DOI:10.1145/3396261

Connor Bolton, Kevin Fu, Josiah Hester, and Jun Han

▸ **Peter G. Neumann,** Column Editor

Inside Risks

How to Curtail Oversensing in the Home

Limiting sensitive information leakage via smart-home sensor data.

Future homes will employ potentially hundreds of Internet of Things (IoT) devices whose sensors may inadvertently leak sensitive information. A previous *Communications* Inside Risks column ("The Future of the Internet of Things," Feb. 2017) discusses how the expected scale of the IoT introduces threats that require considerations and mitigations.[2] Future homes are an IoT hotspot that will be particularly at risk. Sensitive information such as passwords, identification, and financial transactions are abundant in the home—as are sensor systems such as digital assistants, smartphones, and interactive home appliances that may unintentionally capture this sensitive information. IoT device manufacturers should employ sensor sensor permissioning systems to limit applications access to only sensor data required for operation, reducing the risk that malicious applications may gain sensitive information. For example, a simple notepad application should not have microphone access. However, even if this least-privilege approach was enacted across all IoT systems (a difficult task), sensor systems still gather much more information than intended or required by an application—for example, how motion sensors can capture nearby sounds, including words and keystrokes. We call this *oversensing*: where authorized access to sensor data provides an application with superfluous and potentially sensitive information. Manufacturers and system designers must employ the principle of least privilege at a more fine-grained level and with awareness of how often different sensors overlap in the sensitive information they leak. We project that directing technical efforts toward a more holistic conception of sensor data in system design and permissioning will reduce risks of oversensing.

Risks of Oversensing

Oversensing unintentionally leaks potentially sensitive information, such as login information, user location, and identification information, through sensor data. Smart-device manufacturers prevent malicious applications from trivially obtaining such information through sensor permissioning systems, attempting to limit applications to only necessary sensors. However, oversensing subverts permissioning systems; one sensor's data may allow an adversary to access sensitive information that should require a different permission. For example, motion sensor access may provide information on nearby speech. Intelligent attackers may then covertly decipher the sensitive information using tools such as machine learning.

We see primarily two mechanisms of oversensing: first, individual sensors provide information outside of their normal permission context or more information than necessary for the application goal, and second, synthesis of multiple sensor's data to reveal new information that is beyond the scope of existing permissions. A common source of individual oversensing involves cases where a sensor converts physical stimuli not explicitly specified in its de-

sign. For example, speech should not affect accelerometer output; however, deficiencies in signal processing such as aliasing and ineffective low-pass filters mix speech content with measured acceleration. Thus, an application with permission to use motion sensor data may also receive some speech information. Additionally, we regard a sensor providing more information than required by an application to also be oversensing. For example, consider an application that requests unfettered camera access to count the number of human occupants in a room. The application needs only the number of occupants, but receives additional unnecessary, potentially sensitive information from the video. Furthermore, attackers may synthesize multiple co-located sensors to reveal emergent information that would be outside any of the original sensor permissions. For example, access to multiple co-located accelerometers, geophones, microphones, or gyroscopes may enable an adversary to triangulate events in that room, such as keystrokes.

CREDIT TK

Vulnerabilities

Applications will soon have access to networks of sensors within smart-homes, with research showing that many of those common sensor systems already leak information via oversensing. Vulnerabilities may reside in hardware or software design. Sensor hardware designed to sense one stimulus may allow other physical stimuli to affect measurement output. Even minute alterations in output may result in privacy complications, as adversaries may use advancements in machine learning to extract private information without significant engineering. Permissioning system design may allow applications access to more sensed information than required. The research community demonstrates several examples of how a malicious application may masquerade as a benign application can receive sensor access, then use oversensed information from that sensor to access information from an elevated privilege. For example, an application could masquerade as a game to gain a motion sensor privilege, and then use that motion sensor privilege for eavesdropping on nearby speech, which should be a separate privilege.

Illustrative Oversensing Risks

It is no surprise that adversaries with sensor access obtain information exceeding permissions or specifications, since current research in activity recognition and synthetic sensors do this by design. Selected sensors and examples of how access to each can lead to unintended consequences include the following schemes:

Motion Sensors. Adversaries with motion sensor permissions can infer sensitive information such as nearby speech, keystrokes, and location.[1] Most motion sensors are designed to capture signals with significantly lower frequency than speech, indicating that speech would not be sensed. However, typical digital filters may not eliminate all traces of higher-frequency information as needed for privacy and security. Many filters are designed to ensure the desired signal is distinct, meaning sensitive information may remain in the filtered signal in an altered state. The

very presence of this sensitive information, even altered, may be all that is necessary for an attack to render the defense ineffective.

