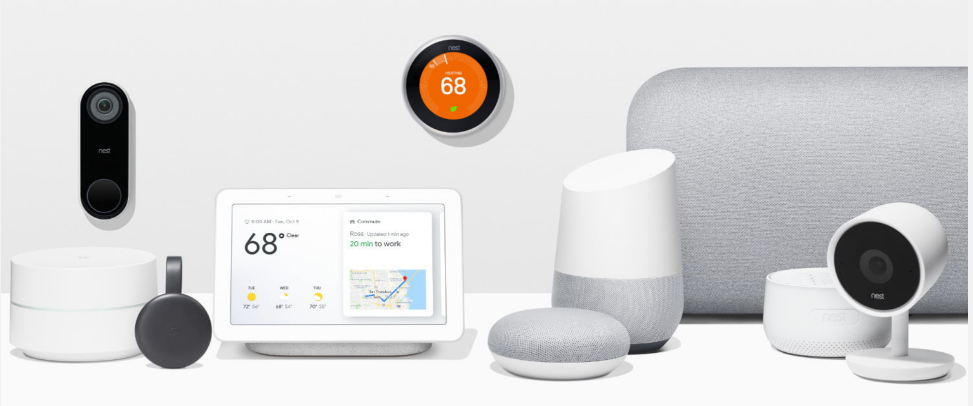
Are You Spying on Me? Large-Scale Analysis on IoT Data Exposure through Companion Apps

Yuhong Nan[1*], **Xueqiang Wang**[2*], Luyi Xing[3], Xiaojing Liao[3], Ruoyu Wu[4], Jianliang Wu[4], Yifan Zhang[3], and XiaoFeng Wang[3]

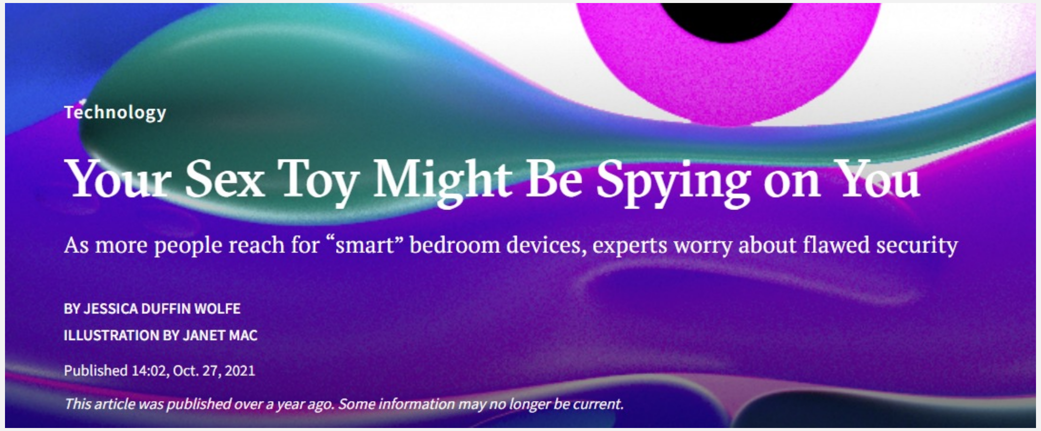
Sun Yat-sen University[1], University of Central Florida[2], Indiana University Bloomington[3], Purdue University[4]



Home automation

Health monitoring

Are You Spying on Me?
Large-Scale Analysis on IoT Data Exposure through Companion Apps



Technology

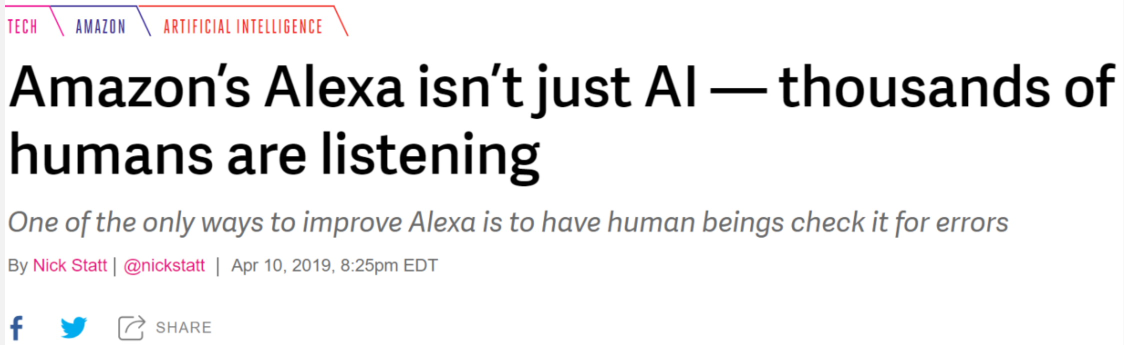
Your Sex Toy Might Be Spying on You

As more people reach for “smart” bedroom devices, experts worry about flawed security

BY JESSICA DUFFIN WOLFE
ILLUSTRATION BY JANET MAC

Published 14:02, Oct. 27, 2021

This article was published over a year ago. Some information may no longer be current.

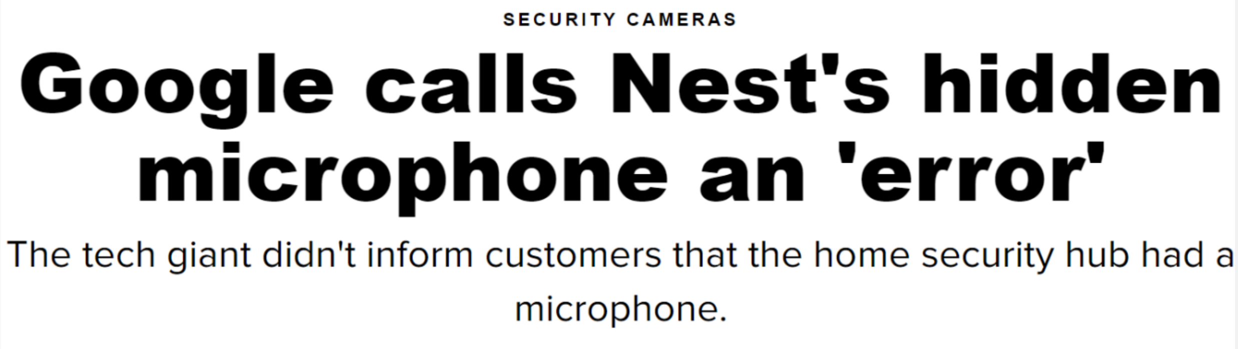


TECH | AMAZON | ARTIFICIAL INTELLIGENCE

Amazon’s Alexa isn’t just AI — thousands of humans are listening

One of the only ways to improve Alexa is to have human beings check it for errors

