IoTFlow: Inferring IoT Device Behavior at Scale through Static Mobile Companion App Analysis

David Schmidt, Carlotta Tagliaro, Kevin Borgolte, Martina Lindorfer

ACM Conference on Computer and Communications Security, 2023

Introduction

- Existing approaches :

re-identify / multiple / require physical devices

- IoTFlow :

discover automatically / individual / without requiring

- evaluate IoTFlow on 9,889 companion apps
- study the differences between the companion apps and popular apps

Introduction

- verify the accuracy of IoTFlow
- compare it with dynamic analysis

- Approach :

(1) identifie communication trigger points
(2) Value Set Analysis (VSA)
(3) Data-flow Analysis (DFA)
(4) assesse the corresponding impact

MOTIVATION

- Large-scale IoT Device Behavior Analysis.
- IoT Control Infrastructure.

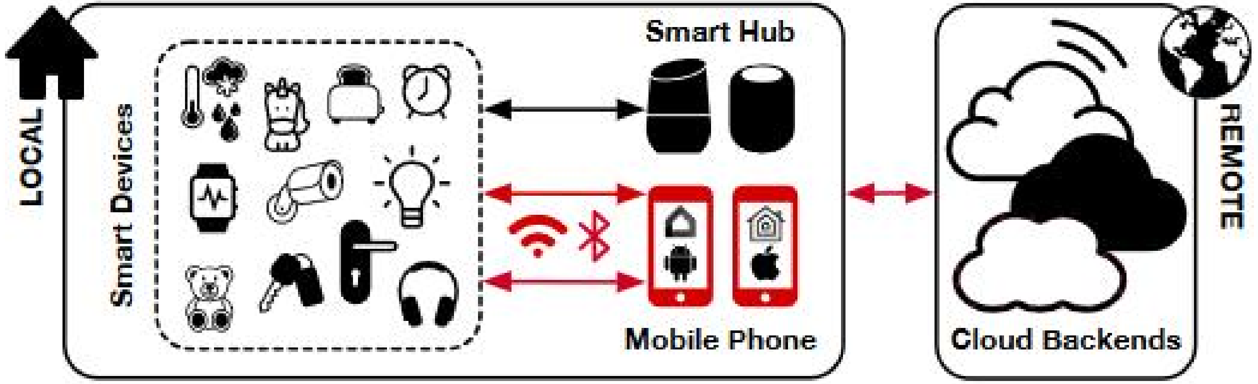


Figure 1: Overview of the IoT ecosystem and its command and control scenarios, including apps as intermediaries.

- General-purpose Apps vs. Companion Apps.

IOTFLOW

- Value Set Analysis (VSA)
- Data-flow Analysis (DFA)

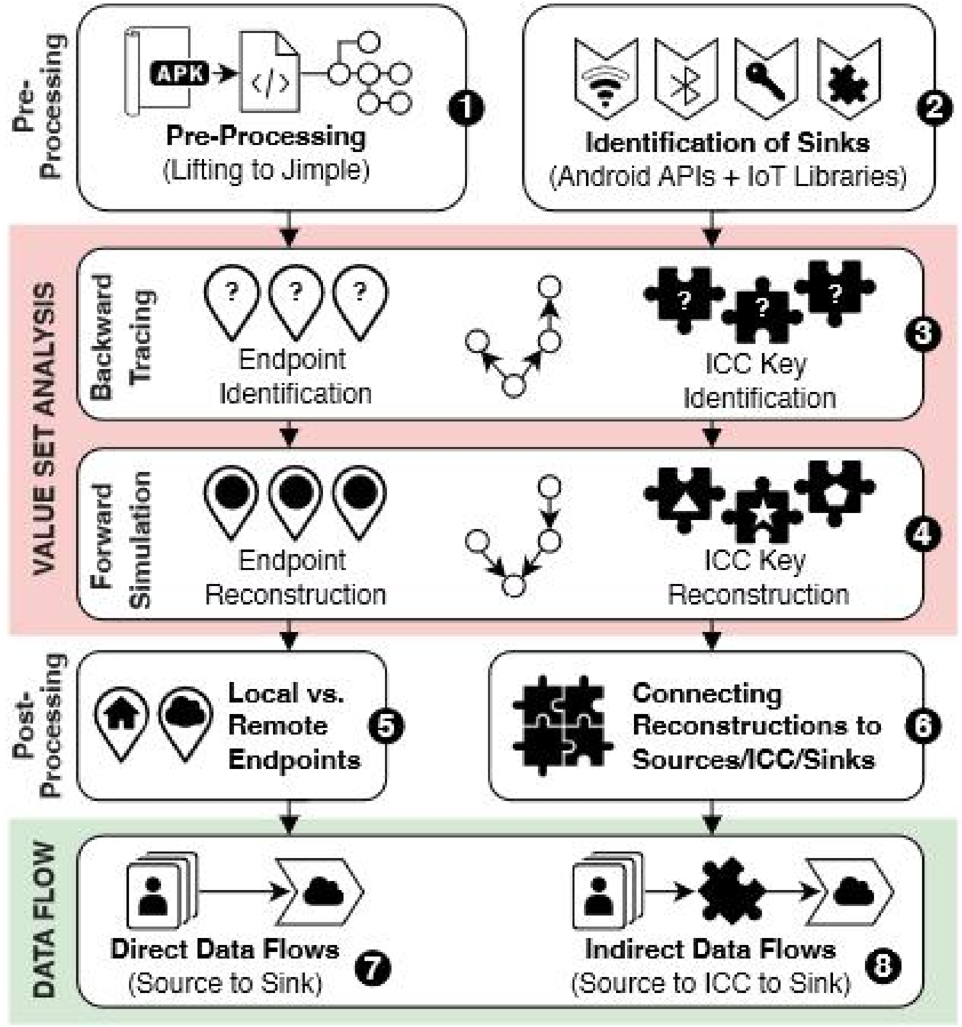


Figure 2: Overview of IoTFlow. We use VSA to reconstruct endpoints, cryptographic data, and ICC keys for the flow analysis. We use flow analysis to find data leaks, and connect request/response data with endpoints. With the ICC information of the VSA, we support data flows involving ICC.

