Mitigating Security Risks on Virtual Personal Assistant Systems

A mini-project report submitted

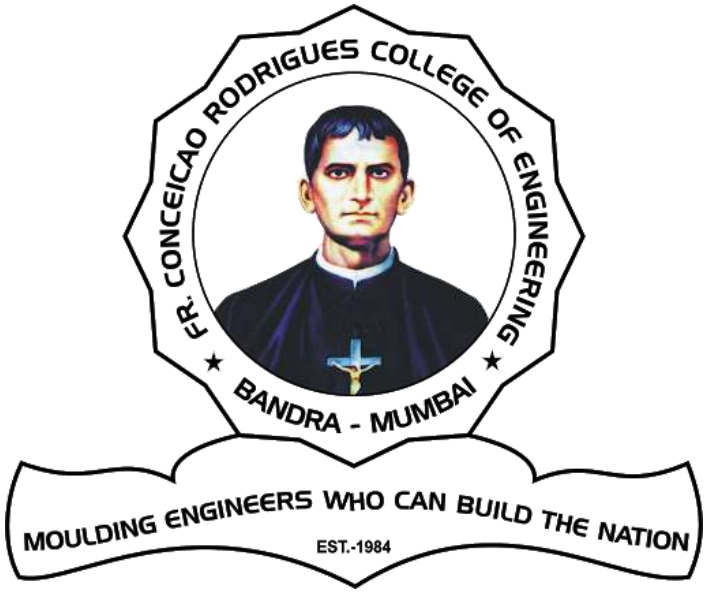
Cryptography and Network Security (Semester V)

by

Mareena Fernandes	(8669)
Fatima Pereira	(8690)
Hanita Rego	(8693)
Ashally Tuscano	(8710)

Under the guidance of

Prof. Prajakta Bhangale



DEPARTMENT OF INFORMATION TECHNOLOGY

Fr. Conceicao Rodrigues College of Engineering

Bandra (West), Mumbai – 400 050

Approval Sheet

Project Report Approval

This project report entitled by **Mitigating Security Risks on Virtual Personal Assistant Systems** by **Mareena, Fatima, Hanita and Ashally** is approved as mini project in Third Year Engineering, Information Technology.

Examiners

1. _____
2. _____

Date:

Place:

Abstract

Significant advances have been made in VPA technology due to the market leaders such as Amazon's Alexa, Apple's Siri, Google Home, and Microsoft's Cortana. VPA-enabled apps are becoming widespread to the point that they will be embedded within most services which however is known to be vulnerable, lacking proper authentication (from the user to the VPA).

A new authentication challenge, from the VPN service to the user, has emerged with the rapid growth of the VPA ecosystem, which allows a third party to publish a function (called skill) for the service and therefore can be exploited to spread malicious skills to a large audience during their interactions with smart speakers like Amazon Echo and Google Home.

VPA systems will also become increasingly dependent upon advanced data analytics to process the huge amount of data from a variety of stheces to learn about users and organize information. In this report, there is a study that concludes such remote, large-scale attacks are indeed realistic. We discovered two new attacks: voice squatting in which the adversary exploits the way a skill is invoked (Example: ‘open capital one’), using a malicious skill with a similarly pronounced name (Example: ‘capital won’) or a paraphrased name (Example: ‘capital one please’) to hijack the voice command meant for a legitimate skill (Example: ‘capital one’), and voice masquerading in which a malicious skill impersonates the VPA service or a legitimate skill during the user’s conversation with the service to steal her personal information.

Table of Contents

Chapter 1

Introduction

Voice assistant device are getting very popular especially echo devices and google home devices. We can do a lot of other tasks like to play music, to control the smart home devices or to send out some money using PayPal. Actually, we can also use it to access the medical information. These devices allow the user to command the system with voice only: for example, one can say "what will the weather be like tomorrow?" "set an alarm for 7 am tomorrow", etc., to get the answer or execute corresponding tasks on the system. In addition to their built-in functionalities, VPA services are enhanced by ecosystems fostered by their providers, such as Amazon and Google, under which third-party developers can build new functions (called skills by Amazon and actions by Google1) to offer further helps to the end-users, for example, order food, manage bank accounts and text friends.

These devices quickly-gained popularity, however, could bring in new security and privacy risks, whose implications have not been adequately understood so far.