Microphones. Recent research shows how adversaries may employ acoustic data from microphones, a common sensor permission, to create a copy of a physical key and infiltrate a home.[4] This work enables a nearby attacker to utilize a smartphone microphone to capture the emitted sound when a victim inserts a key. This is possible because the time interval between audible clicks is correlated to the key's cut depths. This is a prime example of oversensing, as users and manufacturers would not expect that an inherent acoustic noise would enable a nearby attacker to infer one's key secrets, and eventually create a duplicate key to gain access to their house.

Power Consumption. While access to device power consumption data may seem innocuous, power consumption may leak sensitive information such as location.[3] Given prior knowledge of the area, the power draw of the smartphone can be used as a proxy for the location of the phone with some prior knowledge of the area. Essentially, wireless connections such as 4G, WiFi, and so forth, require a variable power draw with signal strength. Then with a bit of machine learning, the changes within the rate of consumption of the power draw can be used to discern the relative location of the smartphone and its owner. Users are unlikely to anticipate the mismatch between the permission granted (power consumption) and what an attacker may acheive (location inference).

Yet, these examples of individual sensor oversensing are likely only the beginning of risks, as future adversaries may synthesize sensor data to achieve further capability. "Synthetic sensors"—a combination of different sensors that synergize to enable new activity recognition capabilities—have shown to be effective in non-security and privacy settings. These typically use common sensors that are included in a variety of smart-devices. In the context of security, this suggests adversaries may be able to use multiple permissions to obtain access to these same sensors, and then create synthetic sensors of their own. Moreover, adversary capability to exploit oversensing will grow as machine learning improves. Sensor-related research such as activity recognition and synthetic sensing continue to improve largely due to advancements in machine learning. This benign research may enable more sophisticated attacks through the creation of adversarial synthetic sensors. Additionally, usability improvements to machine learning toolkits allow non-expert to easily train complex models. From this we conclude that these attacks will only grow in capability.

Each of the aforementioned oversampling examples is currently difficult to detect. Adversaries have two options: to process data onboard the device or send it back for offline processing. With offline processing there is little a defender may do to detect privacy violations, however even with onboard processing the black box nature of machine learning makes discerning each application's purpose for sensor data difficult. Thus, adversarial applications may easily masquerade as benign.

Confronting Oversensing

Fixing oversensing is difficult, as one must minimally affect benign applications while balancing the effectiveness of the mitigation with ease of deployment. Higher-capability sensors are desirable to benign application designers; therefore, mitigations are burdened with maximal preservation of the original sensing characteristics. Deployment of any mitigation is further complicated by the ubiquity of devices already in existence with many different operating systems, hardware suppliers, purposes, and protocols. Software mitigations may not easily fix an oversensing problem that is essentially baked in on manufacture or accidentally by design. Yet, hardware solutions can be tricky to design and cannot be easily updated after deployment. Despite the difficulty, there are steps taken from established fields—such as operating system design and cryptography—that researchers, operating system designers, and manufacturers can employ to begin to confront oversensing.

Principle of Least Privilege. Manufacturers and operating system designers should further integrate the operating systems concept of the principle of least privilege (POLP) into sensor system design. Often applications receive more information from sensor data than needed for full functionality. For example, an application that receives a raw camera video stream to detect presence or count people in a room is overprivileged for the task. The application's only required information is the number of people in a room, yet it will have access to additional information such as facial data, documents, or anything else the camera captures.

Software mitigations may not easily fix an oversensing problem that is essentially baked in on manufacture or accidentally by design.

To employ POLP for sensors, permissioning systems must err on the side of fine-grain permissions and be far more strict in giving full sensor access, even for seemingly innocuous sensors. Taking the preceding example, the optimal solution is if a "number of people on camera" permission were available. However, providing such individualized permissions for the near limitless number of applications may be difficult. Thus, a balance between fine grain and coarse grain will be needed for practical deployment, but should err on the side of fine grain permissions. Systems should minimize the number of applications that need permissions to raw data. Common machine learning features should be supported as a permission.

Development of application specific hardware may enable a realistic, fine-grain sensor permissions that could serve as the basis for practical POLP in sensor systems. A sensor processing chip could provide a large variety of common sensing-based calculations while simultaneously easing the computational burden on the processor. More infrequent permissions could be supported by operating systems as these could be calculated on demand in software.

