Ataques a Sistemas Baseados em Voz: Uma Ameaça em Desenvolvimento

Jorge Gregorio

Instituto de Computação
Universidade Federal Fluminense (UFF) – Niterói, RJ – Brasil

jgregorio.reis@gmail.com

Abstract. *Technology has been constantly changing since its emergence, we went from analog to digital, the invention of mobile devices and added to telephone communications, televisions gained Smart versions to access the internet and today almost everything is connected to this huge computer network. But something has been conquering this space, voice-based systems. In order to further facilitate our daily lives, it is integrated into cars, cell phones, and smart homes, but this evolution has also brought a new risk, the possibility of inaudible attacks for such devices. This work aims to show possible pre-existing attacks and some methods to minimize these risks.*

Resumo. *A tecnologia está em constante mudança desde seu surgimento, passando de sistema analógico para digital, a invenção de dispositivos móveis e adicionados nas comunicações telefônicas, as televisões ganharam versões Smart para acessarem a Internet e hoje quase tudo está conectado a essa enorme rede de computadores. Mas algo vem conquistando esse espaço, os sistemas baseados em voz. Com objetivo de facilitar ainda mais o nosso cotidiano, ele está integrado em carros, celulares e casas inteligentes, mas esta evolução acarretou também um novo risco, a possibilidade de ataques inaudíveis para tais dispositivos. Este trabalho pretende mostrar possíveis ataques já existentes e alguns métodos de minimizar estes riscos.*

1. Introdução

Com avanço tecnológico, o sistema baseado em voz está sendo cada vez mais implementado em dispositivos que estão conectados à Internet. Não somente dispositivos celulares, mas como televisores que antes eram somente analógicos, mas hoje se transformaram não só em digital, mas podem ser conectados na Internet, proporcionando acesso imediato a conteúdos que antes eram de difícil acesso. Há as Cidades inteligentes que estão se tornando cada vez mais tecnológicas, com monitoramento em tempo real, acionamentos e denúncias por meios da tecnologia, interligando todos os serviços públicos, como polícia, ambulâncias, bombeiros, todos centralizados em um único local. Existem modelos de casas inteligentes que são projetadas para prover conforto, qualidade de vida e segurança para seus proprietários. Alguns desse benefícios antes só poderiam ser executados de forma tátil, ou até mesmo presencial. Mas com esse avanço tecnológico tais ações podem ser executadas por comandos de voz, que podem ativar uma smart TV, ligar um som, trancar as portas e janelas de uma casa, efetuar uma ligação para emergência entre outras tarefas, proporcionadas por essa tecnologia que se expande a cada ano que passa. Mas toda essa vantagem, também acarretou desvantagens que são conhecidas como vulnerabilidades e podem ser exploradas por atacantes a fim de obter benefícios próprios.

Este artigo tem o objetivo de mostrar algumas dessas vulnerabilidades existentes, métodos de ataques e defesas.

O restante do texto está estruturado da seguinte maneira. A Seção 2 apresenta as Vulnerabilidades que podem ser exploradas por atacantes, a Seção 3 adiciona as informações sobre os métodos de ataque, a Seção 4 apresenta métodos de defesa que podem minimizar os riscos existentes e na Seção 5 é mostrado a conclusão da pesquisa.

2. Vulnerabilidades dos Sistemas Existentes

Atualmente existem vários sistemas baseados em voz, todos com o mesmo objetivo, prover facilidade e agilidade para ações humanas. Mas cada uma dessas tecnologias possui suas características próprias, ou seja, podem oferecer o mesmo objetivo, mas com algum recurso exclusivo. Nesta pesquisa serão abordados os principais sistemas baseados em voz e as vulnerabilidades que podem ser exploradas para realizar ataques.

No Alexa[1] e no Assistant, não existia a autenticação para realizar um comando, ou ao menos confirmar um código PIN antes de realizar compras, ou qualquer comando sensível. Mas foi necessária essa implementação, após serem informados por pesquisadores de possíveis ataques, mas ainda sim existe uma brecha que envolve o usuário. Também ainda existe falha para uma análise mais profunda de aplicativos que possam ser integrados aos sistemas de vozes, uma vez que você autoriza essa integração, o aplicativo consegue executar comandos em background. Uma outra brecha relatada no artigo publicado pelos pesquisadores [Zhang et al.,2018], diz respeito a quantidade de ruídos presente no mesmo ambiente do sistema de voz, podendo ser informada alguma palavra que ele não consegue distinguir e acaba executando tal ação.