By Nick Statt | @nickstatt | Apr 10, 2019, 8:25pm EDT



SECURITY CAMERAS

Google calls Nest's hidden microphone an 'error'

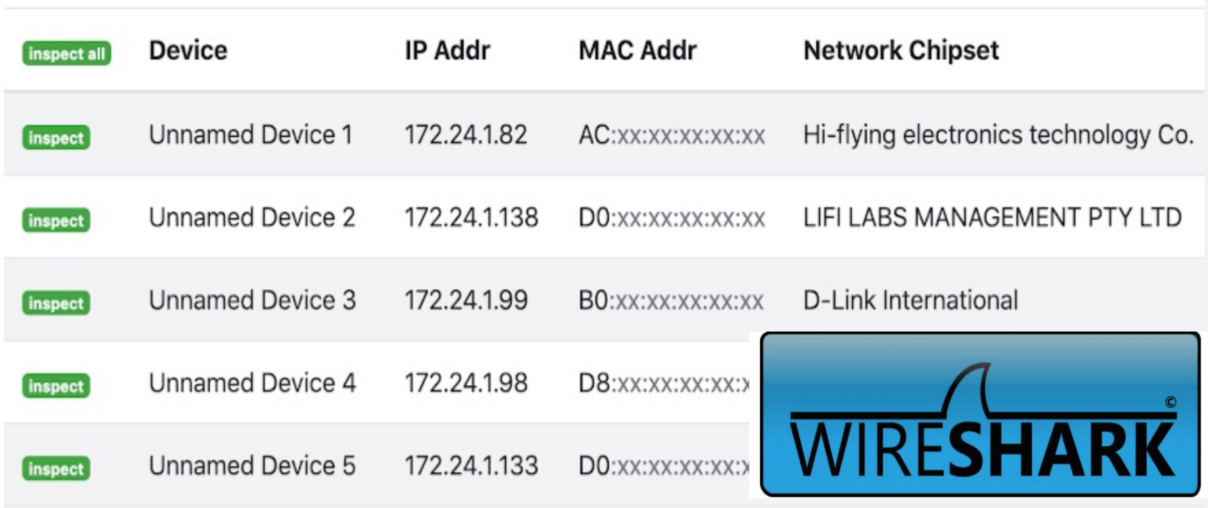
The tech giant didn't inform customers that the home security hub had a microphone.

Key questions to address privacy concerns

- What types of data are being collected by IoT devices?

- How is the data collected, and to which party is it shared?

Prior research & limitations



inspect all	Device	IP Addr	MAC Addr	Network Chipset
inspect	Unnamed Device 1	172.24.1.82	AC:xx:xx:xx:xx:xx	Hi-flying electronics technology Co.
inspect	Unnamed Device 2	172.24.1.138	D0:xx:xx:xx:xx:xx	LIFI LABS MANAGEMENT PTY LTD
inspect	Unnamed Device 3	172.24.1.99	B0:xx:xx:xx:xx:xx	D-Link International
inspect	Unnamed Device 4	172.24.1.98	D8:xx:xx:xx:xx:x	
inspect	Unnamed Device 5	172.24.1.133	D0:xx:xx:xx:xx:x	

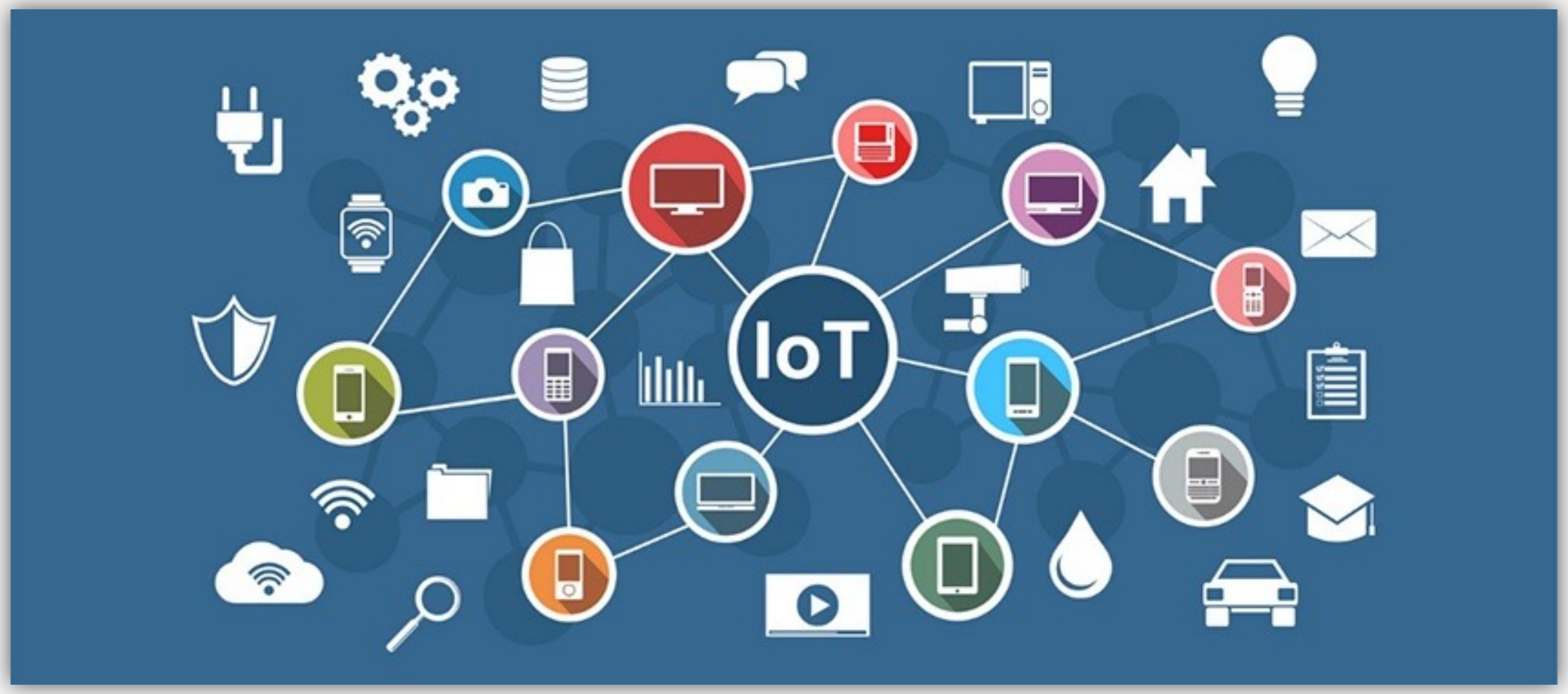
Monitoring network traffic
[IMC19, CCS19, PETS19]