Design Patterns. Hardware and software design patterns for oversensing resiliency should be heavily researched and adopted industry-wide. Similar to cryptography, implementing oversensing mitigations can be technically challenging, since small implementation details can render mitigations effectively useless. The difficulty lies in how an attacker goal may be achievable using only minute traces of unintended data remaining in a signal. Furthermore, determining if this is the case can be difficult, as tools to measure oversensing do not currently exist.

This new context demands that well established fixes proposed purely for performance must be carefully applied to not impact privacy. For example, consider the removal of speech information from motion sensor data. A commonly suggested method to remove the speech from the motion sensor data is to use low-pass filtering, which has existed since before the transistor. However, digital filters will encounter the problem of aliasing, which will leave traces of the original speech signal—*even digital anti-aliasing filters included on some motion sensor chips*. These filters are designed for performance, not privacy. As such they are designed to attenuate the signal so that legitimate motion is clear, but they are *not* designed to completely remove traces of all other signals. Methods such as machine learning may be able to discern remaining traces, rendering the defense ineffective.

Lessons from cryptography may inspire successful mitigation design patterns. Adding randomness to sensor output may render minute traces of sensitive information unidentifiable. Another idea may be to use non-linear data transforms on sensor data before providing it to applications, such as the aforementioned processed data for specific privileges. Dedicated sensor processing hardware may alleviate the difficulty of efficiently and verifiably implementing these design patterns—similar to dedicated crypto-hardware.

Thoughts for the Future

Anticipating Standards and Regulatory Oversight. Economic and market incentives, consumer outrage, technical difficulty for mitigation implementation, coupled with political forces, make it likely that some level of regulatory oversight or at least industry standards will happen; how should we as a community respond to this? If history is any indication, we ignore the effect of market incentives and consumer outrage in creating regulatory frameworks at our peril. The General Data Protection Regulation (GDPR) and similar centralized privacy legislation require protection of user data, with companies having to drastically change storage guidelines and collection policy to be in compliance. In the U.S., YouTube, TikTok, and others ran afoul of the Children's Online Privacy Protection Act (COPPA) by collecting data on minors; they were fined, and had to institute software changes. IoT devices are new enough to have escaped significant regulatory scrutiny in this oversensing realm, but a parade of works like those cited in this column (and subsequent coverage in the media) have increased user awareness and market incentives for changes and standardization. As a word of caution, regulation and oversight for IoT privacy is inevitable; its coming will be made less disruptive and more beneficial should researchers, manufacturers, and developers aid the process ahead of the attacks.

Additionally, there are technical benefits for industry-wide standards or oversight due to the difficulty in oversensing mitigation implementations and the vast heterogeneity and rapid development of the IoT landscape. As previously stated, technical solutions to oversensing are challenging in a similar manner to cryptography implementations. A single, minute flaw can lead to wholly ineffective defenses in both cases, and an incredibly diverse set of hardware and applications must be able to use the systems correctly. We should learn from the more established field of cryptography, and similarly discourage in-house solutions to this deceptively technically difficult problem.

Similarly, regulation can aid users in understanding privacy concerns and reducing safety and security risks posed by the diverse set of applications strewn through the IoT landscape. There is a need for formalized oversensing risks, sensor permissions, protocols, and related language that is universal for all

Calendar of Events

VOLUTPAT ORNARE ARCU

Donec sit amet neque nec odio pharetra semper. Suspendisse dictum ligula eu diam. Pellentesque convallis porttitor eros. Nunc placerat accumsan ante. Etiam scelerisque nisl non ligula. Quisque vitae lacus. Pellentesque in augue. Integer laoreet nisl nec ipsum. Ut massa orci molestie quis, blandit et cursus et lorem. Donec congue massa quis metus.

DONEC EU MAGNA

Nunc aliquet ante eget lectus. Vestibulum scelerisque dignissim nisi. Phasellus id elit suspendisse aliquet. Aenean semper, magna quis interdum sagittis, arcu odio tincidunt lacus, non tristique diam arcu sed nibh. Vestibulum non eros vitae dolor dignissim volutpat.

Suspendisse elementum, felis vel hendrerit congue, neque urna consectetuer nisl, ac vehicula nisi leo id arcu. Aenean aliquam. Sed suscipit. Quisque semper justo sed leo. Aenean porta, diam non pellentesque pulvinar, ipsum orci ultrices dui, in elementum velit mauris sit amet dolor.

