

Exfiltration de Données Sans Connectivité Réseau via Codes QR, Une Menace Sous-Estimée"

|  |
| --- |
| Par Alain NICOLAS  2022/2023  Université de Technologie TROYES  Thèse professionnelle de Mastère Spécialisé® Expert en Cybersécurité    Tuteur : Alain CORPEL |

Résumé exécutif

Un executive summary, également appelé résumé exécutif, est un résumé succinct et condensé d'un document ou d'un rapport qui vise à fournir aux décideurs et aux dirigeants une vue d'ensemble rapide des informations clés, des principales conclusions et des recommandations essentielles. Il est généralement utilisé dans un contexte professionnel ou organisationnel pour permettre aux destinataires occupés de comprendre rapidement les aspects les plus importants d'un document sans avoir à lire l'intégralité du contenu.

L'executive summary se trouve généralement au début du document, après la page de titre, mais avant le contenu détaillé. Il est rédigé de manière concise et précise, et il est souvent limité à une ou deux pages, selon les exigences spécifiques.

Dans un executive summary, vous devez inclure les éléments suivants :

1. Contexte et objectifs : Présentez le contexte de la problématique abordée dans le document et expliquez les objectifs principaux de l'étude ou du rapport.

2. Méthodologie : Résumez brièvement les méthodes de recherche ou les approches utilisées pour collecter les données et analyser les informations.

3. Principales conclusions : Résumez les résultats et les principales conclusions de l'étude ou du rapport. Mettez en évidence les découvertes les plus importantes, les tendances clés ou les faits saillants.

4. Recommandations : Présentez de manière concise les principales recommandations ou actions proposées à la suite des conclusions. Indiquez les mesures spécifiques à prendre ou les décisions à prendre pour résoudre la problématique ou pour mettre en œuvre les améliorations suggérées.

5. Implications et impacts : Discutez brièvement des implications stratégiques, des avantages potentiels ou des impacts attendus des recommandations. Mettez en évidence les bénéfices attendus et les résultats escomptés.

L'executive summary doit être rédigé de manière claire, concise et persuasif, en utilisant un langage simple et accessible. Il doit capter l'attention des lecteurs et leur donner une compréhension rapide des points clés du document, afin qu'ils puissent prendre des décisions éclairées en fonction des informations présentées.

Il est important de noter que l'executive summary est généralement destiné à un public spécifique, souvent des décideurs ou des hauts responsables, et il doit donc être adapté à leurs besoins et à leurs préoccupations spécifiques.

Mots-clés : exfiltration de données, télétravail, QR codes, capteurs de la caméra, prévention des fuites de données, sécurité des données sensibles.

Remerciements

Francis, Pascal et Jean Christophe

Ma famille

BNPP

Reza

Alain Corpel

Attack Model nom du graphe

Conclusion: https://www.slideshare.net/Checkmarx/mind-the-airgap-checkmarx-research-into-nfc-and-smart-bulb-data-exfiltration

https://blogs.cisco.com/security/sensitive-data-exfiltration-and-the-insider

Liste des tableaux et figures

Liste des abréviations

Les abréviations présentes dans le texte sont suffixées par un astérisque \*. Vous pouvez vous référer à la liste ci-dessous pour obtenir sa signification.

|  |  |
| --- | --- |
| **Abréviation** | **Signification acronyme** |
|  |  |
| **bps** | bit per second |
| **CC** | Critères Communs |
| **CCPA** | California Consumer Privacy Act |
| **DLP** | Data Loss Prevention |
|  |  |
|  |  |
| **POC** | Proof Of Concept |
|  |  |
|  |  |
| **SMC** | Small Message Criterion |
|  |  |
| **RGPD** | le Règlement général sur la protection des données |
| **TCSEC** | Trusted Computer System Evaluation Criteria |

Glossaire

Les termes présentes dans le texte sont suffixés par un astérisque #. Vous pouvez vous référer au glossaire suivant pour obtenir leur définition.

|  |  |
| --- | --- |
| **Termes** | **Définitions** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| **bit** | Le bit est atomique, la plus petite unité de stockage informatique qui prend la valeur 0 ou 1 |
| **bit per second** | Le bps est une unité de mesure informatique usuellement utilisée dans les mesures de débits de liaisons informatiques. |
| **Byte** | Un octet soit 8 bits |
| **Code QR** | Un code QR (en anglais QR code), en forme longue quick response code (« code à réponse rapide »), est un type de code-barres à deux dimensions constitué de modules-carrés noirs disposés dans un carré à fond blanc. Ces points définissent l'information que contient le code. Ce dessin, lisible par machine, peut être visualisé sur l'écran d'un appareil mobile ou imprimé sur papier. |
| **Critères Communs** | Les critères communs (CC) sont un ensemble de normes (ISO 15408) internationalement reconnu dont l'objectif est d'évaluer de façon impartiale la sécurité des systèmes et des logiciels informatiques. Également dénommés Common Criteria, ce référentiel est né d'un partenariat entre le Canada, les États-Unis et l'Europe. |
| **Data Loss Prevention** | Une pproche de la cybersécurité visant à protéger les données sensibles et confidentielles d'une organisation, en prévenant leur divulgation, leur vol ou leur perte non autorisée. |
|  |  |
| **Proof Of Concept** | Le proof of concept (POC ou « preuve de concept », pour la définition de POC en français) sert à démontrer la faisabilité d’un produit, d’une méthode ou d’une idée. En effet, pour que les parties prenantes et investisseurs donnent suite à votre projet en toute confiance, vous devez prouver le bien-fondé de votre idée. |
| **Rançongiciel** | (ou ransomware en anglais) Type de logiciel malveillant qui chiffre les fichiers d'un système informatique et demande une rançon pour les déchiffrer. Les cybercriminels utilisent des rançongiciels pour bloquer l'accès aux fichiers d'un utilisateur et exigent un paiement en échange d'une clé de déchiffrement. Les rançongiciels peuvent se propager par e-mail, sites web infectés ou vulnérabilités logicielles. Les attaques de rançongiciels représentent une menace majeure en cybersécurité, avec des conséquences financières et opérationnelles significatives pour les individus et les organisations. |
| **Trusted Computer System Evaluation Criteria** | Les Trusted Computer System Evaluation Criteria, expression abrégée en TCSEC, sont un ensemble de critères énoncés par le département de la Défense des États-Unis qui permettent d'évaluer la fiabilité de systèmes informatiques centralisés  On parle parfois de l'Orange Book (livre orange), d'après la couverture du livre. Il s'agit d'un cahier des charges définissant quatre degrés de sécurité, de A à D ; A étant le niveau le plus sécurisé. En général, le niveau C2 est suffisant. |
| **Tunneling** | Le tunneling est une technique de cybersécurité qui permet d'encapsuler et de transmettre des données sécurisées à travers un réseau non sécurisé. Il crée un "tunnel" virtuel qui garantit la confidentialité et l'intégrité des données lorsqu'elles sont transférées d'un point à un autre. Cette méthode est couramment utilisée pour établir des connexions sécurisées à distance, comme les tunnels VPN ou SSH. |
|  |  |

Résumé (Executive Summary) :

Mots-clés : exfiltration de données, télé-travail, QR codes, webcam, prévention des fuites de données, sécurité des données sensibles, airgap.