Fig. 1: Examples of Virtual Personal Assistant

Problem Statement

Understanding and Mitigating Security Risks of Voice-Controlled Third-Party Functions on Virtual Personal Assistant Systems. With respect to increasing technology and security people find loophole to misuse data, to avoid which we need to calculate the risks and understand to extent it can affect the individual using the technology.

Chapter 2

Literature Survey

Security risks in VPA voice control:

Today's VPA systems are designed to be primarily commanded by voice. The attacks here include obfuscated voice commands or even completely inaudible ultrasound to attack speech recognition systems. These attacks impersonate the authorized user to the voice-controlled system, since no protection is in place to authenticate the user to the system. The emergence of the VPA ecosystem brings in another authentication challenge: it also becomes difficult for the user to determine whether she is indeed talking to the right skill and the VPA itself as she expects. The problem comes from the fact that through the skill market, an adversary can publish malicious third-party skills designed to be invoked by the user's voice commands (through a VPA device such as Amazon Echo or Google Home) in a misleading way, due to the ambiguity of the voice commands and the user's misconception about the service. As a result, the adversary could impersonate a legitimate skill or even the VPA (potentially in a large scale) to the user.

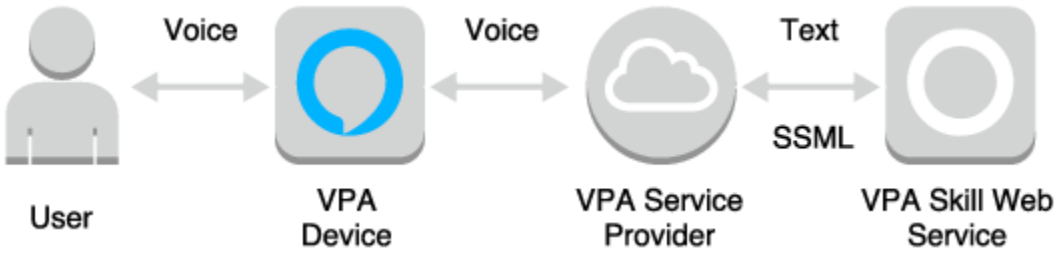


Fig. 2: Infrastructure of VPA System

Voice-based remote attacks:

This study has Two threats never known before, called voice squatting attack (VSA) and voice masquerading attack (VMA). In a VSA, the adversary exploits how a skill is invoked (by a voice command), and the variations in the ways the command is spoken

(Example: phonetic differences caused by accent, courteous expression, etc.) to cause a VPA system to trigger a malicious skill instead of the one the user intends. For example, one may say "Alexa, open Capital One please", which normally opens the skill Capital One, but can trigger a malicious skill Capital One Please once it is uploaded to the skill market

A VMA aims at the interactions between the user and the VPA system, which is designed to hand over all voice commands to the currently running skill, including those supposed to be processed by VPA system like terminating the current skill and switching to a new one. In response to the commands, a malicious skill can pretend to yield control to another skill (switch) or the service (terminate), yet continue to operate stealthily to impersonate these targets and get sensitive

information from the user Example: “play some sleep sounds”. These expressions allow the adversary to mislead the service and launch a wrong skill in response to the user’s voice command, such as ‘some sleep sounds’ instead of sleep sounds.

Mitigation:

This expression describes how a name is pronounced, allowing us to measure the phonetic distance between different skill names. Those sounding similar or having a subset relation are automatically detected by the scanner. This technique can be used to vet the skills uploaded to a market. Example: a skill with an invocation name “me a dog fact” looks suspiciously related to the popular skill “dog fact”. In this study, we developed a novel technique that automatically identifies those similar to system utterances, even in the presence of obfuscation attempts (Example: changes to the wording), and also captures the user’s skill switching intention from the context of her conversation with the running skill.

Fig. 3: Mitigation - Risks

Chapter 3

Objective

To study the potential of both VSA and VMA in real-world settings, four skills were published on Alexa to simulate the popular skill "Sleep and Relaxation Sounds" (the one receiving most reviews on the market as of Nov. 2017) whose invocation name is "sleep sounds" as shown in the figure. The attack skills provide only legitimate functions like playing sleep sounds just like the popular target. Although their invocation names are related to the target, their welcome messages were deliberately made to be different from that of the target, to differentiate them from the popular skill as shown in the figure. Also, the number of different sleep sounds supported by the skills is much smaller than the target (9 versus 63). Also, to find out whether these skills were mistakenly invoked, another skill we registered as a control, whose invocation name "incredible fast sleep" would not be confused with those of other skills. Therefore, it was only triggered by users intentionally.