Em abril de 2019 foi publicado um artigo no site WIRED[2] [Greenberg ,2019] relatando uma nova vulnerabilidade descoberta por Takeshi Sugawara da Universidade de Electro-Comunicações de Tóquio em sistemas de vozes. Ele mais um grupo de pesquisadores provaram que equipamentos, como Alexa e Google Assistant, podem responder comandos realizados por lasers de longa distância. Na próxima seção será descrito como essas vulnerabilidades podem ser exploradas.

3. Métodos de Ataques

Os atacantes utilizam métodos para explorar quaisquer vulnerabilidades existentes em dispositivos conectados à Internet. Isso não seria diferente com os sistemas baseados em voz, que precisam se conectar para executar os comandos que são enviados por seus usuários. Uma das vulnerabilidades básicas é a repetição de voz, explorada no artigo publicado pelos autores [Gong e Poellabauer, 2018] mostrando que esse tipo de ataque básico, consegue alcançar vários dispositivos através de áudio oculto (áudios maliciosos, que podem expressar comandos legíveis para qualquer dispositivo IoT, mas não detectáveis pelo usuário) emitido pelo *Youtube*. Neste mesmo artigo um ataque baseado em SO, pode ser explorado de forma perigosa e prática, um *malware* pode ser instalando no dispositivo alvo para que ele execute comandos inaudíveis caso consiga elevar seu privilégio para administrador.

Um método chamado de *Voice Squatting*, foi testado em uma pesquisa realizada no artigo publicado pelos pesquisadores [Zhang et al.,2018], que consiste em enviar comandos com palavras que possam ser parecidas com as palavras existentes no sistema alvo, com o propósito de confundir os dispositivos. Isso é possível porque os sistemas de voz ainda não conseguem identificar 100% das palavras que são enviadas por usuários. Este tipo de ataque pode não gerar perda para o usuário, mas uma avaliação negativa para o fabricante, gerando uma vantagem para seus concorrentes. Um segundo método analisado nesse mesmo artigo, conhecido como *Voice Masquerading* tem por

objetivo induzir o usuário a fornecer informações sensíveis. O *Voice Masquerading* pode ser usado de duas formas: *In-communication skill switch* e *Faking termination*. O atacante utiliza o *In-communication* de forma a forçar uma execução do dispositivo alvo, para que ele responda qualquer chamada quando for solicitada e assim conseguir informações que possam ser úteis para seu propósito. Esse ataque de acordo com artigo é executado no *Google Assistant*, emulando sua própria voz através SSML, uma linguagem baseada em XML para aplicativos de voz. *Faking termination* é um ataque que se baseia em conseguir execução de comandos enquanto o *Google Assistant* ou *Alexa* aguarda alguma chamada do usuário legítimo. Isso ocorre porque de acordo com a pesquisa do artigo citado dos pesquisadores [Zhang et al.,2018], o usuário ainda não está totalmente alinhado de como essas tecnologias funcionam, assim quando são utilizadas as palavras "pare" ou "cancele" ambos os sistemas mantêm a porta de comunicação aberta, aguardando algum outro comando. Nesse momento o atacante pode enviar chamadas e conseguir controle sobre o dispositivo.

Esses comandos muitas vezes podem ser inaudíveis ao ouvido humano, mas potencialmente perigosos para os dispositivos que podem executar tarefas que não sejam solicitadas pelo usuário original. No artigo [Roy et al.,2018], foi realizada uma pesquisa sobre ataques que podem ser realizados de forma imperceptível. No início, ataques desse tipo eram limitados a 1,5 metros, chagando no máximo a 3 metros de alcance, mas de acordo com [Roy et al.,2018], a pesquisa conseguiu mostrar um ataque com cerca de 7,5m de distância. O *LipRead* foi implementado em vários alto-falantes em conjunto de um amplificador e todos ligados a um computador. Para realizar o ataque, é preciso digitar o comando e este por sua vez é convertido pelo MATLAB e enviado de forma inaudível pelo *LipRead* através dos alto-falantes.