# Introduction

## Importance de la Donnée

Pour la grande majorité des entreprises de taille significative,les données constituent un **actif majeur**.

La protection des données d'entreprise est d'une importance capitale dans le contexte actuel de la digitalisation. Elle doit permettre de garantir la confidentialité, l'intégrité et la disponibilité des informations sensibles appartenant à l'entreprise, ainsi que d’être conforme aux lois et réglementations en vigueur.

La protection des données d'entreprise est essentielle pour plusieurs raisons :

⇒ La confidentialité des informations : Les entreprises collectent et stockent souvent des données sensibles telles que des informations financières, des données clients, des secrets commerciaux et des propriétés intellectuelles telles que du code informatique. La protection de ces données confidentielles est essentielle pour éviter les fuites d'informations et les atteintes à la réputation de l'entreprise.

⇒ La conformité réglementaire : Les réglementations telles que le RGPD\* en Europe et le CCPA\* aux États-Unis imposent des obligations légales en matière de protection des données. Le non-respect de ces réglementations peut entraîner des amendes financières et des dommages à la réputation de l'entreprise.

⇒ La gestion des risques : Les cyberattaques et les violations de données peuvent avoir des conséquences financières désastreuses pour une entreprise. Les coûts associés à la remédiation d'une violation de données, tels que les enquêtes forensiques, les notifications aux personnes concernées et les mesures correctives, peuvent être extrêmement élevés. Une bonne protection des données aide à réduire les risques et les impacts financiers potentiels.

⇒ La confiance des Clients et partenaires : Les clients et les partenaires commerciaux attachent de plus en plus d'importance à la protection de leurs données personnelles. Une entreprise qui démontre un engagement envers la sécurité des données peut gagner la confiance de ses clients et de ses partenaires, renforçant ainsi sa réputation et sa compétitivité.

⇒ L’avantage concurrentiel : Dans un paysage commercial de plus en plus concurrentiel, une solide posture en matière de protection des données peut constituer un avantage concurrentiel pour une entreprise. Les clients sont de plus en plus conscients des risques liés à la protection des données et préfèrent souvent faire affaire avec des entreprises qui accordent la priorité à la sécurité de leurs informations.

## L’exfiltration de la donnée

Cette augmentation de la quantité et de la valeur des données a également ouvert la porte à de nouveaux risques et menaces.

Les entreprises sont confrontées à une multitude de menaces potentielles telles que les cyberattaques, les vols de données, les rançongiciels(Voir Glossaire) et les violations de la vie privée.

Les techniques d'exfiltration de données sont diverses et en constante évolution, allant des attaques de type "man-in-the-middle" aux logiciels malveillants sophistiqués, en passant par l'utilisation abusive de services cloud. Cependant, toutes ces techniques ont un point commun : elles exploitent généralement la connectivité réseau des systèmes pour transférer les données.

Mais qu'en est-il si cette connectivité réseau n'était pas nécessaire pour l'exfiltration de données ?

## Objectif de cette thèse professionnelle

Dans le contexte actuel de l'augmentation du télétravail, cette thèse se propose d’explorer un scénario mettant à l'épreuve les mécanismes traditionnels de prévention des pertes de données (DLP). Imaginez un collaborateur en fin de mission, un "insider", qui cherche à exfiltrer des données sensibles - que ce soit du code informatique, des documents sensibles, ou d'autres informations de valeur - avant de quitter l'entreprise. Ce collaborateur travaille à distance, à l'abri du regard direct de ses collègues ou supérieurs, ce qui lui donne plus de liberté pour mener à bien son acte malveillant.

Il dispose de deux devices : l'une, un ordinateur portable Entreprise, la source des informations qu'il cherche à exfiltrer. L'autre est une device personnelle, qui lui permet de recevoir et de stocker les données exfiltrées. L'objectif du collaborateur est de contourner les systèmes de DLP en place, qui sont généralement conçus pour surveiller et contrôler les transferts de données via les réseaux.

L’objet vecteur d’information est le code QR, si familier, pour transférer les données sans faire appel à la connectivité réseau.

Cette méthodologie présente une menace significative, car elle pourrait passer inaperçue par les systèmes de DLP classiques. En développant et en démontrant un proof of concept, nous cherchons à comprendre cette menace et à proposer des contre-mesures pour la combattre."

À travers un état de l’art puis une analyse de cette menace basée sur mon proof of concept, ma contribution est de mettre en lumière cette technique d'exfiltration de données, de la démontrer techniquement ,évaluer le débit et de proposer des contre-mesures pour aider les entreprises à se protéger contre elle.

# Contexte et état de l’art

## La technologie code QR

### Histoire et cas d’utilisations

Les codes Quick Response (QR) ont été créés par Denso Wave, une filiale de Toyota, en 1994 et étaient initialement utilisés pour suivre l'inventaire lors de la fabrication de pièces de véhicules. Ils peuvent être décrits comme un code-barres matriciel bidimensionnel ayant la capacité de stocker des données (Tiwari, 2016). De nos jours, les codes QR sont de plus en plus populaires car leur précision, leur vitesse de lecture et leur commodité se sont révélées supérieures à celles des codes-barres linéaires unidimensionnels (Narayanan, 2012). Les codes QR peuvent contenir beaucoup plus de données et peuvent être facilement scannés avec un smartphone. Cela a incité les entreprises à les intégrer dans leurs stratégies de marketing, où ils peuvent être utilisés pour promouvoir l'entreprise et fournir au client des informations supplémentaires sur un produit ou un service. Les applications des codes QR sont nombreuses et sont utilisées dans plusieurs industries car elles peuvent facilement encoder une URL, un numéro de téléphone, une adresse e-mail, du texte, et même donner accès à un réseau Wi-Fi.

### Fonctionnement du QR Code

Le système de codes QR est composé d'une procédure d'encodage et de décodage. La procédure d'encodage consiste à convertir du texte en utilisant l'un des quatre modes disponibles : alphanumérique, numérique, octets ou kanji, en une séquence de bits (qui sont des 1 et des 0).

#### encodage

Lorsque le texte est encodé en utilisant l'un des quatre modes, une séquence de bits est construite qui est divisée en mots de code de données de 8 bits de long. Des mots de code de correction d'erreur sont ensuite générés, qui sont correctement ordonnés aux côtés des mots de code de données et sont ensuite placés de manière spécifique dans la matrice 2D du code QR.

