Cours : thème 1 - Question 2

Question 2 : Les évolutions technologiques sont-elles exemptes de risques pour l'organisation ?

Notions abordées :
- Protection des données personnelle : identité numérique, popularité et e-reputation, Loi Informatique et Libertés ;
- Sécurisation et fiabilisation des données : sauvegarde, tolérance aux pannes, authentification, droits d'accès, mot de passe fiable hachage ;
- Sécurisation des échanges : cryptage ou chiffrement, cryptage symétrique et asymétrique, authenticité et certificat électronique, exemple de cryptage.

1. Protection des données personnelles

1.1. Identité numérique

Définition



Chaque personne dispose d'une identité constituée de caractéristiques. Une « personne physique » sera par exemple caractérisée par des attributs tels que : son sexe, son prénom, son nom, son n° de sécurité sociale, sa taille, son poids… Une organisation, c'est-à-dire une « personne morale », sera plutôt caractérisée par : sa raison sociale, son n° de SIRET, son code APE, l'adresse de son siège social… Cette identité est propre à chaque personne et la caractérise. L'identité numérique peut se définir comme la transposition de cette identité réelle à une ou plusieurs réalités virtuelles. Une personne peut ainsi être amenée à disposer de plusieurs identités numériques : une ou plusieurs sur les réseaux sociaux, les blogs, les jeux en ligne, les sites en tout genre…

Problème

A partir du moment où les personnes et leurs identités numériques circulent sur internet, on peut se poser de multiples questions concernant les risques encourus :
- Peut-on usurper l'identité de quelqu'un et se faire passer pour lui ? Est-ce puni par la loi ?
- Si une organisation est présente sur internet, sa notoriété peut-elle être entachée par les avis des internautes ?
- Mes informations personnelles demeurent-elle confidentielles ?
- Peut-on mettre quelqu'un sur écoute ou, autrement dit, peut-on intercepter les échanges de données entre deux personnes et/ou deux systèmes informatisés ?
- Etc.

1.2. La popularité et l'e-reputation

Définition	La popularité d'une organisation est de plus en plus liée à celle qu'elle acquiert sur internet, qualifiée d'e-reputation. Cette dernière n'est pas une denrée palpable mais elle est propre à l'organisation. C'est l'avis général que les gens se font d'elle.
Problème	Les technologies de l'information et de la communication ayant toujours plus d'influence sur la popularité d'une organisation, on peut craindre : • que l'inactivité d'une organisation sur la toile ait une influence néfaste sur son image ; • que des internautes malavisés nuisent à la réputation de l'organisation.
Solution	• Le community management est un métier consistant justement à entretenir la popularité d'une organisation sur la toile en passant du temps à interagir avec les internautes, à communiquer et diffuser de l'information, etc. (*cf. cours Q7, nouvelles façons communiquer, d'échanger et de travailler*). • Le droit permet à une organisation de mener une action en diffamation ou en dénigrement si des individus nuisent sciemment à la réputation de l'organisation. • L'usurpation d'identité est quant à elle punie par 2 ans d'emprisonnement et/ou 20 000€ d'amende.

1.3. La CNIL et la Loi Informatique et Libertés

La Loi Informatique et Liberté du 6 janvier 1978 définit un cadre juridique visant à protéger en outre les internautes. Elle a permis d'instaurer la CNIL (Commission Nationale de l'Informatique et des Libertés), cette dernière étant une autorité indépendante et impartiale chargée de veiller au respect de cette loi.

Cette loi confère ainsi plusieurs droits aux individus :

Droit d'information	Toute personne a droit de savoir si des informations concernant son identité sont conservées dans un fichier. Toute personne a le droit de savoir de quel fichier il s'agit et quelle en est la finalité.
Droit d'opposition	Une personne à le droit de s'opposer, à savoir de refuser, à ce que des informations concernant son identité soient collectées.
Droit d'accès	Toute personne a le droit d'accéder aux informations concernant son identité qui seraient stockées au sein d'un fichier, c'est-à-dire de consulter ces informations.
Droit de rectification	Toute personne a le droit de modifier ou de faire modifier les informations relatives à son identité dans le fichier qui contient ces informations.
Droit d'oubli	Toute personne a le droit de demander la suppression des données personnelles stockées le concernant.

