Injection de commandes vocales sur ordiphone

José Lopes Esteves et Chaouki Kasmi
jose.lopes-esteves@ssi.gouv.fr
chaouki.kasmi@ssi.gouv.fr

ANSSI

Résumé Les interférences électromagnétiques intentionelles (IEMI) présentent un risque non-négligeable pour la sécurité des systèmes d'information (SSI). Jusqu'à présent, les risques en *confidentialité*, *intégrité* et *disponibilité* ont largement été abordés dans la littérature. A notre connaissance, l'utilisation des sources de forte puissance se limite à un impact sur la *disponibilité*. Dans cet article, nous introduisons une nouvelle technique d'injection de commandes à distance sur ordiphone par l'utilisation d'IEMI.

1 Introduction

L'intérêt des sources à énergie dirigée dans les applications militaires a largement été démontré afin de perturber voire détruire des systèmes électroniques. Le défi de générer une grande quantité d'énergie a lentement évolué vers la définition de formes d'onde efficaces afin d'abaisser le niveau des champs émis [13]. Les récents progrès dans le domaine de la radio logicielle [4] ont fourni la possibilité de générer des formes d'onde complexes [6]. La combinaison d'émetteurs [4] et d'amplificateurs de faible coût permet de concevoir des systèmes d'émission reconfigurables pour un budget relativement faible.

Ces dernières années, la classification fine des effets induits par ce type de sources sur des systèmes d'information a été étudiée [14]. Des résultats expérimentaux [14] ont mis en évidence la susceptibilité de la carte son des ordinateurs et des ordiphones, c'est à dire la possibilité d'induire des courants (i.e. tensions) parasites numérisés et démodulés par la carte. L'introduction de services de commandes vocales dans les ordinateurs et les ordiphones devient naturellement une cible adéquate pour ce type d'injection. Nous proposons dans cette étude de considérer les IEMI comme un nouveau vecteur d'injection de commandes sur ordiphones et ordinateurs (*n.b. la sécurité des systèmes de reconnaissance vocale n'est pas l'objet de ce travail*).

Cet article est organisé comme suit : dans la Section 2, les services de contrôle vocal disponibles dans la majorité des ordiphones sont brièvement

présentés. Ensuite Section 3, la conception d'une source de forte puissance intelligente conçue pour interagir avec l'interface de commande vocale est détaillée. Enfin, Section 4, une analyse de risques et des contre-mesures sont proposées.

2 Contrôle vocal et ordiphones

La commande vocale permet l'utilisation d'un terminal mobile en « mains libres ». Ce service intégré aux ordiphones est étendu aux objets connectés, ordinateurs de bureau et véhicules. Parmi les interpréteurs vocaux les plus répandus, on peut citer Google Voice Search (Google [7]), Siri et Voice Control (Apple [2]), S-Voice (Samsung [15]), Speech [11] et Cortana [10] (Microsoft). Lorsque le service de commande vocale n'est pas activé en permanence il est généralement activable par une pression longue sur un bouton (bouton matériel sur l'ordiphone ou la commande du casque) ou en démarrant manuellement les applications susmentionnées. L'utilisateur a ensuite accès à un grand nombre de fonctionnalités (variable selon l'appareil et le système d'exploitation) par la voix comme : l'accès à des services de téléphonie (envoyer des messages textes ...), l'accès à des services Internet (visiter des pages Web ...) et l'accès à des services locaux (démarrer des applications ...).

Il est à noter que les fabricants d'appareils mobiles tendent à activer progressivement les interpréteurs vocaux par défaut, ce qui expose de façon permanente l'interface vocale. Notons également que certaines de ces fonctionnalités peuvent être accessibles sans que l'ordiphone ne soit déverrouillé (c'est notamment le cas pour les services de téléphonie). Cette accessibilité pré-authentification des services de commande vocale a donné lieu à certaines attaques démontrant l'accès à des données sensibles [12], menant certaines entreprises à en interdire l'usage [3].

Selon les cas, les interpréteurs vocaux sont des services locaux ou distants. Généralement, les capacités des interpréteurs locaux sont plus limitées. Par exemple, Voice Control et Speech sont disponibles hors connexion, alors que Google Voice Search et Siri ne fonctionnent qu'avec une connexion internet. Cela signifie que l'utilisation de ces services sur des réseaux non maîtrisés peut donner lieu à des attaques modifiant la réponse des interpréteurs vocaux afin d'exécuter des commandes malveillantes. Ce point ne fait pas l'objet de la présente étude, mais constituerait un axe de recherche intéressant. L'utilisation de ces services comme canal caché pour l'exfiltration d'information a également été envisagée [9].