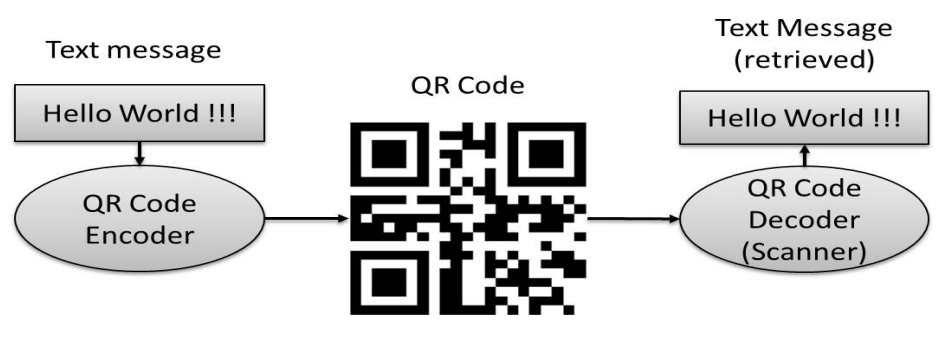
Ensuite, le masquage de données est utilisé pour alterner la couleur de chaque module de noir à blanc ou vice versa pour permettre aux scanners de lire facilement le code QR (les modules sont les motifs carrés noirs et blancs qui composent le code QR). Il existe 8 modèles de masque qui peuvent être choisis, le plus adapté doit être calculé en vérifiant le score de pénalité que chacun aurait s'il était utilisé, et celui avec le score de pénalité le plus bas doit être sélectionné (Garg, 2015).

Les pixels d'information de version et de format sont ensuite ajoutés au code QR dans les zones qui ne sont pas occupées par les modules de données.

Figure 1: Modèles de masques (Assaad, 2019)

#### décodage

Lorsque le code QR est scanné, leur arrangement se traduit de nouveau en la forme originale des données, c'est le processus de décodage (Scott, 2020). Cela se produit lorsque le scanner analyse les modules blancs et noirs comme un tableau 2D composé de bits de 1 et de 0. Les informations de version et de format sont déterminées, puis le motif de masquage est libéré. Les mots de code de correction d'erreur et de données sont restaurés et toute erreur est corrigée à l'aide des mots de code de correction d'erreur. Enfin, les données sont décodées en utilisant le mode qui a été utilisé au départ pour encoder les données.

Figure 2: Encodage et décodage d'un code QR (Tiwari, 2016)

### Structure d’un code QR

La structure d'un code QR comprend :

- Un motif de synchronisation - Permet au décodeur de déterminer la largeur d'un module de données unitaire.

- Une zone de repos - Il s'agit de l'espace vide qui entoure le code QR et qui permet de l'identifier par rapport à son environnement.

- Un motif d'alignement - Ce motif permet de décoder le code QR sous tous les angles, en corrigeant la distorsion du code QR. Les codes QR de version 1 n'ont pas de motif d'alignement.

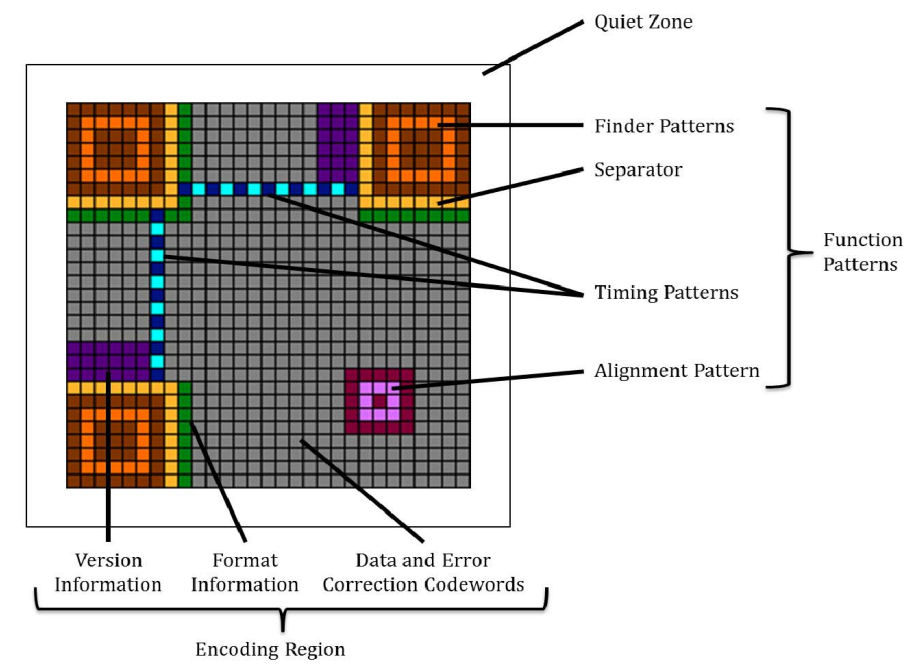
- Un motif de position - Il y a 3 motifs de position situés sur les codes QR qui aident les décodeurs à indiquer la bonne direction du code QR.

- Données et correction d'erreur - Stocke les informations du code QR et les mots de code de correction d'erreur.

- Un motif de version - Ce motif identifie la version du code QR.

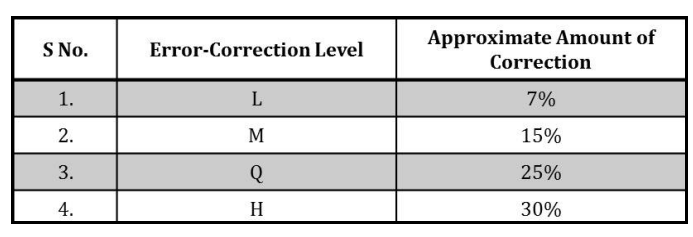
- Format - Ce motif contient des informations sur la correction d'erreur et le masque de données.

- Séparateurs - Il s'agit d'une zone d'espace blanc qui sépare le motif de position de la zone d'encodage. La zone d'encodage étant la zone qui contient les informations de format, la version, **et les données**.

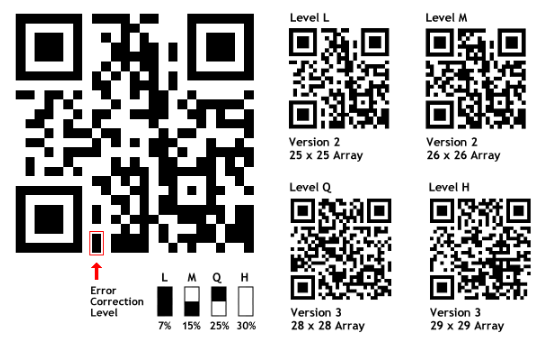
Figure 3: Structure d'un code QR (Tiwari, 2016)

#### Correction d’erreur

Les codes QR contiennent une fonction de correction d'erreur qui leur permet d'être lus même si le code QR est endommagé ou sale. Les bits de données que nous encodons sont utilisés pour créer des mots de code de correction d'erreur. Cela est réalisé en utilisant les codes de Reed-Solomon, qui sont des codes algébriques qui effectuent une correction d'erreur vers l'avant. Des blocs de données numériques sont pris par un encodeur Reed-Solomon et ajoutés ensemble avec des bits supplémentaires. Chaque bloc est traité par le décodeur Reed-Solomon et une tentative est faite pour récupérer les données originales et corriger les erreurs trouvées. La catégorie et le nombre d'erreurs qu'il tente de corriger sont basés sur les propriétés du code de Reed-Solomon utilisé (Anon., 2012). Il existe 4 niveaux différents de correction d'erreur que les codes QR peuvent utiliser.