Attack Skill		**Victim Skill**
Skill Name	**Invocation Name**	**Target Invocation Name**
Amazon		
Smart Gap	smart gap	smart cap
Soothing Sleep Sounds	sleep sounds please	sleep sounds
Soothing Sleep Sounds	soothing sleep sounds	sleep sounds
My Sleep Sounds	the sleep sounds	sleep sounds
Super Sleep Sounds	sleep sounds	sleep sounds
Incredible Fast Sleep	incredible fast sleep	N/A
Google		
Walk Log	walk log	work log

Table 1: Skill names, invocation names of the attack skills we registered on Amazon and Google as well as the target invocation name of the victim skills

Chapter 4

Methodology

Findings:

Three weeks of skill usage data was collected. The results are shown in the table below.

Skill Invocation Name	# of Users	# of Requests	Avg. Req/User	Avg. Unknown Req/User	Avg. Instant Quit Session/User	Avg. No Play Quit Session/User
sleep sounds please	325	3,179	9.58	1.11	0.61	0.73
soothing sleep sounds	294	3,141	10.44	1.28	0.73	0.87
the sleep sounds	144	1,248	8.49	1.11	0.33	0.45
sleep sounds	109	1,171	10.18	1.59	0.51	0.82
incredible fast sleep	200	1,254	6.12	0.56	0.06	0.11

Table 2: Real-world attack skills usage. The usage data are total number of unique users, total number of requests sent by these users, average number of requests sent per user, average number of requests unknown to the skills sent per user, average number of instant quit sessions (quit immediately after invocation without further interaction) per user, and average number of no-play-quit sessions (quit without playing any sleep sounds) per user.

Analysis:

- 325 users took the skill as the target, the attack skill got a higher number of unknown requests, higher chance of quitting the current session immediately without further interacting with the skill or playing any sleep sounds.
- Compared with the control, it was invoked by more users, received more requests per user, also much higher rates of unknown requests and early quits.
- Out of the 9,582 user requests collected, 52 was for skill switch i.e. trying to invoke another skill during the interactions with the skill, and 485 tried to terminate the skill using StopIntent or CancelIntent, all of which could be exploited for launching VMAs.
- It was also found that some users strongly believed in the skill switch that they even cursed Alexa for not doing that after several tries.

Finding Voice Squatting Skills:

To better understand potential voice squatting risks already in the wild and help automatically detect such skills, a skill-name scanner was developed and used to analyze tens of thousands of skills from Amazon and Google markets. The data was collected from amazon.com and its companion App as well as Google Assistant app. 23,758 skills from Amazon and 1,001 skills from Google Assistant app were retrieved. Competitive Invocation Name (CIN) is a name with a similar pronunciation as that of a target skill or uses different variations (Example: “sleep sounds please”)

of the target's invocation utterances. The scanner takes two steps to capture the CINs for a given invocation name: utterance paraphrasing and pronunciation comparison.

Fig. 4: Voice Squatting

Utterance Paraphrasing:

Identifies suspicious variations of a given invocation name. A simple yet effective approach alternative was taken, which creates variations using the invocation commands collected from the survey study. Specifically, 11 prefixes of these commands were gathered, Example: "my" and 6 suffixes, Example: "please", and applied them to a target skill's invocation name to build its variations recognizable to the VPA systems. Each of these variations can lead to other variations by replacing the words in its name with those having similar pronunciations, Example: replacing word "please" with word "plese".

