"Alexa, Can I Trust You?"

Hyunji Chung, Michaela Iorga, and Jeffrey Voas, NIST
Sangjin Lee, Korea University

Several recent incidents highlight significant security and privacy risks associated with intelligent virtual assistants (IVAs). Better diagnostic testing of IVA ecosystems can reveal such vulnerabilities and lead to more trustworthy systems.

Intelligent virtual assistants (IVAs) have opened up a new world where you can ask a machine questions as if it's a human and request it to perform certain tasks. For example, upon waking up: "Hey, what's on my schedule for today?" Before you leave the house for work: "What's my commute time?" At dinner: "Have one large pepperoni pizza delivered from Luigi's." When you go to sleep: "Turn off the bedroom lights." Ideally, such interactions should be solely between you and the device assisting you. But are they? How do you know for sure?

IVAs are becoming increasingly popular: according to Gartner, the IVA market will reach $2.1 billion by 2020.[1] However, recent news reports have revealed that popular voice-activated assistants such as Google Home, Apple's Siri, and Amazon Alexa aren't always reliable or trustworthy.

For example, in January 2017, a 6-year-old Dallas girl sharing her love of dollhouses and cookies with the family's new Amazon Echo Dot prompted Alexa to order—much to her parents' surprise—a $160 KidKraft Sparkle Mansion and four pounds of sugar cookies. After reporting the story, the anchor of a San Diego TV morning show remarked, "I love the little girl saying 'Alexa ordered me a dollhouse.'" Several Echo owners watching the broadcast reported that, after hearing the anchor's comment, their own devices also tried to order pricey dollhouses.[2]

The following month, during the Super Bowl, a Google Home ad using the system's voice-search-activation phrase "OK, Google" reportedly set off many viewers' own devices.[3] Capitalizing on the incident, in April, Burger King ran an ad for the Whopper in which an actor playing an employee at one of its restaurants says that 15 seconds isn't enough time to describe the sandwich and instead asks Google, which cites the definition from Wikipedia—prompting viewers' devices to repeat the question and thus essentially extend the ad.[4] Ironically, after publically exploiting the system's vulnerability, the marketing stunt backfired—someone altered the Wikipedia entry for the product to say that it contained cyanide and caused

 Published by the IEEE Computer Society

EDITOR **JEFFREY VOAS**
NIST; j.voas@ieee.org

cancer[5]—and became a sobering lesson that a hijacked IVA could cause real harm.

Here we explore the nature of IVAs and some of the security and privacy concerns associated with this emerging technology. Are IVAs secure? Are they recording our conversations? If so, where is this voice data stored? The presence of IVAs in homes makes this a public-facing challenge, and one that attracts instant—and unwelcome—media attention when problems arise.

INTELLIGENT VIRTUAL ASSISTANTS

IVAs evolved from chatbots, software agents programmed to converse with humans through either text or voice (en.wikipedia.org/wiki/Chatbot). The first chatbot, ELIZA, was developed by Joseph Weizenbaum at MIT 16 years after Alan Turing first proposed his test of artificial intelligence in 1950. ELIZA used natural-language processing to recognize key words in typed input and generate pre-scripted responses that to some users resembled human understanding. PARRY, introduced in 1972 by psychiatrist Kenneth Colby, convinced a number of trained experts that it was a real person with paranoid schizophrenia.

Over time, chatbots such as Alice (the inspiration for the film *Her*), Jabberwacky, and Cleverbot incorporated increasingly sophisticated algorithms to create more natural and complex dialogue. Motivated by research indicating that most users prefer to interact with human-like programs, simple chatbots are now integrated in many phone systems and web applications for customer service, information retrieval, marketing, education, entertainment, and other purposes.

IVAs extend chatbot functionality to Internet of Things (IoT) devices. Thus, they respond to text and voice commands to answer questions, play music and videos, purchase items, make recommendations, provide directions, turn on lights, open garage doors, and so on (en.wikipedia.org/wiki/Virtual_assistant_(artificial_intelligence)). We use the term intelligent virtual assistant, but other names are also commonly used such as smart assistant, intelligent personal assistant, digital assistant, and personal virtual assistant. Regardless of the terminology, the system's "brain"—the intelligence that converts human voice to text, performs linguistic analysis, and carries out the requested action—is a cloud-hosted service; the devices themselves run agent programs and, whether communicating with the service by default or configured to do so, have no embedded intelligence.