2. Sécurisation et fiabilisation des données

2.1. Sauvegarde et pertes de données

Problème	Tout matériel informatique a une durée de vie limitée. Il finit par être inutilisable, soit que le temps ait fini par l'user soit encore qu'il ait été détruit suite à un événement fortuit : piratage, chute, incendie, dégât des eaux, etc. Doit-on se résoudre à perdre toutes ses données lorsqu'un système n'ai plus opérationnel ? Pareillement, un utilisateur peut supprimer ses données par inadvertance. Faut-il encore faire avec ? Après tout « tant pis »… Ou peut-on au contraire envisager d'avoir un système plus fiable ?
Solution	La mise en place d'un système de sauvegarde permet de prévenir les risques de pertes de données. Bien entendu, au sein d'une organisation, il serait difficile de pratiquer des sauvegardes sur tous les matériels informatiques, d'où l'utilité de centraliser les données sur un ou plusieurs serveurs de données (=serveur de fichiers). De la sorte, on pourra se cantonner à mettre en place un dispositif de sauvegarde sur ce ou ces serveurs de données, ce qui s'avère être une solution plus facilement maintenable et administrable. En cas de perte de donnée, les dispositifs de sauvegarde permettent de pratiquer une restauration des données, à savoir qu'ils permettent de recouvrer les données issues d'une sauvegarde.

2.2. Tolérances aux pannes

Problème	Le système d'information (SI) d'une entreprise est nécessaire à son activité. Toute panne, matérielle ou logicielle, risque d'occasionner des dysfonctionnements plus ou moins graves. Que se passe-il par exemple si les salariés d'une entreprise n'ont plus accès à internet ? Sans conteste, l'entreprise perd du chiffre d'affaires et, si le problème persiste, risque jusqu'à faire faillite.
Solution	La tolérance aux pannes consiste à mettre en place des dispositifs matériels et logiciels afin d'assurer la continuité de service, c'est-à-dire afin de s'assurer que le système d'information (SI) continue à fonctionner, même en cas de panne matérielle ou logicielle. Il s'agit souvent de la mise en place de systèmes redondants : on double les batteries, on double les câbles, on utilise des systèmes de sauvegarde redondants (exemple : RAID 10), etc. La redondance permet d'éviter les défaillances. Par exemple, si un système tombe en panne, le système alternatif prend le relai.

2.3. Sécurisation des informations

2.3.1. Authentification et droits d'accès

Problème	Chacun devrait normalement avoir accès aux informations qui le concerne. Il devrait être le seul à pouvoir traiter ses informations personnelles. Par exemple, un utilisateur peut être amené à manipuler des informations confidentielles ou encore intimes, aussi ne devraient avoir accès à ces informations que les personnes concernées.
Solution	L'accès à des ressources informatiques nécessite souvent la mise en place d'un système d'authentification qui permet à un utilisateur d'accéder à ses informations après qu'il a saisi son identifiant et son mot de passe. Son identifiant pourra être, selon les cas : une adresse email, un pseudonyme voire une adresse MAC ou une adresse IP.

Remarque ! Un identifiant se doit d'être unique et propre à l'utilisateur. Aussi, l'adresse IP n'est pas un identifiant fiable dans la mesure où plusieurs utilisateurs peuvent potentiellement avoir la même adresse IP (*cf. cours Q6 : réseau*), voire un utilisateur peut faire croire qu'il a une certaine adresse IP (*spoofing*).

2.3.2. Mot de passe fiable

Problème	Lorsqu'on a recours à l'authentification pour accéder à des données personnelles, on peut craindre qu'un utilisateur, un pirate, parvienne à s'introduire dans le système à la place d'un autre utilisateur en se faisant passer pour lui, c'est-à-dire en lui prenant son identité. Il s'agit d'une forme d'usurpation d'identité. Afin de prévenir ce risque, les utilisateurs doivent en outre disposer d'un mot de passe fiable et confidentiel.
Définition	Tout d'abord, qu'est-ce qu'un mot de passe fiable ? Un mot de passe fiable est un mot de passe qui n'est pas sensible au *brute forcing* (force brute). L'attaque par force brute (*force-brute attack* ou plus simplement *brute forcing*) est un type d'attaque consistant à tester des combinaisons jusqu'à trouver la bonne. A titre informel, une variante du *brute forcing* classique, beaucoup plus efficace, est l'attaque par dictionnaire.
Solution	Il convient pour les usagers d'utiliser des mots de passe suffisamment longs et constitués de suffisamment de types de caractères (chiffres, lettres minuscules/majuscules, caractères spéciaux). Il faut encore que les utilisateurs fassent preuve de précaution et gardent leurs mots de passe secret.

