Security and Privacy in Speech Communication

Technological Perspective

Tom Bäckström

Aalto University

Fall 2021

Speech-Privacy in the News

… week 32, 2019

The Washington Post — Democracy Dies in Darkness

'Alexa, delete what I just said': How to manage voice recordings on your smart devices

THE VERGE

Microsoft contractors are listening to select Skype calls and Cortana recordings

Microsoft does not explicitly say humans listen to the recordings

BUSINESS INSIDER

A new lawsuit accuses Apple of violating user's privacy by allegedly allowing Siri to record without consent

Lisa Eadicicco Aug. 8, 2019, 10:05 AM

TheStreet

Will Privacy Concern Slow Progress of Amazon's Alexa and Other Voice Assistants?

By Annie Gaus

CNBC — TECH

Amazon's Alexa comes under scrutiny of Luxembourg privacy watchdog

PUBLISHED FRI, AUG 9 2019 • 1:55 AM EDT

ADWEEK — VOICE

How Chatbots and Voice Assistants Will Be Affected by Data Privacy Laws

There are legal, business and ethical concerns

By Audrey Wu | August 7, 2019

TC

Amazon's lead EU data regulator is asking questions about Alexa privacy

Natasha Lomas @riptari / 6 days ago

TECH TIMES

Companies Listening To Voice Assistant Recordings: How To Stop It From Your Device

By Athena Chan Tech Times

The Tribune

Life Style

Posted at: Aug 12, 2019, 8:45 AM; last updated: Aug 12, 2019, 8:45 AM (IST)

Alexa, are you listening?

While it's a good thing that new age home assistants are listening to our commands, but are they listening to our conversations too? We read the fine print

cnet — SECURITY

Alexa privacy fears spark questions for Amazon in Europe

A European privacy watchdog is taking a keen interest in what happens to personal data collected through Alexa.

BY KATIE COLLINS | AUGUST 9, 2019 10:36 AM PDT

Happed previously

- Social media had Cambridge Analytica.
- Speech operated devices and services:
  - NSA/CIA eavesdrops non-US calls on Skype (caused a raid at the Brazilian home of journalist who covered Edward Snowden).
  - Amazon, Google, Microsoft and Apple had employees listen to conversations of their smart speakers.
  - Amazon smart speaker has called and transmitted all local conversations to a random person.
  - Recordings of smart devices have been used to catch criminals.
  - Smart devices have automatically called emergency services.
  - You can eavesdrop on your room-mates by browsing through their voice history of shared device (through your phone, even when you are not at home).
  - etc…

Motivation

- Speech operated devices have not yet had their Cambridge Analytica.
  - Can we fix privacy before it happens?
- European Union has introduced legislation, the General Data Protection Regulation (GDPR).
  - Partial solution to real problem.
  - Does not state specifics.
  - Applicable only within the EU.
- The research community has started to address the issue.
  - ISCA Special Interest Group "*Security and Privacy in Speech Communication*". `spsc-sig.org`

Definitions

- *Privacy* = Free from public attention
- *Security* = Free from threat or danger

⇒ The two concepts are so close to each other that it usually best to always consider them together.

- More detailed definition is very difficult.
    - Leads to a philosophical discussion about ethics and morals.

Information content in speech

- Literal, intended text content
- Accent; geographical, ethnic and cultural background (conscious?)
- Gender and gender identity (conscious?)
- Health (conscious and unconscious!)
- Other?
- Unconscious choices of words
- Speaking style (conscious and unconscious)
- Emotion (conscious and unconscious)
- Speaker identity
- Age
- Environment (background noise and reverberation)

Information content in speech 2

- Speaking partner (Individual info of both)
- Relationship between speakers
- Power structure between speakers
- Level of intimacy/distance
- Level of familiarity
- Level of match (differential) in reference groups
- Level of privacy in conversation
- Importance of topic for relationship
- Other?

Possible exposure

- Intended recipient (aware)
- Unintended but inconsequential recipient (aware and unaware)
  - E.g. person at the next table at the cafe, during casual conversation
- Undesirable recipient (unaware), unintentional listening
  - E.g. person at the next table at the cafe, during *private* conversation
- Undesirable recipient (unaware), intentional listening = malicious eavesdropping
  - E.g. hiding to overhear conversation, or secretly recording/analyzing conversations in the cloud
- Unintended but beneficial recipient, intentional listening = Public good
  - E.g. law enforcement, security monitoring (detect shouting, fire, glass breaking, person falling etc.)

Type of information and exposure

- Both lists of information and exposure types are open-ended.
- The number of combinations with information types and exposure types is large! (At least 100)
  - Hard for user to keep track of everything.
- With human discussion partner:
  - Well-developed culture and habits which dictates behavior, i.e. how to act such that level of privacy is reasonable.
- Machine-in-the-loop:
  - Intuition does not work; we do not have a pre-existing culture wrt privacy, which takes machine into account.
  - None of us have a clear picture of the risks or consequences, wrt privacy.

