Your Privilege Gives Your Privacy Away: An Analysis of a Home Security Camera Service

Jinyang Li*†, Zhenyu Li*†§, Gareth Tyson‡, Gaogang Xie*†
*ICT-CAS, †University of Chinese Academy of Sciences, ‡QMUL, §Purple Mountain Laboratories, China
{lijinyang,zyli,xie}@ict.ac.cn, gareth.tyson@qmul.ac.uk

***Abstract*—Once considered a luxury, Home Security Cameras (HSCs) are now commonplace and constitute a growing part of the wider online video ecosystem. This paper argues that their expanding coverage and close integration with daily life may result in not only unique behavioral patterns, but also key privacy concerns. This motivates us to perform a detailed measurement study of a major HSC provider, covering 15.4M streams and 211K users. Our study takes two perspectives: (*i*) we explore the per-user behaviour, identifying core clusters of users; and (*ii*) we build on this analysis to extract and predict privacy-compromising insight. Key observations include a highly asymmetrical traffic distribution, distinct usage patterns, wasted resources and fixed viewing locations. Furthermore, we identify three privacy risks and explore them in detail. We find that paid users are more likely to be exposed to attacks due to their heavier usage patterns. We conclude by proposing simple mitigations that can alleviate these risks.**

***Index Terms*—Home security camera; Privacy; IoT; Usage pattern;**

I. INTRODUCTION

The majority of Internet traffic is now video, dominated by the likes of Netflix [21] (entertainment content), YouTube [22] (user generated uploads), Periscope [27] (personal social streaming) and Twitch [24] (live e-sports). However, the advent of low-cost Internet-enabled cameras has resulted in the arrival of a new, rather different, type of video streaming service. IoT Home Security Cameras (**HSCs**) are a growing area of online video, with players such as 360 [2], Nest [9], Netgear [10], Hikvision [6] and XiaoMi [12] dominating the market. Although a few years ago these were considered a luxury, they have since entered the mainstream and, with that, we have witnessed emerging privacy and security concerns [33], [16], [1]. Their growth should not be underestimated: The global HSC market is expected to reach $1.3 billion by 2023 [7].

These HSC services differ from traditional video streaming platforms in several ways. First, HSCs are standalone Internet-connected devices, without the need for an attached computer. Their content is streamed directly to a cloud platform, and consequently all content is made remotely accessible, often without any local storage. Second, HSCs are unicast in nature, where content is only available to the owner of the camera. By definition, the content is therefore private and past research into delivery optimization of public content (*i.e.,* caching) may therefore be ineffective. Third, HSCs follow an on-demand model, where video is only streamed when a user requests it, or when motion is observed. This, we argue, may constitute a privacy leak in itself, particularly as HSCs are often installed in intimate locations. For example, an attacker with access to passive network data may be able to infer the camera owner's household activity by inspecting HSC traffic [29], [19]. For most people it is difficult to verify if protections are in place, and therefore this offers strong motivation to study the behaviors of cameras in-the-wild [38].

Thus, we argue that these novel characteristics warrant further investigation. Interesting questions include: (*i*) What are the usage patterns of HSCs? (*ii*) How often do motion-triggered cameras upload videos, and what percentage of them will be watched? (*iii*) How predictable are motion-triggered uploads, and user access patterns? (*iv*) Are there any privacy risks, and could a tractable adversary exploit them? (*v*) What mitigation would address these privacy concerns?

To answer these questions, we perform the first empirical study of a mainstream HSC system in China. We have obtained a unique dataset from a major HSC provider for a week period. It covers 15.4M streams from 211K active users (§II). This data contains a mix of free and premium (paid) users, allowing us to explore a wide diversity of HSC behaviors. We begin by exploring the behaviors of these user types, and perform fitting to identify core user archetypes (§III). We observe a platform dominated by uploads, with motion-triggered cameras streaming a large volume of unwatched content. We then explore a set of intuitive privacy attacks and characterize their efficacy (§IV), discovering a subset of highly regular users for whom we can effectively predict their activities. To briefly summarise, key findings include:

1) Premium users constitute just 59% of all accounts, yet contribute more than 95% of total traffic, dominated by motion-triggered on-demand *uploads*. Despite this, 60% of these videos go unwatched. This waste is largely attributed to a handful of very heavy premium users ($\sim$1/4).
2) As well as dominating upload traffic, premium users are more active than "normal" free users in *viewing*. About 10% of users appear to utilize the HSCs as a regular surveillance service and generate a huge amount of viewing traffic. Such users tend to view their HSC streams from 1 or 2 key (network) locations, and often these are at a different location to the camera.
3) We identify three major privacy risks based on traffic monitoring: (*i*) the traffic surge risk, (*ii*) traffic regularity risk, and (*iii*) traffic rate change risk. While the first risk has been examined in previous studies [19], the

last two are newly explored in this work. These allow us to predict the daily patterns of the camera uploads, and even identify activities on the camera feed. We propose methodologies to infer privacy-compromising information, and explore the risks with both controlled experiments and our dataset.

4) We find that premium users are more vulnerable to privacy risks due to their heavier usage and the exclusive availability of the motion detection mode (this is not available for normal users). The accuracy of predicting the patterns of premium users' upload streams is as high as 0.75 (3× the accuracy for live streams by non-premium users). We propose counter measures to mitigate the risks.

II. Background & Dataset

A. Primer on Home Security Cameras

We first briefly explain the operating procedures of typical HSC services. Upon purchase, the owner of a HSC first binds the camera to their account. The camera receives commands from the servers that are hosted in a cloud operated by the HSC provider. After setup, the user can remotely request a live stream or an archived replay via the cloud servers. It is worth noting that users *never* connect directly to the camera — all video traffic is forwarded via the servers.

Most HSC cameras support two modes of streaming. These two modes are common in major HSC providers, including Nest, Netgear, Hikvision, and 360:

- *Live streaming mode*: The user is able to login and initiate a live stream from the camera in realtime, via the cloud server as an intermediary. The video will *not* be stored anywhere by default.
- *Motion detection mode*: When a motion is detected, an app notification is sent, and the user is then given the option of viewing the stream in real time. Again, nothing will be stored by default.

As the above modes may be inconvenient for users who cannot immediately view streams in realtime, some HSC services offer a third feature for premium users (who pay a fee). This motion detection mode automatically uploads and stores motion-triggered streams to the cloud servers hosted by the HSC provider. These streams contain video footage from a few seconds before the motion begins, until a few seconds after. Users can then replay the video at any time, and a video will be saved for at least 7 days. We term this *replay mode*. Note that other platforms offer 24/7 recording functions too; for example, Nest offers this for premium users (via *Nest Aware* [8]).

