État de l'art des techniques de stéganographie audio

KAMIL MOHOBOOB, École Nationale Supérieure d'Ingénieurs de Bretagne Sud

Cacher de l'information dans un signal audio est plus compliqué que pour une image, à cause de la capacité de l'oreille humaine à percevoir plus facilement une dégradation de la qualité. Par rapport à la popularité des images, la meilleure capacité offerte par les fichiers audio, la popularité du format FLAC qui utilise la compression sans perte, ajoutée aux plateformes d'échange et de streaming de musique, font de la stéganographie audio une alternative intéressante.

Dans cet état de l'art, nous expliquons le principe des différentes techniques de stéganographie audio.

Additional Key Words and Phrases: steganographie audio, survey

ACM Reference Format:
Kamil Mohoboob. 2020. État de l'art des techniques de stéganographie audio . 1, 1 (May 2020), 16 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Table des matières

Author's address: Kamil Mohoboob, contact@natsec.fr, École Nationale Supérieure d'Ingénieurs de Bretagne Sud, Rue André Lwoff, Vannes, France, 56000.

XXXX-XXXX/2020/5-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

La stéganographie est la discipline qui s'attache à dissimuler une information dans une autre de la façon la moins perceptible possible. Pour cela, les techniques de stéganographie utilisent les libertés permises par les spécifications d'un format de fichier. Elles profitent surtout des limites des systèmes de perception humains pour altérer une information sans qu'un observateur ne le remarque.

Contrairement à un message chiffré, qui s'il est intercepté, indique une volonté de communiquer en excluant l'intercepteur, la stéganographie dissimule l'existence du message et donc l'intention de communiquer.

Dans un premier temps, nous expliquerons les propriétés qui permettent d'atteindre les objectifs recherchés par les principales applications de la stéganographie, ainsi que les modes de transmission, et nous justifierons le choix de l'audio plutôt que l'image pour dissimuler de l'information. Ceci nous permettra d'établir un graphe thématique du sujet.

Ensuite, nous verrons les utilisations notables de la stéganographie dans l'Histoire.

Dans un troisième temps, nous présenterons les auteurs et les travaux de référence du domaine.

Nous expliquerons le fonctionnement des différentes techniques de stéganographie audio et nous synthétiserons leurs avantages et inconvénients dans un tableau comparatif.

Enfin nous parlerons des perspectives d'évolution des techniques et des modes de transmission dans ce domaine.

2 PRINCIPES DE LA STÉGANOGRAPHIE

Les explications qui suivent décrivent le graphe thématique 2 établis à la fin de cette section.

2.1 Applications

De nos jours, la stéganographie est toujours utilisée pour **communiquer secrètement**, même si le chiffrement permet le plus souvent de répondre à ce besoin. En effet, malgré que le chiffrement assure la confidentialité du message, son utilisation peut avoir l'inconvénient de divulguer l'intention de communiquer, elle peut même attirer l'attention puisqu'un message chiffré peut indiquer un message plus important.

Cela revient à transporter un coffre à la vue de tous. Dans certains contextes, le déplacement d'un coffre d'un point A vers un point B est une information plus importante de le contenu du coffre. Par exemple, dans un régime autoritaire qui restreint la liberté d'expression et qui surveille les communications, deux personnes qui communiquent auront beau garantir la confidentialité de leurs échanges avec du chiffrement, la circulation d'un message chiffré peut indiquer un souhait d'exclure l'observateur, et devenir suspect.

Malgré le principe de Kerckhoffs qui dit que la sécurité d'un système de chiffrement ne doit reposer que sur le secret de la clé, une autre critique du chiffrement est que si l'algorithme de chiffrement est vulnérable par erreur

ou par conception, ou si l'observateur a les ressources/connaissances nécessaires pour réussir une attaque sur le message, alors la sécurité du chiffrement ne repose plus uniquement sur le secret de la clé. Dissimuler son message permettrait de restreindre fortement son exposition et ainsi d'augmenter sa confidentialité.

Si on ne fait pas confiance aux observateurs sur la ligne, ou si on ne fait pas confiance aux algorithmes de chiffrement, ces deux critiques permettent de justifier une préférence pour la stéganographie.

Une autre application de la stéganographie est l'**exfiltration d'information**.

En sécurité informatique, une APT est une attaque ciblée et sophistiquée menée par un groupe organisé, motivé et avec des ressources. Le but de ce type d'attaque est de s'installer durablement dans le réseau de la cible, pour effectuer des opérations d'espionnage ou de sabotage. Dans le cas où l'attaquant souhaite exfiltrer des informations sensibles, il aura souvent recours à la stéganographie.

Le volume de données à exfiltrer pouvant être important, une consommation excessive de la bande passante ou des connexions suspectes (connexions régulières à des serveurs FTP) peuvent alerter la cible qu'une attaque est en cours. La stéganographie peut permettre à l'attaquant d'exfiltrer de l'information en restant discret sur une longue période.

On peut aussi imaginer de faire rentrer de l'information discrètement sur le réseau de la cible, en dissimulant des malwares dans des fichiers audio qui ne seront pas filtrés par les moyens de protection de la cible. De cette manière, la charge utile de l'attaquant devient moins détectable par les antivirus. Ainsi en octobre 2019, la société Cylance [3] rapportait que le groupe WaterBug aurait dissimulé du code malveillant dans des fichiers audio en utilisant la technique LSB.

Enfin une autre application répandue de la stéganographie est le **watermarking**.

Le watermarking consiste à marquer une information de manière à pouvoir l'authentifier, la tracer ou attribuer sa propriété intellectuelle. Les protections classiques pour empêcher la reproduction de biens matériels comme les billets de banque fonctionnent en utilisant des techniques de fabrication difficiles à reproduire fidèlement. En informatique, si on peut lire une donnée, on peut la stocker, et la recopier à l'identique. Il est donc très difficile d'empêcher la copie d'une information.

Au lieu de rendre plus difficile la reproduction d'information, le watermarking consiste à incorporer une information dans la donnée à protéger de manière plus ou moins visible. En recherchant cette information plus tard dans une autre donnée, on peut savoir si la donnée est une copie identique. Les moyens vont de la fausse entrée volontairement insérée dans une œuvre [5] à l'utilisation de techniques de stéganographie.

Des utilisations notables sont la vérification de contenu protégé par le droit d'auteur et la collecte de données pour le suivi des habitudes des consommateurs [6]. En mars 2019, les équipes de recherche d'Amazon ont publié une nouvelle méthode [25] qui améliore l'identification en temps réel d'un signal marqué dans un milieu acoustique défavorable, notamment en intérieur.

2.2 Objectifs recherchés et propriétés des techniques

Une technique de stéganographie possède trois propriétés qui répondent à des objectifs différents. On peut représenter la relation entre ces trois propriétés comme un triangle 1 dans lequel la maximisation d'une propriété se fait au détriment des deux autres.