IOTFLOW

- VSA:reconstruct values at specific program points

① Pre-Processing

② Identification of Sinks

③ Backward Tracing

④ Forward Simulation

⑤ Local vs. Remote Endpoints

IOTFLOW

- DFA:trace data flows from IoT devices and sensitive Android methods

⑥ Connecting Reconstructions

⑦ Direct Data Flows (Source to Sink)

⑧ Indirect Data Flows (Source to ICC to Sink)

INSIGHTS INTO THE IOT ECOSYSTEM

- Dataset
- Performance

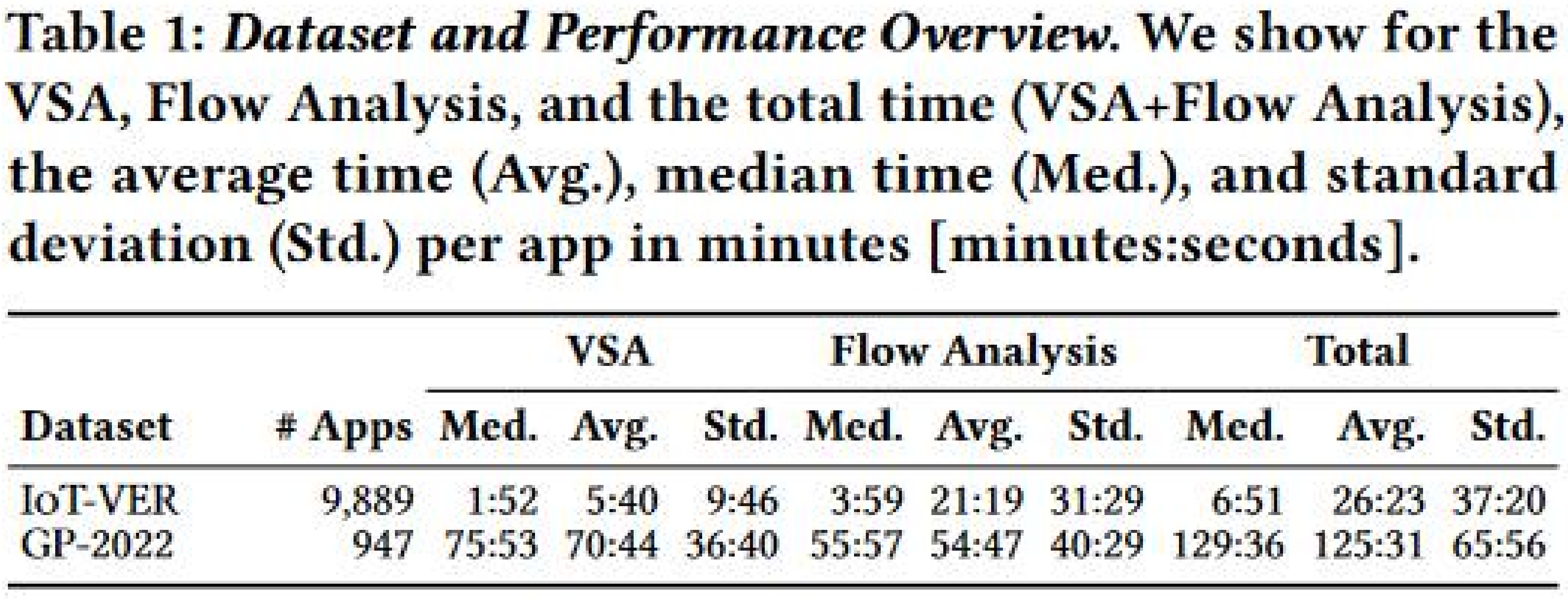


Table 1: *Dataset and Performance Overview.* We show for the VSA, Flow Analysis, and the total time (VSA+Flow Analysis), the average time (Avg.), median time (Med.), and standard deviation (Std.) per app in minutes [minutes:seconds].

		VSA			Flow Analysis			Total		
Dataset	**# Apps**	**Med.**	**Avg.**	**Std.**	**Med.**	**Avg.**	**Std.**	**Med.**	**Avg.**	**Std.**
IoT-VER	9,889	1:52	5:40	9:46	3:59	21:19	31:29	6:51	26:23	37:20
GP-2022	947	75:53	70:44	36:40	55:57	54:47	40:29	129:36	125:31	65:56

INSIGHTS INTO THE IOT ECOSYSTEM

- How Companion Apps Communicate(RQ1):