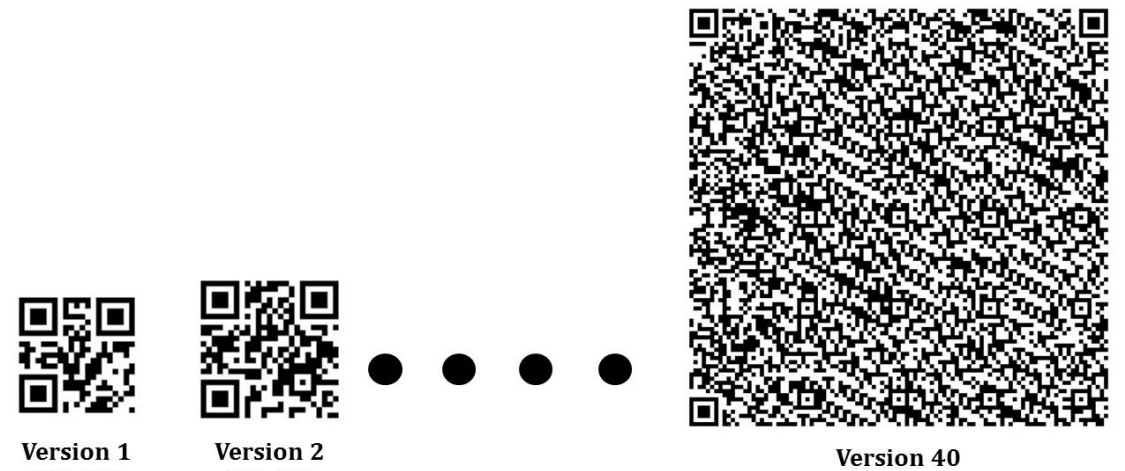
Tableau 1: Niveau de correction d'erreur et % de correction d'erreur

Chaque niveau ajoute une quantité différente de données de sauvegarde en fonction du niveau de dommages que le code QR est censé subir. Plus le pourcentage de correction d'erreur utilisé est élevé, plus le code QR devient grand car davantage de rangées et de colonnes sont ajoutées pour soutenir l'ajout des données de sauvegarde (admin, 2011).

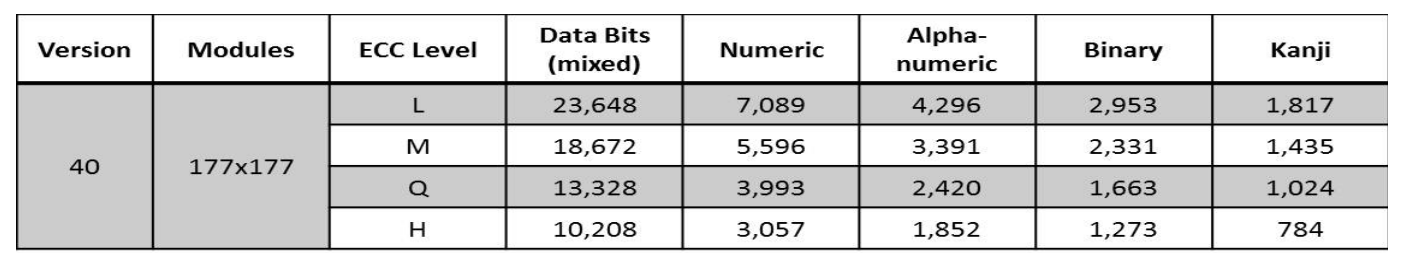
Tableau 2: Même information encodée dans les 4 niveaux de correction d'erreur. (admin, 2011)

#### Tailles, version et capacité

La taille d'un code QR varie car il existe différentes versions qui peuvent être créées. Ces versions vont de 1 à 40. Chacune de ces versions a un nombre contrasté de modules. Les versions 1 se composent de 21 rangées de pixels par 21 colonnes de pixels, tandis que le plus grand code QR possible est la version 40 qui se compose de 177 rangées de pixels par 177 colonnes de pixels. Avec le plus grand code QR possible composé de 177x177 modules, cela signifie qu'il peut comporter jusqu'à 31 329 carrés qui peuvent encoder 3 Ko de données. Cela donne à chaque mode un nombre maximal de caractères possibles qui peuvent être encodés, comme le montre le tableau ci-dessous.

Figure 4: Aperçu des différentes versions de code QR (Tiwari,2016)

Une augmentation du niveau de correction d'erreur réduit la quantité de données utiles qui peuvent être encodées dans le code QR. Voici un tableau récapitulatif pour la version 40, celle que je vais utiliser dans mon POC.

Tableau 3: Capacité en datas de la version 40 du code QR en fonction du type de data et du niveau de correction d'erreur (Tiwari, 2016)

## Risques de cybersécurité potentiels liés à l'utilisation malveillante de la technologie des codes QR.

Les codes QR sont largement utilisés dans divers domaines, tels que le marketing, la publicité, le suivi des produits, les billets d'événements, les cartes de visite électroniques, les paiements mobiles, etc. Leur popularité est due à leur capacité à stocker des informations de manière compacte et à leur facilité d'utilisation grâce aux smartphones et aux applications de lecture de QR code.

L’ère du Covid-19 lui aura permis un regain d’utilisation en 2020. Les règles sanitaires préconisant de toucher le moins de surface possible afin de limiter la propagation du coronavirus, le QR Code se révèle comme un moyen pratique de donner accès à des données « à distance »1.

Les smartphones actuels permettent de flasher facilement les codes QR soit avec une application dédiée soit avec la lecture directement intégrée dans la caméra.

L'augmentation des usages liés à cette technologie attire les cybercriminels qui remplacent des codes légitimes par des codes malveillants2 permettant d’opérer les types d’attaques suivants :

* Phishing : Les codes QR malveillants peuvent être utilisés pour diriger les utilisateurs vers de faux sites web conçus pour voler des informations sensibles. Les attaquants peuvent inciter les utilisateurs à fournir leurs identifiants de connexion, leurs informations bancaires ou d'autres données personnelles via des pages web frauduleuses.
* Malware : Les codes QR peuvent être utilisés comme vecteurs de distribution de malwares. En scannant un QR code compromis, un utilisateur peut involontairement télécharger et installer un logiciel malveillant sur son appareil. Ce logiciel peut permettre à l'attaquant de prendre le contrôle de l'appareil, d'accéder à des données sensibles ou de surveiller les activités de l'utilisateur.
* Attaques de paiement : Les codes QR peuvent être manipulés pour effectuer des paiements frauduleux. Les attaquants peuvent remplacer un QR code légitime par un code malveillant, conduisant ainsi à un transfert d'argent non autorisé depuis le compte de l'utilisateur vers celui de l'attaquant.
* Ingénierie sociale : Les codes QR peuvent être utilisés dans des attaques d'ingénierie sociale pour tromper les utilisateurs et les inciter à effectuer des actions indésirables. Par exemple, un QR code malveillant peut diriger les utilisateurs vers une fausse promotion ou une fausse loterie, les incitant à partager des informations personnelles ou à effectuer des paiements.
* Vulnérabilités des lecteurs de codes QR : Les lecteurs de codes QR peuvent présenter des vulnérabilités de sécurité, notamment des failles de lecture ou des erreurs de traitement. Les attaquants peuvent exploiter ces vulnérabilités pour exécuter du code malveillant sur les appareils des utilisateurs ou pour collecter des informations sensibles.
* Fuites d'informations sensibles : Les codes QR peuvent contenir des informations sensibles telles que des informations d'identification personnelle, des numéros de carte de crédit ou des données confidentielles. Si ces codes QR sont interceptés ou exploités de manière malveillante, ces informations peuvent être divulguées et utilisées à des fins nuisibles.