La propriété la plus importante est la capacité. Cette propriété représente la **quantité d'information** qui peut être dissimulée. Le fichier porteur dispose d'une quantité limitée d'espace pour stocker ses propres informations. Il peut stocker une quantité tout aussi limitée d'information supplémentaire avant que celle-ci n'impacte la qualité du fichier porteur. La capacité peut être exprimée comme un rapport entre la taille du fichier porteur et la quantité d'information dissimulée. Avec des fichiers audio, on peut aussi exprimer la capacité en quantité d'information dissimulée par seconde du fichier.

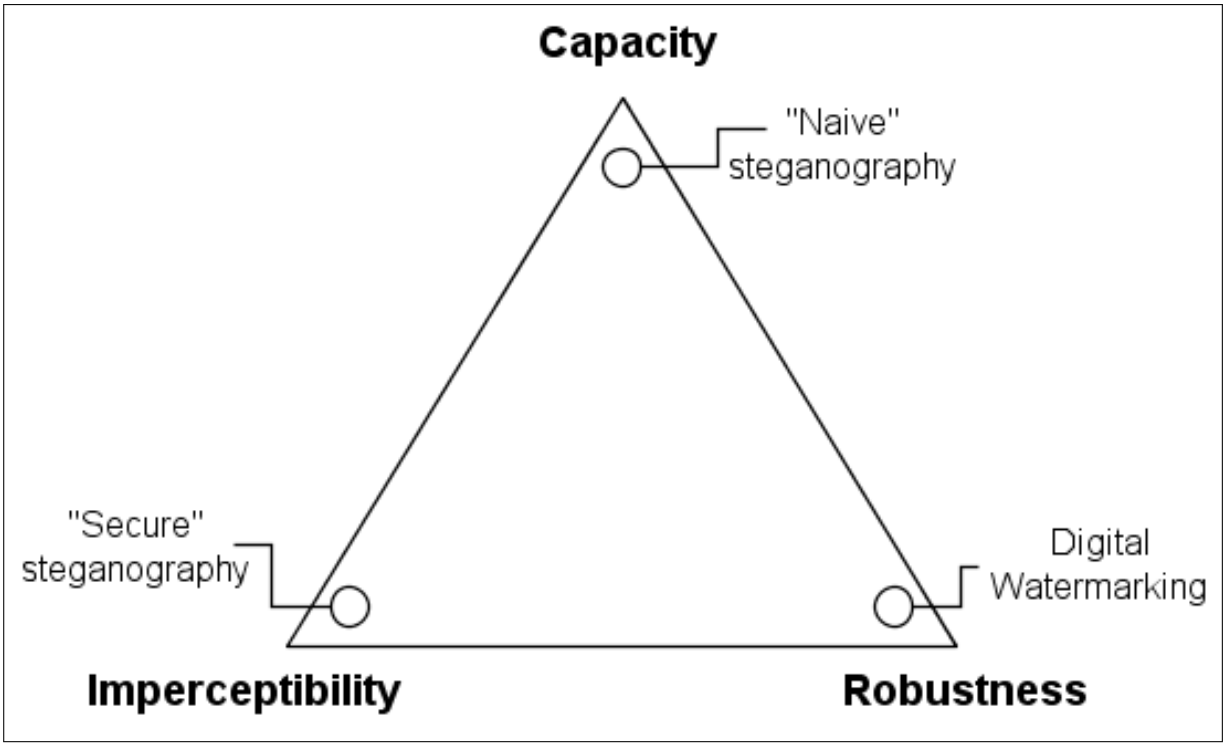

Fig. 1. Triangle de la stéganographie.

L'autre objectif recherché est la **résistance à la détection**. Plus une technique a la propriété d'être transparente, plus il est difficile de détecter que le fichier contient une information. Les limites du **système auditif humain** sont exploitées pour dissimuler l'information. Si une technique n'est pas assez transparente, des artefacts sont perceptibles. Il existe aussi des **méthodes statistiques** pour détecter des anomalies dans la représentation d'une information, une bonne technique sera une technique qui saura aussi résister à une étude statistique.

Enfin le troisième objectif recherché dans la stéganographie est la **résistance à la destruction**. Une technique robuste permet de résister aux traitements du signal qui pourraient altérer le fichier porteur et détruire le message. Certaines techniques de watermarking qui sont très visibles ne font que sacrifier la transparence au profit d'une grande robustesse. Toute modification du fichier porteur peut altérer son contenu.

La modification la plus courante pour détruire un message caché est la **compression**, elle va enlever les informations redondantes et les nuances imperceptibles pour l'oreille humaine. Le format MP3 utilise notamment un modèle psychoacoustique humain pour déterminer les informations qui peuvent être détruites sans impacter la qualité perçue du son.

Une autre modification est l'**ajout de bruit** uniforme dans le signal pour écraser les bits de poids faibles. Comme les techniques les plus simples utilisent ces régions pour y encoder de l'information, l'ajout d'un bruit suffit à détruire le message de manière systématique sans pour autant affecter la qualité perçue du signal.

Pour différencier la détection d'un message et sa destruction, on parle aussi d'attaque passive et active.

2.3 Modes de transmission

Il y a plusieurs façons de transmettre un message qui contient une information dissimulée.

La méthode la plus simple consiste à **envoyer directement** le fichier audio au destinataire, si la technique utilisée est suffisamment transparente, une oreille humaine ne remarquera pas la présence de l'information.

Une autre méthode de transmission est le **dead drop**, son principe est le suivant : L'émetteur vient déposer un objet dans un lieu public, s'en va, et le destinataire vient le récupérer plus tard. De cette manière, les deux parties ont transmis un objet sans se rencontrer. L'effet est encore plus efficace pour transmettre une information car plusieurs personnes peuvent venir récupérer l'information sur un **serveur de fichier**, il devient alors difficile de savoir qui était le vrai destinataire du message.

Les **plateformes de streaming** de musique permettent d'obtenir un fonctionnement similaire. De plus, comme la qualité est un critère important pour les consommateurs et les artistes, ces plateformes vont conserver le fichier original mis en ligne, alors que les plateformes de partage d'image ont tendance à compresser et à dégrader les images pour gagner de la place.

Enfin une autre méthode de transmission est la **conversation téléphonique**. Elle est plus compliquée à mettre en œuvre car elle ne peut être utilisée que pour les techniques utilisant les codecs et nécessite des conditions particulières. En revanche elle est très transparente car l'existence du message est éphémère et la détection du message nécessiterait des moyens importants. Il faudrait pouvoir intercepter le signal et disposer du bon décodeur.