B. Dataset Description

Our work relies on a 7 day dataset of log entries (from April 2018) shared by a major Chinese HSC service. The anonymous HSC provider serves hundreds of thousands of users per day and supports all the above features. Note that there were no national holidays or unusual events during the data collection period, giving us confidence that this is representative of a 'typical' week. The dataset covers *all* cameras that were connected to the Internet via a major ISP in China. Every individual log related to these cameras is included in our dataset.[1]

Within the logs, one video view or upload corresponds to one *stream*. A service log is generated for every 30-second *segment* for each stream, so it is reasonable for 1 stream to be related to more than 1 log. In total, we obtain 96,515,229 logs of 15,432,950 streams from 211,806 unique active users who have uploaded at least once (either live stream or replay videos). Of these users, 124,985 (124K) are premium, accounting for 59% of all active users. The remaining non-premium users are referred to as *normal users* throughout this paper.

Each log entry, which corresponds to a 30-second segment, includes three main categories of information:

1) *User-specific information*: the anonymized user ID that is uniquely bound to a registered account, as well as this user's camera(s); the IP address (anonymized using `Crypto-PAn`[3]); the anonymized BGP prefix of the IP address, which is obtained by querying `Team Cymru` [11].
2) *Stream-specific information*: the anonymized stream ID; stream type (*up* for video uploading, *down* for video viewing).
3) *Segment-specific information*: the average bit rate of this segment (kbps); data volume (KB); and timestamps that mark the start and end of the segment.

Alongside our partner HSC service, we have inspected two other popular HSC providers: Nest (popular in western countries) and XiaoMi (popular in China). We confirm that they offer similar services and that they all rely on Variable Bit Rate (VBR) encoding (which is one of the reasons contributing to the privacy issue discussed in §IV).

Ethical issues: We took a number of steps to ensure ethical use of the data shared with us. We have no access to the content of video streams, and can only observe metadata (*e.g.* stream duration). The logs used are routinely gathered for operational purposes, and no extra data collection was triggered. All user information, including user ID, IP address, BGP prefix and even the stream ID, is fully anonymized. We are unable, and not allowed, to link logs to users. Later on (§IV-D), we also leverage volunteers for controlled experiments. The volunteers were aware that we only collect traffic rate information. In addition, the cameras were placed at working areas (rather than homes). Finally, we have reported all potential privacy risks to the service provider, and assisted them in implementing fixes.

III. Exploring User Behavior

Before investigating privacy issues, it is first necessary to understand user behavior. Here, we present a characterization of typical usage patterns in the examined HSC service.

[1]This includes cases where a user is viewing the camera feed from a different ISP.

A. Basic Characterisation

TABLE I: Data volume distribution.

	Normal user	Premium user		
	live stream	live stream	replay	All
Up	1.37%	12.96%	65.89%	80.22%
Down	1.36%	12.77%	5.65%	19.78%
All	2.73%	25.73%	71.54%	100%

We first inspect the data volumes uploaded/downloaded by each user and traffic type. Note that all upload (*up*) streams are initiated by a camera uploading data to the server, whereas all download (*down*) streams are initiated by a user viewing the video. Recall that the *replay* mode is only available to premium users: the replay-up streams are exclusively triggered by motion seen by the camera, while the replay-down streams are triggered by premium users watching the relayed videos.

Table I summarizes the results. The platform is dominated by traffic generated by services supported for premium users. These premium accounts tend to be heavy users: they generate 97% of the traffic. This is caused by the dominance of motion-triggered automatic uploads — replay-up streams contribute over 2/3 of the total workload. As a striking contrast, only ∼5% of download streams come from this source (replay). In fact, we see that a remarkable 60.24% of video uploaded is never downloaded, suggesting a significant waste in both network and storage resources. This is particularly as replay-up streams, on average, last longer and have larger volumes (median around 4MB). In contrast, the replay-down streams are shorter and with smaller sizes (0.65MB). This drives the asymmetry of the workload in Table I.

B. Characterising Live Stream Mode

We next inspect the generation and consumption of live content by users. Note that these include *both* normal users and those with premium accounts.

Overview of Live Users: We first count the number (termed *frequency* hereafter) and the *total* duration of the live-down streams generated per user.[2] As expected, premium users are more active than normal ones: the median frequency of streams is 7 for premium users *vs.* 2 for normal users. This observation is also mirrored when inspecting duration: the median total duration of normal users and premium users are 90 seconds and 435 seconds respectively. Nevertheless, some normal users generate over 100 streams and watch over 5 hours during the observation period.

Clustering Live Stream Users: The above suggests a diversity of users groups. Thus, we proceed to identify core behavioral types within the user population. To this end, we fit the *frequency* and *total duration* statistics of all users to a 3-component Gaussian Mixed Model (GMM) [5].

We experimented with a number of configurations from 2 to 5 GMM components, and selected 3 based on the balance between relatively small AIC (Akaike Information Criterion) and not too small components ($\alpha < 0.001$). Table II presents the clusters identified, alongside their fitting results. The results expose three main sub-populations, shared across both normal and premium users. We term these *light* (L), *medium* (M) and *heavy* (H). Light users use the live streaming service rarely. In contrast, medium users tend to check their camera feeds daily. The heavy users deviate significantly from the average, with extremely regular viewing patterns. It seems likely that heavy users use HSCs as a (potentially commercial) surveillance camera service, and the cameras likely cover high value regions (*e.g.* in a shop).

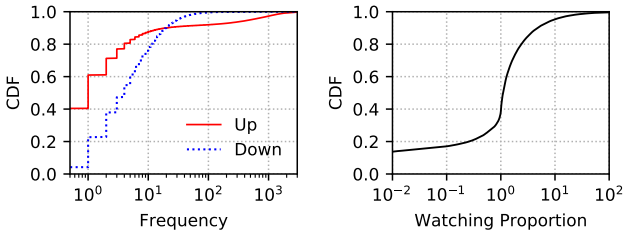
[2]Note that upload and download streams are approximately symmetrical in terms of frequency and duration in the case of live streaming.

TABLE II: User clustering for live streaming.

	no.	freq.	dura.(s)	α	feature
Normal	#1	1.5	41.7	0.49	Light
	#2	4.8	287.1	0.42	Medium
	#3	16.9	3,719.1	0.09	Heavy
Premium	#1	3.4	129.5	0.45	Light
	#2	18.4	1,422.8	0.44	Medium
	#3	64.6	22,217.2	0.11	Heavy

C. Characterising Replay Mode

We next inspect users of the video *replay* service. This is *only* available to premium users (59% of population). When activated, the replay mode automatically uploads all motion-triggered content to the cloud for later on-demand access.