## Méthodes d'exfiltration de données

L'exfiltration de données se produit lorsqu'il y a une copie, un transfert ou une récupération non autorisés de données d'un serveur ou de l'ordinateur d'un individu3 .

Les organisations possédant des données de grande valeur sont particulièrement exposées à ce type d'attaques, qu'elles soient le fait d'acteurs extérieurs ou d'initiés de confiance.

Les menaces internes constituent l'un des facteurs majeurs contribuant à l'exfiltration de données, qu'elles soient causées par des erreurs accidentelles ou des actes malveillants. Les menaces d'initiés malveillants se réfèrent à des individus en position de confiance au sein d'une organisation, qui cherchent délibérément à extraire des données dans le but de nuire à l'entreprise dans leur propre intérêt ou dans celui d'autres personnes.

L'exfiltration de données est une préoccupation majeure pour les organisations de nos jours. Selon une récente étude de McAfee, 61 % des professionnels de la sécurité ont fait face à une violation de données dans leur entreprise actuelleErreur : source de la référence non trouvée. Cette statistique souligne l'importance croissante de prendre des mesures préventives et des mesures de protection pour contrer ces menaces internes et minimiser les risques liés à l'exfiltration de données.

### Méthodes courantes

#### Utilisation des protocoles de communication légitimes :

Cette méthode consiste à utiliser des protocoles de communication légitimes, tels que HTTP, DNS, FTP, SSH pour exfiltrer des données en encapsulant les informations sensibles dans les requêtes ou les réponses de ces protocoles. (Cidon, 2018).

#### Tunneling :

Les attaquants utilisent des techniques de tunneling pour transférer les données hors du réseau, en établissant des tunnels sécurisés tels que des tunnels VPN ou SSH, afin de masquer le trafic de données et de contourner les systèmes de détection. (Brown, 2016)

#### Phishing :

Technique d'ingénierie sociale où des attaquants se font passer pour une entité légitime afin de tromper les utilisateurs et les inciter à divulguer des informations sensibles, telles que des identifiants de connexion ou des informations financières. (Jakobsson, 2019)

#### Pharming :

Attaque visant à rediriger le trafic d'un site web légitime vers une fausse version contrôlée par un attaquant. Cela permet de collecter des informations confidentielles telles que les identifiants de connexion. (Cohen, 2017)

#### Social Engineering :

Méthode utilisée pour manipuler les individus afin d'obtenir des informations confidentielles ou de les persuader d'effectuer des actions indésirables. Cela peut inclure des techniques de manipulation psychologique pour inciter les personnes à divulguer des informations sensibles. (Hadnagy, 2018)

#### Shoulder surfing :

Technique où un attaquant observe discrètement l'écran d'un utilisateur pour collecter des informations sensibles, telles que des mots de passe ou des informations confidentielles affichées à l'écran. (Roush, 2018)

#### Privilege escalation :

Méthode consistant à obtenir des privilèges plus élevés sur un système ou un réseau que ceux initialement attribués, ce qui permet d'accéder à des informations sensibles ou de contourner les restrictions de sécurité. (Sepehrdad, 2020)

#### Botnets :

Réseau d'ordinateurs infectés par des logiciels malveillants et contrôlés à distance par un attaquant. Les botnets peuvent être utilisés pour exfiltrer des données sensibles en utilisant les machines compromises comme points de sortie. (Antonakakis, 2017)

#### Rootkits :

Logiciels malveillants conçus pour accéder ou contrôler un système de manière furtive, en masquant leur présence et en modifiant les fonctionnalités du système d'exploitation. Les rootkits peuvent être utilisés pour exfiltrer des données de manière invisible. (Choo, 2021)

#### Spyware :

Logiciels malveillants conçus pour collecter des informations sur un utilisateur ou un système sans le consentement de l'utilisateur. Les données collectées peuvent être exfiltrées à des fins malveillantes. (Eckert, 2020)

#### Utilisation de périphériques externes :

Les attaquants exploitent des périphériques externes tels que des clés USB, des disques durs externes ou des cartes mémoire pour copier et transporter les données sensibles hors du réseau. (Conti, 2010)

#### Le cas de la Stéganographie :

Elle constitue une technique intéressante car elle permet de dissimuler des informations sensibles au sein d'autres médias ou supports de communication, tels que des images, des fichiers audio ou des vidéos, sans éveiller les soupçons. Cela peut être utilisé pour exfiltrer des données de manière furtive, en les cachant au sein d'un média apparemment anodin. La stéganographie est une technique qui permet de cacher des informations sensibles au sein d'autres données, telles que des images, des fichiers audio ou des vidéos, sans éveiller les soupçons. Elle peut être utilisée à la fois comme méthode d'exfiltration et de protection de l'information.

Exfiltration de l'information :

Lorsqu'il s'agit d'exfiltrer des données, la stéganographie permet de dissimuler des informations confidentielles au sein de médias apparemment innocents. Par exemple, des fichiers sensibles peuvent être dissimulés dans des images ou des documents apparemment normaux. Cela permet aux attaquants de contourner les mécanismes de détection et de passer inaperçus lors de la transmission des données hors du réseau. Les destinataires autorisés peuvent extraire les données cachées en utilisant une clé ou une méthode spécifique.

Protection de l'information :

La stéganographie peut également être utilisée comme moyen de protection de l'information. En cachant les données sensibles au sein de supports apparemment anodins, elle rend difficile la détection et la compréhension de ces données par des personnes non autorisées. Cela peut être utile pour protéger des informations confidentielles lorsqu'elles sont stockées ou transmises. Par exemple, en utilisant la stéganographie, il est possible de dissimuler des messages ou des fichiers importants dans des médias publics tels que des images sur Internet. (Johnson, 1998)

### Méthodes plus atypiques : les canaux cachés

Ici, ce sont des canaux de communication dissimulés qui permettent à des utilisateurs ou à des programmes de transférer des données de manière secrète et non autorisée. Les « covert channels » peuvent être utilisés pour exfiltrer des données sans être détectés. (Vasilomanolakis, 2020)

Ici, ces sont les éléments matériels, physiques qui à travers leur fonctionnement produisent des ondes, des vibrations, des changements d’états qui sont utilisés pour encoder et transporter l’information. Ce sont des techniques plus sophistiquées mettant en relation différents capteurs.