2.4 Les techniques dans l'audio

Les images sont le type de fichier le plus utilisé pour faire de la stéganographie. Cela peut s'expliquer par la popularité des images sur Internet et par le fait que par rapport à un stimulus auditif, un humain a plus de mal à détecter un stimulus visuel [23]. Ces deux avantages se traduisent par une meilleure transparence de la stéganographie dans l'image.

Les fichiers audio ont aussi des avantages. Comme ils sont moins utilisés en stéganographie, ce sont potentiellement des supports plus discrets. Les fichiers audio sont plus volumineux que les images (jusqu'à 100 fois plus) et permettent donc de dissimuler plus d'informations. Comme un défaut de qualité audio est plus remarquable, les utilisateurs ont tendance à préférer les formats sans compression avec perte, les fichiers auront tendance à être moins traités, ce qui les rend plus robustes.

Les techniques de stéganographie dans l'audio ont donc suffisamment d'avantages pour compenser leurs faiblesses et peuvent devenir plus intéressantes à utiliser.

Avant l'utilisation d'une technique, la **compression** de l'information à dissimuler permet de gagner en capacité [8]. Son **chiffrement** permet de gagner en transparence [11] car il permet d'éliminer des motifs reconnaissables.

Les techniques de stéganographie audio peuvent être regroupées en trois familles.

Les techniques dans le **domaine temporel** dissimulent les informations dans les échantillons du son. Elles sont relativement faciles à implémenter et permettent d'obtenir une grande capacité. La technique LSB offre la meilleure capacité sans perte d'information. Cependant il est facile de détruire le message et pour gagner en robustesse, les variantes et autres techniques diminuent en capacité.

Les techniques dans le **domaine fréquentiel** sont généralement plus transparentes et plus robustes. La technique utilisant les **coefficients cepstraux** est la technique la plus robuste car elle peut résister à 4 traitements destructeurs différents.

Enfin les techniques qui utilisent les **codecs** cachent l'information à la volée dans un signal transmis en temps réel.

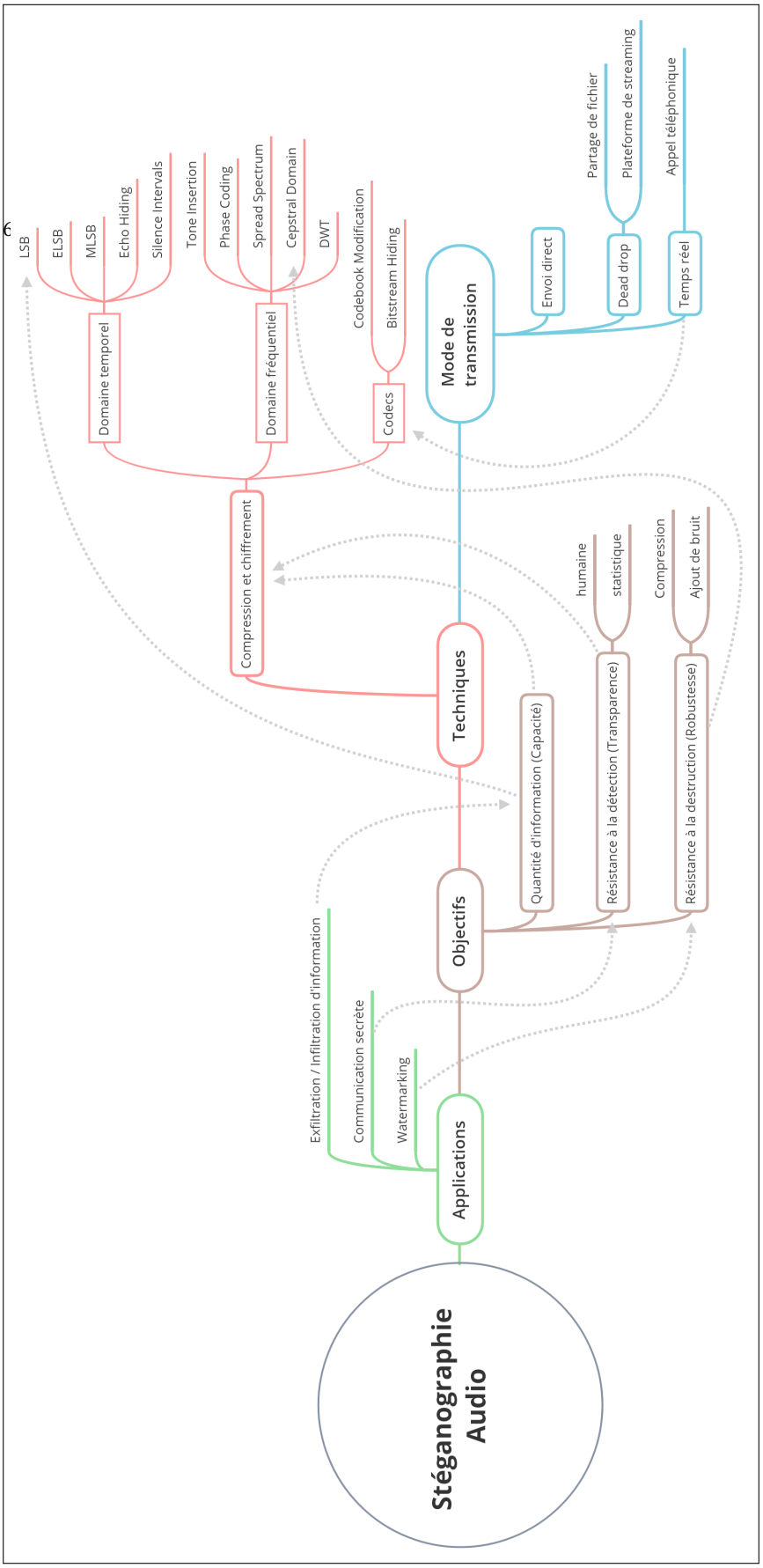

Fig. 2. Graphe thématique de la stéganographie dans le domaine audio.

3 HISTOIRE DE LA STÉGANOGRAPHIE

Le terme stéganographie vient du grecque et signifie *écriture cachée*. Les premières traces de l'utilisation de la stéganographie datent de la Grèce antique.

Au VI^{e} siècle av. J.-C, pour transmettre un message, le dignitaire mède Harpage tua un lièvre et cacha un message dans son corps. Il le donna à un messager qui voyagea sous l'identité d'un chasseur.

Pour informer ses amis qu'il était temps de se révolter contre les mèdes et les persans, un habitant d'Histiée rasa la tête d'un de ses esclaves de confiance et y tatoua un message pour l'envoyer transmettre le message.

Le roi sparte Demaratus utilisait les tablettes de cire pour y cacher de l'information. Une tablette de cire est constituée de deux panneaux de bois reliés comme un livre. On inscrivait son message sur une couche de cire déposée sur le bois. Le destinataire, après avoir lu le message, faisait fondre la cire pour pouvoir réutiliser la tablette. Demaratus, plutôt que d'inscrire son message sur la cire, l'inscrivit directement sur le bois avant de le couvrir d'une couche de cire vierge.