(a) Frequency (b) Watching Proportion

Fig. 1: Per-user characteristics of replay mode.

Overview of Replay Users: Figure 1(a) presents the number of replay streams per user. It shows that viewing replay content is significantly more frequent (per-user) than uploading. This was surprising as Table I found that the majority of data is generated by replay uploads. However, the reason is clear: whereas, by volume, the majority of data uploaded is replay traffic, this is driven by a small subset of streams. A remarkable 40% of (premium) users *never* upload a video for replay during the week, leaving a small subset of users with exceedingly high upload rates — 97% of the total replay upload traffic is generated by the top 5% of premium users. This indicates a mix of camera installations, with many fitted in locations with very limited motion and others in high activity zones.

This disparity raises the question of whether users actually view the videos uploaded. To measure this, we define the *Watching Proportion* as the ratio of a user's total replay-

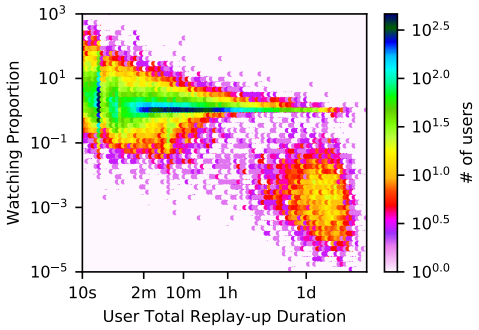


Fig. 2: Distribution of premium users, based on their watching proportion and total replay-up duration. The scales of axes and color bar are logarithmic.

down duration to total replay-up duration.[3] Figure 1(b) depicts the distribution of watching proportions. While the median is around 1, it is less than 10^{-2} for 13.8% of users, meaning that they watch far less than their cameras upload. Nevertheless, a number of users have a watching proportion in excess of 1, indicating that they repeatedly watch the same streams.

Figure 2 further examines the correlation between the uploaded volume and the watching proportion. We can observe two typical patterns. Most users are within the first group (upper left), where the total uploaded duration is relatively low, yet the watching proportion is relatively high (around 1). The remaining users are within the second group (lower right), where the total replay-up duration is relatively high, but the watching proportion is relatively low. The actions of these users result in over 60% of network and storage resources being wasted, which creates strong motivation for a more informed upload strategy that moves beyond motion triggered uploads alone.

Clustering Replay Users: The above shows a wide range of behavioral types. Thus, we repeat our earlier user clustering process. Here, we use total replay-up duration and watching proportion of active premium users to fit a 3-component GMM.[4] The fitting result is shown in Table III. This exposes three broad categories of users, which we index as *light* (L), *medium* (M) and *heavy* (H) for watchers (W) or uploaders (U), respectively. Thus, each user cluster is tagged with both the watching and upload behaviours.

TABLE III: Clustering active premium users.

no.	up dura.(s)	watch. prop.	α	feature
#1	46.8	20.6	0.11	LU-HW
#2	298.7	1.4	0.65	MU-MW
#3	94,823.3	0.5	0.24	HU-LW

The majority (around 2/3) of premium users fall into the Medium Uploader and Medium Watcher category (MU-MW). They keep the best balance between upload and watching rates. The next most populated group are those who are Heavy Uploaders but Light Watchers (HU-LW). Such users are most costly to the system, as they consume large amounts of network and storage, yet do not benefit from them. At the opposite extreme, the smallest group are those that have a low rate of uploads, but a high rate of watching (LU-HW).

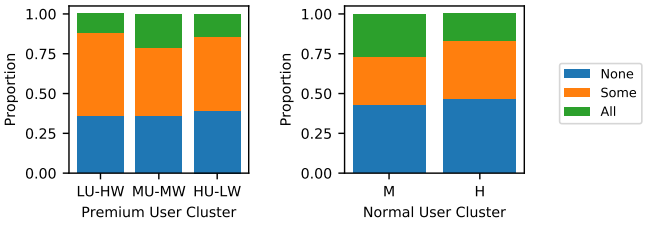
[3]We only include premium users who have uploaded replay video for at least 10 seconds.

[4]We decided the component number here using the same approach as mentioned in §III-B.

D. Characterising Viewing Locality

We next proceed to explore *where* streams are uploaded and consumed from. This is particularly important for QoE improvement, *e.g.* via caching or pre-fetching.

On-site* vs. *Off-site Access: Since we are interested in the network footprint of users, we use the BGP prefix of the user's IP address to represent the user's location. We represent proximity as a binary metric where we test if the camera and viewer are located within the same prefix. We refer to accesses that occur from the same prefix as the camera as *on-site*, and similarly, we denote accesses that occur from a different prefix as the camera as *off-site*.



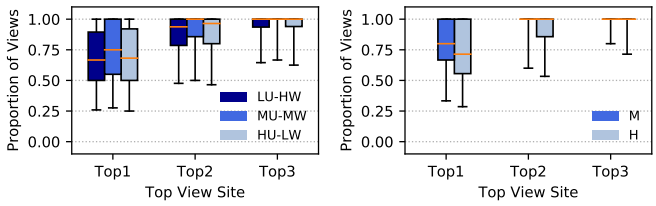
(a) Premium user (b) Normal users

Fig. 3: The proportion of users who have watched a video from the same BGP prefix as the camera for none, all or some of the views in each user component.

To inspect the locality patterns of the different user groups, we take the premium user categories from Table III, as well as the normal user categories from Table II. We exclude any users who have fewer than 3 views in the observation period, to avoid bias caused by the sparse sampling of irregular users. Figure 3 presents the breakdown of on-site *vs.* off-site views for each group of users. *None* indicates that no view come from the same prefix; *some* indicates that a fraction (> 0) of views come from the same prefix; and *all* indicates that all user views emanate from the same prefix as the camera.

Users exhibit similar behaviors across all usage groups. About 30% of examined users consume no streams on-site. This is rational, as there is perhaps little sense in accessing camera feeds from the same site in many cases. The remaining users may experience local access under several situations where a single site covers a large area (*e.g.* factory) or where users employ cameras for monitoring local activities (*e.g.* sleeping children).

User Mobility: We finally inspect how mobile users are, *i.e.,* whether users always view from the same location. To this end, we compute for each user the proportion of views that happened at the top k locations,where $k \in \{1, 2, 3\}$. Users are again grouped based on the earlier clustering results.