Direct Device Communication

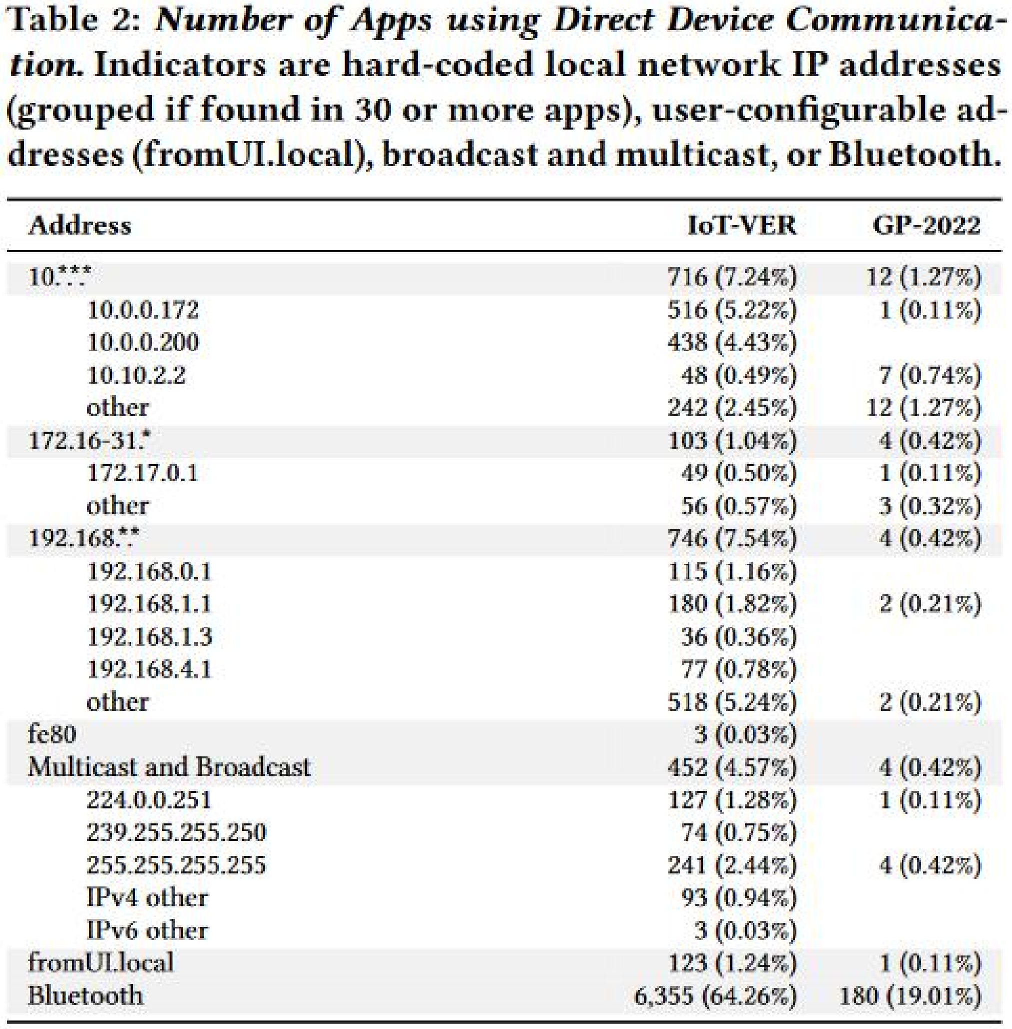


Table 2: ***Number of Apps using Direct Device Communication.*** Indicators are hard-coded local network IP addresses (grouped if found in 30 or more apps), user-configurable addresses (fromUI.local), broadcast and multicast, or Bluetooth.

Address	IoT-VER	GP-2022
10.*.*.*	716 (7.24%)	12 (1.27%)
10.0.0.172	516 (5.22%)	1 (0.11%)
10.0.0.200	438 (4.43%)	
10.10.2.2	48 (0.49%)	7 (0.74%)
other	242 (2.45%)	12 (1.27%)
172.16-31.*	103 (1.04%)	4 (0.42%)
172.17.0.1	49 (0.50%)	1 (0.11%)
other	56 (0.57%)	3 (0.32%)
192.168.*.*	746 (7.54%)	4 (0.42%)
192.168.0.1	115 (1.16%)	
192.168.1.1	180 (1.82%)	2 (0.21%)
192.168.1.3	36 (0.36%)	
192.168.4.1	77 (0.78%)	
other	518 (5.24%)	2 (0.21%)
fe80	3 (0.03%)	
Multicast and Broadcast	452 (4.57%)	4 (0.42%)
224.0.0.251	127 (1.28%)	1 (0.11%)
239.255.255.250	74 (0.75%)	
255.255.255.255	241 (2.44%)	4 (0.42%)
IPv4 other	93 (0.94%)	
IPv6 other	3 (0.03%)	
fromUI.local	123 (1.24%)	1 (0.11%)
Bluetooth	6,355 (64.26%)	180 (19.01%)

INSIGHTS INTO THE IOT ECOSYSTEM

- How Companion Apps Communicate(RQ1):