Um outro tipo de ataque que atualmente é mais reconhecido se chama *DolphinAttack*, mas ao contrário do *LipRead* ele possui curto alcance e precisa atender a alguns requisitos, para que consiga atingir seu alvo. O *DolphinAttack* permite que o atacante possa controlar o dispositivo alvo de forma remota, desde que esteja a 1,5 metros de distância do atacante sendo necessário também que esteja desbloqueado. Com esses requisitos atendidos, o atacante consegue efetuar comandos de voz com frequências acima de 20Khz percebidas de forma clara pelos dispositivos, que executam as tarefas como se estivessem sendo enviadas pelo usuário original. Assim ele pode realizar chamadas, ativar câmeras e quaisquer aplicativos que são controlados por voz. Recentemente foi identificado um novo ataque contra dispositivos controlados por voz, um pesquisador chamado Takeshi Sugawara da Universidade de Electro-Comunicações de Tóquio que atua na segurança cibernética, conseguiu descobrir esse novo método de ataque. De acordo com Sugawara é possível executar comandos de voz com feixe laser apontado diretamente para o dispositivo, que interpreta esse feixe como uma voz, podendo atingir até 100 metros de distância. Essa vulnerabilidade foi explorada em vários dispositivos de voz, como *Amazon Alexa*, *Google Home*, *Siri* e *Samsung Galaxy S9*. Uma pequena demonstração está na figura abaixo, observando atentamente pode se perceber o feixe de um laser atingindo o dispositivo na janela.

Fonte: wired.com

FOTOGRAFIA: UNIVERSIDADE DE ELECTRO-COMUNICAÇÕES; UNIVERSIDADE DE MICHIGAN

Figura-1

"É possível fazer os microfones responderem à luz como se fosse som", diz Sugawara pesquisador da Universidade de Electro-Comunicações de Tóquio. "Isso significa que qualquer coisa que atue nos comandos de som funcionará nos comandos de luz". Conforme Sugawara, não se faz necessário uma precisão sobre o ponto do alto-falante, bastando somente inundá-lo com luz e conforme a intensidade os comandos são compreendidos. O atacante, obtendo controle sobre o dispositivo, pode explorar a vulnerabilidade invocando comandos, para abrir portões de garagem, porta da casa ou qualquer ação que esteja vinculada nos dispositivos alvos.

4. Métodos de Defesa

Com todas essas ameaças existentes, foi necessário implementar métodos de defesa, para minimizar os riscos e tentar ao máximo prover privacidade e segurança para cada usuário que utiliza dispositivos controlados por voz. Alguns pesquisadores recomendam que seja habilitado toda e qualquer medida de segurança em dispositivos móveis ou até

aqueles que são utilizados em residências para que estas sejam transformadas em casas inteligentes.

Para dispositivos moveis é muito importante habilitar bloqueio de tela, duplo fator de autenticação para contas com privilégios de administrador, caso exista, para habilitar a autenticação para comandos de voz. Existem métodos de defesas específicos em alguns casos, para aumentar ainda mais a segurança em dispositivos baseados em voz, com mencionado no artigo publicado pelos pesquisadores [Gong e Poellabauer, 2018], para segurança contra-ataques baseados em SO, é necessário separar entrada e saída de voz. Essa defesa foi proposta no artigo publicado pelos pesquisadores [Petracca et al., 2015], que criaram um aplicativo Android que realiza essa execução de forma automática, após ser instalado no dispositivo. Ainda de acordo com o artigo, ele é capaz de prevenir contra 6 tipos de ataques diferentes baseados em SO, rastreando e controlando o fluxo

de canais para evitar tais ataques, ou seja, ele alterna entre entrada e saída de áudio para que ambos não sejam executados simultaneamente.

Mas o Android torna-se ineficaz para ataques baseados em hardware, que são muito explorados pelo *DolphinAttack*. Uma defesa contra esse tipo de ataque citada no próprio artigo publicado em 2015 pelos pesquisadores Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, Wenyuan Xu†, propõe uma alteração no microfone, para que estes não detectem mais sons acima de 20KHz, por padrão. Mas para alterar esse padrão, não depende do usuário, mas sim do fabricante que precisaria atribuir a função de filtrar sons acima desse nível.

