7	341562514	Banda	Sales	658874527	Errazuriz
8	730922093	Bates	Sales	222355712	Cambrault
9	271997715	Bell	Shipping	535847535	Vollman
10	591388917	Bernstein	Sales	491468339	Russell
11	804367677	Bissot	Shipping	306289564	Fripp
12	781651266	Bloom	Sales	222355712	Cambrault
13	677021755	Bull	Shipping	306289564	Fripp
14	168025776	Cabrio	Shipping	306289564	Fripp
15	222355712	Cambrault	Sales	661272028	King
16	656815628	Cambrault	Sales	491468339	Russell
17	326663086	Chen	Finance	676071097	Greenberg
18	377578286	Chung	Shipping	767169700	Kaufling
19	599127479	Colmenares	Purchasing	223670049	Raphaely
20	263658281	Davies	Shipping	217192383	Mourgos
21	542323343	De Haan	Executive	661272028	King
22	662036427	Dellinger	Shipping	306289564	Fripp
23	611923906	Dilly	Shipping	767169700	Kaufling
24	371097973	Doran	Sales	381048770	Partners
25	805452976	Ernst	IT	491928244	Hunold
26	658874527	Errazuriz	Sales	661272028	King

	EMPLOYEE_SID	LAST_NAME	DEPARTMENT_NAME	MANAGER_SID	MANAGER_NAME
1	+n"nt4)E5	Abel	Sales	s!T6$I60s	Zlotkey
2	GB@S/p^x.	Ande	Sales	Kb11{O~N(	Errazuriz
3	X^a;,U%	Atkinson	Shipping	Q}6Chz?DG	Fripp
4	1>1hYC-g!	Austin	IT	I[*:rLwCF	Hunold
5	Xm\NjmAWf	Baer	Public Relations	V/qT2wdF*	Kochhar
6	%@]A-`b`o	Baida	Purchasing	Nxjd)#1)&	Raphaely
7	Xu!>%-$pK	Banda	Sales	Kb11{O~N(	Errazuriz
8	dSA#Y;vNg	Bates	Sales	}:45`Sp!l	Cambrault
9	P+ bD@Zul	Bell	Shipping	8 r1CZc|H	Vollman
10	,Oza'qBnJ	Bernstein	Sales	I?sCsl("\	Russell
11	?Bt55af~4	Bissot	Shipping	Q}6Chz?DG	Fripp
12	h]qC:d'zu	Bloom	Sales	}:45`Sp!l	Cambrault
13	GO;_g'G\h	Bull	Shipping	Q}6Chz?DG	Fripp
14	m;pa)F{<F	Cabrio	Shipping	Q}6Chz?DG	Fripp
15	}:45`Sp!l	Cambrault	Sales	ZjW]0IX9d	King
16	csnDOW|k=	Cambrault	Sales	I?sCsl("\	Russell
17	!zn ^+a5x	Chen	Finance	(8}/,:E/^	Greenberg
18	r?"p=@?Q+	Chung	Shipping	X~]2vt[nh	Kaufling
19	)iID'{,/Y	Colmenares	Purchasing	Nxjd)#1)&	Raphaely
20	SIorLE{:8	Davies	Shipping	=[{ch=%3K	Mourgos
21	9,W?Jyv:d	De Haan	Executive	ZjW]0IX9d	King
22	PAl6Hq+@n	Dellinger	Shipping	Q}6Chz?DG	Fripp
23	SHp91=5sa	Dilly	Shipping	X~]2vt[nh	Kaufling
24	^"wj"]AKu	Doran	Sales	`YG,TX$UG	Partners
25	HN[g7G{'$	Ernst	IT	I[*:rLwCF	Hunold
26	Kb11{O~N(	Errazuriz	Sales	ZjW]0IX9d	King