(a) Premium user (b) Normal user

Fig. 4: The proportion of views that happened at the top k locations for each user, where $k \in \{1, 2, 3\}$.

Figure 4 presents the fraction of views from the top k locations per-user as a box plot. The majority of users view primarily from their top 1 or 2 locations. The median fraction of views in the top 1 location and top 2 locations for premium users is about 0.7 and 0.95, respectively. Although normal users tend not to watch the live streams on-site (see Figure 3), they are more likely to view the content at a single location than premium users: over 40% of normal users watch all streams at the top locations, while this number for premium users is only about 30%. This indicates that users may not move often (*e.g.* staying at their office). These observations imply the possibility of predicting users' next viewing location and perform pre-loading of material for improved QoE.

E. Takehome Messages

We make three notable observations: (*i*) *Wasted resources*: premium users generate the majority of the workload (97.27%), largely due to the exclusive availability of replay mode. This results in 60% of the uploaded videos going unwatched. The waste is attributed to a handful ($\sim$1/4) of heavy premium users, who have cameras with high levels of motion triggered uploads (see Table III). (*ii*) *Distinct usage patterns*: premium users are more active than normal ones; these account types have a mix of access patterns. (*iii*) *Watching locality*: users tend to view streams from 1 or 2 key locations, with a sizeable portion of users watching streams from individual remote sites, suggesting a surveillance use case.

The above indicates that a set of simple innovations could streamline HSC operations. Most notably, HSC services could benefit from on-demand (rather than real time) uploads for the replay mode. This is because the majority of motion triggered uploads go unwatched, and therefore local recording (with on-demand uploads) could offload a significant volume of unwatched uploads from the network. There are a number of ways this could be implemented without negatively impacting customer experience. For example, HSCs could upload the first 1 minute of a stream by default, allowing users to 'preview' the content. If a user wishes to continue viewing (this applies to less than 20% of streams), the remaining video can be requested from the HSC. This could also be mixed with more sophisticated methods of delivery, whereby only users predicted to consume content have their previews uploaded. The predictability of *where* users view content from means these videos could even be pre-fetched (*e.g.* to the top 1 or 2 locations). This will reduce the startup delay, and thus improve the QoE.

IV. Exploring Privacy Risks

HSCs are always-on sensors designed to actively detect movements in private places. This differs substantially from other traffic-inference attacks (*e.g.* monitoring a user's Netflix usage) as variations in the camera feed (*e.g.* bit rate changes) can expose specifics about behavior, such as exposing Activities of Daily Living (ADL). In this section, we study potential privacy risks and solutions.

A. Adversary Model

We adopt a similar attack model to [19], where the attacker has similar capabilities to an ISP. Our adversary is able to monitor all network traffic in and out of home gateway routers (*e.g.* via WiFi sniffing). The attacker can utilize metadata including IP packet headers and traffic rate, which are routinely gathered by major ISPs for operational purpose. Due to encryption, the payload data is unavailable though. Note that this adversary model is quite feasible: Identifying HSC traffic from bulk traffic flowing through the home gateway router is well studied [19], [26], [15], [31]. Based on our adversary model, we identify 3 major privacy risks:

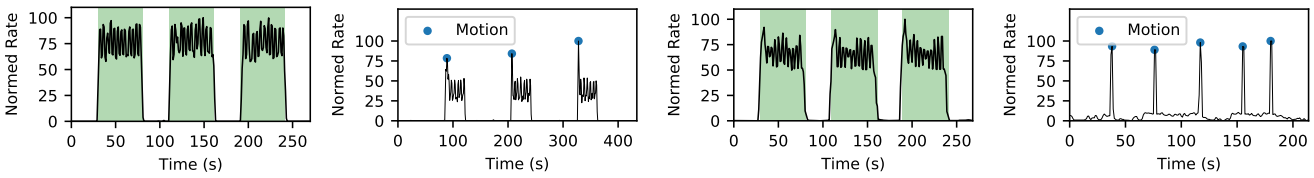
1) *Traffic surge risk*: If the traffic rate of a camera surges from its base rate, this indicates that the video is being uploaded. In the case of motion detection mode, this indicates activity near the camera zone.
2) *Traffic regularity risk*: After a period of observing surges, an attacker may be able to infer a user's daily patterns. For example, a camera consistently uploading motion-triggered video at 18:00 might indicate that family members arrive home at that time.
3) *Rate change risk*: The different activity patterns of the photographed subject will result in different HSC traffic rates. Based on these rate variations, an attacker may be able to infer the intensity of activity being undertaken, and even the types of activity.

We also assume a targeted attacker, who has an approximate understanding of the camera's context (*e.g.* if it is mounted in a house), and who owns it. Although the examined HSC is based in China, we note that these attacks are equally applicable to other cameras that use Variable Bit Rate (VBR) encoding, including Nest [19].

B. Traffic Surge Risk

A *traffic surge* is a point in time where the bitrate of a camera's feed increases dramatically. This creates a privacy risk that is inherently part of the transmission schemes of HSC services (since the camera only uploads when some certain functions are activated and stays idle the rest of time). This constitutes the foundation stone of all subsequent attacks.

Methodology: A traffic surge may be triggered by one of two events: (*i*) a user viewing the stream live; or (*ii*) the motion capture triggering an upload for later replay (in the case of premium users). In both cases, once an upload is



(a) Examined HSC live-up (b) Examined HSC replay-up (c) Nest live-up (d) Nest motion detection

Fig. 5: Normalized traffic rates for the examined HSC and Nest HSC in live streaming mode and motion detection mode.

triggered, a significant surge in traffic is observed. Trivial peak identification across the bit rate time series can therefore be used to verify if the camera is recording.

It is noteworthy that motion-triggered uploads correspond to the real time movements in private places, while live uploads may be triggered by the viewing requests of users. As such, an attacker who can differentiate between live and motion triggered uploads is threatening. Inspection of the traffic reveals that it is possible to differentiate between live and motion triggered uploads. This is because motion triggered uploads always start with an initial peak, due to the uploading of a motionless video at the beginning of the transmission. We can identify peaks from the traffic rate time series, $S = \{s_1, s_2, ..., s_n\}$, where s_i represents the bit rate observed at the time point i. The maximum value within S is denoted as p. We assume the traffic rates follow a Gaussian distribution, then we construct a second set with p removed, $S' = \{s|s \in S, s \neq p\}$. We calculate the standard deviation (σ) and the average of S' (*i.e.,* $\bar{S'}$). Values less than $1.96\sigma + \bar{S'}$ occupy approximately 95% of the probability space [14]. If the maximum value within S is greater than $1.96\sigma + \bar{S'}$, this time series indicates a replay upload. Otherwise, it indicates a live stream. We have also verified the presence of this traffic surge risk in two other HSCs: XiaoMi and Nest.