Um segundo método de defesa contra o *DolphinAttack*, seria utilizar SVM, conforme explicado no artigo pesquisadores [Yan et al., 2017]. Essa técnica consiste em uma análise do sinal modulado enviado para o dispositivo, exibindo a diferença entre as frequências originais e as que estão sendo utilizadas para ataque. Mas para que esta defesa seja eficiente, é necessário adicionar palavras maliciosas nas configurações da SVM (*supported vector machine*), como se fosse uma blacklist. Alguns ataques podem causar danos não somente para o usuário, mas para o próprio fabricante desses dispositivos. Como mencionado no artigo dos pesquisadores [Zhang et al.,2018], o ataque *Voice Squatting* pode denegrir a imagem da empresa e isso não depende das configurações de segurança do usuário, mas de um filtro para melhor interpretação dos comandos que são invocados para tais dispositivos.

Para ataques como *Voice Masquerading* também descrito no artigo publicado por [Zhang et al.,2018], os pesquisadores implementaram um detector de reposta de entrada e saída de voz, assim ele consegue enviar um alerta para o usuário quando existe algo suspeito. O detector foi implementado em dois módulos SRC (*Skill Response Checker*) que captura os comandos suspeitos e o UIC (*User Intention Classifier*) que analisa os comandos em uma base de dados interna.

Para o outro método de ataque por laser, ainda não foi identificado nenhuma medida de segurança. Pesquisadores recomendam que os dispositivos sejam mantidos longe de janelas, ou que algo sólido seja adicionado à frente de qualquer um, minimizando a falha existente. As empresas têm trabalhado junto com os pesquisadores para melhor entender essa brecha em seus sistemas, para manter a privacidade e segurança de qualquer usuário.

5. Considerações Finais

Tudo que hoje está conectado na Internet, seja aplicativo, dispositivos de voz ou moveis, até mesmo eletrodomésticos, estão sujeitos a ataques de usuários mal-intencionados. Tudo isso traz vulnerabilidades que podem ser exploradas por alguém em algum momento, como em 2011 quando uma das maiores empresas de tecnologia, a Sony, sofreu um ataque de SQL *Injection* causando prejuízo de US$ 600 mil. Alguns desses ataques ainda possuem limitações de alcance, principalmente contra celulares que exigem desbloqueio de tela ou uma distância que pode variar entre 1,5 m até 3,0 m.

Mas algo chama atenção nesses tipos de ataques, se puderem ser realmente executados pelo Youtube, seria desastroso para usuário e fabricantes, que precisariam “blindar” rapidamente qualquer porta aberta em seus dispositivos. Uma outra ameaça com alto grau de criticidade, está na exploração por lasers, onde o atacante consegue invocar comandos somente acertando o microfone de um dispositivo alvo, este método com certeza merece uma atenção especial, pois qualquer usuário que adquira um *Alexa* ou

Google Home, acaba integrando tudo que for *smart* a essas dispositivos. Um atacante precisaria somente estar na linha de visão do dispositivo para controlar e invocar comandos nítidos e abrir por exemplo a porta de uma garagem ou até mesmo a porta da casa do proprietário.

Mesmo com as pesquisas e algumas comprovações de vulnerabilidades, esses ataques ainda não são considerados uma ameaça de alto risco, mas merecem uma atenção dos fabricantes para que brechas existentes sejam corrigidas a fim de amadurecer a tecnologia. Para empresas como *Google Assitant* ou *Siri* existe o requisito de autenticação por voz, para comandos de privilégio elevado, como por exemplo efetuar uma compra. Antes que a transação seja efetuada, é solicitado um código PIN para aprovação, mas que pode haver uma brecha caso o usuário não diga a palavra "STOP" após a realização da compra.

São vários os métodos de ataque existentes. Ataques como MITM, pode ser modificado para conseguir informações privilegiadas de usuários. Uma variante desse ataque foi criada com nome de *Skill-In-The-Middle*, que consegue manipular informações solicitadas pelo usuário fornecendo repostas falsas podendo induzir o usuário a erros desastrosos.

As precauções de segurança devem ser de ambas as partes, cabe ao usuário evitar aplicativos que não sejam comercializados fora das lojas correspondentes a seus dispositivos e mesmo assim pesquisar antes de utilizar para ter uma ideia do que está adicionado ao seu celular ou qualquer outro aparelho.