Crowd source
[SEC19,Ubicomp2020]

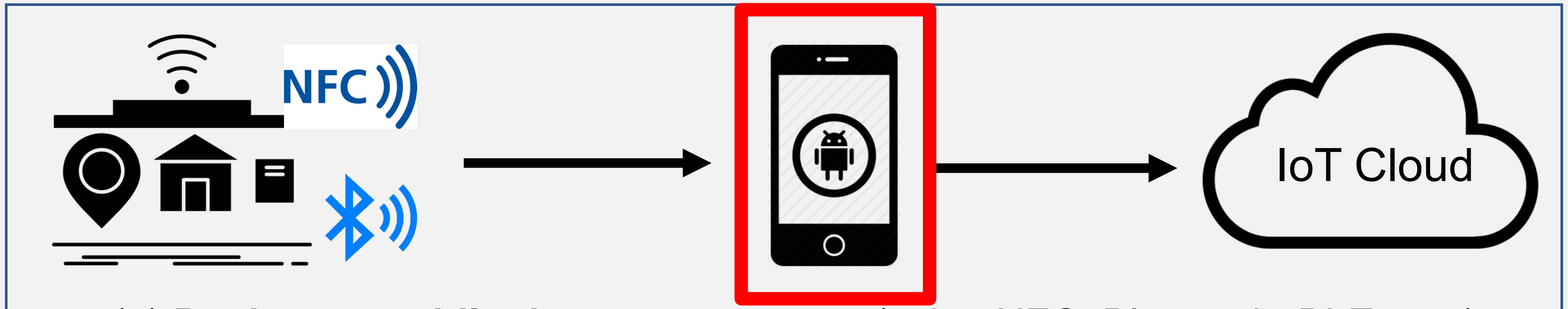
- ➢ Small device set, lab-environment ☹
- ➢ Data encryption, not scale-well ☹

- ➢ Collecting real user data and leads to big privacy concern ☹

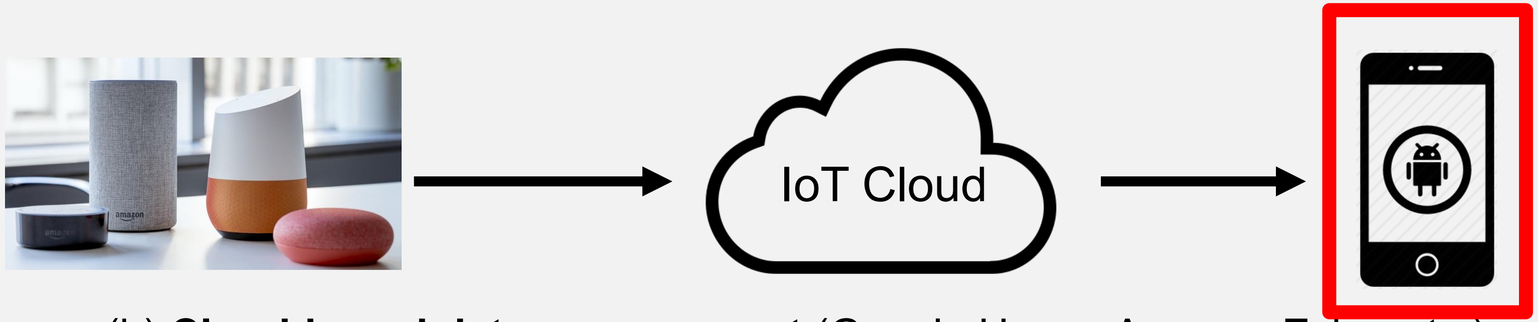
Our goal: Large-scale, fine-grained understanding of IoT device data exposure in the wild.