URL Protocol Schemes

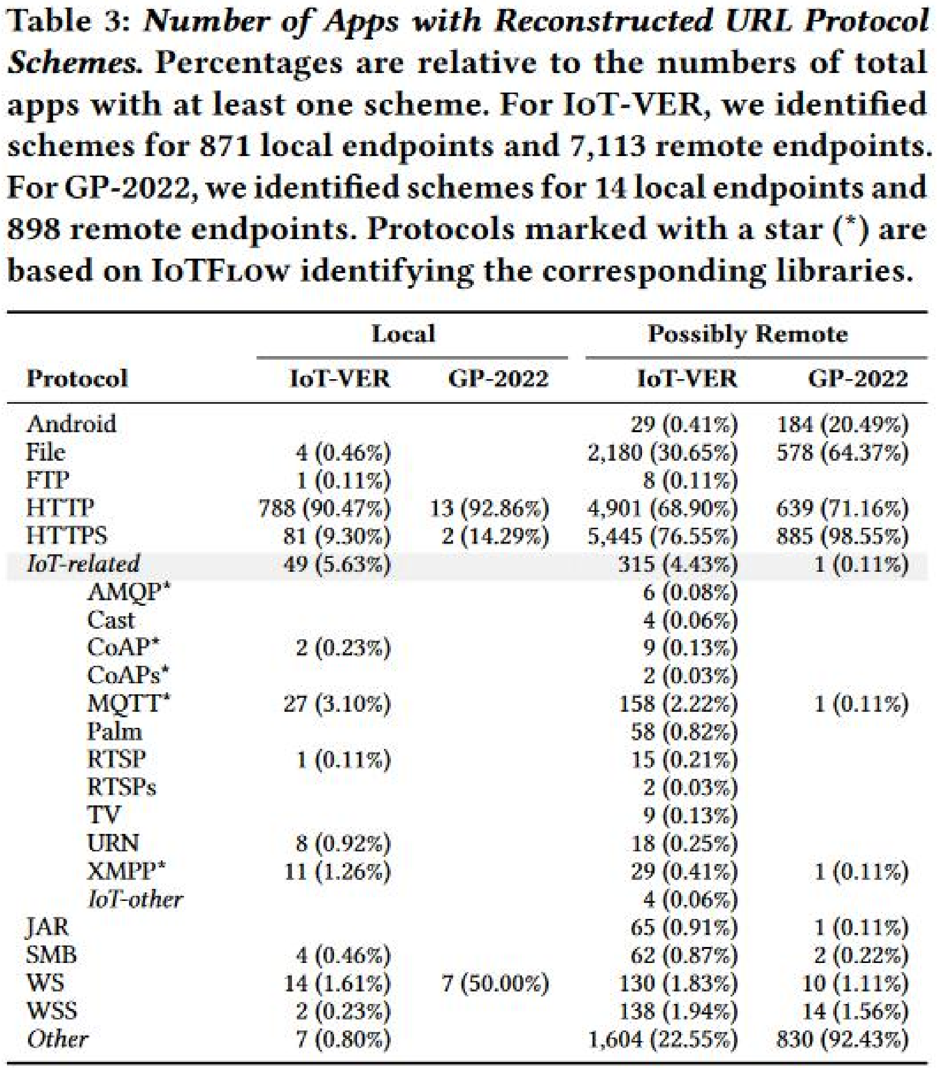


Table 3: ***Number of Apps with Reconstructed URL Protocol Schemes.*** **Percentages are relative to the numbers of total apps with at least one scheme. For IoT-VER, we identified schemes for 871 local endpoints and 7,113 remote endpoints. For GP-2022, we identified schemes for 14 local endpoints and 898 remote endpoints. Protocols marked with a star (*) are based on IoTFlow identifying the corresponding libraries.**

	Local		Possibly Remote	
Protocol	**IoT-VER**	**GP-2022**	**IoT-VER**	**GP-2022**
Android			29 (0.41%)	184 (20.49%)
File	4 (0.46%)		2,180 (30.65%)	578 (64.37%)
FTP	1 (0.11%)		8 (0.11%)	
HTTP	788 (90.47%)	13 (92.86%)	4,901 (68.90%)	639 (71.16%)
HTTPS	81 (9.30%)	2 (14.29%)	5,445 (76.55%)	885 (98.55%)
IoT-related	49 (5.63%)		315 (4.43%)	1 (0.11%)
AMQP*			6 (0.08%)	
Cast			4 (0.06%)	
CoAP*	2 (0.23%)		9 (0.13%)	
CoAPs*			2 (0.03%)	
MQTT*	27 (3.10%)		158 (2.22%)	1 (0.11%)
Palm			58 (0.82%)	
RTSP	1 (0.11%)		15 (0.21%)	
RTSPs			2 (0.03%)	
TV			9 (0.13%)	
URN	8 (0.92%)		18 (0.25%)	
XMPP*	11 (1.26%)		29 (0.41%)	1 (0.11%)
IoT-other			4 (0.06%)	
JAR			65 (0.91%)	1 (0.11%)
SMB	4 (0.46%)		62 (0.87%)	2 (0.22%)
WS	14 (1.61%)	7 (50.00%)	130 (1.83%)	10 (1.11%)
WSS	2 (0.23%)		138 (1.94%)	14 (1.56%)
Other	7 (0.80%)		1,604 (22.55%)	830 (92.43%)

INSIGHTS INTO THE IOT ECOSYSTEM

- How Companion Apps Communicate(RQ1):