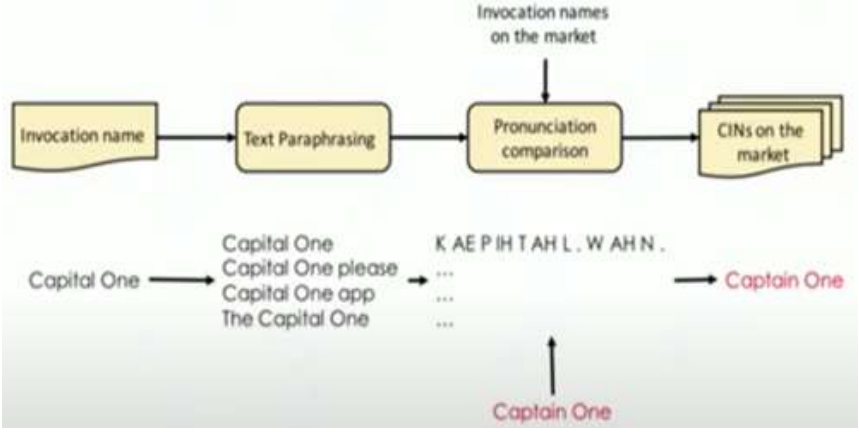


Fig. 5: Utterance Paraphrasing

Pronunciation Comparison:

Finds the similarity in the pronunciation between two different names. To identify the names with similar pronunciation, the scanner converts a given name into a phonemic presentation using the ARPABET phoneme code. The approach uses CMU pronunciation dictionary (including 134,000 words) to find the phoneme code for each word in the name. Among 9,120 unique words used to compose invocation names, 1,564 are not included in this dictionary. To get their pronunciations, an approach proposed in the prior research was followed. After turning each name into its phonemic representation, the scanner compares it with other names to find those that sound similarly. To this end, *edit distance* is used to measure the pronunciation similarity between two phrases, i.e., the minimum cost in terms of phoneme editing operations to transform one name to the other. 9,181 pairs of alternative pronunciations were collected from the CMU dictionary.

The measurement study done on Alexa and Google Assistant skills using the scanner produced results as follows (given in the table below):

# of Skills	# of unique invocation names	Transformation cost	Skills has CIN* in market			Skills has CIN in market excluding same spelling			Skills has CIN in market through utterance paraphrasing		
			# of skills	Avg. CINs per skill	Max CINs	# of skills	Avg. CINs per skill	Max CINs	# of skills	Avg. CINs per skill	Max CINs
19,670	17,268	0	3,718(19%)	5.36	66	531(2.7%)	1.31	66	345(1.8%)	1.04	3
		≤ 1	4,718(24%)	6.14	81	2,630(13%)	3.70	81	938(4.8%)	2.02	68

* Competitive Invocation Name

Table 3: Squatting risks on Alexa skill markets

Squatting risks on skill markets:

- As shown in table above, 3,655 (out of 19,670) Alexa skills have CINs on the same market, which also include skills that have identical invocation names (in spelling). After removing the skills with the identical names, still 531 skills have CINs, each on average related to 1.31 CINs. The one with the most CINs is "cat fax": it was found that 66 skills are named "cat facts" and provide similar functions.
- Interestingly, there are 345 skills whose CINs apparently are the utterance paraphrasing of other skills' names. Further, when raising the threshold to 1 (still well below what is reported in the experiment), it was observed that the number of skills with CINs increases dramatically, suggesting that skill invocations through Alexa can be more complicated and confusing than thought.
- By comparison, Google has only 1,001 skills on its market and does not allow them to have identical invocation names. Thus, we are only able to find 4 skills with similarly pronounced CINs under the threshold 1.

- The study shows that the voice squatting risk is realistic, which could already pose threats to tens of millions of VPA users in the wild. So, it becomes important for skill markets to beef up their vetting process (possibly using a technique similar to the scanner) to mitigate such threats.

Case Study

From the CINs discovered by the scanner, a few interesting cases were found. Particularly, there is evidence that the squatting attack might already happen in the wild: as an example, relating to a popular skill “dog fact” is another skill called “me a dog fact”. This invocation name does not make any sense unless the developer intends to hijack voice commands intended for “dog fact” like “tell me a dog fact”. Also, intriguing is the observation that some skills deliberately utilize the invocation names unrelated to their functionalities but following those of popular skills. Prominent examples include the “SCUBA Diving Trivia” skill and “Soccer Geek” skill, all carrying an invocation name “space geek”. This name is actually used by another 18 skills that provide facts about the universe.

Chapter 5

Defending against voice Masquerading

The scheme consists of two components:

The Skill Response Checker (SRC) and the User Intention Classifier (UIC). SRC captures suspicious responses from a malicious skill such as a fake skill recommendation that mimics the service utterances produced by the VPA system. UIC looks at the other side of the equation, checking the voice commands issued by the user, to find out whether she attempts to switch to a different skill in a wrong way, which can lead her right into the trap set by the malicious skill.