Observation 1 - Two typical data management modes in IoT ecosystem

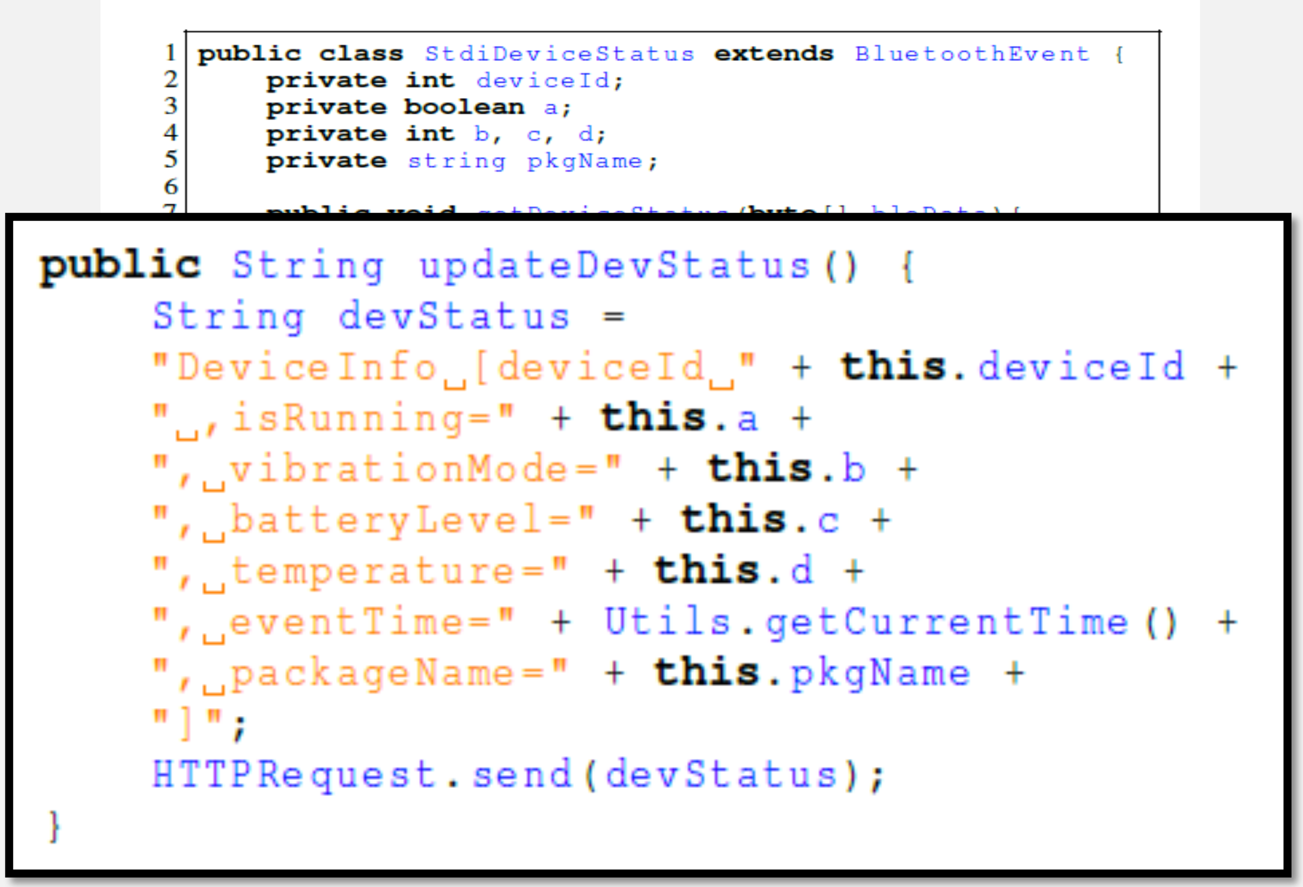


(a) **Device-to-mobile data management** (using NFC, Bluetooth, BLE, etc.)



(b) **Cloud-based data management** (Google Home, Amazon Echo, etc.)

Observation 2 – Semantics-rich IoT companion apps



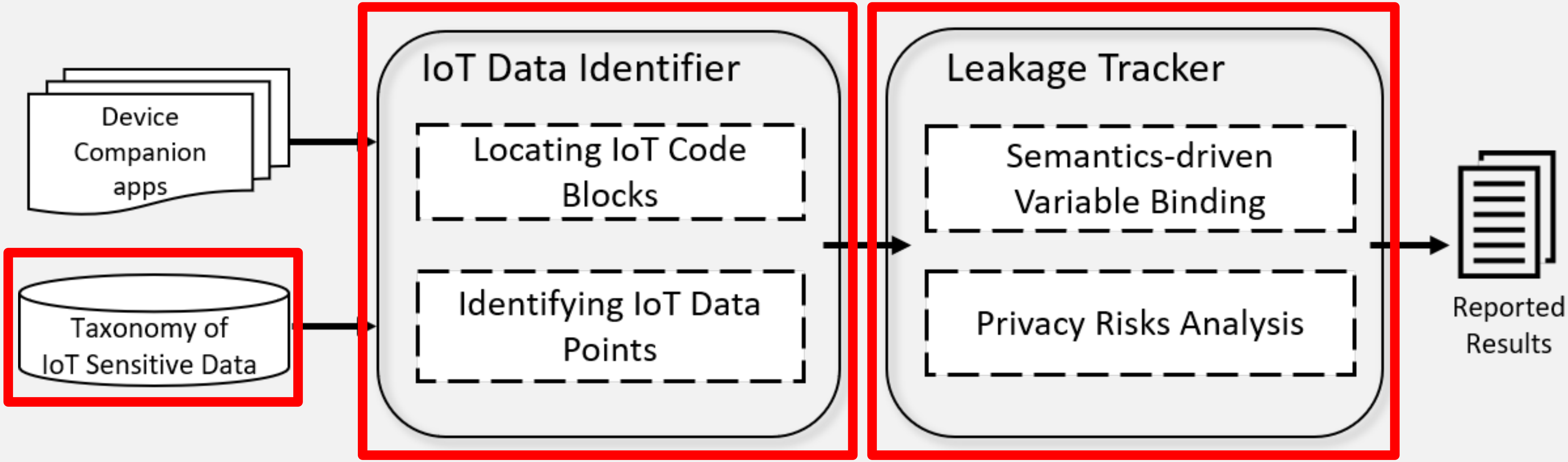
```
public class StdiDeviceStatus extends BluetoothEvent {
    private int deviceId;
    private boolean a;
    private int b, c, d;
    private string pkgName;

    public void getDeviceStatus(byte[] bleData){
```

```
public String updateDevStatus() {
    String devStatus =
    "DeviceInfo␣[deviceId␣" + this.deviceId +
    "␣,isRunning=" + this.a +
    ",␣vibrationMode=" + this.b +
    ",␣batteryLevel=" + this.c +
    ",␣temperature=" + this.d +
    ",␣eventTime=" + Utils.getCurrentTime() +
    ",␣packageName=" + this.pkgName +
    "]";
    HTTPRequest.send(devStatus);
}
```

- ➢ IoT Code Block
  - ➢ **A cluster of texts** (in a method) which includes meaningful labels describing IoT device data.
- ➢ IoT Data Point
  - ➢ **Individual IoT data** if the text label indicates information related to IoT Devices.

IoTProfiler



Device Companion apps

Taxonomy of IoT Sensitive Data

IoT Data Identifier

Locating IoT Code Blocks

Identifying IoT Data Points

Leakage Tracker

Semantics-driven Variable Binding

Privacy Risks Analysis

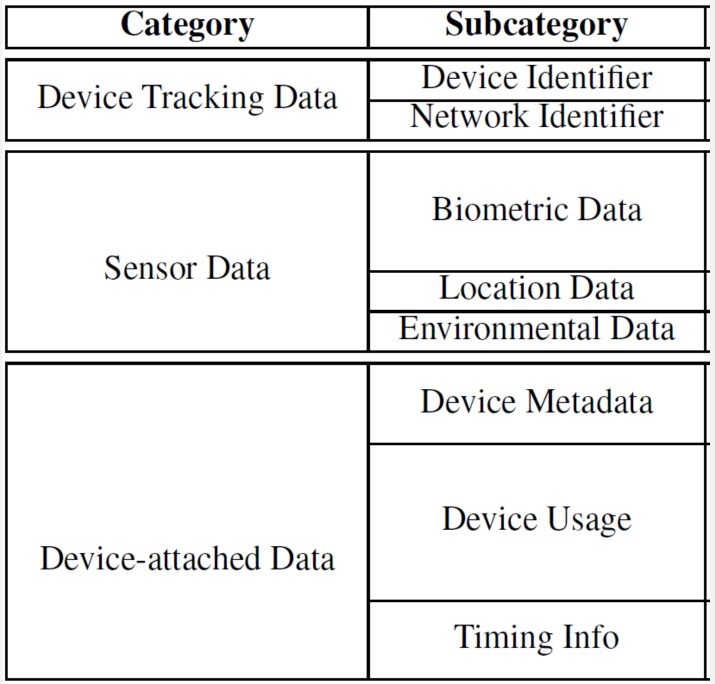
Reported Results

A taxonomy of privacy-sensitive IoT data

- Challenge: Diversity of privacy-sensitive IoT data.