Énée le tacticien faisait des petits trous au-dessus des lettres d'un message anodin. Cette technique fut aussi utilisée pendant la Renaissance et jusqu'à la Première Guerre mondiale.

Pendant la Seconde Guerre mondiale, les avancées en photographie ont permis l'utilisation de micro points. On les appelle ainsi parce que ce sont des images qui font la taille d'un point. Ils étaient utilisés par les allemands pour transmettre des documents techniques qui étaient censurés par des embargos.

Toutes ces techniques constituent une forme de stéganographie que l'on pourrait qualifier de technologique. Une autre forme de stéganographie dite linguistique consiste à utiliser le langage et son interprétation pour véhiculer un sens différent en fonction des interlocuteurs.

Ainsi pendant la Première Guerre mondiale, les espions allemands utilisaient de fausses commandes de cigares pour représenter différents types de bateaux anglais. Une commande de 5000 cigares à livrer à Portsmouth signifiait que 5 croiseurs étaient à Portsmouth.

Un autre exemple est l'américain Jeremiah Denton. Pendant la Guerre du Viêt Nam, cet aviateur a été retenu comme prisonnier de guerre par l'armée populaire vietnamienne. Lorsqu'il est contraint de participer à une interview de propagande qui vise à montrer que les prisonniers américains sont bien traités, il forme le mot *torture* en morse en clignant des yeux.

Avec l'arrivée de l'informatique, les techniques de stéganographie ont été transformées pour s'adapter à cette nouvelle forme de communication.

4 AUTEURS DE RÉFÉRENCE

Pour déterminer les travaux de référence, nous n'utiliseront que des articles parus dans des revues scientifiques, et en particulier le corpus retenu pour faire cet état de l'art.

Pour établir les travaux de référence, nous avons compté les auteurs de chaque article cité dans le corpus.

La Table 1 montre les dix auteurs les plus cités.

L'auteur le plus cité dans le corpus est Nedeljko Cvejic. L'article le plus cité de cet auteur concerne l'amélioration de la technique du LSB. La majorité des articles du corpus parlent de variations de la technique du LSB, il est donc logique que cet article revienne souvent dans ce corpus. Nedeljko Cvejic étudie la capacité des algorithmes et des techniques de dissimulation de l'information afin d'améliorer la robustesse des procédures de watermarking audio. En pratique, la tâche principale de Cvejic est de développer des algorithmes pour le watermarking et la stéganographie des signaux audio.

Citations dans le corpus	Auteur	H-index de l'auteur	Article le plus cité dans le domaine de la stéganographie	Citations de l'article le plus cité
17	Nedeljko Cvejic	16	Increasing the capacity of LSB-based audio steganography, 2002	233
12	Fatiha Djebbar	7	Comparative study of digital audio steganography techniques, 2012	152
11	Tapio Seppänen	43	Increasing the capacity of LSB-based audio steganography, 2002	233
11	Wang Jun-ji	-	Multiple bits reversible data hiding algorithm in image interpolation space, 2019	-
9	Hassan Shirali-Shahreza	16	A new approach to Persian/Arabic text steganography, 2006	218
9	Xinpeng Zhang	43	Reversible data hiding in encrypted image, 2011	664
8	Walter Bender	-	-	-
8	Daniel Gruhl	33	Techniques for data hiding, 1996	4160
8	Habib Hamam	22	Comparative study of digital audio steganography techniques, 2012	152
7	Kaliappan Gopalan	10	-	-

Table 1. Top 10 des auteurs les plus cités dans le corpus.

Dans les 10 articles les plus cités du corpus apparaît un auteur dont l'article le plus cité a été cité plus de 4000 fois. Ce nombre se démarque par rapport au nombre de citations des autres articles. Son auteur s'appelle Daniel Gruhl. C'est un diplômé du MIT qui travaille aujourd'hui pour IBM sur l'intelligence artificielle et l'analyse automatisée de texte, il a commencé sa carrière en étudiant la stéganographie. Son article *Techniques for data hiding* paru en 1996, coécrit avec Walter Bender, Norishige Morimoto et Anthony Lu ; présente les principes de traitement du signal qui peuvent être utilisés pour cacher de l'information de manière imperceptible dans du texte, des images ou du son.

Parmi les 14 sources de cet article, 9 sont des articles parlant de traitement du signal sans évoquer la dissimulation d'informations. 2 articles parlent d'applications du traitement du signal pour marquer des documents, et seulement 3 sources parlent de dissimulation d'information. Parmi ces 3 sources, une est des notes de conférence non publiées de Walter Bender.

La date de parution, le faible nombre d'articles parus auparavant sur le même domaine, et le nombre de fois qu'il a été cité plus tard permettent de considérer la publication *Techniques for data hiding* comme un travail de référence dans le domaine de la stéganographie. Même si l'article ne traite pas spécifiquement du sujet dans le domaine audio, il a été suffisamment influent pour donner des pistes de recherche à d'autres auteurs, et donner naissance à de nouvelles techniques dans l'audio.

5 TRAVAUX DE RÉFÉRENCE

Dans ce résumé des travaux de référence, on traitera d'abord les articles qui sont eux-mêmes des états de l'art du sujet. On traitera ensuite les articles abordants des techniques en détail, dans leur ordre d'apparition historique. Pour chaque technique, on résume aussi les améliorations successives.

Avant toute utilisation d'une technique de stéganographie, Valarmathi et al. [11] proposent de chiffrer l'information, ce qui permet d'augmenter la transparence et la robustesse de la technique. En effet, une information

chiffrée a un aspect plus aléatoire qui ressemble à du bruit. L'information sera plus difficile à dissocier d'un bruit réel dans le son.
Begum et al. [8] proposent aussi de compresser le texte à encoder avec des algorithmes de compression par dictionnaire, ce qui permet d'augmenter le SNR [1] du signal.