Skill Response Checker (SRC):

To defend against such attacks, the core idea is to control the avenues that a malicious skill can take to simulate either the VPA system or a different skill, allowing the user to be explicitly notified of VPA system events (Example: a context switch and termination) when a security risk is observed. For this purpose, SRC maintains a set of common utterance templates exclusively used by the VPA system to capture the similar utterances generated by a running skill. Whenever a skill's response is found to be similar enough to one of those utterance templates, an alarm is triggered and actions may be taken by the VPA system to address the risk, Example: reminding users of the current context before delivering the response. A challenge here is how to reliably measure whether a given response is similar enough to one of those templates, as the attacker could morph (rather than copy) the target system utterance.The threshold is determined by looking at the SRs between legitimate skill responses and the templates.

User Intention Classifier (UIC):

UIC further protects the user attempting to switch contexts (which currently is not supported by the VPA) from an impersonation attack. For this purpose, it aims at automatically detecting such erroneous commands from the user, based upon the semantics of the commands and their context in the conversation with the running skill. If such attempts can be perfectly identified by the VPA, it can take various actions to protect the user, Example: reminding her that she is talking to the skill, not the VPA, or following the instructions to terminate the skill, which closes the surface for the impersonation attack. However, accurately recognizing the user's intention (for context switch) is nontrivial. The challenges come from not only the variations in natural-language commands (Example: "open sleep sounds" vs. "sleep sounds please") but also the observations that some context-switch like commands could be legitimate for both the running skill and the VPA: for example, when interacting with Sleep Sounds, one may say "play thunderstorm sounds", which can be interpreted as commanding the skill to play the requested sound, as well as asking the VPA to launch a different skill "Thunderstorm Sounds".

Chapter 6

Features

Virtual personal assistants (VPA) (Example: Amazon Alexa and Google Assistant) today mostly rely on the voice channel to communicate with their users, which however is known to be vulnerable, lacking proper authentication (from the user to the VPA). A new authentication challenge, from the VPA service to the user, has emerged with the rapid growth of the VPA ecosystem, which allows a third party to publish a function (called skill) for the service and therefore can be exploited to spread malicious skills to a large audience during their interactions with smart speakers like Amazon Echo and Google Home. In this paper, we report a study that concludes such remote, large-scale attacks are indeed realistic. We discovered two new attacks: voice squatting in which the adversary exploits the way a skill is invoked (Example: “open capital one"), using a malicious skill with a similarly pronounced name (Example: “capital won") or a paraphrased name (Example: “capital one please") to hijack the voice command meant for a legitimate skill (Example: “capital one"), and voice masquerading in which a malicious skill impersonates the VPA service or a legitimate skill during the user's conversation with the service to steal her personal information. These attacks aim at the way VPAs work or the user's misconceptions about their functionalities, and are found to pose a realistic threat by the experiments (including user studies and real-world deployments) on Amazon Echo and Google Home. The significance of the findings has already been acknowledged by Amazon and Google, and further evidenced by the risky skills found on Alexa and Google markets by the new squatting detector we built. We further developed a technique that automatically captures an ongoing masquerading attack and demonstrated its efficacy.

Fig. 6: Features of VPA

At a high level, we found from real-world conversations that if a user intends to switch context, her utterance tends to be more semantically related to the VPA system (Example: "open sleep sounds") than the current skill, and the relation goes the other way when she does not. Therefore, we designed UIC to compare the user's utterance to both system commands and the running skill's context to infer her intention, based upon a set of features. Some of these features were identified through a semantic comparison between the user utterance and all known system commands. To this end, we built a system command list from the VPA's user manual, developers' documentation and real-world conversations collected in the study. Against all commands on the list, an utterance's maximum and average SRs are used as features for classification.

Chapter 7

Limitations

To evaluate the VMA defense (SRC and UIC), we tried the best to collect representative datasets for training and evaluation, and the good experimental results strongly indicate that the defense is promising for mitigating real-world VMA risks as described above. In the meantime, we acknowledge that the datasets might still not comprehensive enough for covering all real-world attack cases, and evasion attacks could happen once the approach is made public. Note that these are the problems for most machine learning based detection systems, not limited to the approach. We believe that VPA vendors are at a better position to implement such defense in a more effective way, leveraging the massive amount of data at their disposal to build a more precise system and continuing to adapt the defense strategies in response to the new tricks the adversary may play.

