How Private Are News Applications? An Investigation Into the Tracking Behaviour of Popular Android News Applications

Ruairí Gielty, Trinity College Dublin, Ireland

19th April 2022

Abstract—The user tracking behaviour of 7 popular Android applications is examined, with the goal of assessing potential privacy concerns present in each. The applications examined are: The Irish Times, Irish Independent, RTE News, BBC, The Guardian, CNN, and The New York Times.

When logged into these applications, the user behaviour/interactions recorded by each app may be linked to the real-world identity of the user. The Irish Times, Irish Independent, RTE, Guardian, CNN, and New York Times applications make use of tracking services which collect data which can be partially or fully linked to the real-world identity of the user. The functionality to opt-out of the collection of data which is linked with the identity of the user is only provided in the Irish Independent and CNN applications.

Non-anonymous account IDs are transmitted in tracking requests alongside anonymous Android device identifiers. In doing this, the potential exists for the anonymous identifier to be de-anonymised, and linked to the real-world identity of the user. With the exception of the BBC and Guardian applications, all apps transmit non-anonymous identifiers alongside anonymous identifiers when the user is logged in. Identifier de-anonymisation of the Google Advertising ID is a particularly troubling privacy concern, since this ID is used extensively for tracking/advertising purposes in Android applications.

The potential for cross-application tracking across a number of these applications is identified. Services such as Google Analytics and the Facebook SDK are used widely across these applications for the purposes of user tracking. Via the use of a system-wide device identifier, these services are capable of tracking user behaviour across independent applications.

CONTENTS

I Acknowledgements
II Introduction
III Background
   III-A What is Privacy?
   III-B GDPR and the ePrivacy Directive
   III-C Related Work and Motivations
IV Experiment Design
   IV-A Hardware and Software used
   IV-B Decrypting HTTPS Connections with Mitmproxy
   IV-C Frida, and the JADX decompiler
   IV-D Persistent Identifiers
   IV-E Experiment Protocol
      IV-E1 Fresh-Open
      IV-E2 Browse Application
      IV-E3 Login
      IV-F Google Mobile Ads SDK
      IV-G Google Analytics
V Experiment Results
   V-A Irish Times
      V-A1 Fresh-Open
      V-A2 Browse application
      V-A3 Login
   V-B Irish Independent
      V-B1 Fresh-Open
      V-B2 Accept Privacy/Cookie Policy
      V-B3 Reject Privacy/Cookie Policy
      V-B4 Login
   V-C RTE News
      V-C1 Fresh-Open
      V-C2 Browse Application
      V-C3 Login
   V-D BBC News
      V-D1 Fresh-Open
      V-D2 Accept Privacy/Cookie Policy
      V-D3 Reject Privacy/Cookie Policy
   V-E Guardian
      V-E1 Fresh-Open
      V-E2 Accept Privacy/Cookie Policy
      V-E3 Reject Privacy/Cookie Policy
      V-E4 Login
   V-F CNN
      V-F1 Fresh-Open
      V-F2 Accept Privacy/Cookie Policy
      V-F3 Reject Privacy/Cookie Policy
      V-F4 Login
   V-G New York Times
      V-G1 Fresh-Open
      V-G2 Accept Privacy/Cookie Policy
      V-G3 Reject Privacy/Cookie Policy
      V-G4 Login
VI Privacy Concerns and Evaluation
   VI-A Compliance with user privacy choices
      VI-A1 Behaviour during fresh-open experiments
      VI-A2 Behaviour after rejecting privacy policy
   VI-B Advertiser Behaviour
I. ACKNOWLEDGEMENTS

I would like to thank my supervisor Douglas Leith, for his valuable guidance and assistance over the duration of this research.

I would also like to thank my parents Claire and Gerard, for their continual support and encouragement.

II. INTRODUCTION

Seven popular Android news applications are examined in this research, with the goal of assessing potential privacy concerns present in each. The applications of three Irish publications are examined: The Irish Times, Irish Independent, and RTE News. The apps of two British publications are examined: BBC News and The Guardian. Two applications from American publications are also examined: CNN and The New York Times.

In Section III, relevant background information is presented to the reader. An explicit definition of privacy is detailed, in the context of the research undertaken. The relevance of the GDPR and the ePrivacy Directive in this work is also outlined. This section also describes the motivation of this research and discusses related work in the field.

Section IV describes the experiments which were undertaken during this research and explains the technologies and hardware used in these experiments.

In Section V, the results of the experiments conducted on all 7 Android news applications are detailed.

Section VI evaluates the results of these experiments, and presents consequent privacy concerns to the reader.

III. BACKGROUND

A. What is Privacy?

A primary objective of this research is to evaluate potential privacy concerns in popular Android news applications. Therefore, it is important to formulate a definition of user privacy in the context of this work.

A basic definition of privacy is the right to use these applications without personal behaviour and interactions being tracked (or the right to opt out of these practices). User tracking is not an concern for consumers who read physical newspapers, so why should it be a potential issue for those who use the corresponding Android application of some publication?

It is important to note that the transmission of user/device data to servers is not necessarily a privacy concern. For example, information such as device model and screen resolution can be used to serve content which can be suitably rendered on a given device. Such information is common across many devices, thus it cannot reasonably be linked to a specific user.

Privacy concerns in the context of these news applications emerge with the use of long-lasting device/user identifiers, which can be used to track user interactions and behaviour in an application over an extended period of time. The use of these identifiers is most troubling when the potential exists for them to be tied to the real world identity of a user. There are a number ways in which user identifiers can be de-anonymised.

Consider the case in which a user is logged into a news application. Applications will almost always require the email of the user to create an account, and may also require a name, address, or payment details. Applications generate account identifiers which can be used to track the user. Any behaviour/interactions tagged with this ID can often be directly linked to the real-life identity of the user, especially in the case in which a full name was provided during account creation. If this account ID is transmitted alongside other ‘anonymous’ user identifiers in the application, they can potentially become de-anonymised via linking to the users account.

There are also businesses which operate with the sole purpose of de-anonymising ‘anonymous’ user identifiers. For example, services exist which take the Google Advertising ID of an Android device, and can potentially supply the name, address, and other personally identifiable information of the owner.

The potential privacy risks of long-lasting user identifiers are heightened when the potential for cross-application tracking exists. Cross-application tracking is made possible when different applications user the same user identifiers. An example of an identifier which could facilitate this is the Google Advertising ID of a device (see Section IV-D), whose value is constant across any applications which access it.

When they can be linked to the real-world identity of a user, long-lasting user identifiers present privacy concerns to users of news applications. The long-term tracking of user behaviour in a news application can reveal detailed information regarding the habits, interests, affiliations, etc. of a user when aggregated. Political views might be inferred from the articles accessed by a user, or even solely from the specific news application which is used. This is by no means a baseless concern. As was exposed in 2018 during the Cambridge Analytica scandal, voter profiles of millions of Americans were created based on their Facebook activity. Hobbies may also be inferred from the sort of content frequently accessed by the user, for example if they often read sports, music, or fashion articles. It is also reasonable to assume that the sex of a user may be identified given the

aggregation of enough app usage/interaction data.

B. GDPR and the ePrivacy Directive

The work described in this paper is primarily focused on determining the user tracking behaviour of Android news applications, with a view of assessing the privacy concerns of each application. These applications are provided within the jurisdiction of the European Union, thus must comply with the General Data Protection Regulation (GDPR), and the ePrivacy Directive.

According to Article 4 of the GDPR\(^3\), personal data is defined as any information relating to an identifiable person. An identifiable person is one who can be identified, either directly or indirectly (via a name, address, identification number, etc.). Article 6 of the GDPR\(^4\) outlines the bases under which personal data may be processed. Within the scope of this work, the processing of personal data is acceptable if either:

1) The user has given explicit consent
2) The data controller\(^5\) has a legitimate interest in processing the personal data of the user.

Note that user consent is not required for the collection of anonymous data, since it cannot be linked to the identity of a person.

The GDPR and the ePrivacy Directive also describe the conditions under which cookies may be stored on a device\(^6\). The only cookies which may be set on a browser/device without explicit consent from the user are "strictly necessary" cookies. These are cookies which are essential for the intended operation of a particular website/application. Other cookies, which may be used for the purposes of user personalisation or tracking, may not be used without prior consent from the user. Accurate information must also be provided to the user regarding the purpose of any cookies used.

C. Related Work and Motivations

Much of the previous research examining consumer privacy, and online tracking behaviour, has focused on the web ecosystem [1]–[4]. A study by Englehardt and Narayanan [1] investigated the Alexa top one million websites\(^7\), with a view of measuring and analysing user tracking across the web. It was determined from the research, that on average, news websites use the most third-party trackers of any category of website. This is followed by the arts, sports, home, gaming, and shopping categories. The authors suggest that the extensive presence of trackers in editorial websites is potentially a result of providing users with free content, resulting in the use of advertising services in order to monetise this content. Note of course, that this study exclusively examines websites, and not mobile applications.

Extensive tracking functionality in news websites is not necessarily indicative of the same behaviour in relative news applications. However, based on this verdict of substantial tracking by news websites, it would be a valuable exercise to analyse such behaviour in Android news applications, on which no research has been previously conducted.

In accordance with GDPR and the ePrivacy Directive, privacy/cookie consent banners are used on almost all websites accessed by European consumers. These banners are used to comply with GDPR and the ePrivacy Directive, and more importantly, to ensure that user privacy is respected. However, as recent research has illuminated, these consent banners are used as mere smokescreens at an alarming frequency, providing users with only the illusion of choice. A study by Bilge et al. [3], published July 2019, evaluated the behaviour of privacy/cookie banners after the introduction of GDPR. The paper outlines several alarming issues with regards to the compliance of many websites with the privacy/cookie banner choices submitted by the user. The study examined a selection of 2,000 of the Alexa top one million websites. It was determined that 92\% of these websites track users without explicit consent. It was also found that rejecting the privacy/cookie policy is often ineffective, as on average, the number of cookies set via websites were not seen to decrease after rejecting the policy.

With regards to the mobile application landscape, previous studies have examined the prevalence of privacy/cookie policies [5], [6], other studies have examined the compliance of a variety of mobile apps with the users privacy policy choices [7], [8]. No studies exist which explicitly evaluate the privacy policy compliance of Android news applications.

The tracking behaviour, and the consequent privacy issues present in a large proportion of Android applications remains largely in the dark. The vast majority of tracking services used by Android apps operate in the background, with no visual indication to the user that this tracking is taking place. Furthermore, almost all network connections which are made by Android applications are encrypted via HTTPS, increasing the opaqueness under which user tracking takes place.

A recent study by D. Leith and S. Farrell on GAEN contact tracing Android apps [9] examined the tracking potential of applications including CovidTracker, CoronaWarn, and RadarCOVID. This paper revealed excessive user tracking via the Google Play Services component used in the studied applications, allowing for location tracking via IP address, as well as the collection of user phone number and email address, among other sensitive data.