Risk Exploration: To measure the efficacy of this risk, we connect the examined HSC and a Nest HSC via a WiFi access point, and perform packet capture. We leave the cameras dormant before starting to view the stream after 25 seconds. Figure 5(a) and Figure 5(c) present the time series of the normalised bit rate for the examined HSC and Nest HSC respectively. When we start to watch a live stream, both cameras switch from motion detection mode (white) to live streaming mode (green): this is shown by the sudden spike in bitrate. When we finish watching, both cameras switch back to motion detection mode and stop the transmission. We obtained similar results for XiaoMi HSC.

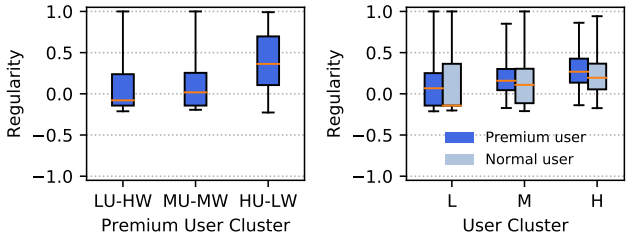
To confirm our ability to differentiate live and motion-triggered uploads, we repeat the above setup with premium user accounts and periodically simulate motion in front of the cameras. Figure 5(b) presents the result for the examined HSC, confirming that motion triggers an immediate upload. For the Nest HSC in Figure 5(d), we can also observe clear traffic spikes when the motions were triggered. Importantly, the traffic peaks at the beginning of transmissions are notable and trivial to detect using our methodology (*i.e.,* we obtain 100% accuracy).

C. Traffic Regularity Risk

The above shows that a passive attacker can infer (*i*) if a camera is uploading; and (*ii*) if that content is being streamed for motion capture replay. We next explore if this can be exploited to identify regular patterns in a user's behavior. For example, if a camera in a house regularly initiates a motion-triggered stream at 7AM, an attacker could infer that this is the time the owner awakes. Such information could be used to enable physical attacks, *e.g.* burglary.

Methodology: To test if such regularity can be inferred from network traffic, for each user, we define the *Regularity Value* (RV) as follows. We first count the upload duration per hour across the observed period. This yields one 24-element vector per day. We then filter out the days without any uploading. For each of the remaining vectors, we compute the moving average with a window size of 3 hours, in order to compensate for variations in a user's daily activity (*e.g.* viewing a stream at 10:03 rather than 9:59). We then calculate the pairwise Pearson Correlation Coefficient between all possible pairs of a user's daily vectors; we refer to the final per-user average as the regularity value.

The regularity value ranges from from -1 to 1. If the value is positive, the upload patterns of the user are regular (the closer to 1, the stronger the regularity). This indicates the daily patterns of the user can be inferred. Note that the regularity value is not an attack in itself — it quantifies how susceptible users in our dataset are to this type of analysis by an attacker.



(a) Replay video (b) Live streaming

Fig. 6: Distr. of regularity value of different clusters of users.

Risk Exploration: We next test the regularity of users in our dataset. We only inspect those users who have uploaded for at least 2 days within the observed week. This includes the 67% of all users who have performed live uploads, and the 51%

of premium users who have motion-triggered replay uploads. Figure 6 presents the distribution of regularity values across normal and premium users. Again, we separate these users into their categories as identified via our earlier GMM clustering.

Figure 6(a) reveals significant diversity across the different clusters for motion-triggered replay predictability. The majority of LU-HW users and MU-MW users show little-to-no regularity. Their median regularity is near to 0, indicating that their daily patterns are difficult to predict. This is partly driven by their very nature, which consists of limited usage. In stark contrast, HU-LW users show stronger regularity, with the 75-th percentile as high as 0.69. In the case of live videos (Figure 6(b)), in all clusters, the regularity of premium users is higher than that of normal users. Furthermore, as the usage frequency becomes higher (for both premium and normal users), the regularity becomes higher. This intuitive finding indicates that heavier users are easier to predict.

To find a reasonable threshold ($thresh$) for what might be considered *strongly predictable*, we use the approach suggested in [28]. We fit all positive user regularity values into a 2-component GMM, and we define the intersection of the 2 resulting components as the threshold. The resulting value of $thresh$ is 0.34 in both the cases of normal and premium live uploading, and 0.35 in the case of replay uploading. Consequently, 17.4% of replay-up premium users, 18.7% of live-up premium users and 11.5% live-up normal users can be considered as strongly predictable.

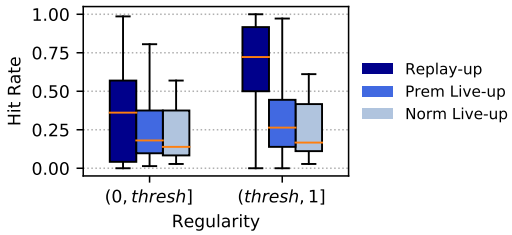


Fig. 7: Distribution of accuracy of upload behavior prediction (hit rate) for users of different ranges of regularity value.

Predicting Activity: We next confirm that this regularity can be exploited to predict upload patterns. We extract all users who have uploaded video data on *all* 7 days: 9,912 (8%) premium users for replay-up and 14,826 (7%) users for live-up. We then use their first four days to fit a Holt-Winters model [4] to predict the following three days time series (seasonal period was set to 24). Note that the time series are binary (0 for no upload in hour, 1 otherwise). We compute the prediction accuracy as $hit\ rate = 1/n * \sum_{i=1}^{n} hr_i$, where hr_i is 1 if the prediction for the i-th time slot is correct (0 otherwise), and n is the number of hours under prediction (*i.e.,* $3 * 24$).

Figure 7 presents the distribution of hit rates across all users. We perform separate predictions for replay, premium live and normal live streams. We then split these users into the two categories of regularity: $0 < RV \leq thresh$ and $RV > thresh$. Unsurprisingly, the more regularly the user uploads, the higher the hit rate is. The predictions are most accurate for replay video uploading, where the hit rates are as high as 0.75 (3x the accuracy in the cases of live uploads). This is likely because it depends solely on motion, rather than user behavior. This confirms that, particularly for heavy users, motion-triggered uploads *do* have the capacity to allow attackers to predict future activity. This could be an effective tool for identifying the best time for physical attacks, *e.g.* burglary.