Justification : tous les ordinateurs, même les supercalculateurs, ont des limites de performance. Un ordinateur du marché, relativement puissant, peut approximativement tester 5 millions de possibilités par seconde. Or, le nombre P de possibilités de mots de passe dépend de la longueur N maximum du mot de passe et du nombre C de caractères possibles : $P \approx C^P$.

Preuve (à titre informel) : $P = C^P + C^{P-1} + C^{P-2} + \cdots + C + 1 = \frac{1-C^{P+1}}{1-C} \approx C^P$

Longueur maximum	Types de caractères	Nombre de possibilités	Temps de brute forcing
5	Chiffres Total : 10	$\approx 10^5 = 100\,000$	$\approx 20\ milisecondes$
5	Chiffres et minuscules Total : $10 + 26 = 36$	$\approx 36^5 = 60\,466\,176$	$\approx 12\ secondes$
5	Chiffres et minuscules/majuscules Total : $10 + 26 \times 2 = 62$	$\approx 62^5 = 916\,132\,832$	$\approx 3\ minutes$
10	Chiffres Total : 10	$\approx 10^{10} = 10\ milliards$	$\approx 33\ minutes$
10	Chiffres et minuscules Total : $10 + 26 = 36$	$\approx 36^{10} = 3\,656\ billions$	$\approx 23\ ans$
10	Chiffres et minuscules/majuscules Total : $10 + 26 \times 2 = 62$	$\approx 62^{10} = 839\ billiards$	$\approx 53\ siècles$

2.3.3. Hachage

Problème	Le mot de passe est une donnée personnelle confidentielle. Autrement dit, hormis le propriétaire de ce mot de passe, nul ne devrait en avoir connaissance. Or, si le mot de passe est stocké en clair (non crypté) au sein d'une base de données, une personne y ayant accès pourra librement consulter et récupérer cette information confidentielle.
Solution	En pratique, il est recommandé de ne pas stocker en clair les mots de passe des utilisateurs au sein d'une base de données. On préfère stocker « l'empreinte numérique » du mot de passe, qu'on appelle aussi le « haché ». Cette empreinte est calculée grâce à une fonction de hachage, non inversible, qui retourne une valeur de taille fixe. Deux mots de passe ou messages différents ont une empreinte numérique différente. On parle de hachage des mots de passe.

```php
<?php
    // Phrase non cryptée
    $phraseClaire = 'Un cours de SIG avec M. PAQUEREAU';
    // Calcul du haché de la phrase (fonction de hachage : SHA-256)
    $phraseCryptée = hash('SHA256', $phraseClaire);
    // Affichage du haché calculé
    echo $phraseCryptée;
?>
```

Hachage : calcul de l'empreinte numérique unique de la phrase non cryptée

975cd23f268b5e7c7f98bbac783c47bc1b38070415d08474cef8b1f7e87b21c1

Inversion impossible : on ne peut pas retrouver la phrase initiale à partir du « haché » (fonction non inversible)

Remarque ! Lorsqu'un utilisateur saisit son mot de passe au sein d'un formulaire de connexion, on vérifie que le mot de passe (<input name="mdp" type="password" />) saisi est correct comme ceci :