Another study by D. Leith, reports on the tracking functionality of Google’s Dialer and Messages Android applications [10]. This research reports that the data collected by Google in these applications is sufficient to link any two Android devices which communicate using these apps. Also outlined in this research is the potential for user de-anonymisation by linking a phone number with a Google account. Google has said it will makes changes to these applications to combat the privacy

\(^3\)https://gdpr.eu/article-4-definitions/, accessed 22nd March 2022
\(^4\)https://gdpr.eu/article-6-how-to-process-personal-data-legally/, accessed 22nd March 2022
\(^5\)The data controller in the context of this research refers to the publications and third-parties which receive personal data via Android news applications.
\(^6\)https://gdpr.eu/cookies/, accessed 22nd March 2022
\(^7\)https://www.alexa.com/topsites, accessed 31st March 2022
concerns raised in this paper. This highlights the benefits of analysing typically opaque app behaviour, as it may result in changes which benefit the privacy of the end user.

Research by Q. Grundy et al. investigates the data sharing practices of medical applications available on Android devices [11]. The study examined 24 popular medical applications available via the Google Play Store. In these 24 applications, 55 entities were seen to receive/process user data. It was determined that 67% of these entities were used to provide analytics or advertising services. Such use of user tracking services when dealing with sensitive medical data reinforces the need for similar research to be undertaken, as it can shed light on potentially privacy infringing practices, and hopefully result in a change in the excessive tracking utilities commonly employed by Android applications.

The research undertaken in writing this paper aims to provide some illumination into the user tracking behaviour of popular Android news applications, and to evaluate privacy concerns based on the findings.

IV. EXPERIMENT DESIGN

A. Hardware and Software used

Mobile handset: Google Pixel 4a running Android 11. The device is rooted using Magisk v23.0. A Frida Server v15.1.10 is installed on the device.


Wi-Fi access point: Raspberry Pi 4 Model B Rev 1.4 running the Raspbian GNU/Linux 10 operating system. Mitmproxy v5.3.0 is installed, with iptables firewall configured to redirect HTTP(S) traffic to port 8080 (the port on which Mitmproxy listens). Frida v15.1.12 is also installed on the Raspberry Pi.

B. Decrypting HTTPS Connections with Mitmproxy

The vast majority of network connections from applications which are examined in this work are made via the HTTPS (Hypertext Transfer Protocol Secure) network communication protocol, thus are encrypted using SSL/TLS.

Mitmproxy is a HTTPS proxy which is used to intercept and decrypt HTTPS traffic by using what is known as the man-in-the-middle technique. For these experiments, a Raspberry Pi 4 is configured as a W AP (wireless access point), with a mitmproxy listening on port 8080. The Android device connects to this W AP via WiFi, and the firewall of the Raspberry Pi is configured to redirect all HTTP(S) traffic from the device to port 8080, to ensure the proxying is invisible to the device. A custom CA (Certificate Authority) certificate is also installed in the Android device, so that the handset views the proxy running on the Raspberry Pi as a trusted party.

With the setup described above, when network connections are made by processes running on the Android device, the proxy running on the Raspberry Pi acts as a man-in-the-middle between the Android device and the server it is trying to establish a connection with. When the device initiates a connection, it is intercepted by the proxy, which presents a fake certificate to imitate the destination server, allowing mitmproxy to decrypt the traffic. The traffic is then re-encrypted and forwarded to the destination server. In this manner, the mitmproxy proxy pretends to be the server when communicating with the client, and pretends to be the client when communicating with the server, logging the decrypted HTTPS traffic between them as this happens.

C. Frida, and the JADX decompiler

An APK (Android Package) is the file format used to distribute and install Android applications. These files contain all of the necessary data required to install an Android application, including the program code, assets, resources, and the manifest file. The program code contained in an APK file is compiled in DEX (Dalvik Executable) bytecode, a format which is executed by the Android Runtime environment, and is not intended to be easily interpreted by humans.

JADX is a tool which is capable of decompiling the DEX code from an APK file, back to Java source code. Note that it is common practice to obfuscate production source code, in order to make the logic of the application difficult to understand, even once it has been decompiled. JADX is used to decompile the APKs of all apps which are examined in this paper.

Frida is a dynamic instrumentation toolkit, which is capable of modifying the instructions of a program while it is running. Frida provides the functionality to inject JavaScript code into proprietary Android applications. The power of Frida with respect to the experiments which are described in this paper lie in the ability to ‘hook’ functions which are defined in the Java source code of an application. Hooking onto a particular function provides the ability to view the parameters sent in any call to said function, and change the function implementation if desired. Frida also provides the functionality to call and interact with the attributes and methods of instantiated Java objects (or even create new instances of an object).

There are two primary purposes for which the aforementioned tools are used in this work:

1) One purpose is to decode ambiguous data which is captured by the man-in-the-middle proxy during examination of applications. Such data is described as ambiguous because either:
   a. Mitmproxy failed to decode the payload of a request, so it is presented instead as raw binary data.
   b. The URL parameters, or JSON/form keys in a request are ambiguous in describing the data which they are transmitting (for example acronyms).

In these cases, the decompiled source code (produced by JADX) can be examined in order to determine which

---

9https://mitmproxy.org/, accessed 8th Feb 2022
10A Raspberry Pi 4 is a small computer, often used for teaching and research purposes. The official operating system of these computers is the Debian-based ‘Raspberry Pi OS’. See https://www.raspberrypi.org/help/what-%20is-a-raspberry-pi/, accessed 8th Feb 2022
11https://github.com/skylot/jadx, accessed 8th Feb 2022

---
classes/functions are used in generating said requests, and perhaps the exact specifications of the data which is sent in them. Frida can then be used to hook functions which are involved in generating the payloads of these connections, which may also provide valuable insight into the data being shared.

An example of when this technique is used is discussed in Section V-A1, in which connections to the api.splkmobile.com domain are decoded. Figure 4 shows the raw request intercepted by Mitmproxy. By examining the decompiled source code which JADX produced, the functions which generate these requests was determined, an the Frida script in Figure 5 was written. This Frida script is capable of decoding requests made to api.splkmobile.com, and the human-readable output is shown in Figure 6.

2) Another primary purpose of using JADX and Frida, is to bypass any certificate pinning which may be implemented in an Android application. Certificate pinning is a method of preventing man-in-the-middle attacks between a client and server.

This is implemented in many Android applications by hard-coding a list of trusted certificates. When a connection is being established, if a server presents a certificate which does not match any of these hard-coded certificates (which will be the case when the mitmproxy proxy presents a fake certificate), the connection will be refused. The decompiled source code can be examined in order to identify a particular implementation of certificate pinning being used in the app, and a Frida script can be written to patch or bypass this implementation at application runtime.

For all experiments which were undertaken during this research, the application being examined is launched alongside a general purpose Frida script\textsuperscript{13} intended to bypass certificate pinning. The purpose of this script is to bypass some of the most common certificate pinning implementations which are used in Android applications, such as those provided by OkHttp or OpenSSL. This Frida script is by no means unassailable, and will fail to bypass obfuscated or custom certificate pinning implementations. When this is the case, the certificate pinning implementation is manually determined in the source code of the application, and the general-purpose Frida script is updated accordingly to bypass the particular pinning implementation.

An example of custom certificate unpinning which was required in this research is described in section V-C1. The Xtremepush SDK in the RTE application uses an obfuscated certificate pinning implementation, which was not bypassed by the general purpose unpinning script. The source of the certificate pinning was determined via the decompiled source code produced by JADX, and the Frida script shown in Figure 29 was written to bypass the certificate pinning implementation. By running this script alongside the app, requests to the Xtremepush domain are made successfully.

D. Persistent Identifiers

In the experiments which follow, persistent identifiers are defined as any identifiers sent in applications which stay the same across multiples sessions of a particular application, across multiple installations of a particular application, or across multiple applications or services on the Android device. The following list presents the persistent identifiers which were considered for any applications being examined in these experiments.

- **Google Advertising ID**: The Google Advertising ID (GAID) is a unique device identifier provided by Google Play Services. The GAID is tied to a specific Android device, and is used by apps and services for the purposes of delivering targeted advertisements to the user. This ID is resettable via the settings menu of an Android device.
- **Android ID**: An Android ID is a 64 bit number which is unique to every combination of app-signing key, user, and device\textsuperscript{14}. Since the majority of users of Android mobile phones will only have a single user account on their device, the Android ID can essentially be viewed as an ID which identifies a unique device-application pairing (even across installations of the app).
- **Google Android ID**: Confusingly, there is another Android ID, which is used as a device identifier exclusively in connections made to Google owned domains, for example the Google Analytics device registration endpoint \texttt{-android.clients.google.com/c2dm/register3}. This ID is persistent across all Android applications, and is only changed when the device is factory reset.
- **Email address**: The Gmail address which the user is logged into on the Android device.
- **IMEI**: An International Mobile Equipment Identity is a unique, non-resettable string which identifies a SIM slot. Thus an Android device may have as many IMEI identifiers, as the number of SIM slots on the device.
- **Hardware serial number**: This is a non-resettable device identifier set by the phone manufacturer.
- **Device WIFI MAC address**: A Media Access Controller (MAC) address. MAC addresses are unique identifiers which are hardwired on the Network Interface Controller (NIC) of Android devices.
- **Phone number**: The mobile phone number of the SIM card inserted in the device.
- **IMSI**: An International Mobile Subscriber Identity (IMSI) is a 64 bit field which uniquely identifies a SIM card.
- **Miscellaneous user identifiers**: Many of the applications and services which are examined in these experiments generate identifiers which are used to track users across multiples sessions of an application (or across multiple installations of the application in some cases). Also included in this category are cookies which are set as device identifiers.

\textsuperscript{13}https://github.com/httptoolkit/frida-android-unpinning, accessed 18th April 2022

\textsuperscript{14}See the ANDROID_ID constant at https://developer.android.com/reference/android/provider/Settings.Secure, accessed 8th Feb 2022
E. Experiment Protocol

When referring to 'interacting' with a news applications in the following experiment descriptions (and in the proceeding sections of this paper), this interaction consists of browsing through news sections (business, sport etc.) and individual articles, but does not extend to interaction with embedded content such as Youtube videos or Spotify players.

Note also that the experiments were often repeated more than once, with the purpose of determining/verifying the behaviour of a particular application. For example, to determine if some user identifier is persistent across different sessions or installations of the app.

1) Fresh-Open: The first experiment which is undertaken for every application being tested is a ‘fresh-open’. The application is installed via Google Play Store, and is opened on the device for the first time. The application is then left idle for 3 minutes without interaction from the user.

2) Browse Application: Two experiments are conducted for cases in which the application provides the user with a privacy/cookie consent form on application startup. For the first experiment, the default set of permissions requested in the privacy/cookie policy are accepted, and the user interacts with the news application for 5 minutes. In the second experiment, all non-essential options provided in the privacy/cookie policy are rejected (after reinstalling the application), and the user interacts with the application for 5 minutes.