D. Rate change risk

Finally, we explore the potential to identify activity changes on a camera feed via bit rate monitoring, *e.g.* identifying a person shifting from sitting to walking.

Methodology: We take inspiration from Li *et al.* [29], who proved it possible to identify activities by monitoring the bitrate feed from a video stream. Their approach involves first identifying video segments (via change points), and extracting key features. By manually labelling each segment with their associated activities (*e.g.* eating, dressing, styling hair), they then train classifiers to identify activities in other feeds. With these results in mind, we next test the number of potential activity segments that can be extracted from the video streams in our dataset (1 segment maps to one activity [29]). Although we cannot associate these segments with their respective labels (*e.g.* eating), this does offer an upper-bound on how many activities could be extracted. To do this, we convert all streams into a bit rate time series, and then utilize Bayesian Online Change Point Detection (BOCPD) [13] to identify each segment in a camera's feed.

BOCPD assumes that a sequence of observations $(x_1, x_2, ..., x_t)$ contain several non-overlapping partitions ρ [20]. For a given time series, BOCPD computes the *run length*, which represents the number of time steps since the last detected change point (denoted as r_t at time t). The probability distribution of r_t can be computed using a recursive algorithm as follow:

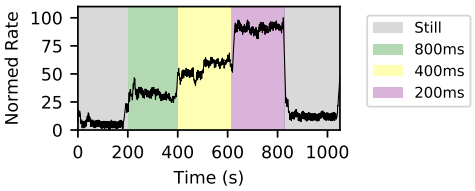
$$P(r_t|x_{1:t}) = \frac{\sum_{r_{t-1}} P(r_t|r_{t-1})P(x_t|r_{t-1}, x_t^{(r)})P(r_{t-1}, x_{1:t-1})}{\sum_{r_t} P(r_t, x_{1:t})} \quad (1)$$

where $x_t^{(r)}$ indicates the set of observations associated with the run r_t. $P(r_t|r_{t-1})$, $P(x_t|r_{t-1}, x_t^{(r)})$ and $P(r_{t-1}, x_{1:t-1})$ are prior, likelihood, and recursive components of the above formulation. The conditional prior has non zero mass under only 2 circumstances:

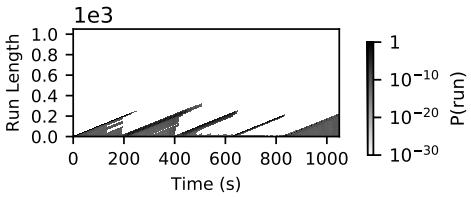
$$P(r_t|r_{t-1}) = \begin{cases} H(r_{t-1}+1) & if\ r_t = 0 \\ 1 - H(r_{t-1}+1) & if\ r_t = r_{t-1}+1 \\ 0 & otherwise \end{cases} \quad (2)$$

$$H(\tau) = \frac{P_{gap}(g = \tau)}{\sum_{t=\tau}^{\infty} P_{gap}(g = t)} \quad (3)$$

In the above, $H(\tau)$ is the *hazard function* [25]. The likelihood term $P(x_t|r_{t-1}, x_t^{(r)})$ therefore represents the probability that the most recent datum belongs to the current run.



(a) Traffic rate timeserie



(b) Change points detection

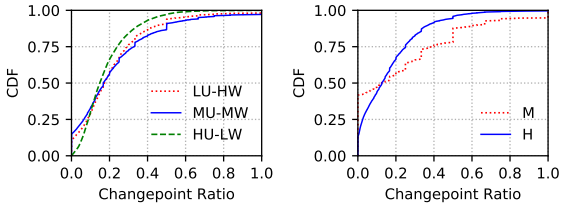
Fig. 8: Traffic rate changes of camera when GIFs shifting period changes and corresponding probability of run length.

Adams and MacKay [13] do not specify an exact method to identify change points after calculating the run length distributions. Thus, we propose a statistical way to identify change points: after obtaining all the run length distributions, we fit all probability values $P(r_t = 0), t \in \mathbb{N}$ into a Gaussian distribution. Since for all Gaussian distributions, 95% of the area is within 1.96 standard deviations (σ) plus the mean (μ) [14], we label any time step t as a change point when $P(r_t = 0) > \mu + 1.96\sigma$ holds. Another benefit of BOCPD is that it is effective in cases where an attacker only has access to periodic (smoothed) bit rate samples (*e.g.* every 30 seconds). In such cases, each sample x can be expanded across the time period by a Poisson distribution of $\lambda = x$, namely $P(\lambda = x)$.

For context, Figure 8 highlights the outcome of this process: we placed an examined HSC in live streaming mode, and presented it with an animated GIF that flickers between all black and all white at different intervals (still, 800ms, 400ms and 200ms). Figure 8(a) highlights the bit rate, whereas Figure 8(b) confirms the correct calculation of the run length. Note that it is out-of-scope to perform the mapping between these runs and their underlying activities, as it is necessary for an attacker to first collect ground truth mapping data for training purposes [29]. Thus, we emphasize such an attack could only be realised by a highly equipped adversary with the ability to contextualize the segments.

Risk Exploration: To gain an idea of the number of *potential* activities that can be extracted from a camera feed, we measure the number of BOCPD segments in each camera's stream. To measure this, we define the *Change Point Ratio* for a stream as $R_u = \frac{1}{n}\sum_{i=0}^{n}\frac{C_i}{P_i}$, where P_i and C_i are the number of data points and the number of identified change points in stream i's rate timeserie respectively; n is the number of up-streams generated by user u.

We report the distribution of change point ratios for each users in Figure 9, where we again separate users into their categories as identified via our earlier GMM clustering. The



(a) Premium user (b) Normal user

Fig. 9: Distribution of users' change point ratio.

change point ratio distribution of the 3 groups of premium users is quite close. The most noteworthy is that under 1% of HU-LW users have never had a change point (this is about 10% in the other two relatively inactive uploading premium user groups). This indicates that the remaining majority of users *do* have activity changes within the streams. Compared with premium users, more normal users have a zero change point ratio (40% medium normal users and 20% high normal users), implying very few activity changes in their streams. This distinct behavior is partly because only live streaming mode is available for normal users, which results in less regular activity than motion-triggered capture. Nevertheless, with appropriate training data, this implies that the attack detailed in Li *et al.* [29] would have widespread applicability.