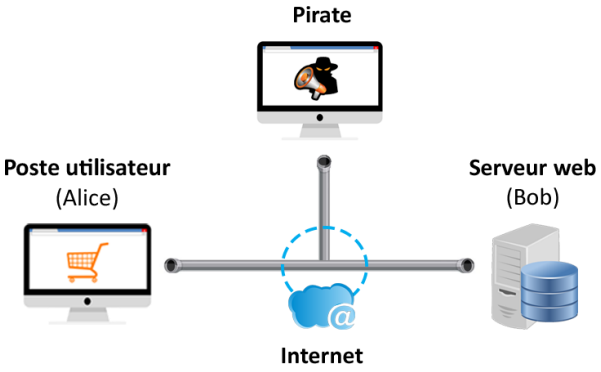
- Dans la base de données, on récupère le mot de passe haché $utilisateur['mdp'] de l'utilisateur ;
- On calcule le « haché » du mot de passe saisi : $mdp = **hash**('SHA256', $_GET['mdp']) ;
- Si $utilisateur['mdp'] est égal à $mdp, c'est que le mot de passe saisi est correct !

3. Sécurisation des échanges

3.1. Cryptage ou chiffrement

Problème

Un risque est celui de l'interception de communication. Intercepter une communication, c'est pour un pirate capter les informations résultant des échanges distants entre deux utilisateurs. L'interception de communication est un risque dès lors que des informations sensibles ou confidentielles sont échangées. Explication et illustration :

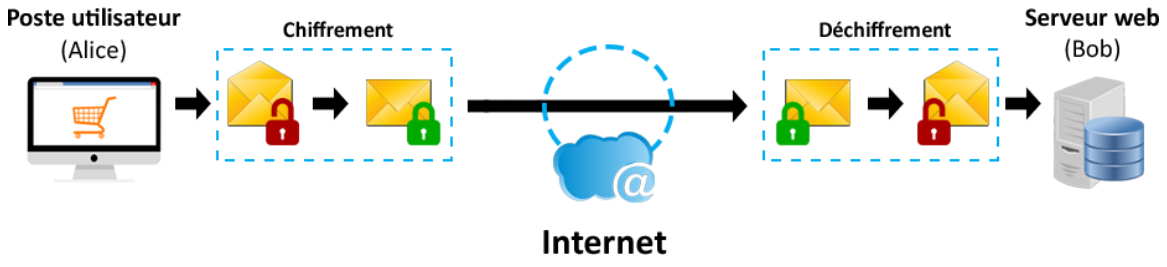


- A partir de son poste informatique, Alice se connecte sur le site e-commerce Bob.
- Alice et Bob échangent donc des messages (requêtes et réponses HTTP, cf. *Cours Q3 - web et PHP*).
- Le pirate intercepte tous les messages échangés.
- Par conséquent, si la communication n'est pas cryptée, le pirate va pouvoir très clairement lire le contenu de tous les messages. C'est pourquoi il convient de crypter/chiffrer la communication. On parle de canal crypté, c'est-à-dire que les messages sont cryptés.

Solution

Dans le cadre d'une communication, les interlocuteurs (Alice et Bob) s'échangent des messages. Le cryptage ou chiffrement d'une communication est un procédé consistant à masquer le contenu des messages circulant entre interlocuteurs :

- L'émetteur, avant de transmettre son message, va le crypter ;
- Il va ensuite transmettre ce message, lequel va circuler sur le réseau ;
- Le récepteur reçoit le message crypté ;
- Il procède au décryptage du message afin de disposer d'un message en clair, c'est-à-dire d'un message lisible, qu'il peut traiter.



3.2. Cryptage symétrique et asymétrique

Un algorithme de cryptage/chiffrement nécessite une clé de cryptage. C'est cette clef de cryptage qui va servir à crypter et décrypter les messages échangés. De manière très générale, il existe deux formes de cryptage :