An alternative experimentation approach is taken when a privacy policy is not presented to the user upon opening the app, and the user must manually navigate to the policy in the settings screen of the application (provided it exists). These experiments aim to emulate how the average user would interact with an application, so instead of accepting/rejecting the privacy form which is located in a section of the app, the user must manually navigate to the policy/cookie policy on application startup.

In the first experiment, the default set of permissions requested in the privacy/cookie policy are accepted, and the user interacts with the news application for 5 minutes. In the second experiment, all non-essential options provided in the privacy/cookie policy are rejected (after reinstalling the application), and the user interacts with the application for 5 minutes.

In the case where a privacy/cookie policy is not presented to the user upon opening the application, the user logs in, and interacts with the app for 5 minutes, without submission of privacy/cookie consent choices.

3) Login: Some applications provide login functionality. If so, additional experiments are conducted.

In the first experiment, the user accepts the default set of permissions requested in the privacy/cookie policy, logs into the application, and interacts with it for 5 minutes. In the second experiment, all non-essential options provided in the privacy/cookie policy are rejected. The user then logs into the application, and interacts with it for 5 minutes.

In the case where a privacy/cookie policy is not presented to the user upon opening the application, the user logs in, and interacts with the app for 5 minutes, without submission of privacy/cookie consent choices.

F. Google Mobile Ads SDK

All of the Android applications which are examined in this paper make use of the Google Mobile Ads SDK to provide advertising functionality. Google Play Services provide this SDK via the com.google.android.gms.ads package.

There are two primary domains which the Google Mobile Ads SDK is seen to make requests to, in providing advertisements in Android applications; doubleclick.net and googlesyndication.com.

Depending on the particular app configuration, third-party advertising domains may also be contacted via doubleclick.net or googlesyndication.com.

G. Google Analytics

Many of the applications examined during this research make use of the services of Google Analytics. Google Analytics provides an SDK for Android applications which is capable of tracking various user behavioural and engagement metrics, in order to provide the app owner with insights into how their application is used.

Google Analytics uses two endpoints to log tracking data: ssl.google-analytics.com/batch, and app-measurement.com/a. Requests to the app-measurement.com/a are transmitted in protobuf format. Mitmproxy fails to serialise to protobuf format. Mitmproxy script written by D. Leith is used to serialise any requests to this endpoint. See Figure 8 for an example of an un-serialised request to the app-measurement.com/a endpoint. The Mitmproxy script provided by D. Leith produces the serialised output shown in Figure 9.
V. EXPERIMENT RESULTS

A. Irish Times

1) Fresh-Open: The Irish Times applications does not present a privacy/cookie consent form to the user when opening the application for the first time.

![Opening screen of the Irish Times application](Image)

Fig. 1. Opening screen of the Irish Times application

The first two requests on opening the Irish Times applications are sent to graph.facebook.com. The initial request is used to fetch remote configuration options for the Facebook SDK used by the application. The next two requests are logging/tracking events.

The first of these requests reports that the Facebook SDK has been initialised in the application.

The second tracking event reports that the Irish Times application has been opened for the first time on the particular Android device. Figure 2 shows this POST request.

Most notably, this request includes the Google Advertising ID of the device in the advertiser_id field.

An unsecure HTTP connection is made to b.scorecardresearch.com. This domain is owned by Comscore, one of the most prevalent advertising/tracking companies in the Android application landscape [12]. This request sends device information and app usage telemetry, along with a long-term device identifier.

Figure 3 shows a condensed summary of the connection made to Comscore's b.scorecardresearch.com endpoint. The data collected in this request includes the event type being logged (event), an event timestamp (ns Ts), a timestamp of when the app was opened for the first time after being installed (ns Ap.Install), the number of times the app has been run (ns Ap.Runs), and the foreground transition count (ns Ap.Fg).

Most significantly in this request, the ns_Ap_i3 value is
an MD5 hash\textsuperscript{22} of the device Google Advertising ID. This identifier is consistent across both sessions and installations on the Android device. This request is sent every time the Irish Times application is launched, and is always sent in plaintext HTTP.

A POST request was made to the https://60b3599d.api.splkmobile.com/1.0/60b3599d/19adb34718a73635bd540bfc3852f78d/0/1 endpoint upon opening the Irish Times application for the first time.

```plaintext
GET https://60b3599d.api.splkmobile.com/1.0/60b3599d/19adb34718a73635bd540bfc3852f78d/0/1
Connection: Keep-Alive
Content-Length: 510

1f 8b 08 00 00 00 00 00 00 00 a5 52 cb 6e db 30
10 fc 95 40 41 6e b6 60 59 6f 9f 62 38 4e 61 34
0c 79 7f 2c 3f a1 b1 42 2b 6f e6 c5 fe d4 8f
bd 91 07 95 f8 8a 47 02 92 49 11 c6 79 ce 09 ab
4a 70 3b 6d 24 a1 73 c5 8d 16 2d c8 b1 11 0c 09
fa a2 f5 be c4 8b 8d 78 c3 f2 22 02 a2 b4 fd db ab
1b 04 04 94 9a 41 d9 de fd fe 40 55 5d 93 00 31
39 f0 22 8c d2 20 83 34 4c c2 b8 e0 71 34 29 76
2c cc e2 e9 2e cd da 19 b5 45 b3 e2 a8 9c dc 08 09
</>
```

Fig. 4. Splunk API request which could not be decoded by Mitmproxy

Mitmproxy failed to decode the data sent in the payload of this request, so the raw hexadecimal content of the request is presented in Figure 4.

By examining the decompiled source code of the Irish Times application, it was determined that the payload of requests made to the https://60b3599d.api.splkmobile.com/1.0/60b3599d/19adb34718a73635bd540bfc3852f78d/0/1 endpoint, are generated by a toJsonLine() method in the com.splunk.mint.ActionEvent class. The Frida script shown in Figure 5 was written to hook onto this method, allowing the payload data being sent in these requests to be intercepted, and printed to the Frida console in JSON format.

```javascript
Java.perform(() => {
    const SplunkActionEvent = Java.use("com.splunk.mint.ActionEvent")
    SplunkActionEvent.toJsonLine.implementation = function()
        console.log(this.toJsonLine())
        return this.toJsonLine()
});
```

Fig. 5. Frida script used determine tracking data sent in requests to 60b3599d.api.splkmobile.com

Figure 6 shows the decoded payload data of the request shown in Figure 4. The uuid value is a unique device identifier. It is persistent across sessions of the Irish Times application, however is not persistent across installs. Requests sent to this endpoint log device details (phone model, battery level, screen orientation, operating system, mobile network provider), and application telemetry (app name, current Android view). These requests are made every time the Irish Times app is opened.

This application uses the services provided by Google Analytics. The connection shown in Figure 7 is made when opening this app for the first time, and registers the device-application pairing. The device value sent in this request is the Google Android ID of the device. This identifier is persistent for this device across any applications which use Google Analytics.

```plaintext
POST https://android.clients.google.com/c2dm/register3
Authorization: AidLogin 4468978717649541595
:7357954099196059186
app: com.irishtimes.newsapp
X-subtype: 212396306073
app: com.irishtimes.newsapp
device: 4468978717649541595
```

Fig. 6. Decoded payload of Splunk analytics call

A Google Analytics request is made to the app-measurement.com/a endpoint, logging a list of app events.

\textsuperscript{22}MD5 is a cryptographically broken hashing function
Mitmproxy fails to serialise the protobuf body of this request (see Figure 8), so the Mitmproxy script referenced in Section IV-G is used, and an excerpt of the output produced is shown in Figure 9.

```
POST https://app-measurement.com/a
<...>
Connection: Keep-Alive
Accept-Encoding: gzip

\xed\x04\x08\x01\x12[\x06\x02_c\x18\x01
\x02_o\x12\x04auto
\x06 \x02_r\x18\x01
\x07 \x03_et\x18\x01
\x08 \x04_pfo\x18\x03
\x08 \x04_sys\x18\x00
\x08 \x04_uwa\x18\x00
<...>
```

Fig. 8. Google Analytics app-measurement.com/a POST request

```
body {
  event {
    event_info {
      setting_code: "screen_class"
      data_str: "IndexActivity"
    }
    event_info {
      setting_code: "screen_name"
      data_str: "Index: Top Stories"
      event_code: "it_custom_screen_name"
      event_timestamp: 1638794830288
    }
    package_name: "com.irishtimes.newsapp"
    app_version: "5.3.16"
    gmp_version: 15300
    gms_version: 214218
    google_ad_id: "4b00817b-5a22-474d-bb17-0745cabe6e6c"
  }
}
```

Fig. 9. Excerpt of serialised app-measurement.com/a request

Figure 9 shows a Google Analytics event, logging that the user has accessed the "Index: Top Stories" page, as indicated by the `data_str` key. All Google Analytics events sent to this endpoint are tagged with the Google Advertising ID of the device, which is the value of the `google_ad_id` key in this request excerpt.

Another tracking/analytics service used by the Irish Times application is provided by Xtremepush. A connection is made to the endpoint https://api.xtremepush.com/push/api/deviceCreate, on first opening the app after installation (see Figure 10). The `device_id` value of this request is the Android ID for this device-app pairing, which is persistent across installations of the application. The response to this request sets an _xpid_2752 cookie, which is an Xtremepush device identifier for the application, its value is also persistent across installations of the Irish Times app.

```
POST https://api.xtremepush.com/push/api/deviceCreate
Content-Type: application/json
Accept-Encoding: gzip

{
  "appkey": "WyLIAnNrFviV_0qF5o6nKejSAisCSIzd",
  "auth": 2,
  "bundle_version": "5.3.16 2858",
  "country": "IE",
  "device_id": "8bacf83ded7da6",
  "device_model": "Pixel 4a",
  "timezone": "Europe/Dublin",
  "type": "android"
}
```

Fig. 10. Xtremepush device registration request

Several requests are made to app.irishtimes.com. These requests fetch app data such as article content and images, and do not transmit any user/device tracking data.

2) Browse application: As was mentioned at the beginning of the previous section, the Irish Times app does not present the user with a privacy/cookie policy on opening the application for the first time. In order to access the privacy/cookie policy, the user must navigate to Settings - Privacy Policy. As was outlined in Section IV-E, instead of accepting/rejecting the privacy form which is located in a section of the app which only a minority of users will access, the application is interacted with, without any privacy form submission.

The Irish Times application makes calls to ping.chartbeat.net every 15 seconds, and every time an article or section is pressed. The Chartbeat service is used to track application/user engagement. A Chartbeat API call is shown in Figure 11.

In this particular application, the Chartbeat SDK is configured to send the following information in every API call:
- A Chartbeat user identifier (u). This ID is persistent across app sessions, but not persistent across new installations of the app.
- The URL of the article/section which the user is viewing (p), and the time spent on this view in minutes (c).
- Whether the user is currently reading (R), writing (W), or idling (I).
- The number of unique pages/android views the user has viewed during the app session (sd).