Pinning and Certificates

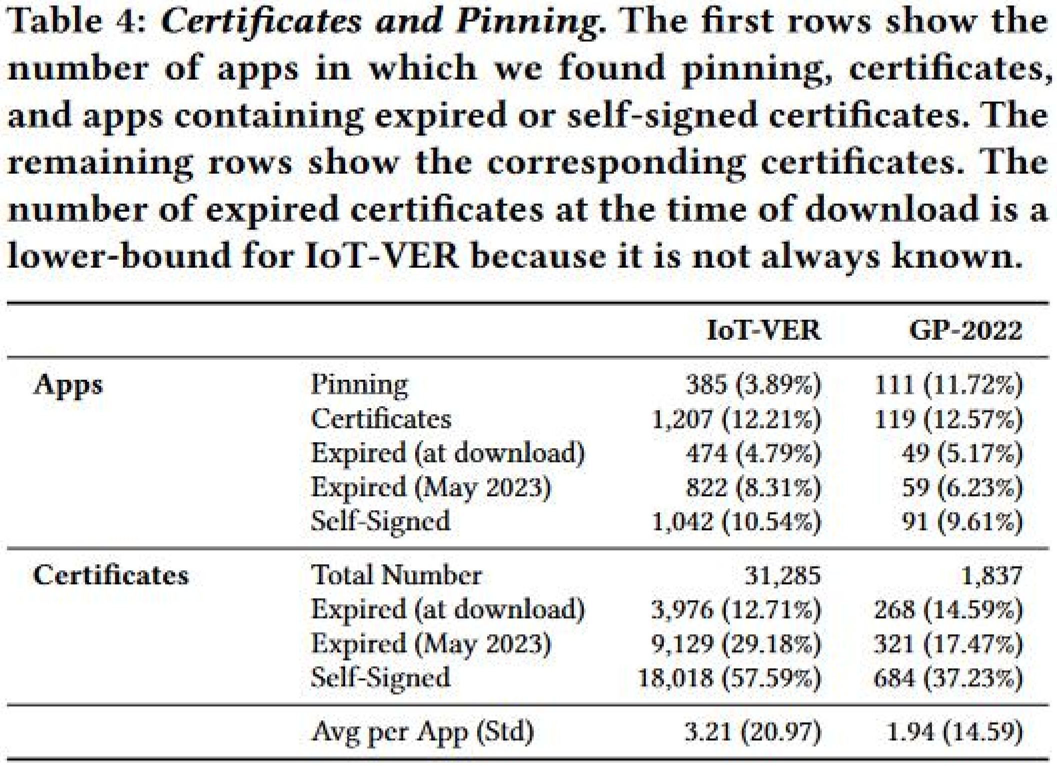


Table 4: *Certificates and Pinning*. The first rows show the number of apps in which we found pinning, certificates, and apps containing expired or self-signed certificates. The remaining rows show the corresponding certificates. The number of expired certificates at the time of download is a lower-bound for IoT-VER because it is not always known.

		IoT-VER	**GP-2022**
Apps	Pinning	385 (3.89%)	111 (11.72%)
	Certificates	1,207 (12.21%)	119 (12.57%)
	Expired (at download)	474 (4.79%)	49 (5.17%)
	Expired (May 2023)	822 (8.31%)	59 (6.23%)
	Self-Signed	1,042 (10.54%)	91 (9.61%)
Certificates	Total Number	31,285	1,837
	Expired (at download)	3,976 (12.71%)	268 (14.59%)
	Expired (May 2023)	9,129 (29.18%)	321 (17.47%)
	Self-Signed	18,018 (57.59%)	684 (37.23%)
	Avg per App (Std)	3.21 (20.97)	1.94 (14.59)

INSIGHTS INTO THE IOT ECOSYSTEM

- With Whom IoT Apps Communicate(RQ2)

Advertisers and Trackers

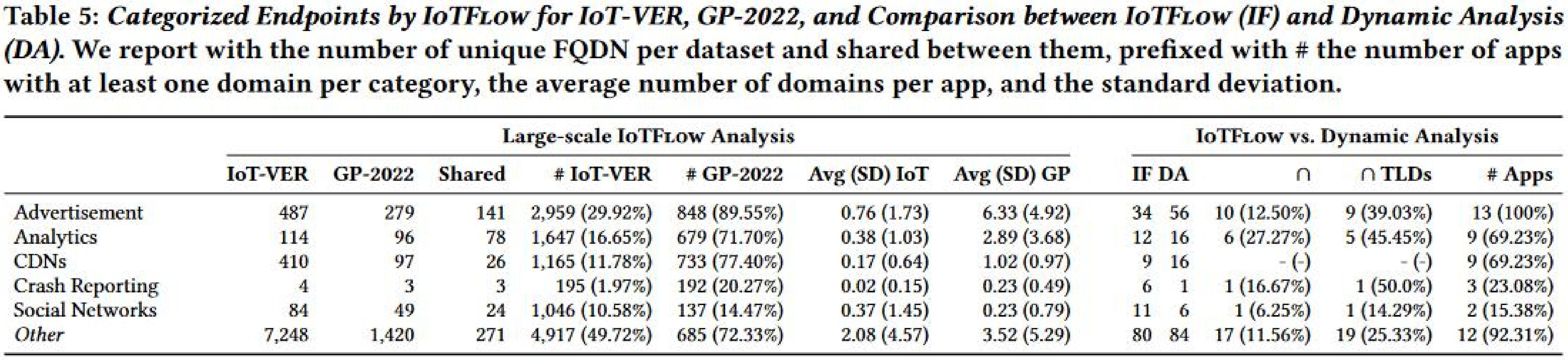


Table 5: *Categorized Endpoints by IoTFlow for IoT-VER, GP-2022, and Comparison between IoTFlow (IF) and Dynamic Analysis (DA).* We report with the number of unique FQDN per dataset and shared between them, prefixed with # the number of apps with at least one domain per category, the average number of domains per app, and the standard deviation.