- Construct an IoT taxonomy by analyzing known IoT reports, papers, documents and industry standards, etc.

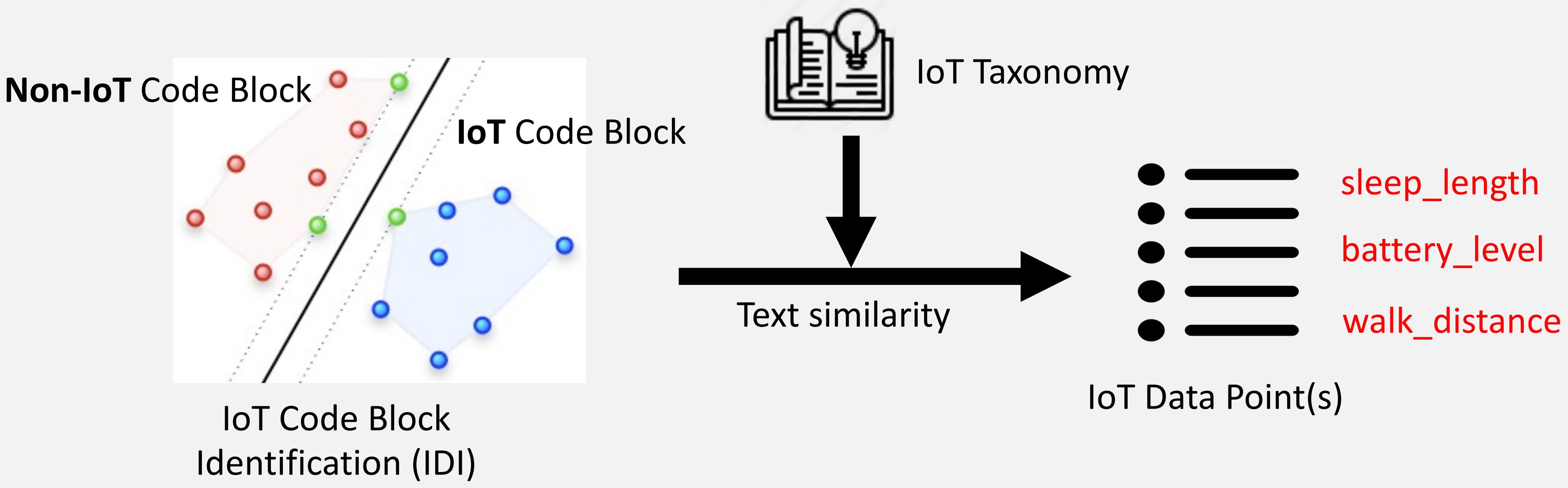
- 550 data items, with 8 sub-categories
  - Released at https://sites.google.com/view/iotprofiler.



Category	Subcategory
Device Tracking Data	Device Identifier
	Network Identifier
Sensor Data	Biometric Data
	Location Data
	Environmental Data
Device-attached Data	Device Metadata
	Device Usage
	Timing Info

IoT data point identification

- Challenge: IoT data handled in companion apps are often kept together with the app local data
- Solution: A two-stage classifier
  - Only considers **clustered IoT data labels**



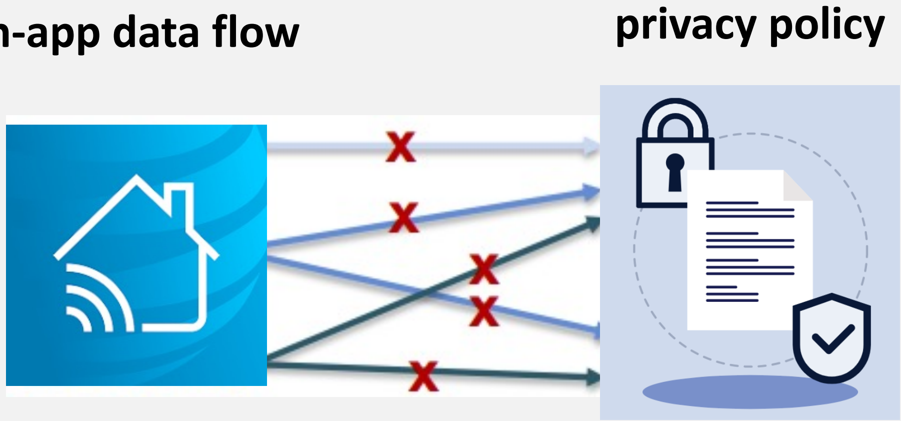
Variable binding and data exposure detection

- **Alias data labels**

```
intent.putExtra("voltage_val", dev.battery_level)
```

- **Link data labels to program variables**

```
JsonObject.put("walk_distance", a)
a = JsonObject.get("walk_distance")
```



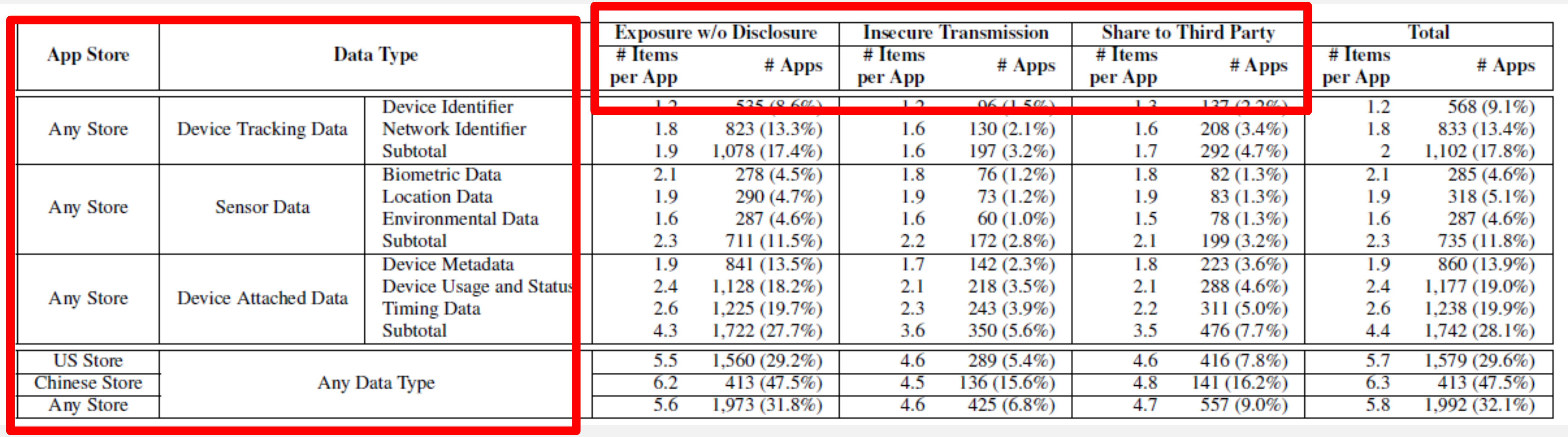
E.g., are these data flows disclosed within the application's privacy policy, or data transmitted securely?