Future Scope

Although the analysis of Amazon and Google skill markets reveals some security risks (in terms of invocation name squatting), we have little idea whether VSA and VMA indeed take place in the real world for collecting sensitive user data, not to mention understanding about their pervasiveness and the damage they may cause. Answering these questions is non-trivial, due to the nature of the skill ecosystem. Each skill market today already hosts a very large number of skills and new ones continue to emerge every day, which makes manual inspection of each skill for malicious activities almost infeasible. Most importantly, a skill's inside logic is invisible to the VPA systems and the user, since they only have their interfaces (in the form of web APIs) registered in the markets by their developers, who implement and deploy the actual programs on their own servers. While this service model gives the developers more flexibility and helps them protect their proprietary code, it prevents a static analysis of skill code to detect malicious activities. Therefore, a potential future direction is to develop a lightweight and effective dynamic analysis system, such as a chatbot, to automatically invoke and communicate with skills, and capture their malicious behaviors during the conversations.

Chapter 8

Conclusion

In this report, the first security analysis of popular VPA ecosystems and their vulnerability to two new attacks, VSA and VMA, through which a remote adversary could impersonate VPA systems or other skills to steal user private information. These attacks are found to pose a realistic threat to VPA IoT systems, as evidenced by a series of user studies and real-world attacks we performed. To mitigate the threat, we developed a skill-name scanner and ran it against Amazon and Google skill markets, which leads to the discovery of a large number of Alexa skills at risk and problematic skill names already published, indicating that the attacks might already happen to tens of millions of VPA users. Further, we designed and implemented a context-sensitive detector to mitigate the voice masquerading threat, achieving a 95% precision.

With the importance of the findings reported by the study, we only made a first step towards fully understanding the security risks of VPA IoT systems and effectively mitigating such risks. Further research is needed to better protect the voice channel, authenticating the parties involved without undermining the usability of the VPA systems.

Chapter 9

References

[1] Alexa skills top 25,000 in the u.s. as new launches slow. https://techcrunch.com/2017/12/15/alexa-skills-top-25000-in-theu-s-as-new-launches-slow/.

[2] Alexa voice service. https://developer.amazon.com/alexa-voice-service.

[3] Amazon has 76% smart home speaker u.s. market share as echo unit sales reach 15m, new study finds. https: //www.geekwire.com/2017/amazon-75-smart-home-speaker-u-smarket-share-echo-unit-sales-reach-15m-new-study-finds/.

[4] Androidviewclient. https://github.com/dtmilano/AndroidViewClient/blob/ master/README.md.

[5] Arpabet. https://en.wikipedia.org/wiki/ARPABET.

[6] The cmu pronouncing dictionary. http://www.speech.cs.cmu.edu/cgi-bin/ cmudict.

[7] Demo. https://sites.google.com/site/voicevpasec/.

[8] Sleep and relaxation sounds. https://www.amazon.com/Voice-Apps-LLCRelaxation-Sounds/dp/B06XBXR97N.

[9] Ssml. https://www.w3.org/TR/speech-synthesis11/.

[10] AGTEN, P., JOOSEN, W., PIESSENS, F., AND NIKIFORAKIS, N. Seven months' worth of mistakes: A longitudinal study of typosquatting abuse. In Proceedings of the 22nd Network and Distributed System Security Symposium (NDSS 2015) (2015).

[11] BANNARD, C., AND CALLISON-BURCH, C. Paraphrasing with bilingual parallel corpora. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (2005), Association for Computational Linguistics, pp. 597–604.

[12] BOWMAN, S. R., ANGELI, G., POTTS, C., AND MANNING, C. D. A large annotated corpus for learning natural language inference. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2015), Association for Computational Linguistics.

[13] BUHRMESTER, M., KWANG, T., AND GOSLING, S. D. Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? Perspectives on Psychological Science 6, 1 (2011), 3–5. PMID: 26162106.

[14] CARLINI, N., MISHRA, P., VAIDYA, T., ZHANG, Y., SHERR, M., SHIELDS, C., WAGNER, D., AND ZHOU, W. Hidden voice commands. In 25th USENIX Security Symposium (USENIX Security 16) (Austin, TX, 2016), USENIX Association, pp. 513–530.