As was mentioned in the previous section, this application uses the Google Analytics SDK. POST requests are occasion-
Fig. 11. Chartbeat engagement API call example

ally made to the ssl.google-analytics/batch endpoint. Sent in these requests are a list of app usage tracking events. Figure 12 displays a sample of analytics events which were sent in one of these ‘batch’ calls. The type of event being transmitted is indicated by the value of the t key.

The first event in Figure 12 logs a UI interaction by the user, as indicated by the ec field. In this particular event, the UI interaction being tracked is the pressing of the login icon by the user. The second event logs that a particular article is being viewed by the user. The third event records the time taken for an article view to load in milliseconds, which is sent in the utt field.

Note that the Google Advertising ID of the users device is sent in every event, as the adid value. The cid value is another user identifier, however it does not persist across sessions of the app.

Table 1 shows a list of advertising related domains

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Uses Google Advertising ID?</th>
<th>Uses Cookies?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Google Mobile Ads SDK first-party hosts</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>doubleclick.net</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>googleads.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Google Mobile Ads SDK third-party hosts</strong></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>adform.net</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>doublever.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>rubiconproject.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>turn.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>everestech.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>advir.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>yahoo.com</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table I shows a list of advertising related domains
which were contacted by the Irish Times application during this ‘Browse application’ experiment. Again, note that no privacy/cookie consent choices were submitted by the user prior to this experiment, as no consent form was presented to the user when opening the application.

3) Login: The Irish Times app allows premium subscribers to login to the application, providing access to premium articles. For this experiment, a premium account was created in order to gauge the behaviour of the app from a tracking and privacy standpoint while a user is logged in. The data required to create a premium account with the Irish Times is an email address, full name, and a phone number. Credit/debit card details are required to pay the monthly subscription fee.

Irish Times premium accounts have a user ID associated with them. When logged into the application, this user ID is shared in Google Analytics calls to ssl.google-analytics/batch, see Figure 13 for an example. Here, the uid value is the Irish Times account ID of the user, and as mentioned in the previous section, the adid value is the Google Advertising ID of the users device.

Account details or identifiers are not shared elsewhere while a user is logged in.

POST https://ssl.google-analytics.com/batch

 Fig. 13. Google Analytics event sending Irish Times account ID and GAID

B. Irish Independent

1) Fresh-Open: The Irish Independent application opens to a privacy/cookie consent form. The user cannot navigate away from this screen without submitting their consent choices.

 Fig. 14. Opening screen of the Irish Indpendent application

Figure 14 shows the opening screen of the Irish Independent application. The user can accept the privacy/cookie policy via the “AGREE & CLOSE” button, or they may customise their privacy/cookie consent choices via the “LEARN MORE” button. The user can opt-in to, or opt-out of the following:

- Store and/or access information on a device
- Advertising and content measurement, audience insights and product development
- Personalised content
- Personalised ads
- Social media
- Actively scan device characteristics for identification
- Use precise geolocation data
- Extended measurement
- Storage and access to geolocation information for targeted advertising purposes

Didomi is a company which provides privacy and consent management solutions. This application uses Didomi to collect user consent/preferences. Two connections are made to sdk.privacy-center.org on behalf of Didomi. These requests fetch the list of vendors which the application wants permission to share user data with, and the purposes for which the vendors desire to use the user data. Figure 15

Fig. 15. Didomi consent management setup
shows an example of the data usage requests of the vendor Airship, in a connection to sdk.privacy-center.org.

```
{
  "id": "airship-9nKg8jy2",
  "legIntPurposeIds": [],
  "name": "Airship",
  "policyUrl": "https://www.airship.com/legal/privacy/",
  "purposeIds": [
    "create_content_profile",
    "geo_ads",
    "improve_products",
    "market_research",
    "measure_content_performance",
    "select_personalized_content"
  ],
  "usesNonCookieAccess": false
}
```

**Fig. 15.** Data usage requests of Airship

Similarly to the Irish Times application, some of the first requests made on opening the Irish Independent app for the first time are made to graph.facebook.com. These requests fetch remote configuration data to initialise the Facebook SDK. Unlike the Irish Times application however, this app does not make a tracking request which logs the first opening of the application (see Figure 2).

This app also uses the services provided by the Xtreme-push SDK. An initial request is sent to api.xtremepush.com which registers the device-application pairing, and sets an _xpid_1417 cookie which acts as a device identifier. This call is functionally identical to the one shown in Figure 10 (sent from the Irish Times application). The _device_id_ value is the Android device ID for this particular combination of app-signing key, user, and device. This ID is used to generate the _xpid_1417 cookie, and as a result, Xtremepush can track the user across both sessions, and new installations of the Irish Independent app.

Note that the Irish Times Xtremepush cookie (_xpid_2752), and the Irish Independent Xtremepush cookie (_xpid_1417) do not hold the same value. Consequently, Xtremepush can track users across sessions and installations of these applications, but they cannot track the same user across both applications.

A POST request is sent to the domain prod-inm.mhtr.be, which is owned by MediaHuis. MediaHuis is a Belgian newspaper publisher who own a number of Irish publications; Irish Independent, Sunday World, Sunday Independent, and Belfast Telegraph. An excerpt of the connection is shown in Figure 16.

```
POST https://prod-inm.mhtr.be/next/v
Accept-Encoding: gzip

{
  "cookieid": "dea9537b-9103-4de6-922a-979e39898e5",
  "device": "Google Pixel 4a",
  "deviceosversion": "11",
  "deviceplatform": "Android",
  "deviceid": "phone",
}
<< 200 OK 46b
Set-Cookie: ARRAffinity=12a11ec6a19b6acbd4d4

```

**Fig. 16.** MediaHuis connection settings device ID cookie

Sessions, but do not persist across new installations of the app.

2) Accept Privacy/Cookie Policy: On accepting the privacy/cookie policy, a POST request is sent to api.privacy-center.org (owned by Didomi), detailing the privacy conditions which the user consented to (enable cookies, geolocation data, measure ad performance etc.), along with the list of third-party vendors which they consented to having this data shared with.

The app makes use of the Cxense SDK, which is used to make analytics requests to the cxense.com domain. Two types of requests are seen to be sent to Cxense. The first is to the api.cxense.com/profile/user/segment endpoint. This request sends a list of user identities to Cxense (see Figure 17). When the user is not logged into the Irish Independent app, the only identifier sent in these requests is the device Google Advertising ID.

```
POST https://api.cxense.com/profile/user/segment?sdk= cxense&sdkp=android&sdkv=2.0.1
Connection: Keep-Alive
Accept-Encoding: gzip

{
  "identities": [
    {
      "id": "4b00817b-5a22-474d-bb17-0745cabee6c9",
      "type": "cx"
    }
  ],
  "siteGroupIds": [
    "1139693563586182179"
  ]
}
```

**Fig. 17.** Cxense user identification call

The second type of call made to Cxense is made to the https://scomcluster.cxense.com/Repo/rep.gif endpoint. Calls to this endpoint are app usage tracking events, which are logged when articles or sections are viewed by the user. Figure 18 shows an example of one such request.
Recorded in these calls are the article/section being viewed (loc), the time of viewing (ltm), along with the device GAID (ckp).

Connections are made to ping.chartbeat.net every 15 seconds, and every time an article/section is pressed. This application uses the Chartbeat service to track application/user engagement. See Figure 11, for an example of an API call made to Chartbeat from the Irish Times application. The Chartbeat API tracks the current article/section which the user is viewing (p), their time spent on the current view (c), the number of unique pages the user has viewed during the session (sd), and ties this information to a Chartbeat user identifier (u), which is persistent across sessions, but not persistent across installations or other applications.

Tracking calls are made to MediaHuis (owner of the Irish Independent publication) on article/section press. Requests to mhtr.be/next/v were described in the previous section (see Fig.16). POST requests are also made to mhtr.be/next/h, which log the user’s attention span on every page they visit (see Fig.19).

Requests made to mhtr.be/next/v track users across sessions using the cookieid field (see Figure 16). Connections to mhtr.be/next/h do not include device identifiers, thus users are not tracked across requests to this endpoint.

As with the Irish Times, the Irish Independent uses the Google Analytics SDK. Batches of events are occasionally sent to ssl.google-analytics.com/batch. The information logged in these events varies slightly from those seen in Figure 12 from the Irish Times app, so an example of some notable data sent in a single event from an Irish Independent batch call is provided in Figure 20. Google Analytics events in the Independent app record which screen has been viewed (cd1), and includes the device Google Advertising ID with each event.

Outbrain provide a block of sponsored advertisements at the bottom of every article in the Independent application.

An example to an advertising request to Outbrain is shown in Figure 22. Included in this request is the URL of the article which the user is viewing (url), whether real-time bidding of advertisements is enabled (rtbenabled), an encoded GDPR consent string which represents the privacy choices made by the user (cnsntv2), along with the device GAID (api_user_id). In the response to this request, this GAID is set as an obuid (Outbrain user ID) cookie on the device.

Teads is an advertising/tracking company whose services are used by some of the biggest publications in the world25.

GET https://odb.outbrain.com/utils/get?URL_PARAMETERS
   ...
   accept-encoding: gzip

URL_PARAMETERS
   ...
   url: https://www.independent.ie/videos/
man-proposes-to-girlfriend-star-of-riverdance---...
doo: false
api_user_id: 4b00817b-5a22-474d-bb17-0745cabee6c9
secured: true
dss: 5.4
dm: Pixel+4a
dosv: 30
app_ver: 6.25.0
app_id: com.feedhenry.fhjjjn7In7fqJG2LHTv3LVLanX
rtbEnabled: true
cnsntv2: CPQzc7kP2jchAHAABAE64CsAP....YAAGhuA

Fig. 22. Advertising request made to Outbrain

including CNN, the BBC, and Forbes. The Teads SDK is used in the Irish Independent application to provide both advertising/tracking services. Tracking calls are made to three Teads endpoints; t.teads.tv/track, studio-t.teads.tv/track, and r.teads.tv/rich.

GET https://t.teads.tv/track?URL_PARAMETERS
   ...
   accept-language: en-US,en;q=0.9
   cookie: optout=0;
yocToken=8uaeol29h2925divnljo2s

URL_PARAMETERS
   ...
   action: adReached
ts: 1638807843048
pageId: 0
pid: 147398
env: sdk-inapp
f: 1
slot: native
vid: 4b00817b-5a22-474d-bb17-0745cabee6c9
fv: 923-cdn-legacy
sv: 4.9.0
inte: classic
referrer: https://sdk.teads.tv/com.feedhenry.fhjjjn7fqJG2LHTv3LVLanX

Fig. 23. Teads tracking call in Irish Independent application

Examples of tracking calls to t.teads.tv/track and studio-t.teads.tv/track are shown in Figures 23 and 24 respectively. The event being tracked in these calls is indicated by the action URL parameter. Event types which were seen to be logged during this experiment include; whether an advertisement was requested from Teads (adPlacement), whether a particular advertisement was viewed by the user (impression), and how much of a particular video advertisement the user has viewed (studioAdVideoStart, studioAdVideoFirstQuartile, studioAdVideoMidpoint).