Please check our paper for further technical details.

Datasets

- IoT companion apps collected in the wild.
  - 6,208 IoT apps as of Aug. 2020
  - https://www.cs.ucf.edu/~xwang/datasets/IoTProfiler-Apps/

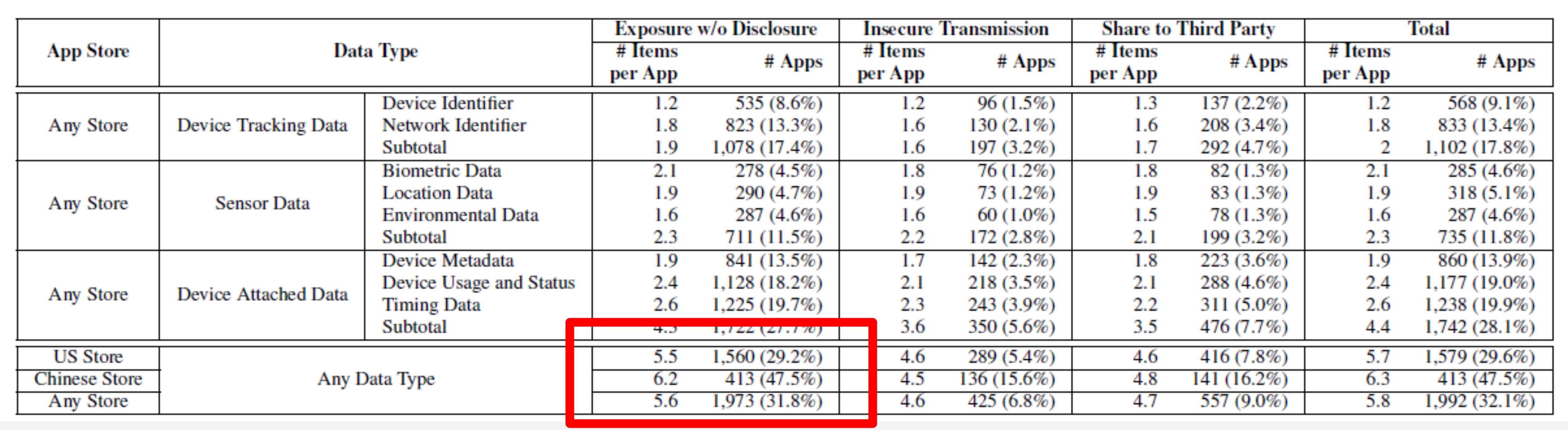
IoT data exposure in the wild



App Store	Data Type		Exposure w/o Disclosure		Insecure Transmission		Share to Third Party		Total	
			# Items per App	# Apps	# Items per App	# Apps	# Items per App	# Apps	# Items per App	# Apps
Any Store	Device Tracking Data	Device Identifier	1.2	535 (8.6%)	1.2	96 (1.5%)	1.3	137 (2.2%)	1.2	568 (9.1%)
		Network Identifier	1.8	823 (13.3%)	1.6	130 (2.1%)	1.6	208 (3.4%)	1.8	833 (13.4%)
		Subtotal	1.9	1,078 (17.4%)	1.6	197 (3.2%)	1.7	292 (4.7%)	2	1,102 (17.8%)
Any Store	Sensor Data	Biometric Data	2.1	278 (4.5%)	1.8	76 (1.2%)	1.8	82 (1.3%)	2.1	285 (4.6%)
		Location Data	1.9	290 (4.7%)	1.9	73 (1.2%)	1.9	83 (1.3%)	1.9	318 (5.1%)
		Environmental Data	1.6	287 (4.6%)	1.6	60 (1.0%)	1.5	78 (1.3%)	1.6	287 (4.6%)
		Subtotal	2.3	711 (11.5%)	2.2	172 (2.8%)	2.1	199 (3.2%)	2.3	735 (11.8%)
Any Store	Device Attached Data	Device Metadata	1.9	841 (13.5%)	1.7	142 (2.3%)	1.8	223 (3.6%)	1.9	860 (13.9%)
		Device Usage and Status	2.4	1,128 (18.2%)	2.1	218 (3.5%)	2.1	288 (4.6%)	2.4	1,177 (19.0%)
		Timing Data	2.6	1,225 (19.7%)	2.3	243 (3.9%)	2.2	311 (5.0%)	2.6	1,238 (19.9%)
		Subtotal	4.3	1,722 (27.7%)	3.6	350 (5.6%)	3.5	476 (7.7%)	4.4	1,742 (28.1%)
US Store	Any Data Type		5.5	1,560 (29.2%)	4.6	289 (5.4%)	4.6	416 (7.8%)	5.7	1,579 (29.6%)
Chinese Store			6.2	413 (47.5%)	4.5	136 (15.6%)	4.8	141 (16.2%)	6.3	413 (47.5%)
Any Store			5.6	1,973 (31.8%)	4.6	425 (6.8%)	4.7	557 (9.0%)	5.8	1,992 (32.1%)

- 50,667 IoT code blocks and 174,943 IoT data points from **5,795/6,208(93.3%)** apps.