3 Injection de commandes vocales sur ordiphone

L'objectif de ce travail est de valider l'hypothèse suivante : *comme les écouteurs agissent comme une antenne FM, il doit être possible de les utiliser comme une interface de couplage pour les signaux HF et donc de porte d'entrée pour l'injection de commandes vocales.*

Dans le but de procéder à cette vérification, le protocole expérimental a consisté dans un premier temps à tester la possibilité d'injecter un signal sonore dans le *front end* audio des ordiphones, via le dispositif suivant : les ordiphones sont placés, avec leur casque d'origine branché, dans une cage de Faraday avec un accès à Internet via un point d'accès sans fil. Une sonde de champ électrique est posée à proximité des cibles afin de mesurer le niveau de champ minimal requis. Le son du microphone sur les ordiphones est enregistré et diffusé sur le réseau local via une application mobile. Les signaux de test sont émis sur une porteuse, dans la bande FM, modulée en amplitude par un signal sonore (de la musique, par exemple) à l'aide d'une USRP [4] couplée à un amplificateur afin d'obtenir des niveaux de champs suffisants.

Ce premier dispositif a permis de valider la première partie de l'hypothèse, puisque le son enregistré par les ordiphones correspondait bien au signal sonore modulant notre porteuse à 103 MHz. Il a été observé que, pour l'induction des signaux, le champ minimal requis autour de la cible est compris entre 25 et 30 V/m pour une fréquence d'injection à 103 MHz. Cela correspond au maximum des niveaux de référence conseillés pour prévenir les risques sanitaires.

L'utilisation de ce vecteur dans le but d'exploiter l'interface vocale recquiert la possibilité d'activer l'interpréteur vocal. Si celui-ci n'est pas activé par défaut, il nous faut émuler un appui long sur le bouton principal du casque. Il a été observé qu'un signal modulé en fréquence permet d'induire dans le câble du microphone du casque un signal électrique équivalent à un appui sur le bouton voulu. Il ne reste plus qu'à injecter un signal sonore correspondant à de la voix et contenant la commande vocale souhaitée. Pour ce faire, la méthode employée lors de l'expérience précédente est utilisée : une porteuse dans la bande FM est modulée en amplitude par le signal vocal.

Dans le cas où le contrôle vocal est configuré dans le mode « actif en permanence », le signal modulé en fréquence n'est plus utile. Il suffit de générer le mot-clé (« *OK, Google* » ou « *Hey, Siri* ») puis d'émettre les différentes commandes souhaitées par l'utilisation des signaux modulés en amplitude.

4 Analyse de risques et contre-mesures

La possibilité d'exécuter des commandes vocales à distance sur un système est considérée comme critique d'un point de vue SSI. En outre, tous les systèmes potentiellement capables de commande vocale peuvent être vulnérables à ce type d'attaque. Le profil de l'attaquant requis pour cette attaque peut être considéré comme « compétent », et implique des équipements radiofréquences « grand public ». Une connaissance de base de la radio logicielle et quelques informations sur le système d'exploitation cible sont nécessaires.

Afin de comprendre l'impact qu'une telle attaque peut avoir sur une cible, certains scénarios d'attaque ont été envisagés comme l'espionnage : l'attaquant active les interfaces sans-fil de la cible pour permettre le suivi de l'utilisateur [5]. Il peut également déclencher un appel vocal à son propre téléphone pour capturer les sons environnants. On peut également évoquer les attaques à but lucratif où l'attaquant cible tous les utilisateurs dans une zone géographique et force ses victimes à utiliser des services de téléphonie surtaxés, ou encore la compromission d'un équipement : l'attaquant peut forcer la victime à visiter une page web malicieuse qui exploite une vulnérabilité pour compromettre le système d'exploitation cible [1].

Quelques contre-mesures simples peuvent être appliquées par l'utilisateur afin de réduire la surface d'attaque comme débrancher le casque lorsque celui-ci n'est pas utilisé, désactiver le contrôle vocal quand il n'est pas utilisé, paramétrer le plus finement possible les fonctions accessibles par commande vocale, notamment lorsque l'ordiphone est verrouillé et activer les notifications de commande vocale (signal sonore, vibration...). Malheureusement, il y a un compromis à faire entre la sécurité et l'ergonomie que proposent ces services. Afin de préserver l'intérêt des utilisateurs pour ce type de service, les éditeurs, les fournisseurs et les fabricants pourront :