Activity Switch Patterns: We are also interested in users' *activity switch patterns*, since people often do things in a certain logical order (*e.g.* washing their hands before dinner). These sequences could therefore add context to any inferences performed by an attacker. We use the difference between average traffic rates of adjacent activities to empirically inspect users' activity switch patterns. Here, we separate activities using BOCPD as discussed above, and take the average bit rate from each segment.

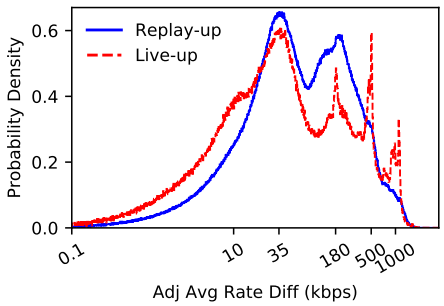


Fig. 10: Distribution of traffic rate difference between adjacent activities. The x-axis is in log scale.

Figure 10 presents the distribution of bit rate changes between consecutive activity segments. As we can see for both replay and live stream types, 4 peaks exist (35kbps, 180kbps, 500kbps and 1,000kbps), which seemingly indicates 4 major activity switch patterns. The four peaks are more pronounced in live streams, whereas there are two more prominent peaks (35kbps and 180kbps) for the replay uploads. Inspired by [29], we conduct a series of short controlled experiments to make

the first attempt to explore the potential corresponding switch patterns of each peak. The experiments were carried out in two illumination environments (natural light and lamplight) by 2 volunteers (1 female and 1 male) from different distances from the HSC (1m, 2m, and 3m). The volunteers performed 3 kinds of activities: (*i*) *gentle*: reading, sitting; (*ii*) *medium*: dressing and stretching; and (*iii*) *dramatic*: walking, exercising. Each activity was repeated for 5 times and lasted 30 seconds. We asked the volunteers to move between each activity, and, in total, the experiments are repeated over 500 times.

We observe that slight activity changes (*e.g.* standing up) will result in a *35kbps* peak, while a *180kbps* peak often happens in the case of distance changes (*e.g.* walking away from the HSC), or relatively dramatic activity switches (*e.g.* getting up from a chair and walking). Finally, *500kbps and 1,000 kbps* peaks only happen in extreme cases of motion, *e.g.* somebody suddenly running into an empty room. Although, we cannot link these activities into the bitrate changes observed in our dataset, our controlled experiments do show the possibility of linking the changes to users' activities, and shed light for future work. Furthermore, it is clear that this approximate taxonomy could be used by attackers to infer a general category of activity taking place on a camera feed.

E. Counter Defences

The root cause of the three risks is that there is a correspondence between the traffic rate and the working state of the camera. Our results show that premium users are most at risk, due to their heavier usage pattern and the motion triggered replay feature available to them. It is therefore necessary to mitigate this correspondence. In the simplest case, this could be done by artificially triggering camera activity to introduce noise to any inferences, for example, by the user directly placing a moving object (*e.g.* clock) in front of the camera for motion triggered recording. This, however, is undesirable for several reasons, not least because it would waste resources: in our controlled experiments, a moving clock generates over 50MB of wasted traffic per hour. Thus, a superior option would be for each HSC to randomly generate streams, thereby undermining the attacks. Such streams could be tagged in order to inform the server to discard them. Notably, the times, duration and traffic rate pattern must be random. HSCs could also perform traffic shaping to flatten spikes in the bit rate [17], [37]. Note that half of the users wait at least 10 minutes before viewing newly uploaded replay videos, suggesting that such traffic shaping could be easily performed without an adverse impact on user experience.

V. RELATED WORK

Privacy Leakages from IoT Traffic: In spite of encryption in IoT, several recent studies have shown the possibility of privacy leakages from application traffic [33]. Li *et al.* [29] showed the possibility of inferring Activities of Daily Living from encrypted surveillance video traffic. Apthorpe *et al.* [19] made similar observations for several IoT devices that include Nest HSC. Wood *et al.* [35] investigated medical IoT devices that may reveal sensitive medical conditions and behaviors. Copos *et al.* [23] presented a scheme that could infer whether a home is occupied by parsing characteristics of the network traffic from smart thermostats. These studies mainly rely on active measurements with a few IoT devices, whereas our work leverages the logs of over 200K home security cameras.

To address the possible privacy leakages in IoT applications, several countermeasures have been proposed. Apthorpe *et al.* [18][17] proposed strategies to protect IoT device consumers from side-channel traffic rate privacy threats, including traffic shaping and tunneling. Zhang *et al.* [37] proposed to reshape packet features through dynamically scheduling packets over multiple virtual MAC interfaces, in order to obscure the features of the original traffic. Some of these solutions, like traffic shaping, can also be applied to HSCs to preserve privacy.

User Behavior Inspection: Xu *et al.* [36] investigated IP cameras without password protection to examine their usage patterns and vulnerabilities. Our results also have similarities with mobile personal livecast systems [30], [32], [34], which also exhibit small streams. Like us, [32] found that many broadcasts also go un-watched, resulting in resource waste.

VI. CONCLUSION

This paper has presented a large-scale study of a major HSC system, highlighting several key findings. Around 95% of replay upload traffic is generated by the top 5% of cameras (largely motion-triggered uploads). This workload results in a significant portion of content going unwatched and therefore wasting resources. These previously unknown patterns contribute to the growing body of work focused on optimizing home IoT devices. We also inspected the privacy implications of using HSCs, driven by the close alignment between real-world activities and subsequent network traffic. We confirmed a range of inferences, and have offered an upper bound for the predictability of user patterns. The susceptibilities of users to these threats differ, and we identified a subset of heavy users who are most at risk. Although we have only empirically tested these concepts on the examined HSC, we note that our privacy attacks are equally applicable to most other HSC services, due to their use of Variable Bit Rate encoding.

In our future work, we will explore how user trends generalize and evolve over a longer duration, and we will expand our controlled experiments (§IV-D) to better understand the inferences that rate changes may enable. We are also working with the service team of the examined HSC to explore the discussed implications. We conclude by stating that HSCs have become a commodity which will likely increase in usage. As they are often placed in intimate locations, it is important we continue to investigate their activities and potential risks.

ACKNOWLEDGMENT

This work was supported in part by National Key R&D Program of China: 2018YFB1800201, the NSF of China (NSFC): 61725206, the Youth Innovation Promotion Association CAS. The corresponding author is Zhenyu Li.

REFERENCES

[1] Hacked cameras, dvrs powered today's massive internet outage. https://krebsonsecurity.com/2016/10/hacked-cameras-dvrs-powered-todays-massive-internet-outage/, 2016.