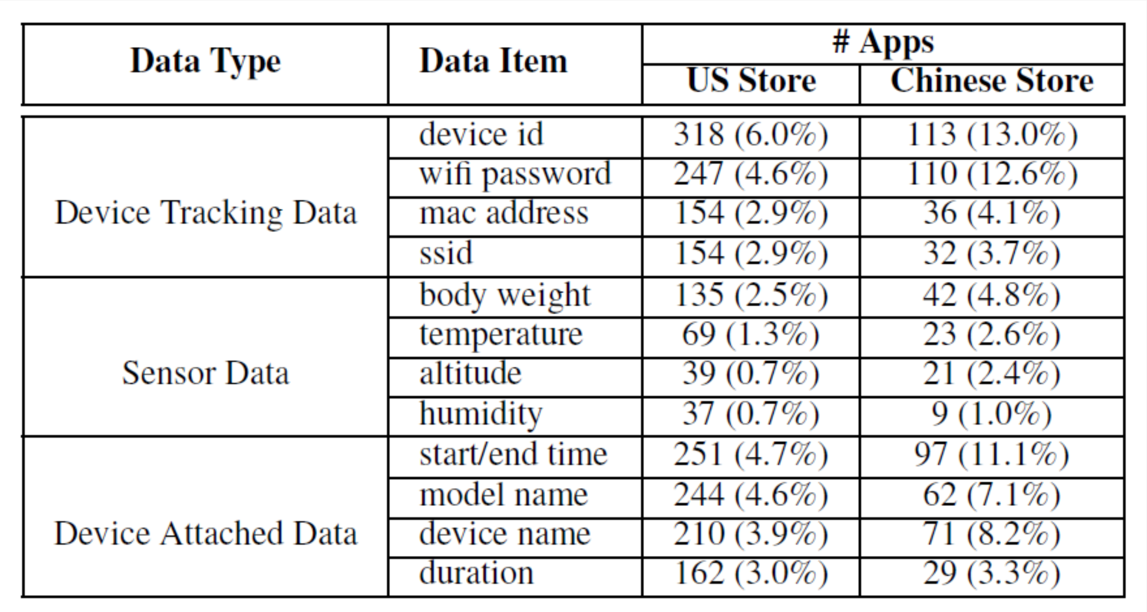
IoT data exposure in the wild



App Store	Data Type		Exposure w/o Disclosure		Insecure Transmission		Share to Third Party		Total	
			# Items per App	# Apps	# Items per App	# Apps	# Items per App	# Apps	# Items per App	# Apps
Any Store	Device Tracking Data	Device Identifier	1.2	535 (8.6%)	1.2	96 (1.5%)	1.3	137 (2.2%)	1.2	568 (9.1%)
		Network Identifier	1.8	823 (13.3%)	1.6	130 (2.1%)	1.6	208 (3.4%)	1.8	833 (13.4%)
		Subtotal	1.9	1,078 (17.4%)	1.6	197 (3.2%)	1.7	292 (4.7%)	2	1,102 (17.8%)
Any Store	Sensor Data	Biometric Data	2.1	278 (4.5%)	1.8	76 (1.2%)	1.8	82 (1.3%)	2.1	285 (4.6%)
		Location Data	1.9	290 (4.7%)	1.9	73 (1.2%)	1.9	83 (1.3%)	1.9	318 (5.1%)
		Environmental Data	1.6	287 (4.6%)	1.6	60 (1.0%)	1.5	78 (1.3%)	1.6	287 (4.6%)
		Subtotal	2.3	711 (11.5%)	2.2	172 (2.8%)	2.1	199 (3.2%)	2.3	735 (11.8%)
Any Store	Device Attached Data	Device Metadata	1.9	841 (13.5%)	1.7	142 (2.3%)	1.8	223 (3.6%)	1.9	860 (13.9%)
		Device Usage and Status	2.4	1,128 (18.2%)	2.1	218 (3.5%)	2.1	288 (4.6%)	2.4	1,177 (19.0%)
		Timing Data	2.6	1,225 (19.7%)	2.3	243 (3.9%)	2.2	311 (5.0%)	2.6	1,238 (19.9%)
		Subtotal	4.3	1,722 (27.7%)	3.6	350 (5.6%)	3.5	476 (7.7%)	4.4	1,742 (28.1%)
US Store	Any Data Type		5.5	1,560 (29.2%)	4.6	289 (5.4%)	4.6	416 (7.8%)	5.7	1,579 (29.6%)
Chinese Store			6.2	413 (47.5%)	4.5	136 (15.6%)	4.8	141 (16.2%)	6.3	413 (47.5%)
Any Store			5.6	1,973 (31.8%)	4.6	425 (6.8%)	4.7	557 (9.0%)	5.8	1,992 (32.1%)

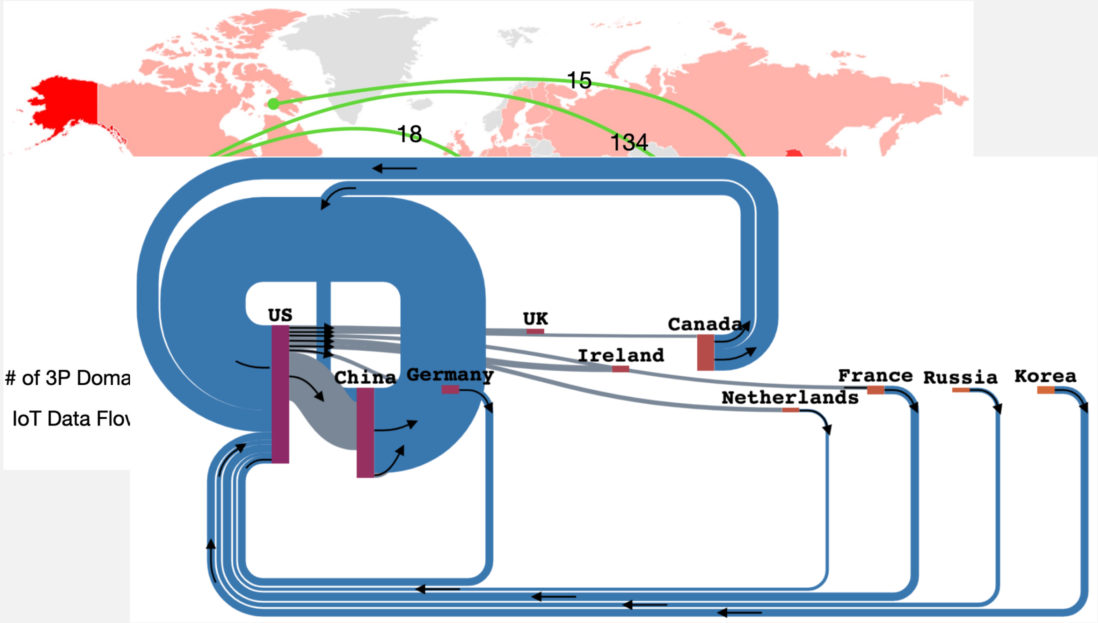
- ➢ 1,973 apps (31.8%) from at least **1,559 unique device vendors** are found to collect sensitive IoT data without proper disclosure.
- ➢ Each app exposes **5.6 IoT data items** on average.