- réduire les fonctionnalités critiques accessibles ;
- réaliser quelques modifications sur le front-end audio afin de réduire la sensibilité de l'interface d'entrée : cela forcerait l'attaquant à émettre des niveaux de champs électromagnétique (EM) plus élevés pour induire des signaux sur celle-ci. Une meilleure protection du câble des écouteurs contribuerait également à cette mesure d'atténuation ;
- intégrer des améliorations sur la reconnaissance vocale : en effet, une meilleure reconnaissance de la voix de l'utilisateur légitime

obligerait l'attaquant à forger des commandes avec la signature vocale de chaque cible. On pourrait également imaginer des techniques pour discriminer une voix acoustique d'une voix numérisée puis rejouée ;
— fournir une meilleure granularité à l'utilisateur pour le paramétrage de la commande vocale : personnalisation du mot-clé (déjà possible sur la plupart des appareils), désactivation de l'interface vocale par défaut (*opt-in*), sélection plus fine des actions accessibles ;
— implémenter, comme proposé dans [8], un service utilisant les capteurs intégrés dans les ordiphones (magnétomètre...) afin d'analyser l'environnement EM avant l'interprétation d'une commande vocale. Cela permettrait de ne pas donner suite à la commande vocale en cas d'activité EM anormale.

5 Conclusion

La commande vocale est une interface homme-machine de plus en plus déployée et connait un fort succès pour des applications diverses. Dans cette perspective, les éditeurs et fournisseurs de services ont tendance à étendre les fonctionnalités accessibles par cette interface. Cela ne doit en aucun cas se faire au détriment de la sécurité.

La contribution de cet article est double : d'une part, une mise en exergue de la criticité de cette interface est proposée. En effet, les fonctionnalités critiques exposées risquent d'être de plus en plus nombreuses. Nous attirons donc l'attention des chercheurs en sécurité sur l'intérêt de développer la recherche sur l'interface vocale ainsi que celle des éditeurs sur la nécessité de considérer cette interface comme une nouvelle surface d'attaque afin de la sécuriser au mieux.

D'autre part, nous montrons qu'au delà des attaques en disponibilité, une utilisation élégante des interférences électromagnétiques intentionnelles peut donner lieu à des attaques plus abouties, comme ici l'exécution de commandes à distance.

En complément des scénarii d'attaque qui sont exposés, une réflexion sur les axes d'amélioration et de recherche pouvant constituer des contremesures est proposée. La communauté scientifique, les éditeurs et les utilisateurs ont tous un rôle important à jouer pour intégrer la commande vocale dans notre quotidien en rendant cette interface pratique, intuitive et sécurisée.

6 Remerciements

Les auteurs tiennent à remercier les éditeurs et les fabricants des dispositifs qui ont été testés, à savoir Apple, Google et Samsung, pour leur réactivité et pour avoir autorisé la publication de cette recherche.

Références

1. A. Moulu. Abusing Samsung KNOX to remotely install a malicious application : story of a half patched vulnerability. `http://blog.quarkslab.com/index.html`, 2014.
2. Apple. Siri. `https://www.apple.com/ios/siri/`, 2015.
3. B. Bergstein. IBM Faces the Perils of Bring Your Own Device. `http://www.technologyreview.com/news/427790/ibm-faces-the-perils-of-bring-your-own-device/`, 2012.
4. Ettus. Universal Software Radio Peripheral. http ://www.ettus.com/.
5. G. Wilkinson. The machines that betrayed their Masters : Mobile Device Tracking and Security Concerns. HackitoErgoSum, 2013.
6. GnuRadio. GNU Radio a free and open-source software development toolkit. `http://gnuradio.org/redmine/projects/gnuradio/wiki`, 2006.
7. Google. Ok Google. `https://support.google.com/websearch/answer/2940021?hl=en`, 2015.
8. C. Kasmi ; J. Lopes-Esteves. Automated analysis of the effects induced by radio-frequency pulses on embedded systems for EMC functional safety. *AT-RASC, URSI*, 2015.
9. L. Caviglione ; W. Mazurczyk. Understanding information hiding in ios. *IEEE Computer Magazine*, 2015.
10. Microsoft. Meet Cortana. `http://www.windowsphone.com/en-us/how-to/wp8/cortana/meet-cortana`, 2015.
11. Microsoft. Speech. `http://www.windowsphone.com/en-us/how-to/wp8/apps/use-speech-on-my-phone`, 2015.
12. N. Gonzalez. Siri exploited again how bypass the lock screen in iOS 8. `http://ios.wonderhowto.com/how-to/siri-exploited-again-bypass-lock-screen-ios-8-protect-yourself-0157749`, 2014.
13. N. Mora ; F. Vega ; G. Lugrin ; F. Rachidi. Study and classification of potential IEMI sources. *System Design and Assessment Note 41*, 2015.
14. C. Kasmi ; J. Lopes-Esteves ; M. Renard. Automation of the immunity testing of cots computers by the instrumentation of the internal sensors and involving the operating system logs technical report. *System Design and Assessment Note 44*, 2014.
15. Samsung. How do I Use Samsung S-Voice. `http://www.samsung.com/us/support/supportOwnersHowToGuidePopup.do?howto_guide_seq=7061&prd_ia_cd=N0000003&map_seq=54784`, 2014.