[15] CHEN, Q. A., QIAN, Z., AND MAO, Z. M. Peeking into your app without actually seeing it: UI state inference and novel android attacks. In 23rd USENIX Security Symposium (USENIX Security 14) (San Diego, CA, 2014).

[16] CONNEAU, A., KIELA, D., SCHWENK, H., BARRAULT, L., AND BORDES, A. Supervised learning of universal sentence representations from natural language inference data. arXiv preprint arXiv:1705.02364 (2017).

[17] DIAO, W., LIU, X., ZHOU, Z., AND ZHANG, K. Your voice assistant is mine: How to abuse speakers to steal information and control your phone. In Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices (2014).

[18] DOWNS, J. S., HOLBROOK, M. B., SHENG, S., AND CRANOR, L. F. Are your participants gaming the system?: Screening mechanical turk workers. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2010), CHI '10, ACM, pp. 2399–2402.

[19] EDELMAN, B. Large-scale registration of domains with typographical errors. https://cyber.harvard.edu/archived content/people/edelman/typodomains/, 2003.

[20] FELT, A. P., AND WAGNER, D. Phishing on mobile devices. 2011.

[21] FENG, H., FAWAZ, K., AND SHIN, K. G. Continuous authentication for voice assistants. In Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (2017).

[22] FERNANDES, E., CHEN, Q. A., PAUPORE, J., ESSL, G., HALDERMAN, J. A., MAO, Z. M., AND PRAKASH, A. Android ui deception revisited: Attacks and defenses. In International Conference on Financial Cryptography and Data Security (2016), Springer, pp. 41–59.

[23] FERNANDES, E., JUNG, J., AND PRAKASH, A. Security analysis of emerging smart home applications. In 2016 IEEE Symposium on Security and Privacy (SP) (2016).

[24] FERNANDES, E., PAUPORE, J., RAHMATI, A., SIMIONATO, D., CONTI, M., AND PRAKASH, A. Flowfence: Practical data protection for emerging iot application frameworks. In 25th USENIX Security Symposium (USENIX Security 16) (Austin, TX, 2016).

[25] HIXON, B., SCHNEIDER, E., AND EPSTEIN, S. L. Phonemic similarity metrics to compare pronunciation methods. In Twelfth Annual Conference of the International Speech Communication Association (2011).

[26] HO, G., LEUNG, D., MISHRA, P., HOSSEINI, A., SONG, D., AND WAGNER, D. Smart locks: Lessons for securing commodity internet of things devices. In Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security (2016).

[27] JANG, Y., SONG, C., CHUNG, S. P., WANG, T., AND LEE, W. A11y attacks: Exploiting accessibility in operating systems. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (2014).

[28] JIA, Y. J., CHEN, Q. A., WANG, S., RAHMATI, A., FERNANDES, E., MAO, Z. M., AND PRAKASH, A. Contexlot: Towards providing contextual integrity to appified iot platforms. In 24th Annual Network and Distributed System Security Symposium, NDSS 2017, San Diego, California, USA, February 26 - March 1, 2017 (2017).

[29] KANG, R., BROWN, S., DABBISH, L., AND KIESLER, S. Privacy attitudes of mechanical turk workers and the u.s. public. In 10th

Symposium On Usable Privacy and Security (SOUPS 2014) (Menlo Park, CA, 2014), USENIX Association, pp. 37–49.

[30] KASMI, C., AND ESTEVES, J. Iemi threats for information security: Remote command injection on modern smartphones.

[31] KHAN, M. T., HUO, X., LI, Z., AND KANICH, C. Every second counts: Quantifying the negative externalities of cybercrime via typosquatting. In 2015 IEEE Symposium on Security and Privacy (2015).

[32] KUMAR, A., GUPTA, A., CHAN, J., TUCKER, S., HOFFMEISTER, B., AND DREYER, M. Just ask: Building an architecture for extensible selfservice spoken language understanding. arXiv preprint arXiv:1711.00549 (2017).

[33] KUMAR, D., PACCAGNELLA, R., MURLEY, P., HENNENFENT, E., MASON, J., BATES, A., AND BAILEY, M.

[34] LI, T., WANG, X., ZHA, M., CHEN, K., WANG, X., XING, L., BAI, X., ZHANG, N., AND HAN, X. Unleashing the walking dead: Understanding cross-app remote infections on mobile webviews. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (2017).