	Large-scale IoTFlow Analysis							IoTFlow vs. Dynamic Analysis				
	IoT-VER	GP-2022	Shared	# IoT-VER	# GP-2022	Avg (SD) IoT	Avg (SD) GP	IF	DA	∩	∩ TLDs	# Apps
Advertisement	487	279	141	2,959 (29.92%)	848 (89.55%)	0.76 (1.73)	6.33 (4.92)	34	56	10 (12.50%)	9 (39.03%)	13 (100%)
Analytics	114	96	78	1,647 (16.65%)	679 (71.70%)	0.38 (1.03)	2.89 (3.68)	12	16	6 (27.27%)	5 (45.45%)	9 (69.23%)
CDNs	410	97	26	1,165 (11.78%)	733 (77.40%)	0.17 (0.64)	1.02 (0.97)	9	16	- (-)	- (-)	9 (69.23%)
Crash Reporting	4	3	3	195 (1.97%)	192 (20.27%)	0.02 (0.15)	0.23 (0.49)	6	1	1 (16.67%)	1 (50.0%)	3 (23.08%)
Social Networks	84	49	24	1,046 (10.58%)	137 (14.47%)	0.37 (1.45)	0.23 (0.79)	11	6	1 (6.25%)	1 (14.29%)	2 (15.38%)
Other	7,248	1,420	271	4,917 (49.72%)	685 (72.33%)	2.08 (4.57)	3.52 (5.29)	80	84	17 (11.56%)	19 (25.33%)	12 (92.31%)

INSIGHTS INTO THE IOT ECOSYSTEM

- With Whom IoT Apps Communicate(RQ2)

Geographic Location

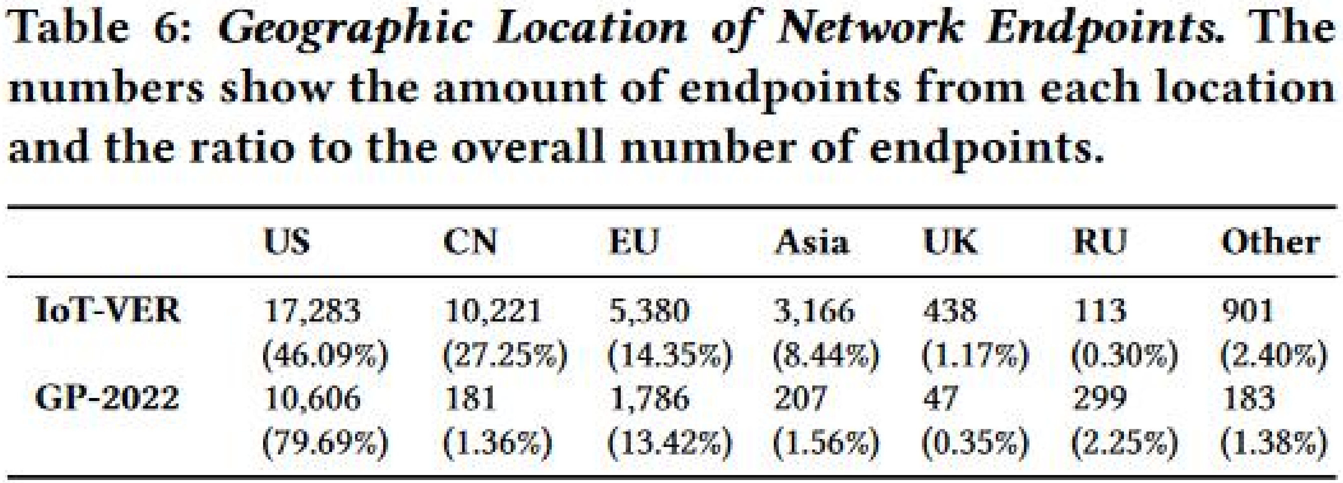


Table 6: *Geographic Location of Network Endpoints.* The numbers show the amount of endpoints from each location and the ratio to the overall number of endpoints.

	US	CN	EU	Asia	UK	RU	Other
IoT-VER	17,283 (46.09%)	10,221 (27.25%)	5,380 (14.35%)	3,166 (8.44%)	438 (1.17%)	113 (0.30%)	901 (2.40%)
GP-2022	10,606 (79.69%)	181 (1.36%)	1,786 (13.42%)	207 (1.56%)	47 (0.35%)	299 (2.25%)	183 (1.38%)

INSIGHTS INTO THE IOT ECOSYSTEM

- With Whom IoT Apps Communicate(RQ2)

Abandoned Domains:
136 abandoned domains in companion apps;
67 domains from 73 apps are available for registration;
73 apps can still be downloaded.

INSIGHTS INTO THE IOT ECOSYSTEM

- What Data Companion Apps Share(RQ3)

Permissions:
permissions with a protectionLevel of :
dangerous(88.67%) & privileged(26.33%)

Data Flows:

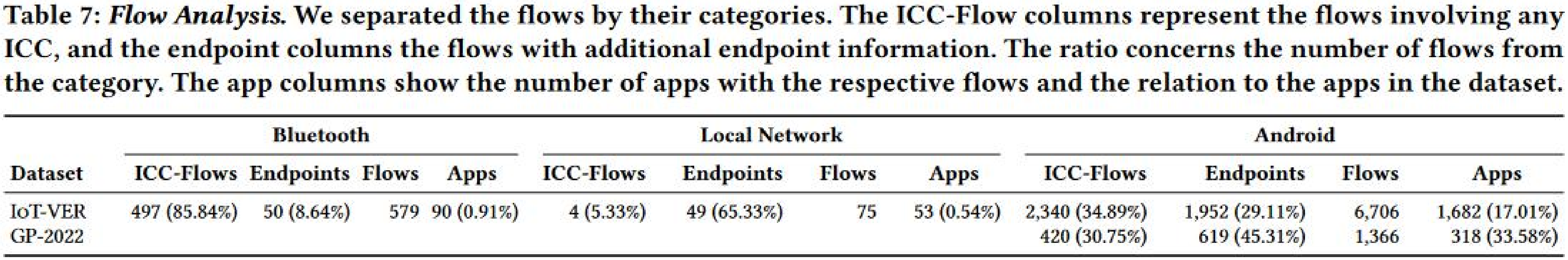


Table 7: ***Flow Analysis.*** **We separated the flows by their categories. The ICC-Flow columns represent the flows involving any ICC, and the endpoint columns the flows with additional endpoint information. The ratio concerns the number of flows from the category. The app columns show the number of apps with the respective flows and the relation to the apps in the dataset.**