En 2012, Jebbar et al. [12] ont fait un état de l'art des techniques de stéganographie audio. Ils classent les techniques en trois grandes familles qui sont : le domaine temporel, le domaine fréquentiel et les techniques tirants parti du codec. Les techniques dans le domaine temporel sont faciles à implémenter, celles dans le domaine fréquentiel tirent avantage de l'effet masque et celles utilisant des codecs permettent une transmission en temps réel.
Nosrati et al. [21] nous introduisent à la stéganographie en nous présentant des méthodes pour cacher de l'information dans du texte, des images, puis des fichiers audio. Les techniques abordées pour le domaine audio (LSB Coding, Phase Coding, Spread Spectrum DSSS et FHSS, Echo Hiding) sont présentées.
Dans un autre article, Nosrati et al. [20] nous présentent un état de l'art des récentes approches de la stéganographie dans le domaine audio. En plus des cinq premières techniques déjà présentées, l'équipe nous présente sept nouvelles approches qui sont :

- la modification des valeurs échantillonnées pour compresser un fichier MP3
- l'insertion de données entre les blocs d'un fichier MP3
- la transformée d'entier réversible (ITSAS)
- l'utilisation du LSB avec XOR et bit de parité
- l'utilisation d'un algorithme génétique (GA) pour améliorer la robustesse du LSB
- deux algorithmes augmentant la transparence du LSB

Singh [24] nous explique le principe de fonctionnement de la technique du LSB qui consiste à utiliser les bits de poids faible d'un signal pour y encoder de l'information. Un point fort de cette technique est sa simplicité de mise en œuvre et sa transparence : les modifications sont suffisamment minimes pour ne pas être perceptibles par une oreille humaine.

Pour améliorer la technique du LSB, Hakeem et al. [13] proposent une approche qui permet d'obtenir une bonne robustesse et une grande capacité. La technique est basée sur le fait qu'un son avec un volume faible sera masqué par un son avec un volume plus élevé. Plus un échantillon a une petite valeur, plus un nombre important de bits du message est encodé dans l'échantillon.

Mohamad et Yasin [17] nous informent de la faiblesse du LSB, qui est sa faible robustesse. Pour pallier ce problème, les auteurs proposent de chiffrer le message à encoder, puis de le chiffrer à nouveau en fonction du contenu du message à l'aide d'un algorithme d'identification des motifs (pattern matching algorithm).

Hartoko et al. [14] présentent le MELSB, une amélioration du ELSB, lui-même une amélioration du LSB. Ces deux techniques permettent de pallier un inconvénient du LSB, qui est la destruction du message lors de l'ajout de bruit. Le MELSB a un meilleur SNR que l'ELSB. De plus le MELSB peut être implémenté sur un réseau de type 802.11n (Wifi) pour des applications en temps réel. Cependant les deux techniques restent vulnérables à la destruction du message par compression du signal.

Nehru et Dhar [19] proposent l'utilisation d'algorithmes génétiques pour améliorer la transparence du LSB, mais ils ne font que décrire brièvement le résultat souhaité d'un tel algorithme. Aucune information nouvelle n'est

1. rapport signal/bruit : propriété d'un signal qui permet de mesurer sa qualité

apportée.

Bassil [9] propose une variante du LSB utilisant deux canaux. La technique consiste à encoder le message avec du LSB qui choisit aléatoirement les échantillons. La sélection aléatoire des échantillons est aussi utilisée pour générer un paragraphe en anglais de manière à ce que les mots et les phrases indiquent les échantillons contenant l'information, ce qui permet de reconstituer le message. L'aspect aléatoire de la technique permet de la rendre indétectable si on ne connaît pas le paragraphe servant de clé pour retrouver le message. Pour rendre la technique encore plus transparente, l'auteur propose une évolution de l'algorithme de génération du paragraphe qui permettrait de rendre les phrases sémantiquement valides.

En 2018, Xin et al. [27] proposent une variante du Low Bit Coding (une amélioration du LSB) qui permet d'augmenter la quantité d'informations dissimulées. Le nombre de bits encodés dans un échantillon varie en fonction des propriétés du signal. Cette technique offre la meilleure capacité par rapport aux autres variantes connues du LSB.

Wheeler et al. [**?**] présentent une méthode différente qui consiste à utiliser les hautes fréquences inaudibles d'un signal audio pour y encoder de l'information. En effet, la fréquence maximale enregistrée dans un fichier audio utilisé pour la diffusion est souvent de 20kHz, mais l'oreille humaine ne perçoit les sons que jusqu'à environ 17kHz. La technique utilise cette bande de fréquence de 3kHz pour y encoder de l'information. Bien qu'elle ne soit pas très robuste car facilement détectable, cette technique fournit une grande capacité de transmission.

Khan et al. [16] partent du chiffrement pour proposer une autre méthode garantissant la confidentialité d'un message en la cachant dans du son. La technique décrite est le Spread Spectrum. Le modèle psychoacoustique humain est utilisé pour déterminer un masque de seuil de fréquence. Ce masque est ensuite utilisé avec une séquence pour étaler le spectre du son sur une bande plus large, le message est fondu dans le bruit. Le type de séquence utilisé à la propriété d'être réversible, on peut reconstituer le message en "désétalant" le signal à l'aide de la séquence qui sert de clé.

Banik et Bandyopadhyay [**?**] utilisent l'effet cocktail party pour cacher de l'information dans l'audio. L'effet cocktail party est une propriété du système auditif humain qui permet de suivre une conversation dans une ambiance bruyante. Les auteurs tentent de reproduire un système permettant de filtrer le bruit d'un signal pour ne conserver que l'information préalablement cachée.

Aswin et Narmadha [2] proposent une technique utilisant le domaine fréquentiel logarithmique.

En juillet 2019, Ye et al. [28] présentent une technique utilisant un réseau antagoniste génératif (GAN) pour générer le fichier contenant le message et un discriminant permettant de retrouver le message.

6 PRÉSENTATION DES TECHNIQUES

Le tableau 2 compare les différentes techniques d'après leurs avantages et inconvénients, et leurs capacités à résister au différents traitements destructeurs.

	Domaine temporel					Domaine fréquentiel					Codecs	
	LSB	**ELSB**	**MLSB**	**Echo Hiding**	**Silence Intervals**	**Tone Insertion**	**Phase Coding**	**Spread Spectrum**	**Cepstral Domain**	**Discrete Wavelet Tranform**	**Codebook Modification**	**Bitstream Hiding**
Principe	Le LSB de chaque échantillon de l'audio est remplacé par un bit de l'information à cacher	Les bits et les échantillions sont choisis aléatoirement	Les bits sont séléctionnés de manière à augmenter le SNR	Intègre les données en introduisant un écho dans le signal de couverture	Utilise le nombre d'échantillons dans l'intervalle de silence pour représenter les données cachées	Insertion de tonalités inaudibles à des fréquences sélectionnées	Modulation de la phase du signal de couverture	Etale le signal sur une bande de fréquence	Modification des coefficients cepstraux pour l'intégration des données	Modification des coefficients d'ondelettes pour l'intégration des données	Modification des paramètres de l'alphabet de codage	Le LSB est appliqué sur le flux de bits résultant du processus de codage
Avantages	Un moyen simple et facile de dissimuler des informations à haut débit	Meilleure robustesse que le LSB	Meilleure capacité que l'ELSB	Résistance aux algorithmes de compression de données avec perte	Résistance aux algorithmes de compression de données avec perte	Imperceptibilité et dissimulation des données intégrées	Robuste contre la manipulation du signal et ne nécessite pas le signal original pour être décodé	Offre une meilleure robustesse	Robuste contre les opérations de traitement du signal	Offre une grande capacité	Robuste	Robuste
Inconvénients	Facile à extraire et à détruire	Détruit par la compression	Détruit par la compression	Faible sécurité et capacité	Faible capacité	Manque de transparence et de sécurité	Faible capacité	Vulnérable à la modification de l'échelle de temps	Distorsion du signal perceptible et faible robustesse	Récupération des données avec perte	Faible vitesse d'encodage	Faible vitesse d'encodage
Capacité (bits/sec)	16k	-	-	50	64	250	333	20	54	70k	2k	1.6k
Transparent	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Résistant à la compression					✓		✓		✓			
Résistant à l' amplification		✓	✓				✓		✓			
Résistant à l' ajout de bruit		✓	✓		✓				✓		✓	✓
Résistant à l'application d'un filtre passe-bas						✓			✓			
Résistant à la requantification							✓					
Résistant au re-échantillonnage												
Résistant au décodage puis re-encodage	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	✓	✓