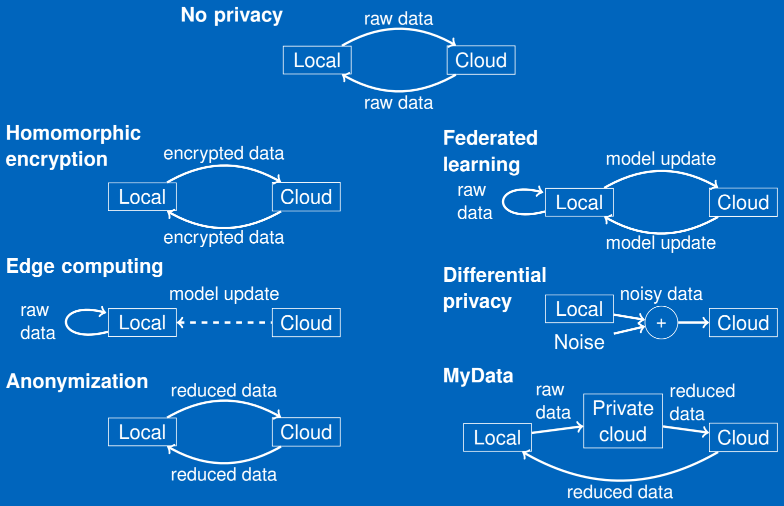
Basic principles

- *Control* – User can at any time choose level of privacy.
- *Transparency* – Level of privacy can be easily observed and checked. Changes in privacy have to be notified.
- *Privacy by design* – Privacy is the default and the system is built ground-up such that it takes privacy into account.
- *Usable privacy* – Reasonable expectations of privacy should not make service unusable. $\Rightarrow$ Privacy is about usability.
  - Every service requires some level of information transfer =leak of information.
  - Service design should cover also privacy.

Activities in Privacy and Security

- *Encryption* – The obvious: Always demand end-to-end and at-rest encryption of all your communication services.
- *Privacy-preserving computations* – Using a cloud services does not mean that you have to reveal all your information (homomorphic encyrption).
- *Federated learning* – Learning in the cloud is possible without leaking private information.
- *Anonymization* – Extract only the information needed and remove everything else.
- *Differential privacy* – Add noise to conceal individuals, but such that ensemble statistics can be deduced.
- *MyData* – All data is stored in private storage (can be cloud) separate from service providers, who must request access when needed.

Activities in Privacy and Security



Activities in Privacy and Security

- *Speaker recognition, -verification, spoofing and voice conversion* – Identify who is speaking and how to hide that.
- *Experience of privacy* – The study of how people perceive the level of privacy in human-to-human communication.
- *Acoustic fingerprint for authentication* – Enable authentication based on physical environment

Encryption

- Should be standard by now! It isn't.
- In-transit; weakest link can reveal all your conversations.
- At-rest; mass-storage is an attractive target for criminals.
- Governments often propose backdoor access;
  - Sooner or later, backdoor-key will be leaked to criminals.
  - = Everyone is exposed, but we have an illusion of security.
  - $\Rightarrow$ Illusion of security is worse than insecurity.
  - You might trust your own government, but do you really trust all other governments as well?
- Meta-data is also sensitive;
  - Frequent calls to a pregnancy clinic or home violence counseling are rather revealing!
- Every lock can be broken with sufficient effort.
  - $\Rightarrow$ Encryption should be treated as a *sufficient* roadblock.

Privacy-preserving computations

- Problem scenario:
  - You do not want to share your data with service provider.
  - Service provider does not want to share model with you.
  - How can we use model on data, when neither trusts one-another?
- The idea of privacy-preserving computations:
  1. Encrypt data on trusted device.
  2. Transmit to untrusted device.
  3. Apply secret processing on encrypted data.
  4. Transmit back to trusted device.
  5. Decrypt processed data on trusted device.

Privacy-preserving computations

- Solution: Homomorphic encryption
  - Enables computations on encrypted signal.
  - Allows only polynomial operations.
    = Any non-linearities need to be rewritten in the form of, or approximated with, polynomial operations.
  - Principal drawback: Significant increase in complexity of computations.

Federated learning

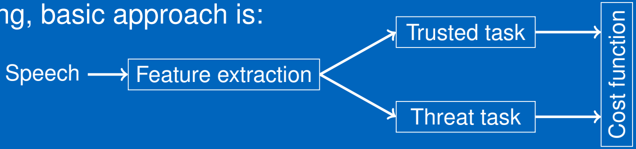
- Current model for digital assistants is based on big-data.
  - Service providers need troves of data to train their models.
  - ⇒ They store large amounts of private data and use it for training models.
  - Model parameters are valuable and secret, they cannot be transmitted to local devices.
- Federated learning is a method where model is stored at a server, but where each client describes how model can be improved (gradient of parameters), such that server can improve model without seeing the data.
  - See also differential privacy.
  - A drawback is that server can train only once; with stored data, server can use data several times.