	Bluetooth				Local Network				Android			
Dataset	**ICC-Flows**	**Endpoints**	**Flows**	**Apps**	**ICC-Flows**	**Endpoints**	**Flows**	**Apps**	**ICC-Flows**	**Endpoints**	**Flows**	**Apps**
IoT-VER	497 (85.84%)	50 (8.64%)	579	90 (0.91%)	4 (5.33%)	49 (65.33%)	75	53 (0.54%)	2,340 (34.89%)	1,952 (29.11%)	6,706	1,682 (17.01%)
GP-2022									420 (30.75%)	619 (45.31%)	1,366	318 (33.58%)

INSIGHTS INTO THE IOT ECOSYSTEM

- What Data Companion Apps Share(RQ3)

Encryption Analysis:

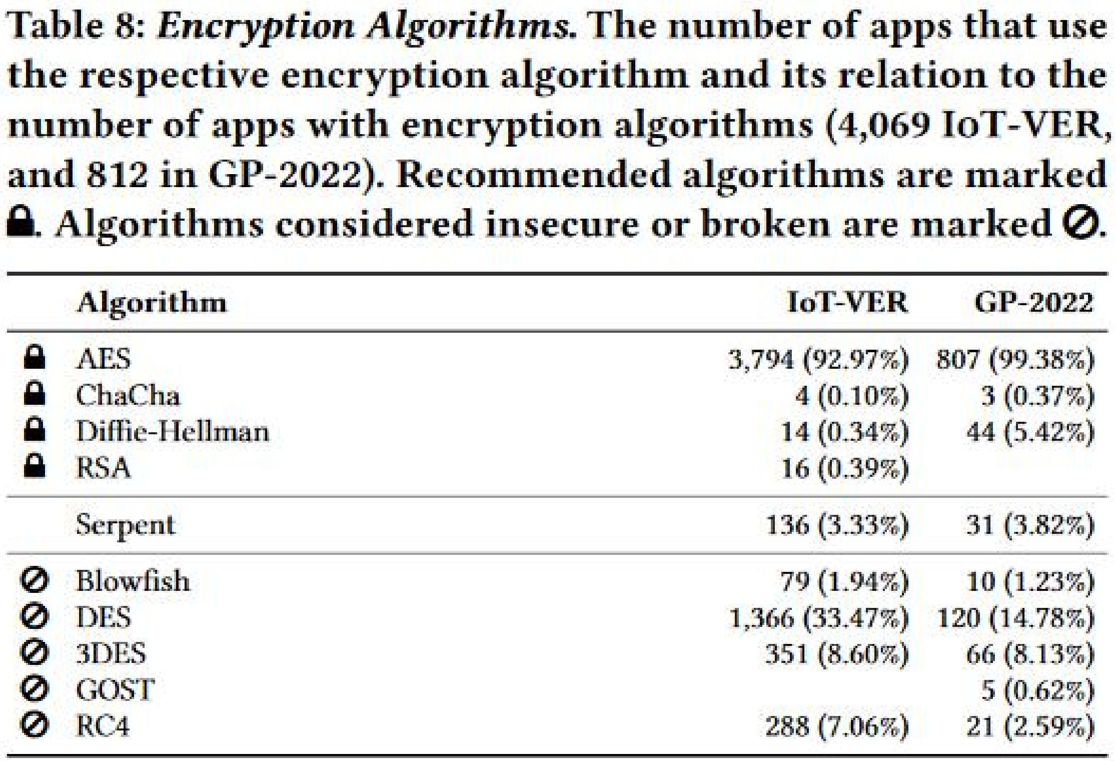


Table 8: ***Encryption Algorithms.*** The number of apps that use the respective encryption algorithm and its relation to the number of apps with encryption algorithms (4,069 IoT-VER, and 812 in GP-2022). Recommended algorithms are marked 🔒. Algorithms considered insecure or broken are marked ⊘.

	Algorithm	IoT-VER	GP-2022
🔒	AES	3,794 (92.97%)	807 (99.38%)
🔒	ChaCha	4 (0.10%)	3 (0.37%)
🔒	Diffie-Hellman	14 (0.34%)	44 (5.42%)
🔒	RSA	16 (0.39%)	
	Serpent	136 (3.33%)	31 (3.82%)
⊘	Blowfish	79 (1.94%)	10 (1.23%)
⊘	DES	1,366 (33.47%)	120 (14.78%)
⊘	3DES	351 (8.60%)	66 (8.13%)
⊘	GOST		5 (0.62%)
⊘	RC4	288 (7.06%)	21 (2.59%)