The Google Advertising ID of the device is shared with all calls to the t.teads.tv/track endpoint, but not in calls to the studio-t.teads.tv/track endpoint. Note however, the yocToken which is included in both Fig.23 and Fig.24. This cookie is included in all calls to these domains, and is a potential session identifier. If this is the case, then although the Google Advertising ID is not sent in requests to studio-t.teads.tv/track, it may possibly be determined via the link provided by this yocToken cookie.

Requests to the r.teads.tv/rich/147398 endpoint track device metrics, and the article currently being viewed by the user. These requests also include the Google Advertising ID of the device. Figure 25 shows an example of one such request.

POST https://r.teads.tv/rich/147398
   ...
   content-length: 1097
   cookie: optout=0
deviceType: Google Pixel 4a
country: US
windowRefererUrl: https://www.independent.ie/business/facebook-irish-boss-steps-down-41122118.html
appVersion: 6.25.0
screenWidth: 392
apiFrameworks: 1,2,7
os: Android
screenHeight: 785
deviceFamily: smartphone
event: sdk-inapp
locale: en_US
omidPn: Teadstv
userId: 4b00817b-5a22-474d-bb17-0745cabee6c9

Fig. 25. Teads tracking call which includes device metrics, article being viewed, and device Google Advertising ID

Table II shows a list of advertising related domains which were contacted by the Irish Independent application during this ‘accept privacy/cookie policy’ experiment.

As can be seen from Table II, the only third-party advertising domain (referred via the Google Mobile Ads SDK) which sent the Google Advertising ID of the device in its connections was adtra.com. Cookies were used by all of these domains, except adsafeprotected.com.

There are three other advertising related domains (contacted independently of the Google Mobile Ads SDK) which were communicated with during this experiment; outbrain.com (See Fig.22), teads.tv (See Fig.23,24), and scorecardresearch.com. The Google Advertising ID of the device is shared with all...
TABLE II
ADVERTISING DOMAINS CONTACTED BY IRISH INDEPENDENT APPLICATION AFTER ACCEPTING PRIVACY POLICY

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Uses Google Advertising ID?</th>
<th>Uses Cookies?</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubleclick.net</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>googlesyndication.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>googleadservices.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>pubmatic.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yahoomobileads.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>casaletmedia.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rubiconproject.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>mookiel.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>adtia.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spotxchange.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>smartadserver.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stickyads.tv</td>
<td></td>
<td></td>
</tr>
<tr>
<td>outbrain.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>teads.tv</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>scorecardresearch.com</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

of these vendors, and cookies are also used by all three.

3) Reject Privacy/Cookie Policy: On rejection of the privacy/cookie policy of the Irish Independent application, a POST request is sent to Didomi’s api.privacy-center.org/events endpoint, detailing the list of data usage purposes which the user rejected to, and the list of vendors which the user refused to share data with (see Figure 26).

When the privacy policy is accepted, tracking calls are made to Chartbeat (Fig.11) every 15 seconds. Tracking calls are also made to Chartbeat and Cxense (Fig.18), every time an article/section is pressed in the application. On rejection of the privacy policy, connections are no longer made to the tracking APIs of these vendors. Google Analytics calls to ssl.google-analytics.com/batch are not made either after rejecting the policy.

App engagement tracking calls to MediaHuis’ mhtr.be domain continue after rejecting the policy (see Fig.16 and Fig.19). These calls include a user identification value (deviceuuid), which remains persistent in these calls across app sessions.

The Google Advertising ID of the device continues to be shared with Outbrain (fig.22). Despite these ad requests still being made, the advertisements returned are not displayed to the user.

Table III shows a list of advertising related domains which were contacted by the Irish Independent application during this ‘reject privacy/cookie policy’ experiment. Compare this table, with Table II, which describes the advertising domain connections made when the privacy/cookie policy is accepted. It can be seen that when the privacy/cookie policy is rejected, the application no longer makes requests to third party advertising vendors via the Google Mobile Ads SDK. However, the application continues to make requests to the advertising services which are provided independently of the Google Mobile Ads SDK, namely outbrain.com, teads.tv, and scorecardresearch.com. The Google Android ID of the device is accessed by all three, while only teads.tv uses cookies.

4) Login: The Irish Independent app allows premium subscribers to login to the application, providing access to premium articles. The data required to create an Irish Independent account is a full name and an email address. Credit/Debit card details are required to pay for the premium subscription.

POST requests made to api.cxense.com/profile/user/segment send a list of user identities. Figure 27 shows an example of one such request when a user is logged into the Irish...
Independent app with a premium account.

```
POST https://api.cxense.com/profile/user/segment
   <...
   Connection: Keep-Alive
   Accept-Encoding: gzip
   
   {  
       "identities": [  
           {  
               "id": "4b00817b-5a22-474d-bb17-0745cabee6c9",
               "type": "cx"
           },  
           {  
               "id": "4672f7b4c1ae48c7a2997d56b14950981",
               "type": "inm"
           }  
       ],
       "siteGroupIds": [  
           "113969563596182179"
       ]
   }
```

Fig. 27. Cxense user identification call when logged in as premium user

Sent in this request is the device Google Advertising ID (cx identity), along with the users Irish Independent account ID (inm identity). This allows the device GAIID to be linked with the users specific Irish Independent account.

Calls to MediaHuis’ prod-inm.mhtr.be/next/v app usage tracking endpoint (see Fig.16 for example) contain an extra field when a user is logged in; identityaccountid. It’s value is the account ID of the Irish Independent user.

Any Google Analytics calls to ssl.google-analytics.com/batch (see Fig.20) contain the users Irish Independent account ID as a field in every reported event if the user is logged in. These events also include the device Google Advertising ID, which facilitates the linking of these two user identifiers.

It is important to note that if the logged in user rejects the privacy/cookie policy, the account ID will no longer be shared as described in the cases above (in which the policy had been accepted).

C. RTE News

1) Fresh-Open: The RTE application does not present the user with a privacy/consent form on opening the application for the first time.

```
Fig. 28. Opening screen of the RTE news application
```

The policy is accessible by navigating to the Account - Privacy Statement screen, or by accessing the ‘Weather’ section.

Similarly to the Irish Times application, two logging events are executed via the Facebook SDK when opening the app for the first time. One of these events logs that the Facebook SDK has been successfully initiated in the app. The other event logs that the app has been opened for the first time on this particular device, and includes the Google Advertising ID along with other device details (model, timezone, screen resolution) in the request. Figure 2 shows the equivalent request made in the Irish Times application.

Another tracking related service which the RTE app shares with both the Irish Times and Irish Independent apps, is the Xtremepush SDK for Android. Due to the use of a custom certificate pinning implementation by Xtremepush in this particular application, connections to this domain were being rejected. It was determined by inspection of the decompiled source code of the RTE app, that the certificate pinning implementation was being provided via the XPushSocketFactory class, thus a Frida script was written to bypass certificate verification checks which are made by this class (see Fig.29).

The connections made to api.xtremepush.com via the Xtremepush SDK are seen to share behaviour with what was observed in the other apps which use this service. A connection to https://api.xtremepush.com/push/api/deviceCreate is made upon first opening the app, in order to register the particular
android device using the application. See Figure 10 for the equivalent connection made when first opening the Irish Times application. Sent in these requests are the Android ID of the app-device pairing (recall that Android IDs are persistent across app installations), along with other device information (model, OS, mobile network carrier). The response to this request includes the Google Android ID, which is used to identify the user in future calls to Xtremepush endpoints from this app.

This is another application which uses the services of Google Analytics. The connection to android.clients.google.com/c2dm/register3 registers the device for Google Analytics services with this particular application, including the Google Android ID of the device in the request. See Figure 7 for the equivalent request made in the Irish Times application.

In an SDK configuration call to the socialize.eui.gigya.com/socialize.getSDKConfig endpoint, the Android ID of this device-app pairing is shared. The request is shown in Figure 30, where the \textit{ucid} value is the Android ID. It is important to recall from Section IV-D, that the Android ID and the Google Android ID are two distinct device identifiers.

The most significant cookie which is transmitted in these requests is \textit{atuserid}. This cookie refers to an AT Internet \textit{user} ID. The cookie is a URL encoded string, which when decoded, produces the following JSON text:

```
{
    "name": "atuserid",
    "val": "e58cc868e13e62c9",
    "options": {
        "end": "2023-03-04T12:29:53.096Z",
        "path": "/",
        "version": 6.26,
        "isConsentEnabled": true,
        "isConsentReviewed": false,
        "isOptedOut": false,
        "isGpcEnabled": 0,
        "datestamp": "Mon, 06 Feb 2023 09:53:07 GMT -0800 (Pacific Standard Time)"
    }
}
```

The \textit{val} field in this cookie is the Android ID of this device-app pairing. This particular cookie was included in over 50 requests to \textit{rte.ie} during the 5 minute period of the experiment.

Tracking calls are made to the \textit{logws1309.ati-host.net/hit.xiti} endpoint on article or section press. Figure 33 is an example of one such tracking event. The tracking functionality of these requests is provided by AT Internet (originating from the \textit{com.atinternet.tracker} package in the application source code).

2) \textbf{Browse Application}: Recall Section IV-E, which outlines the experiment protocols which have been employed during the examinations of these applications. When an application does not present the user with a privacy/cookie form on first opening the app, the user interacts with the application without making any explicit consent submissions. This is to emulate how the average user may interact with the application.

In the RTE app, the policy is accessible by navigating to the \textit{Settings - Privacy Statement} screen, or by accessing the ‘Weather’ section. These are two screens which an average user may not access, thus the RTE app is browsed without submission of a privacy form.

A number of cookies are included in the majority of connections to the \textit{rte.ie} domain during this experiment. See Figure 31, which shows an example of one such request.