#### Signaux Acoustiques :

Cette méthode exploite les variations acoustiques pour exfiltrer des données, en utilisant par exemple les haut-parleurs et les microphones d'un système pour transmettre les informations sous forme de signaux sonores inaudibles. (Asonov, 2004)

Asonov, D., & Agrawal, R. (2004). Keyboard acoustic emanations revisited.

#### Signaux Électromagnétiques :

Les attaquants peuvent utiliser des techniques d'émission électromagnétique pour exfiltrer des données, en captant les signaux électromagnétiques émis par les dispositifs électroniques lors du traitement des données sensibles. (Kuhn, 2013)

Kuhn, M. G. (2013). Compromising electromagnetic emanations of wired and wireless keyboards.

#### Signaux Thermiques :

Cette méthode exploite les variations de température générées par les composants électroniques lors du traitement de données sensibles pour transmettre les informations à travers des variations thermiques détectables. (Genkin, 2014)

Genkin, D., Shamir, A., & Tromer, E. (2014). RSA key extraction via low-bandwidth acoustic cryptanalysis.

#### Signaux Optiques :

Les attaquants peuvent utiliser des techniques optiques pour exfiltrer des données, en exploitant par exemple les variations de luminosité d'un écran ou les clignotements d'un voyant LED pour transmettre des informations discrètes. (Backes, 2010)

Backes, M., Dürmuth, M., & Pinkal, M. (2010). Compromising reflections or how to read LCD monitors around the corner.

#### Signaux Vibratoires :

Cette méthode se base sur la capture des vibrations émises par les appareils lors du traitement des données, en utilisant des capteurs sensibles aux vibrations pour extraire les informations transférées (Xu, 2012)

Xu, H., Zhuang, Y., & Liu, K. (2012). An experimental study of air-gap covert channels in RF-based gesture recognition.

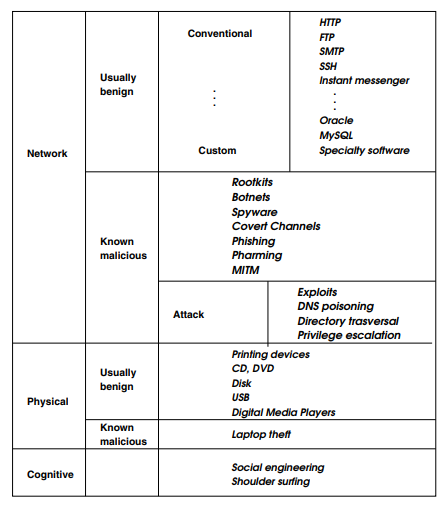
### Proposition de classification

A travers ce panorama des différentes méthodes d’exfiltrations qui se veut le plus exhaustif possible, nous voyons que les possibilités sont énormes et extrêmement variées et que bien souvent une combinaison des techniques est utilisée pour éviter les systèmes de détection DLP.

Plusieurs travaux de classification existe dans la littérature de recherche de cybersécurité.

Les travaux menés par Giani, (2006), propose une discussion autour de la matrice suivante

sur les critères de gravité des conséquences de l’exfiltration, le profil du protagoniste qui déplace les données (Insider, outsider, hacker, Organisation criminelle) et de la méthode de transfert des données ( physique, réseau ou cognititive)

Figure 5 : Classification usuelle des méthodes d'exfiltration (selon Giani, 2006)

J’ai recherché également dans la littérature l’exfiltration par canal discret en mode air-gap, sans connectivité réseau. Les travaux menés par Carrara (2016) propose une analyse multi-critères spécifiquement pour les canaux cachés, discrets (Covert Channel) en mode air-gap selon les critères suivants :

* ***Modèle de « bruit de canal »: bruyant ou sans bruit***

La réception de données via un canal caché peut être bruyante ou sans bruit en fonction de divers facteurs environnementaux. Il est important de connaître ou d'estimer le modèle de bruit du canal pour catégoriser et mesurer efficacement le canal caché. De plus, le modèle de bruit peut influencer les techniques de dissimulation de l'information pouvant être utilisées, car la présence ou l'absence de bruit dans le canal peut limiter certaines techniques.

* ***Modèle de couverture de canal : stéganographique ou ouvert***

Un canal caché stéganographique utilise une source de couverture probabiliste pour établir la communication, tandis qu'un canal caché ouvert utilise soit aucune source de couverture, soit une source de couverture complètement déterministe. La principale différence entre ces deux classes de canaux réside dans le fait que la source de couverture suit une distribution aléatoire dans le cas des canaux cachés stéganographiques, tandis qu'elle suit une valeur déterministe dans le cas des canaux cachés ouverts. Par conséquent, les techniques de dissimulation de l'information basées sur la stéganographie ne conviennent pas aux canaux cachés ouverts.

* ***Modèle de contrôle du canal : partagé ou intermédié***

Dans le modèle de l'attaquant de canal, il peut soit avoir un accès partagé aux symboles transmis, soit être le médiateur du canal. Dans le modèle partagé, les symboles sont directement accessibles à l'attaquant et au récepteur, tandis que dans le modèle médié, les symboles passent par l'attaquant avant d'atteindre le récepteur. Une métaphore utilisée pour illustrer cette distinction est le scénario où les prisonniers communiquent par tapotement sur un radiateur partagé.

* ***Type de modulation : Détectable, indétectable ou canal caché sécurisé indétectable***

Lors de la conception d'un canal caché, il est possible de choisir entre un canal détectable ou un canal indétectable. Le choix du canal indétectable tient compte des capacités d'audit du moniteur de référence du système, tandis que le canal détectable ne fournit aucune défense contre la détection. De plus, le concepteur peut opter pour un canal indétectable sécurisé, qui dépend d'un secret partagé entre l'émetteur et le récepteur, ou pour un canal qui ne l'exige pas. Cependant, l'utilisation d'un canal caché sécurisé indétectable nécessite que Alice et Bob soient en mesure de pré-partager un secret d'une manière quelconque.

* ***Type de modulation : Stockage, synchronisation ou hybride***

Les canaux cachés peuvent être catégorisés en canaux de stockage, de synchronisation ou hybrides, en fonction du temps nécessaire pour transmettre leurs symboles. La capacité d'un canal caché dépend du support utilisé pour communiquer les symboles. Les canaux de stockage cachés ont leur capacité mesurée en bits par utilisation du canal ou en bits par unité de temps, tandis que les canaux de synchronisation cachés ont leur capacité mesurée en bits par unité de temps.