TABLE 2. Comparatif des techniques de stéganographie dans l'audio.

6.1 Domaine temporel

6.1.1 LSB. La technique LSB est la technique de stéganographie la plus facile à implémenter pour cacher de l'information dans du son. Elle permet d'encoder une quantité importante d'information. Il existe de nombreuses variations de la technique qui permettent d'améliorer la robustesse [14, 17], la capacité [13, 27] ou la transparence [9] de l'encodage.

La technique LSB la plus simple consiste à remplacer le bit de poids faible de chaque échantillon du son par un bit du message à cacher. Les modifications sont suffisamment minimes pour ne pas être perceptibles par une oreille humaine. En effet, pour un signal codé sur 8 bits, un bit de poids faible ne représente que 0,4% ($\frac{2^0}{2^8} \times 100$) de l'information de l'échantillon.

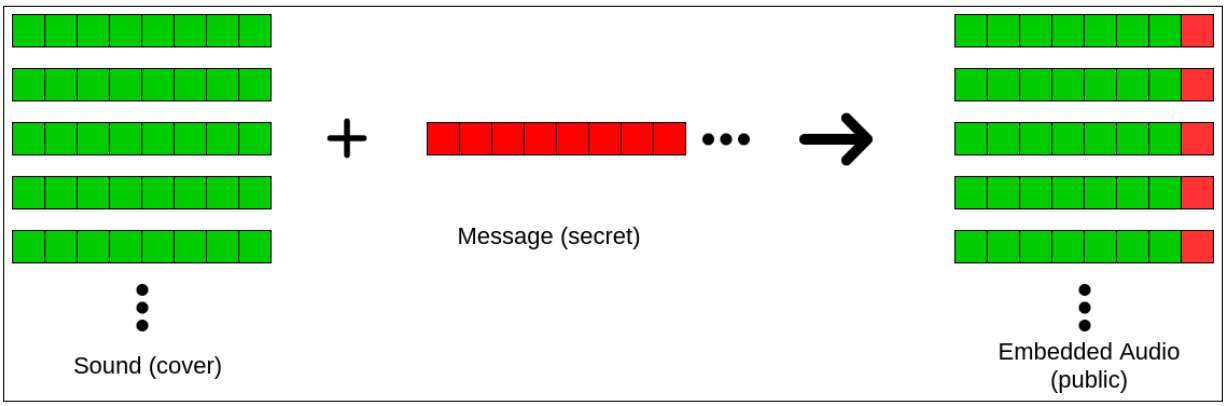

Fig. 3. Schéma de la technique LSB.

L'illustration 3 utilise des échantillons codés sur 8 bits. Les échantillons d'un ficher audio de qualité standard sont codés sur 16 bits.

Il est possible de multiplier la capacité en utilisant plus de bits de poids faible jusqu'à un certain seuil au-delà duquel la qualité audio se dégrade.

6.1.2 Echo Hiding. La technique de l'Echo Hiding [12] consiste à encoder l'information en introduisant un court écho dans le signal audio. Le traitement reproduit le phénomène de réverbération qui peut être ressenti dans un environnement qui réfléchit le son, comme des mûrs de pierre, ou qui l'absorbe, comme le silence de la neige. Pour rester imperceptible, les paramètres tels que l'amplitude, le temps entre le son et l'écho (delay) et le taux d'extinction (decay) de l'écho sont ajustés pour qu'il ne soit pas audible. En dessous d'un délai d'une milliseconde, l'écho n'est pas perceptible par l'oreille humaine. L'inconvénient de cette méthode est sa faible capacité due à la milliseconde d'écart nécessaire pour conserver sa transparence, ce qui explique le nombre limité de travaux sur les applications de cette technique.

6.1.3 Silence Intervals. La technique des intervalles de silence [12] encode de l'information en supprimant les échantillons qui représentent un silence. Elle est donc plus efficace quand elle est utilisée sur de la parole.

La technique consiste à détecter les intervalles de silences à partir d'un seuil de puissance du signal. Pour chaque intervalle de silence détecté, on compte le nombre d'échantillons constituants ce silence. On supprime un nombre d'échantillons x avec $0 < x < 2^n$, n étant le nombre de bits nécessaires pour coder la valeur du message à cacher.

Pour extraire la donnée, on calcule $x = \mathrm{mod}(nb_{chantillons}, 2^n)$.

Les petits intervalles de silences sont ignorés car ils se présentent généralement dans des phrases continues. Les modifier affecterait la qualité de la parole. C'est une des seules techniques qui permet de cacher de l'information en en supprimant et en diminuant l'espace nécessaire pour la stocker.

6.2 Domaine fréquentiel

6.2.1 Tone Insertion. La technique de l'insertion de tonalité [12] utilise la propriété du système auditif humain qui fait que sur deux sons avec deux puissances différentes, nous percevons celui ayant la plus grande.

Pour dissimuler de l'information, la technique choisit deux fréquences f_0 et f_1. Leurs puissances P_{f0} et P_{f1} est un pourcentage de la puissance moyenne du signal P_{moy}. En insérant des tonalités avec des fréquences particulières et une faible puissance dans le signal original, les tonalités restent imperceptibles et permettent de cacher de l'information. Si le rapport entre la puissance moyenne et la puissance de la tonalité 0 est supérieur au rapport entre la puissance moyenne et la puissance de la tonalité 1, alors c'est un bit de valeur 0 qui était caché.

Cette technique a l'avantage de résister à l'application d'un filtre passe-bas et à l'ajout de bruit. Cependant, la technique offre une petite capacité et est facilement détectée par un ordinateur.

6.2.2 Phase Coding. La technique du codage de phase [12] consiste à coder de l'information dans la phase du signal.

Un exemple extrême de déphasage est utilisé pour la réduction de bruit : en décalant toutes les fréquences d'un signal d'une demi-période, et en l'additionnant au signal original, on parvient à réduire voire supprimer le signal original, c'est l'addition destructive.