Anonymization

- Objective: Extract features from signal on local device, transmit to server, which uses its model to extract information.
  - Example: Extract features for phonemic content on local device, such that server can extract text content, but omit all other information.
- Primary issue is that we do not (yet) have methodology for assessing to which extent other information is removed.
  - E.g. we can test whether gender information is preserved.
  - But, we do not know if it is only our gender-predicting model which is insufficient and if a better model could still predict gender.
  - We also do not know whether other categories of information (like accent, age, physical properties) are also removed and we do not have a full list of possible sensitive categories of information.
- Besides, even the text content reveals sensitive information.
  - For example, (unaware) word choices can be very revealing.

Anonymization

In training, basic approach is:



- Underlying assumption: Server has a model which cannot be shared with local device, or, local device does not have capacity to do trusted task.
  - If model could be shared and if local device has enough capacity, then we could do trusted task on local device.
  - $\Rightarrow$ No need to transmit sensitive features.
- Cost function is a balance between best performance in trusted and worst performance in threat tasks.
- Increasing dimensionality of feature vector improves performance on both tasks.

Differential privacy

- Task: Server wants to extract private information, such that answers cannot be connected back to individual.
  - = Extract population statistics, without connection to individual.
- Idea: Dithering = add noise, such that individual answer is unclear, but averages can be extracted.
- Example: What is your gender?
  - `If flip-coin() == Heads`
    - `Answer truthfully`
  - `else`
    - `Answer randomly (50%/50%)`
- Effect of randomness can be canceled from population average.
- 75% chance that individual answer is true.

MyData

- Basic principle: I own and decide what is done with my data.
- Store data in single location of your choice.
  - Gives transparency, control, but also central point of weakness.
- Service providers request access to your data.
- Requires 1. standard APIs and 2. service-providers for private storage.
- Project started in Finland at Ministry of Transport and Communications.
- Now a large world-wide movement.

Speaker identification/verification and spoofing

- Oldest research area within privacy and security.
- Methodology well-developed.
- Deep-fakes are very convincing for humans.
  - Only computers can detect best fakes (spoofs).
- Speaker id will remain difficult;
  - My mother has difficult discerning between me and my brothers.
  - $\Rightarrow$ Most important differences are the hardest.
  - Day-to-day randomness is large (flu, tired, drunk..) and hard to model.

Experience of privacy

- Pioneering work at Aalto.
- Motivation:
  - People have no intuition about privacy and security with devices.
  - What they think about privacy is often incorrect.
  - $\Rightarrow$ Design of privacy in UIs cannot be based on human-to-machine expectations.
- In contrast, human-to-human behavior, wrt privacy, has a long tradition.
  - We whisper our secrets.
  - Private conversations are held in secluded places.
  - We trust our secrets only to our to friends and loved ones.

Experience of privacy
Approach

Solution: Is this environment private? Study human-to-human behavior

- Ask people how they experience different environments (questionnaire).
  - Could you tell a secret to a friend in this environment?
  - How loud could you tell a secret to a friend in this environment? (1=Whisper .. 5=Shouting)
- Automatically analyze acoustic environment to predict response of humans.
  - Machine learning task
- Adapt behavior of voice user interface to reflect and respect current level of privacy.

Acoustic fingerprint for authentication

- In human-to-human communication, people in the same “room” are obviously allowed to participate in a conversation.
  - Room = Bounded acoustic environment
- To make human-to-device communication intuitive, we can use the same model.
  - Devices in the same acoustic environment can interact privately.
  - Higher level of privacy than just “in the same WiFi-network”.
- Solution: Create fingerprint of microphone signal
  - Devices with same fingerprint can interact privately.
  - Use fingerprint as key for encryption.
  - Error correction to fix small differences.
- Problem: Same TV program in different rooms.

Persistent issues

- Function creep
  - Data collected for one purpose, can be also used for other purposes
  - We do not now know how data will be used in the future
  - When giving consent to use data today, we might expose ourselves to unknown risks in the future
- Irrevocable ID
  - The voice is a permanent part of a person.
  - Extremely valuable as identification.
  - Also very risky, since a lost voice fingerprint cannot be revoked.

  $\Rightarrow$ Once the voice fingerprint is compromised, it can never again be used as ID.
  - Applies to all biometric IDs.

Consequences for researchers

- Data collection is essential for all speech research, especially for machine learning.
  - ⇒ Inherent privacy problem!
  - Leading scientist (group leader) is legally responsible, also if some other group uses your open data in a fradulent way.
- Solutions:
  - Limit data collection (amount& type) to the essential.
    ⇒ Balanced data sets (it's good science anyway!).
  - Check consent forms with lawyer.
  - Apply expiry date – use not allowed after.
  - If necessary, limit access with signed contract.

Outlook and summary

- A lot of activities happening in good directions.
- However, mono-cultures vs diversification has not yet been addressed:
  - Big cloud services are inherently attractive targets. Weakest link exposes everything.
  - Big cloud breaches are massively valuable for criminals. We will not know of breaches unless criminal messes up.
    $\Rightarrow$ *How would we know if Big-cloud is already now compromised?*
  - "Local/edge" learning creates diversity, protects against disruption.
  - "Local/edge" processing also gives control/power to user.
  - We need systematic way of creating local diversity.
- Aalto has a leading role in this research.