* ***Mode de modulation : Full Duplex, Half Duplex ou Simplex***

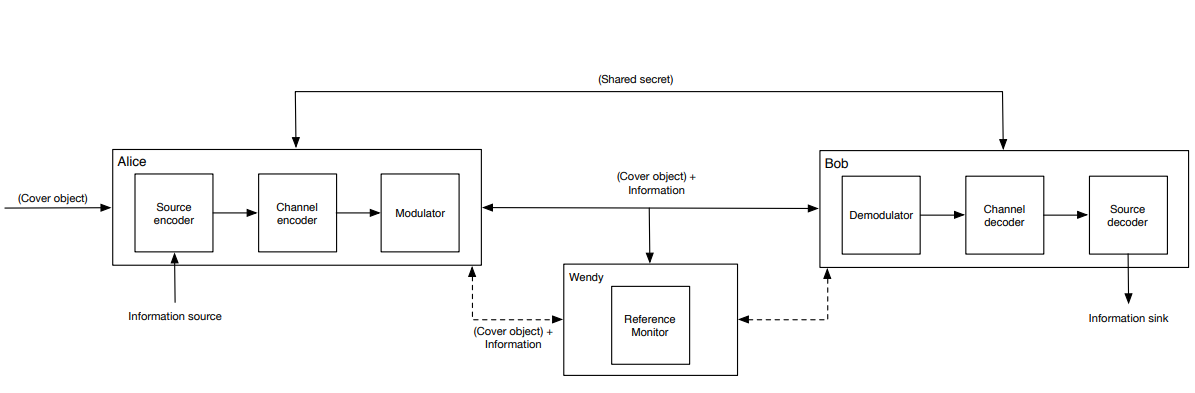
Les canaux cachés peuvent offrir une communication bidirectionnelle ou unidirectionnelle. La communication bidirectionnelle peut être en mode full duplex, permettant des échanges simultanés dans les deux sens, ou en mode half duplex, où les échanges se font dans un seul sens à la fois. Le mode de communication a des implications sur le débit du canal, car les canaux en mode duplex permettent au récepteur d'indiquer quand une retransmission est nécessaire, tandis que les canaux simplex nécessitent l'application de schémas de correction d'erreurs pour assurer une communication sans erreur.

* ***Exploitation cachée : Invasive, semi-invasive ou non invasive***

Les canaux cachés sont souvent accompagnés d'une méthode d'exploitation cachée pour permettre la communication. La classification basée sur l'exploitation cachée permet de mieux comprendre comment le canal est activé et aide à développer des défenses appropriées. Les exploitations cachées peuvent être invasives (nécessitant une modification matérielle), semi-invasives (nécessitant une modification logicielle) ou non invasives (ne nécessitant aucune modification matérielle ou logicielle). Cette classification permet d'étudier la capacité des canaux indépendamment des exploits spécifiques utilisés.

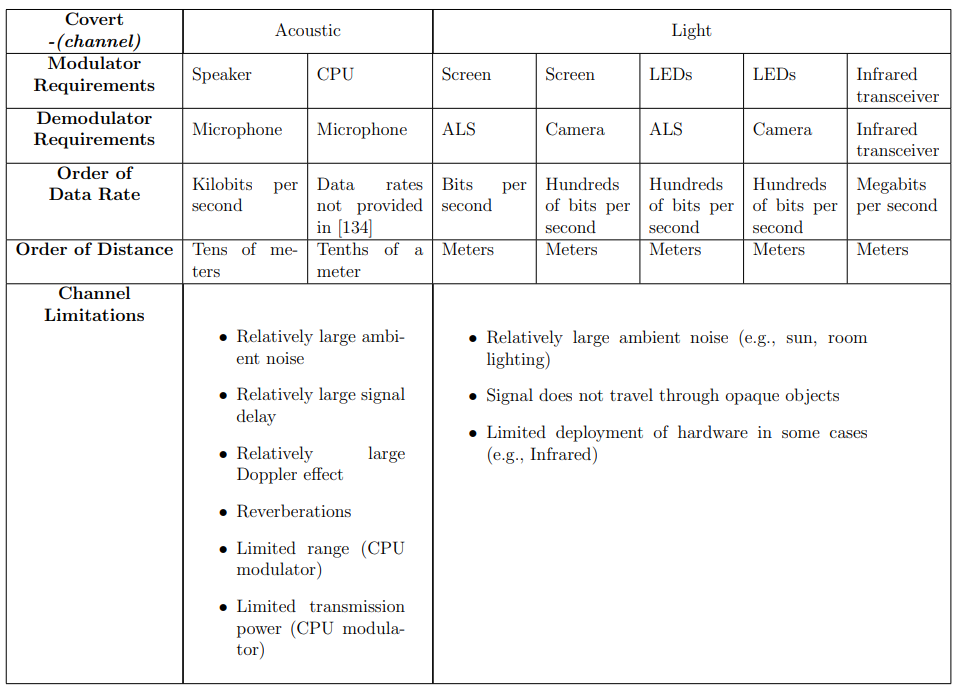
* ***Moniteur de référence : basé sur l'hôte, réseau ou air-gap***

Les canaux cachés peuvent être classés en fonction des moniteurs de référence qu'ils contournent. Certains mécanismes de canaux cachés peuvent contourner les politiques de sécurité de plusieurs moniteurs de référence. Un modèle de canal caché, plus général, est proposé pour comprendre la dissimulation d'information dans les systèmes de communication numériques

Figure 6: Modèle général du canal caché « Covert channel » ( Carrara, 2016)

### Risques et canal caché

Dans le domaine de la cybersécurité, les canaux cachés sont classifiés en fonction de leur bande passante. Le TCSEC\* les classe en trois catégories : les canaux à haute bande passante (plus de 100 bits per second (bps\*) ), les canaux à basse bande passante (entre 1 et 100 bps) et les canaux acceptables (moins de 1 bps ). Le TCSEC recommande de supprimer les canaux à haute bande passante et d'auditer ceux à basse bande passante s'ils ne peuvent pas être supprimés.(Latham, 1986)

Figure 7: Synthèse Acoustic and Light Covert Channel (Carrara, 2016)

En revanche, le CC\* adopte une approche plus pragmatique en laissant la définition de la sécurité aux exigences spécifiques de chaque produit. Il se concentre sur l'analyse des canaux cachés lors de la certification des produits. Les critères de traitement des canaux cachés sont basés sur la quantité d'informations compromises et la capacité d'exploitation des vulnérabilités du système.(21)

Il est important de distinguer les systèmes à source continue, qui produisent des informations sensibles à un débit élevé, des systèmes à source fixe, qui ont besoin de garder une quantité spécifique d'informations confidentielles. Pour les systèmes à source continue, la capacité du canal est mesurée en fonction de la bande passante, tandis que pour les systèmes à source fixe, le critère SMC est utilisé pour évaluer les risques.

L'évaluation des risques des canaux cachés dépend du contexte du système sécurisé. La bande passante est utilisée pour évaluer les canaux cachés dans les systèmes à source continue, tandis que le respect du critère SMC est utilisé pour les systèmes à source fixe. Cependant, il n'existe pas de métrique universellement acceptée pour évaluer les canaux cachés dans les systèmes à source fixe, ni de méthode largement reconnue pour déterminer leur capacité lorsqu'ils sont détectés par audit.(2)

==========

A glisser  ⇒ La recherche sur la façon de séparer les données des systèmes sécurisés peut inspirer des systèmes de protection contre les fuites de données pour empêcher les vulnérabilités révélées d’être exploitées DLP avec une mentalité "black-hat" : trouver des moyens de séparer les données pour trouver les trous. La plupart de ces exploits reposent sur des logiciels malveillants qui compromettent initialement la machine cible, contrairement aux outils d’ex-filtration de validation de principe de Data Border qui sont basés sur un initié voyou avec un accès normal non augmenté souhaitant contourner DLP; Ils illustrent toutefois les types de techniques et de canaux utilisés pour filtrer les données.