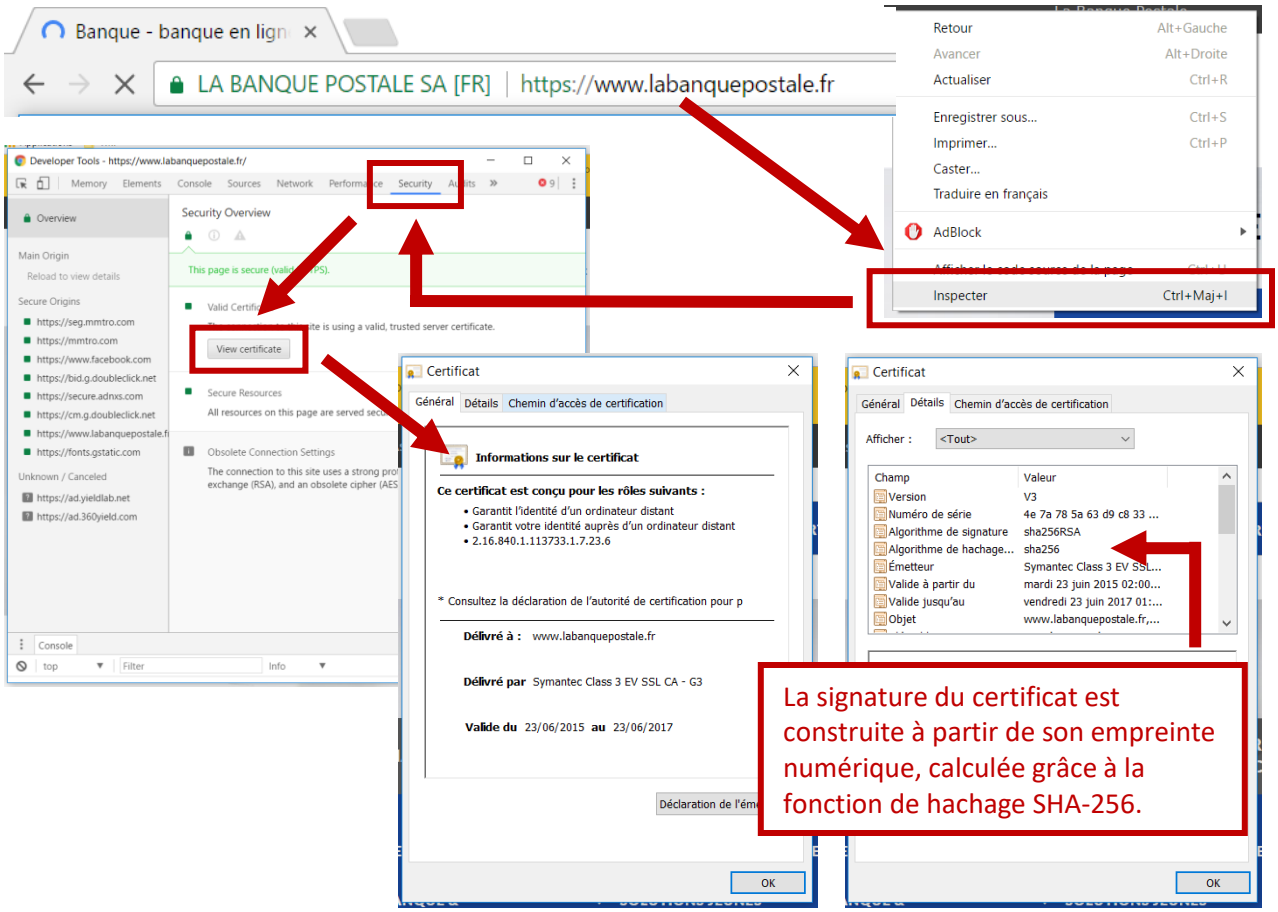
- Le **cryptage symétrique** : la clef de cryptage est partagée par l'émetteur (Alice) et le récepteur (Bob). Elle servira tout à la fois à crypter et à décrypter les messages. Le problème ? Alice et Bob doivent au préalable s'échanger les clefs. Alors, le pirate peut se faire passer pour Bob auprès d'Alice et pour

Alice auprès de Bob (usurpation d'identité). Cette forme de piratage est appelée attaque du *man in the middle* (l'homme au milieu).

- Le **cryptage asymétrique** : la clef de cryptage est constituée d'une clef publique et d'une clef privée. Bob transmet à Alice sa clef publique. Elle permet à Alice de crypter les messages qu'elle transmet à Bob. Bob peut décrypter les messages cryptés grâce à la clef privée, que lui seul connaît. Le problème ? Une fois encore, le pirate peut usurper l'identité d'Alice et Bob et mettre en œuvre une attaque de *man in the middle*. Pour contrer le problème, on utilise des certificats d'authenticité qui permettent de vérifier que la clef publique appartient bien à la bonne personne, ici Bob.