[2] 360 smart camera. http://jia.360.cn/, 2018.

[3] Crypto-pan. https://www.cc.gatech.edu/computing/Networking/projects/cryptopan/, 2018.

[4] Exponentialsmoothing in statsmodels's documentation. https://www.statsmodels.org/dev/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing.html, 2018.

[5] Fit gaussian mixture model to data - matlab fitgmdist. https://ww2.mathworks.cn/help/stats/fitgmdist.html, 2018.

[6] Hikvision ezviz camera. https://www.ys7.com/, 2018.

[7] Home security camera market research report- global forecast 2023. https://www.marketresearchfuture.com/reports/home-security-camera-market-3787, 2018.

[8] Nest aware. https://nest.com/cameras/nest-aware/, 2018.

[9] Nest cam indoor. https://nest.com/cameras/nest-cam-indoor/overview/, 2018.

[10] Netgear arlo camera. https://www.arlo.com/en-us/, 2018.

[11] Team cymru. https://www.team-cymru.com/, 2018.

[12] Xiaomi smart camera. https://www.mi.com/micamera/, 2018.

[13] R. P. Adams and D. J. MacKay. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*, 2007.

[14] D. G. Altman and J. M. Bland. Standard deviations and standard errors. *Bmj*, 331(7521):903, 2005.

[15] S. Aneja, N. Aneja, and S. Islam. Iot device fingerprint using deep learning. *arXiv: Networking and Internet Architecture*, 2019.

[16] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou. Understanding the mirai botnet. In *SEC'17 Proceedings of the 26th USENIX Conference on Security Symposium*, pages 1093–1110, 2017.

[17] N. Apthorpe, D. Y. Huang, D. Reisman, A. Narayanan, and N. Feamster. Keeping the smart home private with smart(er) iot traffic shaping. *CoRR*, abs/1812.00955, 2018.

[18] N. Apthorpe, D. Reisman, and N. Feamster. Closing the blinds: Four strategies for protecting smart home privacy from network observers. *arXiv preprint arXiv:1705.06809*, 2017.

[19] N. Apthorpe, D. Reisman, and N. Feamster. A smart home is no castle: Privacy vulnerabilities of encrypted iot traffic. *CoRR*, abs/1705.06805, 2017.

[20] D. Barry and J. A. Hartigan. Product partition models for change point problems. *Ann. Statist.*, 20(1):260–279, 03 1992.

[21] T. Böttger, F. Cuadrado, G. Tyson, I. Castro, and S. Uhlig. Open connect everywhere: A glimpse at the internet ecosystem through the lens of the netflix cdn. *ACM SIGCOMM Computer Communication Review*, 48(1):28–34, 2018.

[22] A. Brodersen, S. Scellato, and M. Wattenhofer. Youtube around the world: geographic popularity of videos. In *Proceedings of the 21st international conference on World Wide Web*, pages 241–250. ACM, 2012.

[23] B. Copos, K. N. Levitt, M. Bishop, and J. Rowe. Is anybody home? inferring activity from smart home network traffic. In *2016 IEEE Security and Privacy Workshops (SPW)*, pages 245–251, 2016.

[24] J. Deng, F. Cuadrado, G. Tyson, and S. Uhlig. Behind the game: Exploring the twitch streaming platform. In *2015 International Workshop on Network and Systems Support for Games (NetGames)*, pages 1–6. IEEE, 2015.

[25] M. Evans, N. Hastings, and B. Peacock. *Statistical Distributions*. Wiley Series in Probability and Statistics. Wiley, 2000.

[26] X. Feng, Q. Li, H. Wang, and L. Sun. Acquisitional rule-based engine for discovering internet-of-things devices. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 327–341, Baltimore, MD, 2018. USENIX Association.

[27] O. L. Haimson and J. C. Tang. What makes live events engaging on facebook live, periscope, and snapchat. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 48–60. ACM, 2017.

[28] A. Halfaker, O. Keyes, D. Kluver, J. Thebault-Spieker, T. T. Nguyen, K. Shores, A. Uduwage, and M. Warncke-Wang. User session identification based on strong regularities in inter-activity time. In *Proceedings of the 24th International Conference on World Wide Web*, pages 410–418, 2015.

[29] H. Li, Y. He, L. Sun, X. Cheng, and J. Yu. Side-channel information leakage of encrypted video stream in video surveillance systems. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, 2016.

[30] M. Ma, L. Zhang, J. Liu, Z. Wang, H. Pang, L. Sun, W. Li, G. Hou, and K. Chu. Characterizing user behaviors in mobile personal livecast: Towards an edge computing-assisted paradigm. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 14(3):66, 2018.

[31] A. J. Pinheiro, J. de M. Bezerra, C. A. Burgardt, and D. R. Campelo. Identifying iot devices and events based on packet length from encrypted traffic. *Computer Communications*, 144:8–17, 2019.

[32] A. Raman, G. Tyson, and N. Sastry. Facebook (a) live?: Are live social broadcasts really broad casts? In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 1491–1500. International World Wide Web Conferences Steering Committee, 2018.

[33] J. Ren, D. J. Dubois, D. Choffnes, A. M. Mandalari, R. Kolcun, and H. Haddadi. Information exposure from consumer iot devices: A multidimensional, network-informed measurement approach. In *Proceedings of the Internet Measurement Conference*, 2019.

[34] B. Wang, X. Zhang, G. Wang, H. Zheng, and B. Y. Zhao. Anatomy of a personalized livestreaming system. In *Proceedings of the 2016 Internet Measurement Conference*, pages 485–498. ACM, 2016.

[35] D. Wood, N. Apthorpe, and N. Feamster. Cleartext data transmissions in consumer iot medical devices. In *Proceedings of the 2017 Workshop on Internet of Things Security and Privacy*, pages 7–12, 2017.

[36] H. Xu, F. Xu, and B. Chen. Internet protocol cameras with no password protection: An empirical investigation. In R. Beverly, G. Smaragdakis, and A. Feldmann, editors, *Passive and Active Measurement*, pages 47–59, Cham, 2018. Springer International Publishing.

[37] F. Zhang, W. He, and X. Liu. Defending against traffic analysis in wireless networks through traffic reshaping. In *2011 31st International Conference on Distributed Computing Systems*, pages 593–602, 2011.

[38] S. Zheng, N. Apthorpe, M. Chetty, and N. Feamster. User perceptions of smart home iot privacy. *arXiv: Human-Computer Interaction*, 2:200, 2018.