==========

# Méthodologie

## Description détaillée des deux modèles proposés pour l'exfiltration de données via codes QR et webcams intégrées

## Évaluation et mesure des débits attendus pour chaque modèle, en prenant en compte différents formats de données et contraintes techniques

## Scénarios d'expérimentation et paramètres pris en considération

# Résultats

## Présentation des résultats des expérimentations, détaillant les débits obtenus pour chaque modèle et format de données

## Analyse des résultats et discussion de leur signification, mise en évidence des vulnérabilités et des risques associés à l'exfiltration de données via cette méthode

## Identification des vulnérabilités et des risques potentiels

# Contre-mesures

## Proposition de contre-mesures efficaces pour prévenir et détecter l'exfiltration de données via les codes QR

## Discussion des stratégies de sécurité, des bonnes pratiques et des technologies complémentaires pour renforcer la protection des données d'entreprise

## Sensibilisation des employés et formation à la sécurité

# Conclusion

## Résumé des principaux résultats et conclusions de la thèse

## Mise en avant de l'importance de sensibiliser les entreprises à cette vulnérabilité et d'implémenter des mesures de sécurité adéquates

## Mise en avant de l'importance de sensibiliser les entreprises à cette vulnérabilité et d'implémenter des mesures de sécurité adéquates

https://www.futura-sciences.com/tech/actualites/securite-pirater-ordinateur-via-led-son-disque-dur-cest-possible-66433/

https://www.scitepress.org/papers/2014/51204/51204.pdf

Bibliographie :

1 Leila Marchand, 2020 « [Le retour en grâce du QR Code en cinq questions](https://www.lesechos.fr/tech-medias/hightech/le-retour-en-grace-du-qr-code-en-cinq-questions-1258534)» sur [www.lesechos.fr](http://www.lesechos.fr/)

2 Benjamin Hue, 2020 « [QR codes, SMS, dépistage : gare aux nouvelles arnaques qui surfent sur le Covid-19](https://www.rtl.fr/actu/sciences-tech/qr-codes-sms-centres-de-depistage-gare-aux-nouvelles-arnaques-qui-surfent-sur-le-covid-19-7800936467) » , sur www.rtl.fr

3 **Margaret Rouse ,2013** « [Data Exfiltration](https://www.techopedia.com/definition/14682/data-exfiltration) » sur [www.techopedia.com](http://www.techopedia.com/)

4 Kelly Sheridan, 2019 « [Database Leaks, Network Traffic Top Data Exfiltration Methods](https://www.darkreading.com/vulnerabilities-threats/database-leaks-network-traffic-top-data-exfiltration-methods)» sur [www.darkreading.com](http://www.darkreading.com/)

5 Cidon, I., Gavish, S., & Markovitch, S. (2018) « Exfiltration channels for data loss: a comprehensive taxonomy and analysis »

6 Brown, T. W., & Xu, H. (2016) « An Analysis of Data Exfiltration Techniques over Encrypted Channels.»

7 Vasilomanolakis, E., Daubert, J., Frincu, M., & Castillo, C. (2020) « A Survey on Covert Channels and Countermeasures in Intrusion Detection Systems. ACM Computing Surveys (CSUR), 53(5), 1-37»

8 Johnson, N. F., & Jajodia, S. (1998) « Exploring steganography : seeing the unseen»

9 Conti, M., & Sobers, R. (2010). «Investigating data exfiltration techniques on USB devices. IEEE Security & Privacy »

10 Jakobsson, M., & Myers, S. (2019). « Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft. Wiley Publishing. »

11 Cohen, Y., Herzberg, A., & Karnin, E. D. (2017). « Pharming Attacks and Countermeasures: A Survey. ACM Computing Surveys (CSUR), 49(4), 6 »

12 Hadnagy, C. (2018) « Social Engineering: The Science of Human Hacking. Wiley Publishing »

13 Computers & Security, 72, 105-118 » Roush, W. (2018). « A Study of Shoulder Surfing on Mobile Devices.

14 Sepehrdad, P., Ghafarian, M., & Dehghantanha, (2020). A« Systematic Literature Review on Windows Privilege Escalation. Computers & Security, 97, 101945»

15 Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., ... & Levchenko, K. (2017). « Understanding the Mirai Botnet. In Proceedings of the 26th USENIX Security Symposium (pp. 1093-1110) »

16 Choo, K. K. R., & Liu, Q. (Eds.). (2021). « Rootkits and Bootkits: Reversing Modern Malware and Next Generation Threats. Springer »

17 Eckert, C., & Sönmez, T. (2020) «  Cybersecurity: The Insights You Need from Harvard Business Review. Harvard Business Press. »

18 Giani, Berk and Cybenko, (2006) «Data Exfiltration and Covert Channels »

19 Carrara, (2016)« Air-Gap Covert Channels »

20 Donald C Latham.(1986) Department of Defense trusted computer system evaluation criteria. Department of Defense

21 Common Criteris: Part 3: Security assurance components, August 2005 URL <https://www.commoncriteriaportal.org/files/ccfiles/ccpart3v2.3.pdf> (Date last accessed: September 22, 2015).

22 I.S. Moskowitz and M.H. Kang. Covert channels-here to stay? In Computer Assurance, 1994. COMPASS ’94 Safety, Reliability, Fault Tolerance, Concurrency and Real Time, Security. Proceedings of the Ninth Annual Conference on, pages 235–243. IEEE, Jun 1994.

Tiwari, S., (2016). An Introduction to QR Code Technology. [Online] Available at: <https://ieeexplore.ieee.org/document/7966807>

Narayanan, S., 2012. QR Codes and Security Solutions. [Online] Available at: https://www.ijcst.org/Volume3/Issue7/p13\_3\_7.pdf

Garg, G., 2015. How QR Codes work: Everything you need to know and more.. [Online]Available at: https://scanova.io/blog/blog/2015/02/19/how-qr-codes-work/

Assaad, A., 2019. What’s inside The QR code?. [Online] Available at: https://medium.com/analytics-vidhya/whats-inside-the-qr-code-bf8a465378fd

Scott, 2020. How Do QR Codes Work? QR Code Technical Basics.. [Online]Available at: <https://www.sproutqr.com/blog/how-do-qr-codes>

[Anon., 2012. Reed-Solomon Codes. [Online]](https://www.sproutqr.com/blog/how-do-qr-codes-work)Available at: https://www.techopedia.com/definition/25798/reed-solomon-codes[-work](https://www.sproutqr.com/blog/how-do-qr-codes-work)

Anon., 2019. Types of QR Code. [Online] Available at: https://www.qrcode.com/en/codes/

admin, 2011. QR Code Error Correction. [Online]Available at: https://blog.qrstuff.com/2011/12/14/qr-code-error-correction