IOTFLOW VS. DYNAMIC ANALYSIS

- extract requests’ domain names and resource paths

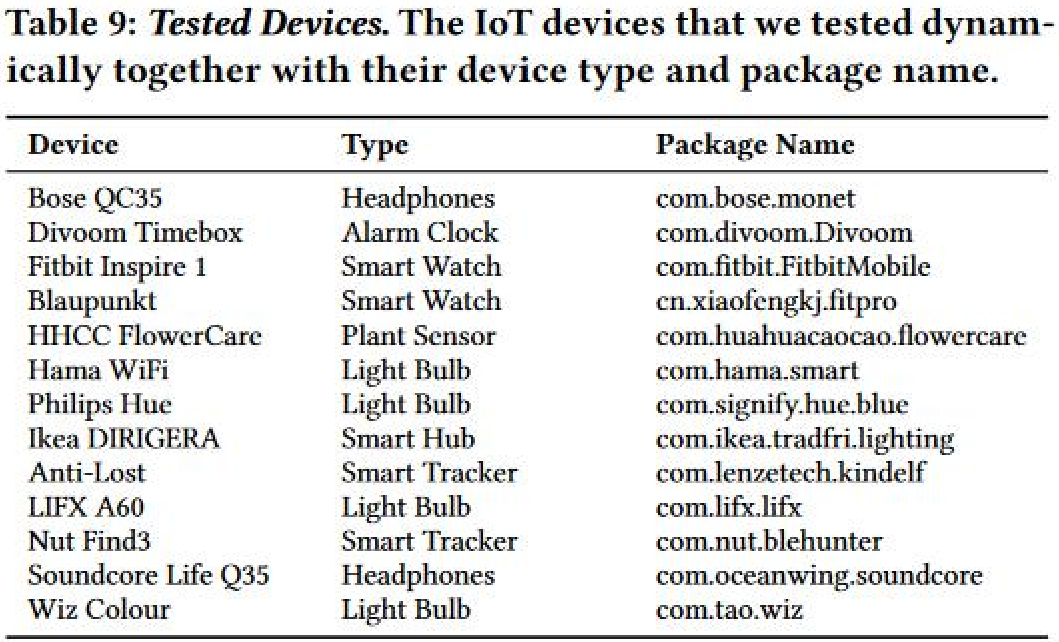
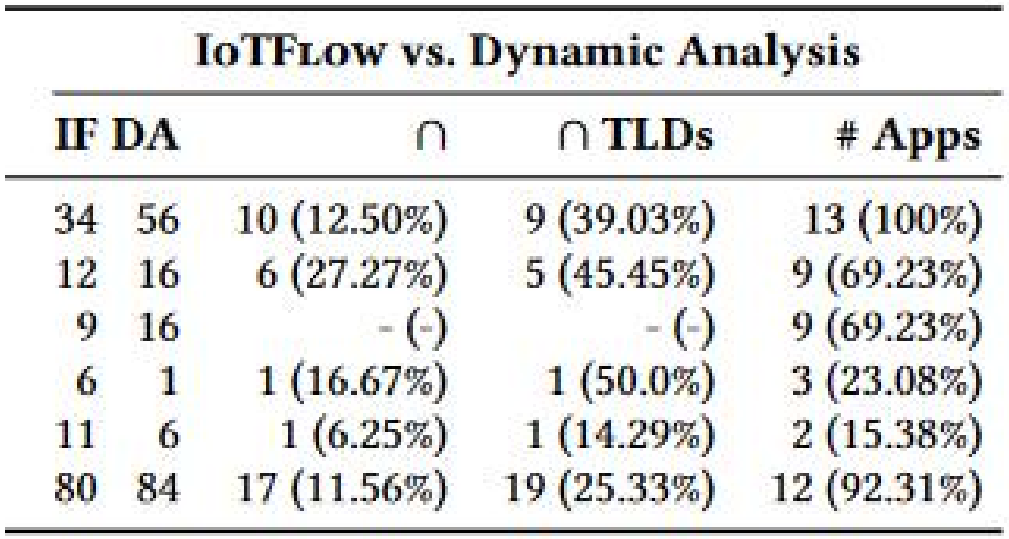


Table 9: *Tested Devices.* The IoT devices that we tested dynamically together with their device type and package name.

Device	Type	Package Name
Bose QC35	Headphones	com.bose.monet
Divoom Timebox	Alarm Clock	com.divoom.Divoom
Fitbit Inspire 1	Smart Watch	com.fitbit.FitbitMobile
Blaupunkt	Smart Watch	cn.xiaofengkj.fitpro
HHCC FlowerCare	Plant Sensor	com.huahuacaocao.flowercare
Hama WiFi	Light Bulb	com.hama.smart
Philips Hue	Light Bulb	com.signify.hue.blue
Ikea DIRIGERA	Smart Hub	com.ikea.tradfri.lighting
Anti-Lost	Smart Tracker	com.lenzetech.kindelf
LIFX A60	Light Bulb	com.lifx.lifx
Nut Find3	Smart Tracker	com.nut.blehunter
Soundcore Life Q35	Headphones	com.oceanwing.soundcore
Wiz Colour	Light Bulb	com.tao.wiz



IoTFlow vs. Dynamic Analysis				
IF	DA	∩	∩ TLDs	# Apps
34	56	10 (12.50%)	9 (39.03%)	13 (100%)
12	16	6 (27.27%)	5 (45.45%)	9 (69.23%)
9	16	- (-)	- (-)	9 (69.23%)
6	1	1 (16.67%)	1 (50.0%)	3 (23.08%)
11	6	1 (6.25%)	1 (14.29%)	2 (15.38%)
80	84	17 (11.56%)	19 (25.33%)	12 (92.31%)

IOTFLOW VS. DYNAMIC ANALYSIS

- IoTFlow Findings :

- 8/13 apps send information via unencrypted HTTP;
- 5/13 apps use hard-coded symmetric encryption keys;
- 2/13 apps send the hardware identifiers (IMEI) to countries outside of the EU;
- 5/13 apps use country-level location information and send this to remote endpoints;
- No apps use hard-coded authentication credentials.

LIMITATIONS

- Impact of Network and Hardware :

- resilience to obfuscation is limited;
- do not support native code;
- not consider code annotations;
- VSA does not supports ICC;
- limit the number of backward steps and set a timeout.

Conclusion

- IoTFlow
- VSA:extract network endpoints and protocols
- VSA with DFA
- analyze 9,889 companion apps and 947 general-purpose apps