Ici, le déphasage est utilisé avec un décalage très faible pour coder de l'information de manière à ce que son impact sur la qualité du son soit imperceptible pour l'oreille humaine.

Malgré que cette technique soit résistante à la compression, le fait que le système auditif humain ne soit pas très sensible aux distorsions de phase permet aussi à un attaquant d'introduire des déphasages pour détruire l'information cachée.

6.2.3 Spread Spectrum. La technique du Spread Spectrum [16] consiste à étaler un signal sur une bande de fréquence plus large pour transmettre l'information. L'énergie totale nécessaire à la transmission reste la même.

Le modèle psychoacoustique humain est utilisé pour déterminer un masque de seuil de fréquence. Ce masque est ensuite utilisé avec une séquence pour étaler le spectre du son sur une bande plus large, le message est fondu dans le bruit. Le type de séquence utilisé à la propriété d'être réversible, on peut reconstituer le message en dé-étalant le signal à l'aide de la séquence qui sert alors de clé.

Cette technique a l'avantage de résister aux interférences. Ces deux avantages du Spread Spectrum en ont fait une technique très utilisée dans les communications militaires car elle permet de résister au brouillage et de fondre le signal dans le bruit de fond.

6.2.4 Cepstral Domain. La technique Cepstral Domain [2] utilise les coefficients cepstraux de la représentation du signal dans le domaine spectral logarithmique.

Cette technique a une capacité d'environ 20 à 40 bps. Elle a l'avantage de résister à la plupart des attaques visant à détruire le message.

6.2.5 Discrete Wavelet Transform. La stéganographie audio basée sur la transformée en ondelettes discrètes [16] dissimule l'information dans les bits de poids faible des coefficients d'ondelette du signal audio.

Pour améliorer la transparence de la technique, des variantes évitent de cacher de l'information lorsqu'il y a un silence dans le signal.

6.3 Codecs

6.3.1 Codebook Modification. Lors de l'envoi d'un signal en temps réel, celui-ci est codé en bit par l'encodeur. Ceci permet d'utiliser des codages plus complexes qu'un simple changement de base 10 en base 2. Ainsi le codage de Gray permet de ne modifier qu'un seul bit à la fois quand un nombre est augmenté d'une unité. Cette propriété est importante pour plusieurs applications.

Pour éviter d'avoir à recalculer en permanence les correspondances $valeur \Leftrightarrow code\ binaire$, il est possible d'enregistrer les correspondances dans un dictionnaire (anglais codebook) pour gagner en débit de transmission.

Le codebook de l'encodeur et du décodeur peut être détourné pour faire de la stéganographie. L'information peut être codée différemment pour transmettre une information différente. En utilisant le même codebook utilisé pour encoder le signal, on peut ainsi retrouver l'information cachée.

6.3.2 Bitstream Hiding. La technique du bitstream hiding consiste à appliquer la technique du LSB sur le résultat de l'encodage du signal plutôt que sur le signal brut.

7 ENJEUX DU DOMAINE

Nous avons vu plus tôt que la stéganographie est principalement utilisée pour communiquer de manière confidentielle.

Le 5 mars 2020, le congrès des États-Unis d'Amérique a proposé au sénat la loi *"Eliminating Abusive and Rampant Neglect of Interactive Technologies"*, ou EARN IT Act. Ce projet de loi demande la constitution d'une commission nationale sur la prévention de l'exploitation des enfants en ligne, pour établir des règles de recherche et de suppression du contenu d'exploitation des enfants. Si les entreprises ne respectent pas ces règles, elles pourraient perdre une certaine protection en vertu de l'article 230 de la Communications Decency Act, qui protège largement les entreprises de toute responsabilité vis-à-vis des publications des utilisateurs.

Pour les détracteurs de ce projet de loi, lorsque les entreprises deviennent responsables de ce que leurs utilisateurs publient, elles modèrent agressivement le contenu en ligne. Comme le projet de loi ne précise pas les mesures qui devront être prises par les entreprises, l'Electronic Frontier Foundation craint notamment que cette commission pourrait décider arbitrairement des conditions pour qu'une entreprise conserve sa protection juridique sur le contenu généré par ses utilisateurs. Elle craint ainsi qu'il soit exigé des entreprises qu'elles affaiblissent volontairement le chiffrement [18] pour permettre la détection de contenu interdit.

Une autre affaire qui pourrait remettre en cause la confiance portée au chiffrement est l'entreprise suisse de cryptographie Crypto AG. Malgré de nombreuses suspicions au cours des quarante dernières années, la déclassification récente (février 2020) de documents, montre qu'après le rachat de la société par la CIA et le BND, l'entreprise a collaboré avec les services de renseignement américains et allemands, en affaiblissant la sécurité des machines de chiffrement qu'ils vendaient à plus de 60 pays et aux Nations Unies [1].

Contrairement à ce qu'on aurait tendance à penser, les États autoritaires ne sont pas les seuls à être tentés d'utiliser la surveillance pour maintenir leurs habitants en relative sécurité.

Dans des sociétés de plus en plus surveillées [4], et où la confiance dans le chiffrement peut être remise en cause [22], la stéganographie devient un moyen de plus en plus considéré pour communiquer de manière confidentielle.

Cependant, là où l'utilisation du chiffrement ne craint pas un gain de popularité, cela a pu se constater avec la démocratisation des messageries chiffrées, l'utilisation de la stéganographie doit rester la plus discrète possible. Une utilisation par un nombre de personnes de plus en plus important de ce moyen de communication irait paradoxalement à l'encontre de ces principes.

Un enjeu majeur de la stéganographie sera donc de parvenir à être utilisée à une plus grande échelle, sans pour autant attirer plus d'attention sur son adoption.

Un autre enjeu de la stéganographie est la transmission du medium porteur de l'information dissimulée, qui est dans notre cas un fichier ou un flux audio. Pour garder une transparence maximum, nous avons vu que la transmission par dead drop était la plus efficace. Cependant, l'émetteur ne maîtrise pas les traitements effectués par les plateformes de partage, qui pourraient détruire le message dissimulé intentionnellement ou non. L'enjeu est donc de trouver la plateforme qui effectuera le moins de traitement possible mais sera suffisamment populaire pour être utilisé par le plus grand nombre de personnes.

Une piste de recherche qui pourrait être intéressante à explorer, serait d'établir un classement des plateformes de partage et de streaming de musique qui altèrent le moins les fichiers partagés. En mesurant les modifications apportées par ces plateformes, on pourrait trouver celle qui conserve le mieux l'intégrité des messages transmis.