Data masking versus data encryption: a comparison of 2 pseudonymization methods

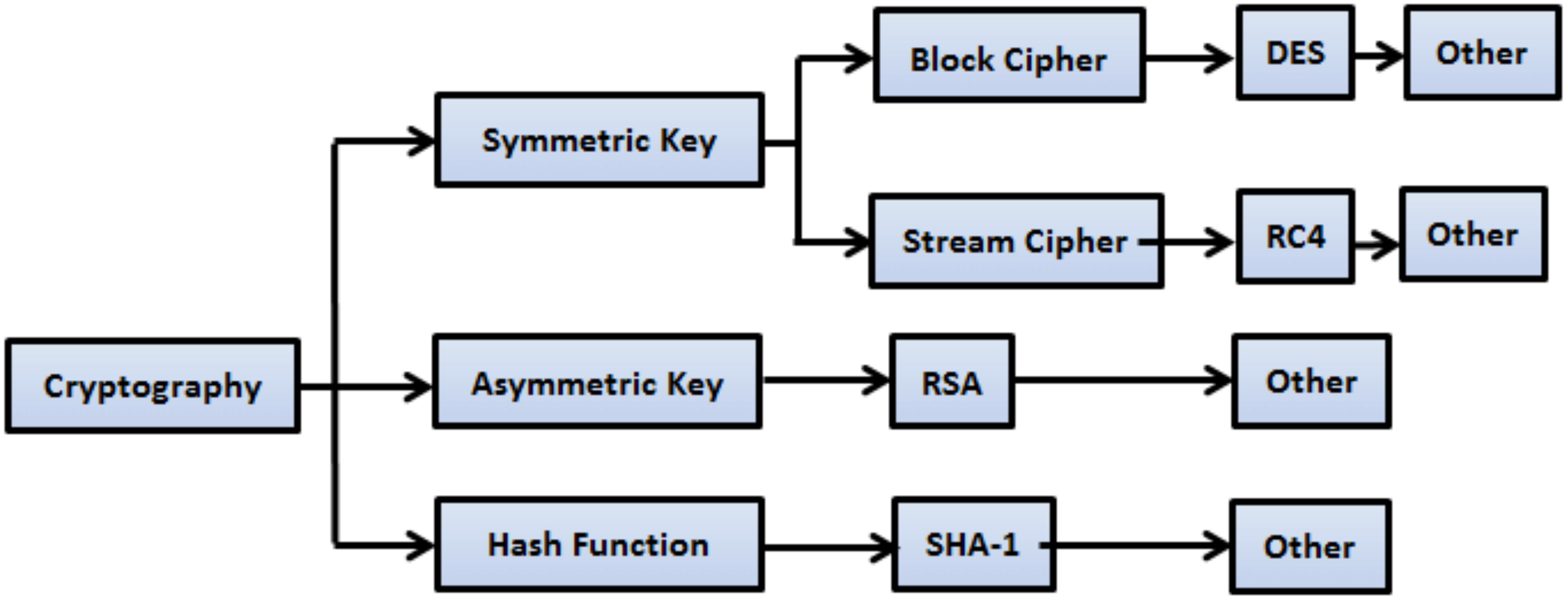
Distinct from data masking, data encryption translates data into another form, or code, so that only people with access to a secret key (formally called a decryption key) or password can read it.

Data masking is a more widely applicable solution as it enables organizations to maintain the usability of their customer data.

Intended objectives	Data Masking	Encryption
Security of data during transfer	X	V
Security of static data	V	V
Continuous availability of data for applications	V	X

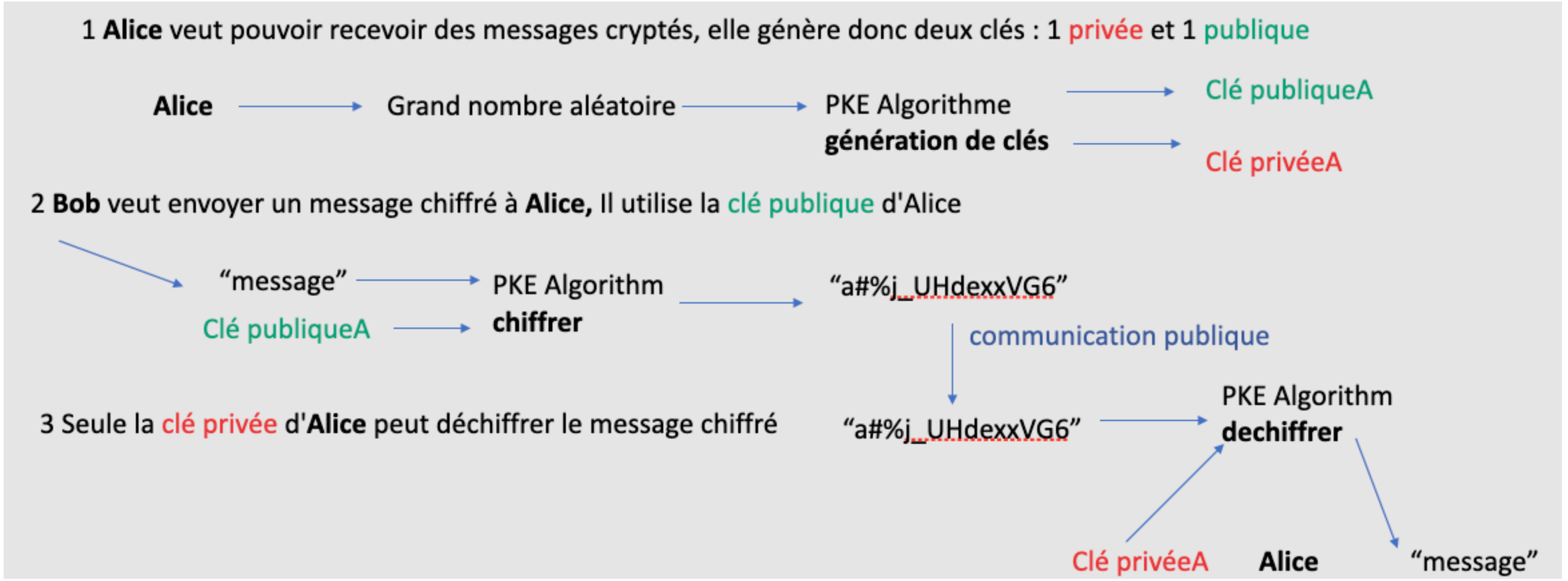
But What Encryption To Use (If any)?