[35] MALLINSON, J., SENNRICH, R., AND LAPATA, M. Paraphrasing revisited with neural machine translation. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers (2017), vol. 1, pp. 881–893.

[36] NIKIFORAKIS, N., BALDUZZI, M., DESMET, L., PIESSENS, F., AND JOOSEN, W. Soundsquatting: Uncovering the use of homophones in domain squatting. In Information Security (2014), S. S. M. Chow, J. Camenisch, L. C. K. Hui, and S. M. Yiu, Eds.

[37] PENNINGTON, J., SOCHER, R., AND MANNING, C. D. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP) (2014), pp. 1532–1543.

[38] PETRACCA, G., SUN, Y., JAEGER, T., AND ATAMLI, A. Audroid: Preventing attacks on audio channels in mobile devices. In Proceedings of the 31st Annual Computer Security Applications Conference (2015).

[39] PRAKASH, A., HASAN, S. A., LEE, K., DATLA, V., QADIR, A., LIU, J., AND FARRI, O. Neural paraphrase generation with stacked residual lstm networks. arXiv preprint arXiv:1610.03098 (2016).

[40] REN, C., ZHANG, Y., XUE, H., WEI, T., AND LIU, P. Towards discovering and understanding task hijacking in android. In 24th USENIX Security Symposium (USENIX Security 15) (Washington, D.C., 2015).

[41] RONEN, E., SHAMIR, A., WEINGARTEN, A. O., AND OFLYNN, C. Iot goes nuclear: Creating a zigbee chain reaction. In 2017 IEEE Symposium on Security and Privacy (SP) (2017).

[42] SHAHRIAR, H., KLINTIC, T., AND CLINCY, V. Mobile phishing attacks and mitigation techniques. Journal of Information Security 6, 03 (2015), 206.

[43] SIKDER, A. K., AKSU, H., AND ULUAGAC, A. S. 6thsense: A contextaware sensor-based attack detector for smart devices. In 26th USENIX Security Symposium (USENIX Security 17) (Vancouver, BC, 2017).

[44] SZURDI, J., KOCSO, B., CSEH, G., SPRING, J., FELEGYHAZI, M., AND KANICH, C. The long "taile" of typosquatting domain names. In 23rd USENIX Security Symposium (USENIX Security 14) (San Diego, CA, 2014).

[45] TIAN, Y., ZHANG, N., LIN, Y.-H., WANG, X., UR, B., GUO, X., AND TAGUE, P. Smartauth: User-centered authorization for the internet of things. In 26th USENIX Security Symposium (USENIX Security 17) (Vancouver, BC, 2017).

[46] VAIDYA, T., ZHANG, Y., SHERR, M., AND SHIELDS, C. Cocaine noodles: Exploiting the gap between human and machine speech recognition. In 9th USENIX Workshop on Offensive Technologies (WOOT 15) (Washington, D.C., 2015).

[47] YAO, K., AND ZWEIG, G. Sequence-to-sequence neural net models for grapheme-to-phoneme conversion. arXiv preprint arXiv:1506.00196 (2015).

[48] YUAN, X., CHEN, Y., ZHAO, Y., LONG, Y., LIU, X., CHEN, K., ZHANG, S., HUANG, H., WANG, X., AND GUNTER, C. A. Commandersong: A systematic approach for practical adversarial voice recognition. arXiv preprint arXiv:1801.08535 (2018).

[49] ZHANG, G., YAN, C., JI, X., ZHANG, T., ZHANG, T., AND XU, W. Dolphinattack: Inaudible voice commands. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (New York, NY, USA, 2017), CCS '17, ACM, pp. 103–117.

[50] ZHANG, L., TAN, S., AND YANG, J. Hearing your voice is not enough: An articulatory gesture-based liveness detection for voice authentication. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (2017).

[51] ZHANG, L., TAN, S., YANG, J., AND CHEN, Y. Voicelive: A phoneme localization-based liveness detection for voice authentication on smartphones. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (2016).

[52] ZHANG, XIANGHANG MI, XUAN FENG, XIAO FENG WANG, YUAN TIAN AND FENG QIAN. Dangerous Skills: Understanding and Mitigating Security Risks of Voice-Controlled Third-Party Functions on Virtual Personal Assistant Systems (2019)