The most significant cookie which is transmitted in these requests is \textit{atuserid}. This cookie refers to an AT Internet\textsuperscript{26} user ID. The cookie is a URL encoded string, which when decoded, produces the following JSON text:

```
{
    "name": "atuserid",
    "val": "e58cc868e13e62c9",
    "options": {  
        "end": "2023-03-04T12:29:53.096Z",
        "path": "/",
        "version": 6.26,
        "isConsentEnabled": true,
        "isConsentReviewed": false,
        "isOptedOut": false,
        "isGpcEnabled": 0,
        "datestamp": "Mon, 06 Feb 2023 09:53:07 GMT -0800 (Pacific Standard Time)"
    }
}
```

\textsuperscript{26}AT Internet is a company which provides web analytics services. See https://www.atinternet.com/en/, accessed 10th Feb 2022
GET https://logws1309.ati-host.net/hit.xiti?
URL_PARAMETERS
User-Agent: Dalvik/2.1.0 (Linux; U; Android 11; Pixel 4a Build/RP1A.200505.006) RTE News/8.1.3
Host: logws1309.ati-host.net
Connection: Keep-Alive
Accept-Encoding: gzip
URL_PARAMETERS
s: 592983
idclient: e58cc86e13e62c9
vtag: 2.14.0
ptag: Android
lng: en-US-
mfmd: [google]-[pixel14a]
manufacturer: Google
model: Pixel 4a
os: [android]-[11]
apid: ie.rte.news
apvr: [8.1.3]
hl: 11x49x39
r: 1080x2160
dg: 5.8
car: Tesco Mobile
cn: wifi
t: 1643629779.8670001029968262
dls: int

Fig. 33. AT Internet tracking call

The idclient value in these requests is the Android ID for the particular app-device pairing, thus is persistent across sessions and installations of the RTE application. The p value is the page name of the article/section which the user has pressed.

The lifecycle JSON field sent in the stc url parameter includes; a session ID which identifies a particular app session (sessionId), a boolean indicating whether this session is the first launch of the app (fs), a boolean indicating whether this session is the first launch after an update (fsau), the number of times the app has been launched (sc), the date on which the app was first launched on the device (fsd), the number of days since last app use on the device (dsls).

Connections are occasionally made to api.xtremepush.com, logging that a specified news section has been pressed in the application. Figure 34 shows an example of one such event.

POST https://api.xtremepush.com/push/api/tagsHit
<...>
Cookie:_xpid_1434=3305760750
Cookie2: $Version=1
Accept-Encoding: gzip

Fig. 34. Call to the api.xtremepush.com/push/api/tagsHit endpoint

As previously mentioned in Section V-C1, the RTE application is another which uses the services of Google Analytics. Batches of events are sent to the ssl.google-analytics.com/batch endpoint. Figure 35 shows an excerpt of one such event.

Fig. 35. Excerpt of Google Analytics event logged by RTE application

These events track every article/section which has been viewed by the user, and associate each event with the Google Advertising ID of the device. In the example shown in Figure 35: adid is the Google Advertising ID of the device, cd3 is the article which was viewed, and t is the event type.

The idclient value in these requests is the Android ID for the particular app-device pairing, thus is persistent across sessions and installations of the RTE application. The p value is the page name of the article/section which the user has pressed.

The lifecycle JSON field sent in the stc url parameter includes; a session ID which identifies a particular app session (sessionId), a boolean indicating whether this session is the first launch of the app (fs), a boolean indicating whether this session is the first launch after an update (fsau), the number of times the app has been launched (sc), the date on which the app was first launched on the device (fsd), the number of days since last app use on the device (dsls).

Connections are occasionally made to api.xtremepush.com, logging that a specified news section has been pressed in the application. Figure 34 shows an example of one such event.

The _xpid_1434 cookie is sent with these requests. Recall that this cookie is a device identifier which is persistent across both sessions and installations of the app, since it is set using the Android ID.

As previously mentioned in Section V-C1, the RTE application is another which uses the services of Google Analytics. Batches of events are sent to the ssl.google-analytics.com/batch endpoint. Figure 35 shows an excerpt of one such event.

These events track every article/section which has been viewed by the user, and associate each event with the Google Advertising ID of the device. In the example shown in Figure 35: adid is the Google Advertising ID of the device, cd3 is the article which was viewed, and t is the event type.

Table IV shows a list of advertising related domains which were contacted by the RTE application during this ‘browse application’ experiment. Recall that no privacy/cookie consent choices were submitted by the user prior to this experiment, as no consent form was presented to the user when opening the application. The Google Advertising ID of the device is not shared with either of the Google owned domains which are used to provide advertising in this app (doubleclick.net).

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Uses Google Advertising ID?</th>
<th>Uses Cookies?</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubleclick.net</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>googlesyndication.com</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>
3) Login: The RTE news application provides the functionality for a user to login. Note that a user will only be able to login to this app if they have previously consented to the use of functional cookies by the application. This app does not provide the user with a privacy/cookie consent form on app startup, so it must be accessed via the Account - Privacy Statement screen.

The information needed in order to create an RTE account is an email address and first name. A unique user ID is generated when an RTE account is created.

When logged in, AT Internet tracking calls to logws1309.ati-host.net/hit.xiti include a new at field, the value of this field is the RTE account ID. Recall from Figure 33, that these calls already include the Android ID as a user identifier. This allows the linking of the Android ID with the RTE account ID.

Requests to rte.ie/bosco/components/player/proxy.html are made in retrieving video players which are embedded in articles. When logged in, calls to this endpoint include both the Android ID and the RTE account ID. Figure 36 shows an example.

```
POST https://www.rte.ie/bosco/components/player/proxy.html
Host: www.rte.ie
Connection: keep-alive
<...>
URL_PARAMETERS
playertype: html5
atiuserid: e58cc868e13e62c9
clipid: 11373689
thumbnail: transparent
pl_src: rnn
pl_cat: Culture, Drama On One, bloody sunday, Thomas Kinsella
pl_pillar: Culture
pl_userid: e63dbd459d284f51bad8b336f3297ab7
pl_ref: http://www.rte.ie/culture/2022/0130/1276736-on-bloody-sunday-listen-to...>
```

Fig. 36. Request made to rte.ie/bosco/components/player/proxy.html, while logged into RTE application

D. BBC News

1) Fresh-Open: The BBC news application opens to a privacy/cookie consent form. The user cannot navigate away from this screen without submitting their consent choices.

![Fig. 37. Opening screen of the BBC news application](image)

Figure 37 shows the opening screen of the BBC application. The user can accept the privacy/cookie policy via the "Consent" button, or reject the policy via the "Do not consent" button. The user can customise their privacy consent choices via the "Manage options" button. The user can opt-in to, or opt-out of the following:

- Store and/or access information on a device
- Select basic ads
- Create a personalised ads profile
- Select personalised ads
- Create a personalised content profile
- Select personalised content
- Measure ad performance
- Measure content performance
- Apply market research to generate audience insights
- Develop and improve products
- Use precise geolocation data

Consent management in the BBC app is provided by Google’s Funding Choices platform. According to the website27 – “Funding Choices is a Consent Management Platform (CMP) that integrates with Google’s advertising services to help you navigate user choice for privacy regulations like GDPR and CCPA”. A POST request to fundingchoicesmessages.google.com/a/consent sends device and application information (including device GAID, app

---

27https://fundingchoices.google.com/start/, accessed 8th Feb 2022
name, screen size and density), and returns a custom consent form, which is displayed to the user (see Figure 38).

Fig. 38. Funding Choices connection to retrieve custom consent form

2) Accept Privacy/Cookie Policy: Calls to the sb.scorecardresearch.com/p2 endpoint are made every time an article/section is accessed by the user. The scorescardresearch.com domain is owned by Comscore. Figure 39 shows an example of one such request.

The ns_alias value is a user identifier, which is persistent across sessions of the application, but not across installations. ns_ap_ev is the event type being reported. The only event type which was reported during this experiment was a 'view' event. The page_title value refers to the page the user is currently viewing, and the section is the section which the article being viewed was accessed from. The bbc_st_or value is the screen orientation of the device, and bbc_st_sr is a boolean which indicates whether the BBC screen-reader is enabled (screen-readers are used to assist the visually impaired).

Tracking calls are also made to the a1.api.bbc.co.uk/hit.xiti endpoint on article or section press. See Figure 40 for an example.

The tracking functionality in these requests is provided by AT Internet, similarly to calls made to logws1309.atti-host.net/hit.xiti in the RTE news application (See Fig.33).

The idclient value is a user identifier which is persistent across sessions, but not across installations. Note that this value, and the ns_alias sent in scorecardresearch.com calls (Fig.39) are identical. The x9 value is the article/section which the user is currently viewing. Also sent in these request are device specifications including manufacturer, model, OS, screen resolution (r), screen size (dg), and carrier (car).

Fig. 39. Comscore application usage tracking call made in BBC app

Fig. 40. Tracking call made to a1.api.bbc.co.uk/hit.xiti

The lifecycle JSON string sent in the stc URL parameter of these requests requests include; a session ID which identifies a particular app session (sessionID), a boolean indicating whether this session is the first launch of the app (fs), a boolean indicating whether this session is the first launch after an update (fsau), the number of times the app has been launched (sc), the date on which the app was first launched (timestamp), and a unique identifier (UUID).
Table V shows a list of advertising related domains which were contacted by the BBC application during this 'accept privacy/cookie policy' experiment. A significant aspect regarding the behaviour of the BBC news application is that it was not seen to share the Google Advertising ID of the device as a user identifier, either in advertising or tracking requests. This is in contrast with the other applications being examined in this paper, which use this particular identifier extensively (see Figures 12, 18, and 35).

### Table V

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Uses Google Advertising ID?</th>
<th>Uses Cookies?</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubleclick.net</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>googleads.com</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table VI shows the advertising related domains which were contacted by the BBC application after rejecting the privacy/cookie policy. The Guardian application does not communicate with any third-party advertising vendors via the Google Mobile Ads SDK, and the only first-party domain which is contacted is doubleclick.net.

### Table VI

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Uses Google Advertising ID?</th>
<th>Uses Cookies?</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubleclick.net</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>

With regards to advertising behaviour of the BBC application when rejecting the privacy/cookie policy, see Table VI. The application does not communicate with any third-party advertising vendors via the Google Mobile Ads SDK, and the only first-party domain which is contacted is doubleclick.net.

E. Guardian

1) Fresh-Open: The Guardian application opens to a privacy/cookie consent form. The user cannot navigate away from this screen without submitting their consent choices.

![Fig. 41. Opening screen of the Guardian application](image.png)

Figure 41 shows to the opening screen of the Guardian application. The user can accept the privacy/cookie policy via the "Yes, I’m happy" button, or they can customise their privacy consent choices via the "Manage my choices" button. The user can opt-in to, or opt-out of the following:

- Store and/or access information on the device.
- Personalised ads and content, ad and content measurement, audience insights and product development.

The Guardian app uses the services of Google Analytics. A request to the `android.apis.google.com/c2dm/register3` endpoint registers the Android device of the user with the service, including the Google Android ID as a device identifier in the request. Figure 7 shows the equivalent registration call made from the Irish Times application.

2) Accept Privacy/Cookie Policy: Privacy management in the Guardian application is provided by Sourcepoint. Upon accepting the privacy/cookie policy of the app, a POST request is sent to `cdn.privacy-mgmt.com/wrapper/tcfv2/v1/gdpr/consent`, which includes the consent choices submitted by the user.