PELLENTESQUE ERAT

Vitae dui semper fermentum. Fusce pede mauris, rutrum at, ullamcorper porta, ultrices ac, felis. Integer nunc enim, bibendum quis, ullamcorper nec, dictum sed, lorem. Morbi lacinia felis vitae massa. Nam tortor magna posuere, adipiscing ac, tincidunt eu, lectus. Nulla tortor nisi, sodales non, luctus non, posuere at, ante. Suspendisse adipiscing sem mollis mi. Duis lobortis commodo orci.

ODIO SED TORTOR

Interdum mollis. Maecenas lobortis, tellus sed mollis nonummy, sapien ante aliquet tellus, et sagittis lacus dolor eu sem. Quisque ut turpis nec risus molestie scelerisque. Nulla placerat. Curabitur sollicitudin quam ut risus.

Mauris aliquet, felis imperdiet adipiscing imperdiet, purus dolor sollicitudin felis, vel convallis ligula lorem scelerisque lorem. Nunc pellentesque. Cras nec lacus. Aenean suscipit sem a orci. Donec feugiat augue at

Sensor systems may leak sensitive information unwittingly by oversensing.

IoT (including smartphones). Each standard permission should be associated with an agency-assigned level of risk, with known attacks and other information made available in a singular public database. This level of risk should be made clear to users in a universal manner proven to be effective by research and case studies across manufacturers and devices. Standardization and regulation is needed as each cognitive difference between systems or methods of displaying information can distract and mislead users.

Developing Tools. There is a need for applications, programs, hardware, and other tools to assist manufacturers, application designers, and users against oversensing. Sensor manufacturers need methods to detect, measure, and otherwise test oversensing to develop mitigations against oversensing threats. For example, a hardware tool that systematically applies a variety of physical stimuli to a sensor and measures the response enables oversensing detection and measurement. Application designers need transparent APIs that give minimal privileges for any given task. Coupling a variety of commonly needed processed sensor data to the APIs may persuade designers to use such APIs instead of creating their own solution. User-focused tools are also needed. For example, tools that automatically detect oversensing risks and make such risks transparent may help users avoid overprivileged applications. High-level optional "safe-sensing" operating modes may give an easily understandable choice between power consumption and privacy to users. The development of these tools, even if imperfect, is needed to enable better mitigation design, more accurate design patterns, and provide higher-quality information for regulatory practices.

Conclusion

Sensor systems may leak sensitive information unwittingly by oversensing. Future homes are particularly at risk due to the abundance of private information coupled with increasing numbers of IoT sensing systems. Sensor and permissioning systems must better implement the principle of least privilege using far more fine-grained permissions. IoT applications should rarely receive raw sensor data, and instead should receive processed information specific to their actual need. The design of privacy-aware hardware accelerators that calculate this fine-grain sensor data may reduce privacy risk while simultaneously easing this computational burden of fine-grain permissions. Both hardware and software should follow standard design patterns to avoid easy-to-make mistakes, similarly to how in-house cryptography implementations are discouraged. With cooperation between academia, manufacturers, and regulatory agencies, the looming threat of oversensing may be mitigated before the problem grows uncontrollably. C

References

1. Ba, Z. et al. Learning-based practical smartphone eavesdropping with built-in accelerometer. In *Proceedings of the Network and Distributed Systems Security (NDSS) Symposium* (Feb. 23–26, 2020, San Diego, CA, USA); DOI: 10.14722/ndss.2020.24076
2. Lindqvist, U. and Neumann, P.G. The future of the Internet of Things. *Commun. ACM 60*, 2 (Feb. 2017), 26–30; DOI: 10.1145/3029589.
3. Michalevsky, Y. et al. PowerSpy: Location tracking using mobile device power analysis. In *Proceedings of the 24th USENIX Security Conference*. (Aug. 12–14, 2015, Washington, D.C., USA).
4. Ramesh, S., Ramprasad, H., and Han, J. Listen to your key: Towards acoustics-based physical key inference. In *Proceedings of HotMobile '20*, (Mar. 3–4, 2020, Austin, TX, USA), ACM; DOI: 10.1145/3376897.3377853

Connor Bolton (mcbolto@umich.edu) is a Ph.D. Candidate of Computer Science at the University of Michigan, Ann Arbor, MI, USA.

Kevin Fu (kevinfu@umich.edu) is an Associate Professor of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, MI, USA.

Josiah Hester (josiah@northwesten.edu) is an Assistant Professor of Electrical and Computer Engineering and Computer Science at Northwestern University, Evanston, IL, USA.

Jun Han (junhan@comp.nus.edu.sg) is an Assistant Professor of Computer Science at the National University of Singapore.

This research supported in part by a gift from Analog Devices, Inc.