IVAs can communicate with multiple compatible IoT devices running a supported OS. Siri works exclusively with Apple products—iPhone, iPad, iPod Touch, HomePod, Mac, Apple Watch, and Apple TV devices. Microsoft Cortana works with Windows 10, Android, Xbox One, Skype, iOS, Cyanogen, and Windows Mixed Reality devices. Alexa works with Amazon's Echo, Fire, and Dash product families and various smart devices running Android and iOS including smartphones, smart speakers and headphones, smartwatches, and smart-home devices including TVs, intercoms, lights, thermostats, and refrigerators. Google Assistant also works with Android and iOS devices. Bixby is a new IVA for Samsung products.

IVA ECOSYSTEMS

To understand IVAs' potential security and privacy threats, we performed cloud-native artifact analysis, packet analysis, voice-command tests, application analysis, and firmware analysis to better understand IVA ecosystems. As Figure 1 shows, such an ecosystem consists of three main components. On the cloud side is the IVA—the software that processes text and voice commands and carries out requested actions. There are two user-side components: IVA-enabled devices—for example, an Echo Dot (Alexa) or a PC running Windows 10 (Cortana)—and companion applications installed on the device that communicate with the IVA.

Requests sent to an IVA, whether in text format (for example, through online chat) or voice format, along with the system's responses are stored in the cloud. These user–IVA "conversations" are usually accessible through a companion app. Obviously, the content of such conversations could contain revealing details—for example, questions about health symptoms. However, user voice recordings themselves also pose a privacy risk because they constitute personally identifiable information—unauthorized entities could use such data to identify the user, maliciously obtain access to systems that implement voice recognition, or simply process data and construct voice artifacts that could be used to impersonate the user.[6]

IVA software can be integrated into IoT device operating systems—for example, the latest versions of iOS and OS X have the Siri agent installed by default, and Windows 10 has the Cortana agent as one of its default processes—or downloaded and installed on compatible devices. Many IVAs enable third-party vendors to link their devices and services to the intelligent assistant, dramatically expanding the IVA's features or "skills." For example, Alexa works with many smart-home devices from brands including ecobee, Philips Hue, Nest, Ring, and Leviton. It also integrates with numerous apps to order food (for example, Domino's Pizza and Wingstop), stream music and video (Pandora and Spotify), get a ride (Uber and Lyft), and check account balances