---

28See https://www.sourcepoint.com/, accessed 21st Feb 2022
Ophan is an in-house analytics/tracking platform which is utilised across the digital outlets provided by the Guardian. In the Guardian application for android, Ophan analytics calls are made to the ophan.theguardian.com/mob endpoint. An example of one of these requests is displayed in Figure 42.

The Content-Type HTTP header specifies that the payload of these Ophan analytics calls use the Apache Thrift framework for the transmission and serialisation of any data being sent in these requests. Mitmproxy does not support the decoding of Thrift, so the payload in Figure 42 appears in a raw text format.

Through investigation of the decompiled Guardian source code, it was discovered that the payload of every Ophan tracking request is generated by the submissionToRequestBody method in the OphanJobService class. The Frida script shown in Figure 43 was written to hook onto any calls to this method, and serialise the analytics data being sent in Ophan requests to an explicit, human readable format. See Fig.44 for the output produced by this script when the request from Fig.42 is sent by the Guardian application.

This script was used to decode Ophan tracking calls which were sent during this experiment, in order to determine the data being sent in them. The following data is included in each analytics call to ophan.theguardian.com; device information (OS, model, manufacturer), and one or more event objects. A Device ID generated by the app is also included in Ophan requests. This identifier is persistent across sessions, but not across installations of the application.

Five event types were seen to be reported during this experiment. Figure 44 shows an AD_LOAD event, which is reported every time an advertisement is loaded in the app. The view_id value corresponds to the particular article on which the advertisement was loaded.

A VIEW event tracks every article opened by the user. AB_TEST events report data related to A/B testing being conducted in the Guardian application. AB_TEST events log a list of tests being conducted and the variation of each test which the user was presented with. Figure 45 shows a decoded call to ophan.theguardian.com which logs both a VIEW, and AB_TEST event.


\*30\ A/B testing is a user experience methodology which involves presenting two variations of some variable (eg. web page, app screen) to users in order to compare their performance

Fig. 42. Apache Thrift encoded tracking request to ophan.theguardian.com

Fig. 43. Frida script used to decode Ophan tracking requests from the Guardian application

Fig. 44. Decoded Apache Thrift tracking data from ophan.theguardian.com request, logging an AD_LOAD event

Fig. 45. Decoded Ophan tracking event logging VIEW and AB_TEST events

The last Ophan event which was reported during this experiment was an INTERACTION, which tracks user interactions with the Guardian application.

22
interactions with the Guardian application. Figure 46 is an example of one such event which was called when pressing the menu button.

Fig. 46. Decoded Ophan INTERACTION tracking event

Similarly to all previously examined applications with the exception of the BBC news app, the Guardian application uses the services of Google Analytics. When the privacy/cookie policy has been accepted, batch calls of events are made to the ssl.google-analytics.com/batch endpoint. The data recorded in these events is similar to that which is sent in Google Analytics requests in other applications (see Figures 12, 20, 35). In this experiment, two event types (t) were encountered: screenview, and event. A sample call from a batch call in the Guardian application is shown in Figure 47.

The Guardian uses the Braze SDK to provide further app engagement tracking services. The privacy policy of the Guardian describes Braze as a life-cycle engagement platform. POST requests are made to the sdk.fra-01.braze.eu/api/v3/data endpoint when the user presses an article. Figure 48 shows an example of one of these requests. This particular request is logging the opening of an article by the user, as indicated by the n key. The keywords key contains values which describe the contents of the article being viewed. The device_id key contains a user identifier. Note that the value of this identifier is the same as those sent in Ophan tracking calls (see Figures 44, 45, 46). Recall that this identifier is persistent across sessions of the Guardian app, but does not persist across new installations.

Table VII shows a list of advertising related domains which were contacted by the Guardian application after accepting the privacy/cookie policy.

The first event in Figure 47 reports that the discover menu button (el) has been pressed from the front screen (cd). The second event logs that an article is being viewed by the user (cd). As with it’s implementation in all previously examined applications, Google Analytics uses the Google Advertising ID of the device as a user identifier in all events logged (adid).

Table VII shows a list of advertising related domains which were contacted by the Guardian application after accepting the privacy/cookie policy.

This app shares the Google Advertising ID of the device in all advertising requests to pubads.g.doubleclick.net/gampad. This ID is shared via the cust_params URL parameter.

31 See https://www.theguardian.com/help/privacy-policy, accessed 22nd Feb 2022
3) Reject Privacy/Cookie Policy: When the privacy/cookie policy of the app is rejected by the user, a POST request is sent to cdn.privacy-mgmt.com/wrapper/tcfv2/v1/gdpr/consent, which includes the consent choices submitted.

If the policy is rejected, the Guardian application does not make tracking requests to ssl.google-analytics.com or sdk.fra-01.braze.eu/api/v3/data, as it does when the policy is accepted (See Figures 47 and 48). However, tracking requests to the ophen.theguardian.com/mob endpoint will be made regardless of the users privacy/cookie consent choices.

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Uses Google Advertising ID?</th>
<th>Uses Cookies?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

As can be seen from Table VIII, the Guardian application does not contact any advertising related domains if the user rejects the privacy/cookie policy.

In the previous experiment, in which the user interacted with the application after having accepted the privacy policy, the Google Advertising ID of the device was shared in requests to both doubleclick.net and ssl.google-analytics.com. Neither of these domains are contacted after having rejected the policy, thus the Google Advertising ID of the device was not transmitted in any HTTP requests during this experiment.

4) Login: The Guardian application allows premium subscribers to login to the application. This removes advertisements in the application, and provides access to other premium features within the app. To create a premium account with the Guardian, users must provide their full name, address, email, and phone number (optional). Users can pay for premium subscription fee via credit/debit card, or PayPal.

A unique user ID is associated with each Guardian account. When logged into the application, the Guardian user ID will be included in all tracking calls to ophen.theguardian.com/mob.
F. CNN

1) Fresh-Open: The CNN application opens to a privacy/cookie consent form. The user cannot navigate away from this screen without submitting their consent choices.

![Privacy Policy](image)

Figure 50 shows the opening screen of the CNN application. The user can accept all permissions requested by the privacy/cookie policy via the "Accept All" button, or can customise their privacy consent choices via the "Manage Cookies+" button. The user can opt-in to, or opt-out of the following:

- Store and/or access information on a device
- Select basic ads
- Select personalised ads
- Create a personalised ads profile
- Create a personalised content profile
- Select personalised content
- Measure ad performance
- Measure content performance
- Apply market research to generate audience insights
- Develop and improve products

The CNN app uses OneTrust\(^{32}\) as a privacy management solution. A request is made to `mobile-data.onetrust.io/bannersdk/v2/applicationdata` to fetch the list of vendors which the application wants permission to share user data with, and the purposes for which the vendors desire to use the user data.

A POST request to `android.clients.google.com/c2dm/register3` initialises the services of Google Analytics for this application. As with all connections to this endpoint seen in previous experiments (see Figure 7), the Google Android ID is sent as a device identifier in this request.

The Android ID of this device-app pairing is shared in requests to the `appcenter.ms` domain, as the value of the `client_unique_id` URL parameter (see Fig.51). The `appcenter.ms` domain is owned by Microsoft.

```
GET https://codepush.appcenter.ms/v0.1/public/codepush/
update_check?URL_PARAMETERS
accept: application/json
<...>
URL_PARAMETERS
deployment_key: H6w9nuOvXtNWkW1a-S5xgefb3FrJhbagGDf
app_version: 6.18.2
client_unique_id: 8590d5b6ebcb97e2
```

Fig. 51. Android ID included in `appcenter.ms` requests

The device Google Advertising ID was not shared during this experiment, and no cookies were set.

2) Accept Privacy/Cookie Policy: The CNN application is yet another which makes use of the tracking/app usage services provided by Chartbeat. Other applications which have been found to use the Chartbeat API during these experiments were the Irish Times and the Irish Independent. In this application, calls are made to the Chartbeat API any time an article or section is pressed by the user. The data tracked and sent in calls to `ping.chartbeat.net` is consistent across these applications, see Figure 11 for reference.

App usage tracking calls are made to `smetrics.cnn.com` both on section/article press, and also on pressing any menu buttons in the CNN application. This tracking functionality is provided by the `com.adobe.marketing.mobile` package in the source code of the application.

Figure 52 shows an analytics request made to `smetrics.cnn.com`. Included in this request is the section which the user is currently viewing (`section`), and the app interaction which is being logged (`action`). In this event, the interaction being tracked is the user pressing the `world` section in the navigation menu. The Google Advertising ID of the device is a user identifier in these requests (`kruxid`).

This application also makes tracking requests to the `collector.cdp.cnn.io/com.snowplowanalytics.snowplow/tp2` endpoint. This functionality is provided by Snowplow Analytics\(^{33}\) via the `com.snowplowanalytics.snowplow` package in the CNN source code. Snowplow specialise in the collection and aggregation of behavioural user data. Figure 53 is an example of a POST request made to the `collector.cdp.cnn.io/com.snowplowanalytics.snowplow/tp2` endpoint.

In Snowplow Analytics requests, the `co` JSON string contains device information such as carrier, model, and Google Advertising ID (`androidIdfa`). The `ue_pr` JSON string

\(^{32}\)https://www.onetrust.com/, accessed 3rd March 2022

\(^{33}\)https://snowplowanalytics.com/, accessed 7th March 2022
Kochava collect app usage and tracking metrics in the CNN application. The data collected by Kochava is sent to the control.kochava.com/track/json endpoint. Figure 54 shows a POST request sent to this endpoint.

In these requests, the `event` value refers to the event being logged, and the `event_data` key contains information regarding the particular event being logged. The event types which were encountered during this experiment were 'Template Load' and 'App Launch'. Also recorded during these events are detailed device metrics, including; device model, orientation, battery level, volume, brightness, and network connection type (WiFi in this case). These requests also have an associated kochava device id, which is used a a device identifier. This ID is persistent across sessions of the CNN application, but not persistent across installations.
AppDynamics\textsuperscript{34} is an application performance monitoring (APM) solution which is used in the CNN application. Connections to mobile.eum-appdynamics.com/eumcollector/mobileMetrics send a batch of events which are recorded via the AppDynamics SDK. Two types of events were seen to be tracked by this SDK during the course of this experiment. A network-request event tracks the majority of HTTP requests which are made from the CNN application. 720 of the 1147 requests which were made by the app during this 5 minute experiment were sent in AppDynamics API calls. A ui event tracks UI changes an Android application, such as the beginning and ending of Android Activities and Fragments\textsuperscript{35}.