IoT data exposure in the wild



Data Type	Data Item	# Apps	
		US Store	Chinese Store
Device Tracking Data	device id	318 (6.0%)	113 (13.0%)
	wifi password	247 (4.6%)	110 (12.6%)
	mac address	154 (2.9%)	36 (4.1%)
	ssid	154 (2.9%)	32 (3.7%)
Sensor Data	body weight	135 (2.5%)	42 (4.8%)
	temperature	69 (1.3%)	23 (2.6%)
	altitude	39 (0.7%)	21 (2.4%)
	humidity	37 (0.7%)	9 (1.0%)
Device Attached Data	start/end time	251 (4.7%)	97 (11.1%)
	model name	244 (4.6%)	62 (7.1%)
	device name	210 (3.9%)	71 (8.2%)
	duration	162 (3.0%)	29 (3.3%)

IoT data exposure for users of different regions

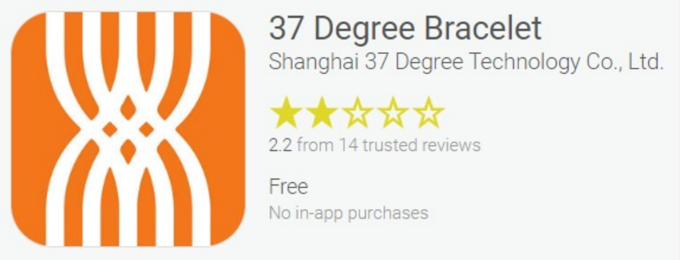


Cross-region IoT data flows

For more findings, please check out our paper.

Example

- ➢ Health monitoring device



37 Degree Bracelet

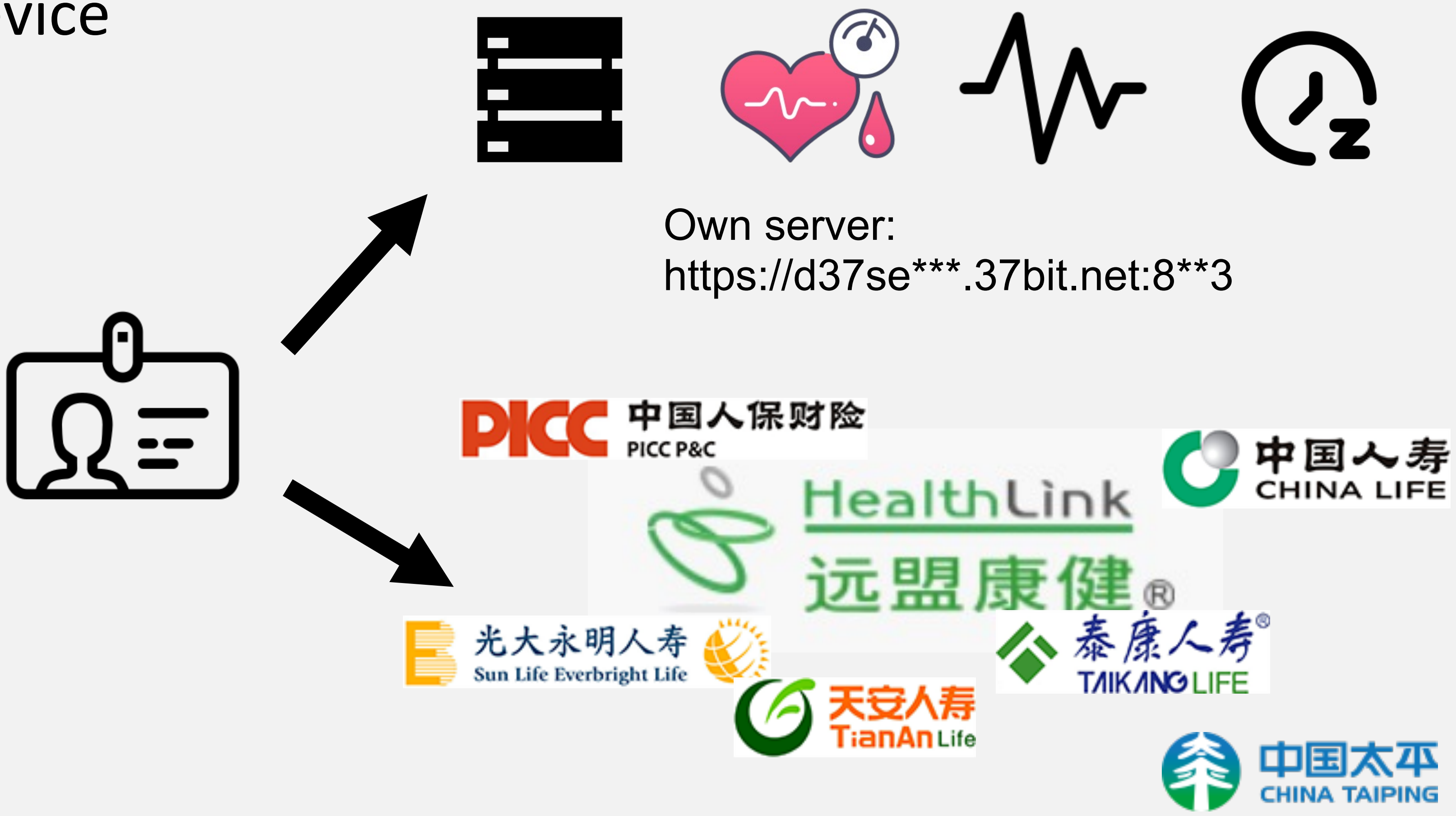
Shanghai 37 Degree Technology Co., Ltd.

★★☆☆☆

2.2 from 14 trusted reviews

Free

No in-app purchases



Own server:
https://d37se***.37bit.net:8**3

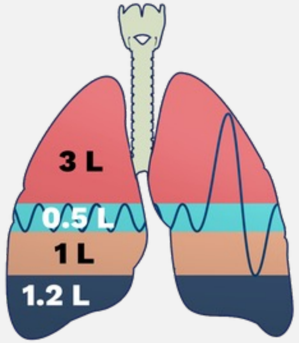
healthlink.cn

Case study

- Cigarette holder – use expected data
  - Harmful substances (e.g., nicotine)
  - Number of cigarettes smoked

- Data without the user's awareness
  - Smoking habits: **smoking times (puff)**, **locations**
  - Health conditions: **breathing capacity**

smoking times (puff)



breathing capacity

Summary

- New techniques to enable fine-grained analysis of IoT data exposure.
- Large-scale understanding of IoT data exposure.

- Potential applications include the auto-generation of privacy labels, which can help IoT apps/devices become privacy compliant.

UNIVERSITY OF CENTRAL FLORIDA

Cyber Security and Privacy Cluster

- Software Supply Chain Security
- Privacy Compliance Automation
- Mobile & IoT Security

Thank You!