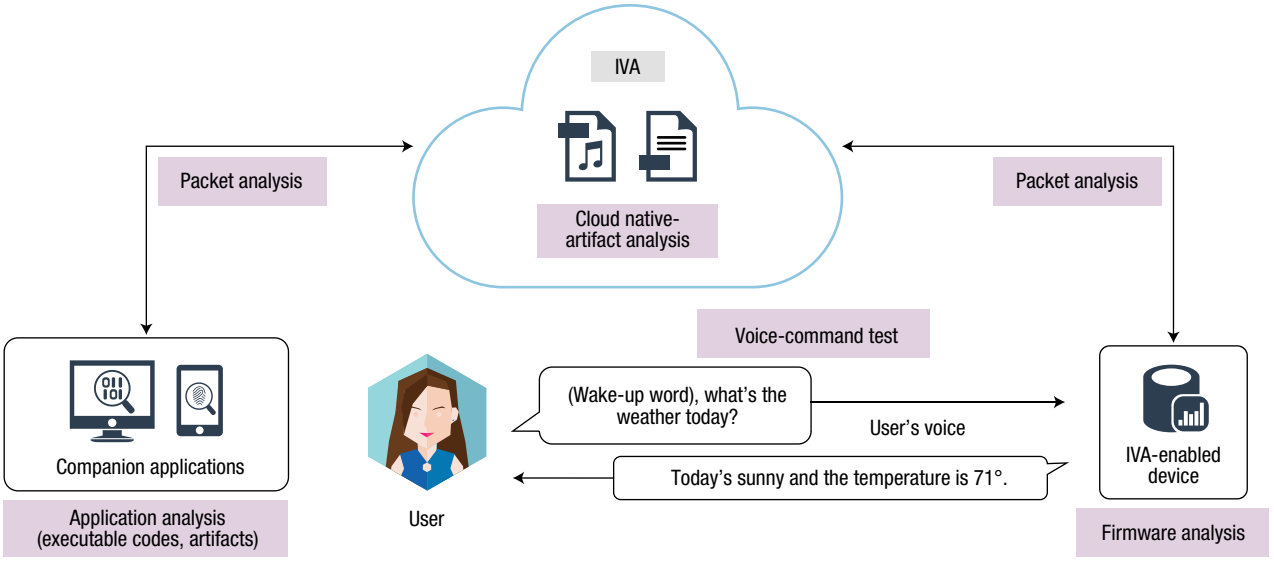

Figure 1. An intelligent virtual assistant (IVA) ecosystem has three main components: the cloud-based IVA, IVA-enabled devices, and companion applications.

and make credit card payments (Capital One). The Alexa Skills Store (www.alexaskillstore.com) currently lists more than 10,000 voice-activated apps.

IVA SECURITY AND PRIVACY RISKS

Given the large ecosystem of IVA-enabled devices and cloud-hosted services from IVA and third-party developers, Figure 2 illustrates four attack vectors that can put system security and user privacy at risk.

Wiretapping an IVA ecosystem

Even if companion apps use encrypted network connections, sniffing the traffic between the apps and the IVA can expose the ecosystem's communication mechanisms (left side of Figure 2a). For example, we used packet interception tools to analyze HTTPS requests and responses and then determine which APIs are used for sending and receiving data to and from the IVA.

In the case of communication between IVA-enabled devices and cloud-hosted services, our analysis revealed that not all network traffic is transmitted over a secure protocol (right side of Figure 2a). For example, many devices don't use encrypted connections to check network connectivity, making it possible to detect IVA devices in a home network. Firmware image data might also be transferred over unencrypted packets, exposing the system to man-in-the-middle attacks and possible malicious modification of images. Even if firmware images aren't altered, the ability to obtain them is a security concern because it provides unauthorized entities a chance to understand an IVA-enabled device's internal operations.[7]

Most communication between IVA-enabled devices and the IVA is encrypted using HTTPS. However, various machine-learning techniques to classify network traffic can still reveal payload sizes, data rates, and other patterns in encrypted traffic that could be used to identify the device's status—for example, idle or in use—or the user's behavior such as turning the device on or off, talking to the assistant, listening to music, and ordering products or services.[8,9]

Compromised IVA-enabled devices

Because IVA-enabled devices are part of the IoT, devices with security vulnerabilities can be compromised like any other computing system connected to the Internet and exploited for nefarious purposes such as distributed denial-of-service (DDoS) attacks. For example, in October 2016, a DDoS attack against the Internet performance management company Dyn exploited vulnerabilities in tens of millions of home IoT devices such as webcams and DVRs to infect them with the Mirai malware and use them as part of a botnet to temporarily cripple Dyn's networks.[10]

Figure 2b shows how a hacker could compromise an IVA-enabled device through its "always on" listening capability, enabling the hacker to monitor all voices and sounds within the device's range in real time. This danger was highlighted by a disturbing incident in Washington State in April 2015, when parents discovered that a stranger had hacked into their three-year-old son's baby monitor by obtaining the companion app's login credentials and was speaking to him at night through the device's speaker as well as operating its camera.[11] Theoretically, an attacker could also remotely control an IVA by talking to the system through another compromised device in the home, such as a smart speaker or intercom.