3.3. Authenticité et certificat électronique

Problème	Dans le cadre d'une communication réseau, le chiffrement des communications ne résout pas le problème d'usurpation d'identité. Et l'on peut dès lors se demander : comment puis-je être sûr de l'identité de mon interlocuteur ? En effet, il ne faudrait par exemple pas qu'un pirate se fasse passer pour le serveur web sur lequel quelqu'un tente d'effectuer un paiement en ligne.
Solution	Le certificat d'authenticité ou certificat électronique est un dispositif permettant de s'assurer de l'identité d'un serveur. Il s'agit en quelque sorte d'une carte d'identité numérique d'un serveur. Il est délivré par une autorité de certification et permet, dans le cadre d'une communication réseau, de vérifier que les données reçues proviennent bien du serveur auquel on a tenté de s'adresser. Le protocole HTTPS, sécurisé, a recours à ces certificats (voir ci-dessous).



3.3.1. Exemple d'algorithme de cryptage

Etudions ensemble un algorithme de cryptage simple. Cet algorithme n'est pas fiable mais c'est un premier pas ! Il s'agit d'un algorithme de cryptage symétrique. L'algorithme que nous allons découvrir est le cryptage par substitution, aussi appelé cryptage César, ce dernier l'ayant utilisé pour crypter manuellement ses propres messages à l'époque romaine. La clef de cryptage de cet algorithme est un nombre entier. Et le principe de l'algorithme est le suivant :

- On associe un nombre à un caractère :

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
U	V	W	X	Y	Z	0	1	2	3	4	5	6	7	8	9	!			
20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37		

+5 (de N à S)

+5 (de U à Z)

Espace (37)

- On se donne une clef de cryptage comprise entre 0 et le nombre de caractères, soit ici entre 0 inclus et 38 exclus, par exemple : $N = 5$
- On prend son message, par exemple : « UN MESSAGE CRYPTE »
- On va crypter un-à-un chacun des caractères du message en décalant chacun d'entre eux d'exactement N caractères.

U	N		M	E	S	S	A	G	E		C	R	Y	P	T	E	*Message en clair*
20	13	37	12	4	18	18	0	6	4	37	2	17	24	15	19	4	
25	18	4	17	9	23	23	5	11	9	4	7	22	29	20	24	9	*Message crypté*
Z	S	E	12	J	X	X	F	L	J	E	H	W	3	U	Y	J	

- Pour décrypter le message, on procèdera en sens inverse, c'est-à-dire qu'on effectue les mêmes calculs, mais avec la clef inverse, soit ici : $N = -5$ ou $N = 38 - 5 = 33$

Afin que vous puissiez tester par vous-même cet algorithme de cryptage/décryptage, en voici une implémentation en PHP :

```php
// Toutes les lettres majuscules et les caractères " " (espace) et "!", soit un jeu de 28 caractères
$lettres = "ABCDEFGHIJKLMNOPQRSTUVWXYZ !" ;
// Phase à crypter
$phrase = "UN SUPER COURS DE CRYPTO AVEC M PAQUEREAU !" ;
// Clef de cryptage
$clef = 9 ;
// longueur de la phrase
$n = strlen($phrase) ;
// cryptage de chacune des lettres de la phrase
for($i=0; $i<$n; $i++){
    // i-ème lettre de la phrase
    $lettre = $phrase[$i] ;
    // récupère le n° du i-ème caractère
    // exemple : 'A' => 0, 'B' => 1, ..., '!' => 27
```

```
        $numero = strpos($lettres, $lettre) ;
        // nouveau n° de la i-ème lettre (substitution)
        $numero = $numero + $clef ;
        // permet de "revenir à 0"
        // exemple : 28 => 0, 29 => 1, 30 => 2, etc.
        $numero = $numero % 28 ;
        // remplace la i-ème lettre de la phrase par sa
        // version cryptée
        $phrase[$i] = $lettres[$numero] ;
}
// affiche la phrase cryptée
echo $phrase ;
```

BWH!BYN HLXB !HMNHL
FYAXHJCNLHVHYJZBN NJBHI

Calcul de la clef de décryptage : dans l'exemple ci-dessus, la clef de cryptage est 9 et le jeu de caractères compte 28 caractères donc la clef de décryptage est :