8 CONCLUSION

Dans cet état de l'art, nous avons présenté le graphe thématique de la stéganographie audio, avant de présenter son histoire. Après avoir expliqué les principes fondamentaux de la stéganographie, nous avons présenté les différentes techniques dans le domaine audio. Dans un travail futur, il pourrait être intéressant d'étudier dans quelles mesures les plateformes de stockage et de streaming modifient les fichiers stockés pour établir la liste de celles qui conservent le mieux l'intégrité d'un message.

RÉFÉRENCES

[1] [n. d.]. The CIA Secretly Bought a Company That Sold Encryption Devices across the World. Then Its Spies Sat Back and Listened. https ://www.washingtonpost.com/graphics/2020/world/national-security/cia-crypto-encryption-machines-espionage/.

[2] [n. d.]. *Licensed Under Creative Commons Attribution CC BY Secured Mobile Communication Using Audio Steganography by Mel-Frequency Cepstrum Analysis*. Vol. 16.

[3] [n. d.]. Malicious Payloads - Hiding Beneath the WAV. https ://threatvector.cylance.com/en_us/home/malicious-payloads-hiding-beneath-the-wav.html.

[4] [n. d.]. Tous surveillés - 7 milliards de suspects. https ://www.arte.tv/fr/videos/083310-000-A/tous-surveilles-7-milliards-de-suspects/.

[5] 2020. Fictitious Entry. *Wikipedia* (March 2020). Page Version ID : 943533827.

[6] Sumit Kumar Arora. 2018. Audio Steganography : The Art of Hiding Secrets within Earshot (Part 1 of 2). https ://medium.com/@sumit.arora/audio-steganography-the-art-of-hiding-secrets-within-earshot-part-1-of-2-6a3bbd706e15.

[7] Barnali Gupta Banik and Samir Kumar Bandyopadhyay. 2018. Blind Key Based Attack Resistant Audio Steganography Using Cocktail Party Effect. *Security and Communication Networks* 15 (2018). https://doi.org/10.1155/2018/1781384

[8] M. Baritha Begum and Y. Venkataramani. 2012. LSB Based Audio Steganography Based On Text Compression. *Procedia Engineering* 2 (Jan. 2012), 703–710. https://doi.org/10.1016/j.proeng.2012.01.917

[9] Youssef Bassil. 1459. *A Two Intermediates Audio Steganography Technique*. Vol. 11.

[10] Walter Bender, Daniel Gruhl, Norishige Morimoto, and Anthony Lu. 1996. Techniques for Data Hiding. *IBM Systems Journal* (1996). https://doi.org/10.1147/sj.353.0313

[11] Cryptography. [n. d.]. *A Novel Approach for Audiography- A Combination of Audio Steganography*. Vol. 1.

[12] Fatiha Djebbar, Beghdad Ayad, Karim Abed Meraim, and Habib Hamam. 2012. Comparative Study of Digital Audio Steganography Techniques. *EURASIP Journal on Audio, Speech, and Music Processing* 3, 1 (Oct. 2012), 25. https://doi.org/10.1186/1687-4722-2012-25

[13] Abdul Hakeem, Mohsin Shah, Zakir Khan, Abdul Qadi, and Noor Amin. 2014. Threshold Based LSB Audio Steganography, Vol. 7. https://doi.org/10.13140/RG.2.1.2948.4000

[14] C. F. S. Hartoko, S. Tjondronegoro, and B. Hidayat. [n. d.]. Audio Steganography Using Modified Enhanced Least Significant Bit In. 9 ([n. d.]).

[15] David Kahn. 1996. The History of Steganography. In *Information Hiding*, Gerhard Goos, Juris Hartmanis, Jan Leeuwen, and Ross Anderson (Eds.). Vol. 1174. Springer Berlin Heidelberg, Berlin, Heidelberg, 1–5. https://doi.org/10.1007/3-540-61996-8_27

[16] Md Shafakhatullah Khan, Asst Professor, V. Vijaya Bhasker, and V. Shiva Nagaraju. [n. d.]. *An Optimized Method for Concealing Data Using Audio Steganography*. Vol. 14.

[17] Fatma Susilawati Mohamad and Nurul Sahira Mohd Yasin. 2018. Information Hiding Based on Audio Steganography Using Least Significant Bit. *International Journal of Engineering & Technology* 8, 4.15 (Oct. 2018), 536–538. https://doi.org/10.14419/ijet.v7i4.15.28363

[18] Joe Mullin. 2020. The EARN IT Bill Is the Government's Plan to Scan Every Message Online. https ://www.eff.org/fr/deeplinks/2020/03/earn-it-bill-governments-not-so-secret-plan-scan-every-message-online.

[19] Gunjan Nehru and Puja Dhar. [n. d.]. *A Detailed Look of Audio Steganography Techniques Using LSB and Genetic Algorithm Approach.* Vol. 10.
[20] Masoud Nosrati, Ronak Karimi, and Mehdi Hariri. 2012. Audio Steganography : A Survey on Recent Approaches. *World Applied Programming, Vol* 5 (2012), 202–205.
[21] Masoud Nosrati, Ronak Karimi, Hamed Nosrati, and Ali Nosrati. [n. d.]. *Taking a Brief Look at Steganography : Methods and Approaches.* Vol. 4.
[22] Bruce Schneier, Matthew Fredrikson, Tadayoshi Kohno, and Thomas Ristenpart. [n. d.]. Surreptitiously Weakening Cryptographic Systems. ([n. d.]), 26.
[23] Jose Shelton and Gideon Kumar. 2010. Comparison between Auditory and Visual Simple Reaction Times. *Neuroscience & Medicine* 1 (Jan. 2010), 30–32. https://doi.org/10.4236/nm.2010.11004
[24] Kamred Udham Singh. [n. d.]. *LSB Audio Steganography Approach.* Vol. 6.
[25] Yuan-Yen Tai and Mohamed F. Mansour. 2019. Audio Watermarking over the Air With Modulated Self-Correlation. *arXiv :1903.08238 [cs]* (March 2019). arXiv:cs/1903.08238
[26] David Wheeler, Daryl Johnson, Bo Yuan, and Peter Lutz. [n. d.]. Audio Steganography Using High Frequency Noise Introduction. 13 ([n. d.]), 5.
[27] Guojiang Xin, Yuling Liu, Ting Yang, and Yu Cao. 2018. An Adaptive Audio Steganography for Covert Wireless Communication. *Security and Communication Networks* 12 (2018). https://doi.org/10.1155/2018/7096271
[28] Dengpan Ye, Shunzhi Jiang, and Jiaqin Huang. 2019. Heard More Than Heard : An Audio Steganography Method Based on GAN. *arXiv :1907.04986 [cs, eess]* 17 (July 2019). arXiv:cs, eess/1907.04986
[29] Mazdak Zamani. 2010. Genetic Based Substitution Techniques for Audio Steganography. 18 (2010).