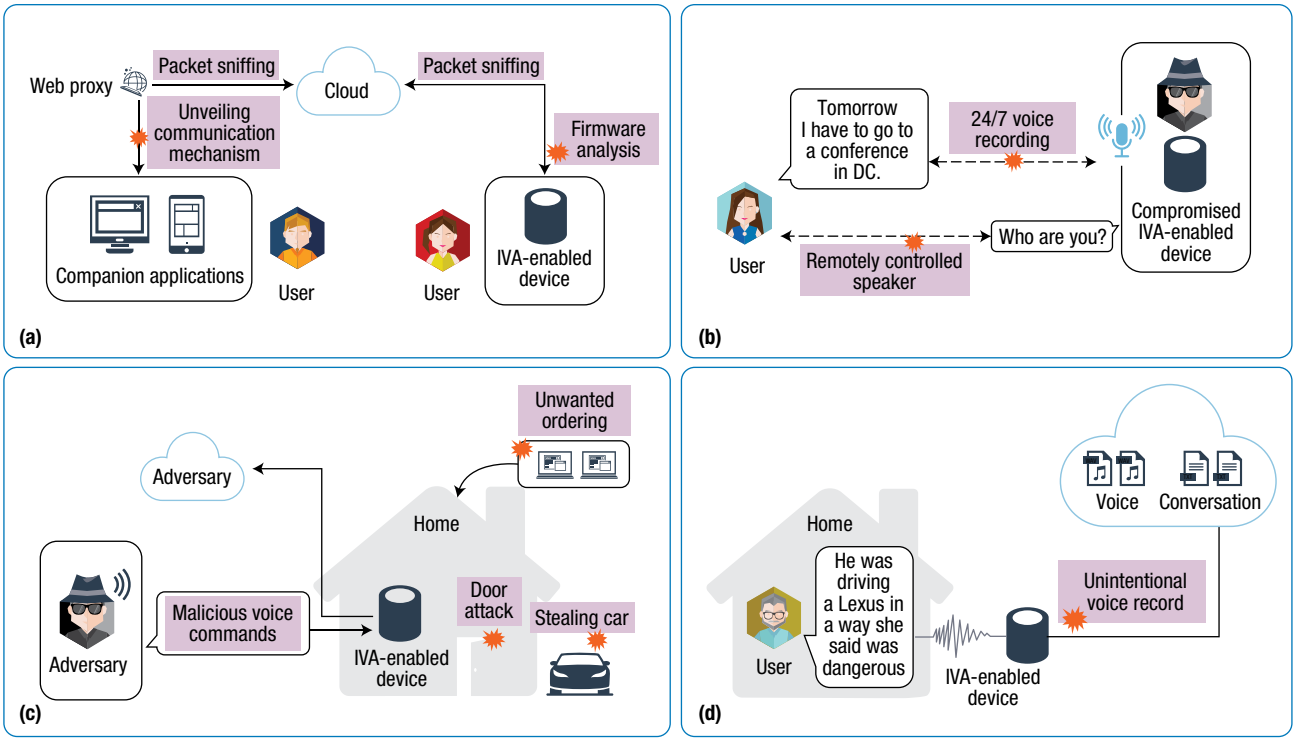

Figure 2. IVA security and privacy risks: (a) wiretapping an IVA ecosystem, (b) compromised IVA-enabled devices, (c) malicious voice commands, and (d) unintentional voice recording.

Malicious voice commands

Figure 2c depicts a third security and privacy risk associated with IVAs: an attacker who impersonates a user and issues malicious voice commands to, for example, unlock a smart door to gain unauthorized entry to a home or garage or order items online without the user's knowledge. Although some IVAs provide a voice-training feature to prevent such impersonation, it can be difficult for the system to distinguish between similar voices. Thus, a malicious person who is able to access an IVA-enabled device might be able to fool the system into thinking that he or she is the real owner and carry out criminal or mischievous acts.

Unintentional voice recording

Finally, as Figure 2d shows, voices within range of an IVA-enabled device can be recorded accidentally and transmitted to the cloud, enabling other parties—including commercial entities with legitimate access to the stored data as well as hackers who might break into the database—to eavesdrop on private conversations. The potential for accidental recording means that users don't necessarily have complete control over their voice data.[12]

As virtual assistants become more intelligent and the IVA ecosystem of services and devices expands, there's a growing need to understand the security and privacy threats from this emerging technology. Several recent incidents highlight significant vulnerabilities in IVAs. Better diagnostic testing can reveal such vulnerabilities and lead to more trustworthy systems.