Nas empresas existem as parcerias com pesquisadores que relatam as falhas para que sejam corrigidos e lançados *patches* de segurança, minimizando os riscos de ataques. O objetivo das empresas deveria ser desenvolver qualquer software ou hardware cumprindo os 3 pilares da segurança (Confidencialidade, Integridade e Disponibilidade), com isso todos ganhariam em privacidade e segurança das informações que são trafegadas pela Internet das coisas.

6. Referências

Yuan Gong, Christian Poellabauer (2018) "An Overview of Vulnerabilities of Voice Controlled Systems" Researchgate.net

Nan Zhang, Xianghang Mi, Xuan Feng, XiaoFeng Wang,Yuan Tian, Feng Qian (2018) "Understanding and Mitigating the Security Risks of Voice-Controlled Third-Party Skills on Amazon Alexa and Google Home" arxiv.org

Arlene Aucella Consultant, 11 Appaloosa Ln., Hamilton, MA 01982 (1986) "Voice:Technology Searching For Communication Need" dl.acm.org

Xuejing Yuan, Yuxuan Chen, Yunhui Long, Xiaokang Liu, Kai Chen, Shengzhi Zhang, Heqing Huang,Xiaofeng Wang,Carl A. Gunter (2018) "CommanderSong: A Systematic Approach for Practical Adversarial Voice Recognition" usenix.org

Nirupam Roy, Sheng Shen, Haitham Hassanieh, Romit Roy Choudhury (2018) "Inaudible Voice Commands: The Long-Range Attack and Defense" usenix.org

X. Yuan et al., "All Your Alexa Are Belong to Us: A Remote Voice Control Attack against Echo," 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, 2018, pp. 1-6, doi: 10.1109/GLOCOM.2018.8647762.

S.Chen et al., "You Can Hear But You Cannot Steal: Defending Against Voice

Impersonation Attacks on Smartphones," 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), Atlanta, GA, 2017, pp. 183-195, doi: 10.1109/ICDCS.2017.133.

Y. Gong and C. Poellabauer, "Protecting Voice Controlled Systems Using Sound Source Identification Based on Acoustic Cues," 2018 27th International Conference on Computer Communication and Networks (ICCCN), Hangzhou, 2018, pp. 1-9, doi: 10.1109/ICCCN.2018.8487334.

J. Shang and J. Wu, "Enabling Secure Voice Input on Augmented Reality Headsets using Internal Body Voice," 2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), Boston, MA, USA, 2019, pp. 1-9, doi: 10.1109/SAHCN.2019.8824980.

R. Zhang, X. Chen, S. Wen, X. Zheng and Y. Ding, "Using AI to Attack VA: A Stealthy Spyware Against Voice Assistances in Smart Phones," in IEEE Access, vol. 7, pp. 153542-153554, 2019, doi: 10.1109/ACCESS.2019.2945791.

N. Zhang, X. Mi, X. Feng, X. Wang, Y. Tian and F. Qian, "Dangerous Skills: Understanding and Mitigating Security Risks of Voice-Controlled Third-Party Functions on Virtual Personal Assistant Systems," 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2019, pp. 1381-1396, doi: 10.1109/SP.2019.00016.

Huan Feng, Kassem Fawaz, and Kang G. Shin. 2017. Continuous Authentication for Voice Assistants. In Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (MobiCom '17). Association for Computing Machinery, New York, NY, USA, 343–355. DOI:https://doi.org/10.1145/3117811.3117823

Richard Mitev, Markus Miettinen, and Ahmad-Reza Sadeghi. 2019. Alexa Lied to Me: Skill-based Man-in-the-Middle Attacks on Virtual Assistants. In Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security (Asia CCS '19). Association for Computing Machinery, New York, NY, USA, 465–478. DOI:https://doi.org/10.1145/3321705.3329842

Yao Wang, Wandong Cai, Tao Gu, Wei Shao, Yannan Li, and Yong Yu. 2019. Secure Your Voice: An Oral Airflow-Based Continuous Liveness Detection for Voice Assistants. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3, 4, Article 157 (December 2019), 28 pages. DOI:https://doi.org/10.1145/3369811

M. Shirvanian, S. Vo and N. Saxena, "Quantifying the Breakability of Voice Assistants," 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom, Kyoto, Japan, 2019, pp. 1-11, doi: 10.1109/PERCOM.2019.8767399.