Cryptography Techniques



Symmetric verses Asymmetric

Public key encryption (PKE)



Electronic Signatures

1 **Alice** veut **signer** un message à **Bob**, et elle utilise sa clé privée

La signature générée pour chaque message est différente

2 **Bob** reçoit un message signé et il veut vérifier qu'il provient d'**Alice** en utilisant sa clé publique

Si le message et la signature ne sont pas cohérents alors c'est pas KO

Homomorphic Encryption

Le chiffrement homomorphe permet aux utilisateurs d'effectuer des calculs sur ses données chiffrées sans d'abord les déchiffrer.

Exemple: Additionner une liste de nombres avec un encodage non sécurisé

5 7 2

5 -> 10 7->14 2->4

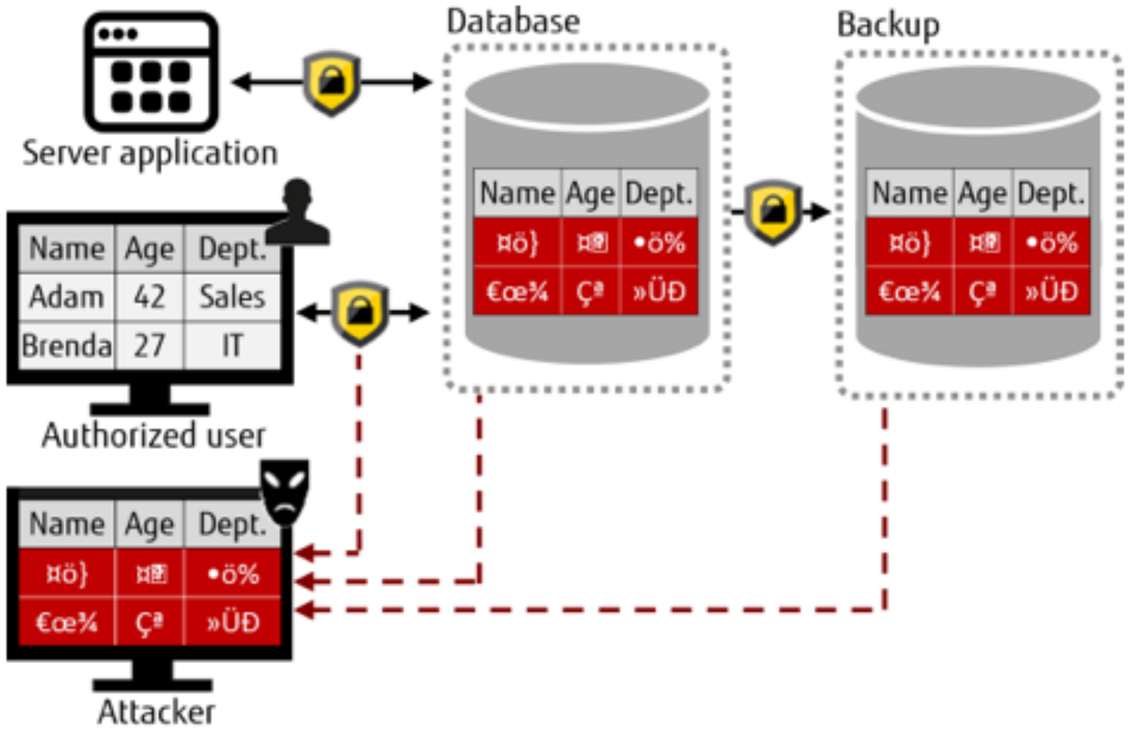
L'algorithme d'addition fonctionne sur les “données chiffrées” 10 + 14 + 4 = 28

Chaque entier individuel n'est jamais déchiffré, seul le résultat final 28 -> 14

Alors pouvons-nous compter les votes **bien chiffrés** sans avoir à déchiffrer les votes individuels ?

Secure Databases

A database is protected from unauthorised access by registering users and making them go through **authentication** and **access controls**. However, this does not protect the underlying operating system files from attackers, who can bypass the database server's authentication and access controls.



The “OSSE” case study - OPTIONAL WORK

- In the cahier des charges - identify data to be encrypted/anonymized
- What are the requirements that help choose the ‘best’ scheme/approach?
- Do your requirements cover the 7 principles of the GDPR?