DISCLAIMER

Certain commercial entities, equipment, or materials identified in this document were used only to adequately describe an experimental procedure or concept. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

REFERENCES

1. "Gartner Says Worldwide Spending on VPA-Enabled Wireless Speakers Will Top $2 Billion by 2020," press release, Gartner, 3 Oct. 2016; www.gartner.com/newsroom/id/3464317.
2. A. Liptak, "Amazon's Alexa Started Ordering People Dollhouses after Hearing Its Name on TV," *The Verge*, 7 Jan. 2017; www.theverge.com/2017/1/7/14200210/amazon-alexa-tech-news-anchor-order-dollhouse.
3. K. Opam, "Google's Super Bowl Ad Accidentally Set off a Lot of Google Homes," *The Verge*, 5 Feb. 2017; www.theverge.com/2017/2/5/14517314/google-home-super-bowl-ad-2017.
4. M. Anderson, "How Burger King Revealed the Hackability of Voice Assistants," Associated Press, 5 May

2017; bigstory.ap.org/2d8036d742504890b2f9edc3f98c77ef.
5. Z. Rodionova, "Burger King Ad Backfires after Asking Google What's in a Whopper and Is told 'Cyanide,'" *The Independent*, 13 Apr. 2017; www.independent.co.uk/news/business/news/burger-king-advert-ask-google-big-whopper-cyanide-cancer-causing-wikipedia-page-us-a7681561.html.
6. E. McCallister, T. Grance, and K. Scarfone, *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*, Special Publication 800-122, NIST, Apr. 2010.
7. "Exploring the Amazon Echo Dot, Part 1: Intercepting Firmware Updates," 2 Jan. 2017; medium.com/@micaksica/exploring-the-amazon-echo-dot-part-1-intercepting-firmware-updates-c7e0f9408b59.
8. T.T.T. Nguyen and G. Armitage, "A Survey of Techniques for Internet Traffic Classification Using Machine Learning," *IEEE Comm. Surveys & Tutorials*, vol. 10, no. 4, 2008, pp. 56–76.
9. C. Gu, S. Zhang, and Y. Sun, "Real-Time Encrypted Traffic Identification Using Machine Learning," *J. Software*, vol. 6, no. 6, 2011, pp. 1009–1016.
10. K. York, "Dyn Statement on 10/21/2016 DDoS Attack," blog, 22 Oct. 2016; dyn.com/blog/dyn-statement-on-10212016-ddos-attack.
11. C. Owens, "Stranger Hacks Family's Baby Monitor and Talks to Child at Night," *The San Francisco Globe*, 3 Nov. 2016; sfglobe.com/2016/01/06/stranger-hacks-familys-baby-monitor-and-talks-to-child-at-night.
12. C. Wood, "Devices Sprout Ears: What Do Alexa and Siri Mean for Privacy?," *The Christian Science Monitor*, 14 Jan. 2017; www.csmonitor.com/Technology/2017/0114/Devices-sprout-ears-What-do-Alexa-and-Siri-mean-for-privacy.

This article originally appeared in Computer, *vol. 50, no. 9, 2017.*

HYUNJI CHUNG is a PhD candidate at the Graduate School of Information Security at Korea University and a guest researcher in NIST's Computer Security Division. Contact her at hyunji.chung@nist.gov.

MICHAELA IORGA is the senior security technical lead for cloud computing at NIST and cochair of its Cloud Computing Security and Cloud Computing Forensic Science working groups. Contact her at michaela.iorga@nist.gov.

JEFFREY VOAS is an IEEE Fellow and computer scientist at NIST. Contact him at j.voas@ieee.org.

SANGJIN LEE is a professor in the Graduate School of Information Security and director of the Digital Forensics Research Center at Korea University. Contact him at sangjin@korea.ac.kr.