Figure 55 shows a POST request made to the mobile.eum-appdynamics.com/eumcollector/mobileMetrics endpoint, and shows an example of both a network-request and ui event.

```
<?>
Accept-Encoding: gzip
Content-Length: 1937
[](1)
{
  "bid": "3d411692-7867-4277-8a6c-69b3c802f122",
  "ca": "Tesco Mobile",
  "cc": 8,
  "cf": "1804800",
  "ct": "wifi",
  "dm": "Google",
  "dmo": "Pixel 4a",
  "event": "Fragment Start",
  "fragmentName": "androidx.lifecycle.y",
  "fragmentUuid": "51708e09-49d0-4cd8-8abb-df289b68d147",
  "is": "AppDynamics.URLConnection",
  "type": "ui"
},
(2)
{
  "bid": "3d411692-7867-4277-8a6c-69b3c802f122",
  "bkgd": false,
  "ct": "wifi",
  "dm": "Google",
  "dmo": "Pixel 4a",
  "event": "Device Discovery Started",
  "httpMethod": "GET",
  "httpStatusCode": 200,
  "is": "AppDynamics.URLConnection",
  "url": "https://c.amazon-adsystem.com/aps_mobile_client_config.json"
}
```
Fig. 55. AppDynamics API call made from the CNN application

The first event shown in this request is a ui event, which is logging the beginning of the lifecycle of a particular Android Fragment in the application. The second event logs a network-request made from the application to https://c.amazon-adsystem.com/aps_mobile_client_config.json. All AppDynamics events are linked with a bid value. This is a device identifier which is persistent across sessions of the application, but not persistent across installations.

A single tracking call was made to the events.claspws.tv/v1/event endpoint during this experiment. The request is shown in Figure 56. The value of the data URL parameter in this request is encoded in base64 format, and when decoded, produces the output displayed in Figure 57.

```
POST https://events.claspws.tv/v1/event?
URL_PARAMETERS
  "event": "DEVICE_DISCOVERY_STARTED",
  "properties": {
    "APP_ID": "vzb2017810",
    "distinct_id": "95.45.132.143",
    "REMOTE_DEVICE_TYPE": "Android",
    "REMOTE_DEVICE_ID": "android:8590d5b36bcb97e2",
    "REMOTE_DEVICE_MAKE": "Google",
    "REMOTE_DEVICE_MODEL": "Pixel 4a",
    "REMOTE_DEVICE_VERSION": "5.1.0",
    "REMOTE_OS_VERSION": "11",
    "GEO_LAT": "UNKNOWN",
    "GEO_LONG": "UNKNOWN",
    "TIMEZONE": "GMT+00:00"
  }
```
Fig. 56. Tracking call to events.claspws.tv/v1/event sent from the CNN application

This request is logging a DEVICE_DISCOVERY_STARTED event. The identifiers which are transmitted in this request are; the Google Advertising ID of the device (IDFA), the external IP address of the device (EXTERNAL_IP_ADDRESS), and the Android ID of this device-app pairing

\textsuperscript{34}https://www.appdynamics.com/, accessed 5th March 2022

\textsuperscript{35}Activities and Fragments are some of the basic building blocks of Android applications, used in creating the user interface. See https://developer.android.com/reference/android/app/Activity, and https://developer.android.com/guide/fragments, accessed 5th March 2022
Similarly to the Irish Independent application, a block of sponsored advertisements provided by Outbrain are displayed at the bottom of each article in the CNN app. GET requests to odh.outbrain.com/utils/get are called to fetch sponsored content from Outbrain when the user reaches the bottom of any given article. See Figure 22 for an example of a request made to this endpoint from the Irish Independent application. The Google Advertising ID of the device is sent in these requests and is consequently set as an obuid cookie in the response from Outbrain.

Table IX shows a list of advertising related domains which were contacted by the CNN application during this experiment (in which the privacy policy has been accepted by the user).

TABLE IX
ADVERTISING DOMAINS CONTACTED BY THE CNN APPLICATION AFTER ACCEPTING PRIVACY POLICY

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Uses Google Advertising ID?</th>
<th>Uses Cookies?</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubleclick.net</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>googlesyndication.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>googleadservices.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Google Mobile Ads SDK first-party hosts</td>
<td></td>
</tr>
<tr>
<td>adnxs.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>casalemedia.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>adsafeprotected.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>rubiconproject.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>yahoo.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>flashalking.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>openx.net</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>teads.tv</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>spotxchange.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Google Mobile Ads SDK third-party hosts</td>
<td></td>
</tr>
<tr>
<td>outbrain.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>vidazoo.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>amazon-adsystem.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>demdex.net</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other advertising providers used by application</td>
<td></td>
</tr>
<tr>
<td>outbrain.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>vidazoo.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>amazon-adsystem.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>demdex.net</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As can be seen from this table, the Google Advertising ID of the device was not shared with any advertising domains (first or third-party) which were contacted via the Google Mobile Ads SDK during this experiment. Cookies were used by doubleclick.net, adnxs.com, casalemedia.com, yahoo.com, openx.net, and spotxchange.com.

Regarding advertising services which operate independently of the Google Mobile Ads SDK, the Google Advertising ID was shared with outbrain.com, vidazoo.com, and amazon-adsystem.com. Cookies were used by both outbrain.com and demdex.net.

3) Reject Privacy/Cookie Policy: When the privacy policy of the CNN application is rejected, a subset of the tracking functionality which was described in the previous section (V-F2) ceases. HTTP requests were not made to ping.chartbeat.net, smetrics.cnn.com (Fig.52), claspws.tv (Fig.56, Fig.57), or collector.cdp.cnn.io (Fig.53) during this experiment.

Connections continue to be made to kochava.com (Fig.54), and eum-appdynamics.com (Fig.55) to collect both app usage and device metrics. These APIs use device identifiers which are persistent across sessions, but not identifiers which are constant across installations (Android ID) or applications (Google Advertising ID).

Table X shows a list of advertising related domains which were contacted by the CNN application during this experiment (in which the privacy policy has been rejected by the user).

TABLE X
ADVERTISING DOMAINS CONTACTED BY THE CNN APPLICATION AFTER REJECTING PRIVACY POLICY

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Uses Google Advertising ID?</th>
<th>Uses Cookies?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Google Mobile Ads SDK first-party hosts</td>
<td></td>
</tr>
<tr>
<td>doubleclick.net</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Google Mobile Ads SDK third-party hosts</td>
<td></td>
</tr>
<tr>
<td>outbrain.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>vidazoo.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>amazon-adsystem.com</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>demdex.net</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When the privacy policy of the CNN application is rejected, a significantly reduced number of advertising domains are contacted via the Google Mobile Ads SDK. The only domain which connections were made to during this experiment was Google’s first-party doubleclick.net. No third-party advertising domains were contacted via the Google Mobile Ads SDK.

The change in behaviour of the advertising services provided independently of the Google Mobile Ads SDK was less drastic. Requests were not made to the demdex.net domain. However, outbrain.com, vidazoo.com, and amazon-adsystem.com are still contacted by the CNN app, and the Google Advertising ID of the device is shared with all three.

4) Login: The CNN application provides login functionality to its users. The only personal data required in order to create an account is an email address. Creating a CNN account is free, thus no payment details (debit/credit card, PayPal) are required.

A unique user identifier is associated with each CNN account. When logged in, this ID is included in all Snowplow Analytics calls made to the collector.cdp.cnn.io/com.snowplowanalytics.snowplow/tp2 endpoint. Figure 58 shows a Snowplow Analytics request made while the user was logged in.

In this request (which is logging a PageView), the cnn_uid value is the account ID of the user. The Google Advertising ID of the device is also transmitted in these requests as the androidIdfa value. The CNN UID is associated with the email address of the user, so in this particular situation, in which the Google Advertising ID is sent along with the CNN UID, there is the potential for linking of the users email address
with the Google Advertising ID of the device.

Note that Snowplow Analytics connections, such as the one seen in Figure 58, are made when a user is logged in and has accepted the privacy policy of the CNN application. In the case where the user is logged in, but has rejected the privacy policy, these requests are not made.

G. New York Times

1) Fresh-Open: The New York Times application opens to a privacy/cookie consent form. The landing screen of the application does not load until privacy consent choices have been submitted by the user.

Fig. 59. Opening screen of the New York Times application

Figure 59 shows the opening screen of the New York Times application. The user can accept the privacy/cookie policy via the "Accept" button, or reject the policy via the "Reject" button.

This is another application which makes use of the services of Google Analytics. Upon opening the application for the first time, a POST request is made to android.apis.google.com/c2dm/register3 to register the Android device of the user with the service. Included in these registration connections is the Google Android ID of the device. Figure 7 shows the equivalent registration call made from the Irish Times application.

A Google Analytics tracking request is made to the app-measurement.com/a endpoint. An excerpt of the decoded body of the request is shown in Figure 60. The event shown in this request is logging that the app has been launched by the user, as indicated by the event_code key. This request is tagged with the Google Advertising ID of the device.

The New York Times application uses the Facebook SDK for Android. As with the Irish Times and RTE applications, a POST request is made to graph.facebook.com which logs a new installation of the application on a particular device. Sent in this request is the Google Advertising ID of the device, along with other device metrics such as model, screen size, carrier, and timezone. Figure 2 is an example of the
Fig. 60. Decoded protobuf body of Google Analytics request to the app-measurement.com/a endpoint
equivalent request made when first opening the Irish Times application.

A tracking request was sent to the sb.scorecardresearch.com/p2 endpoint during this experiment. As previously referenced in Section V-A1, the scorecardresearch.com domain is owned by Comscore, who are one of the most prevalent advertising/tracking companies in the Android application landscape. This request is shown in Figure 61.

Fig. 61. Tracking request made to sb.scorecardresearch.com/p2 when opening the New York Times application for the first time

The ns_ap_ev URL parameter value is the event being logged in the request, which in this case is the New York Times application being opened. This request has an associated user identifier; the ns_ap_i3 URL parameter is the MD5 hash of the device Google Advertising ID.

2) Accept Privacy/Cookie Policy: Upon accepting the privacy policy in the New York Times, a POST request is made to samizdat-graphql.nytimes.com which contains the privacy preferences submitted by the user.

Connections to the samizdat-graphql.nytimes.com and static01.nyt.com domains are made to fetch the data required to render articles/sections in the New York Times application. Requests made to samizdat-graphql.nytimes.com fetch HTML and advertisement configurations, while requests made to static01.nyt.com fetch images, CSS, and JavaScript. All requests to the samizdat-graphql.nytimes.com domain include the Google Advertising ID of the device as an nyt-agent-id header (see Fig.62).

Fig. 62. Requests to samizdat-graphql.nytimes.com include the device Google Advertising ID as a nyt-agent-id header

The New York Times application limits the number of articles which can be viewed by unregistered users, and also limits the number of articles which can be viewed by registered users who do not pay a premium subscription. GET requests to the meter-svc.nytimes.com/meter.js endpoint are used to track the article allowance of any given user. Figure 63 shows one of these requests.

Fig. 63. Request to meter-svc.nytimes.com/meter.js

In these requests, the Android ID of this device-app pairing is a user identifier, and is the value of the rid key. The url
The key is the URL of the article the user wishes to access. The response to these requests determine whether the user will be given access to the article, based on their remaining allowance. The response to this particular request indicates that the user will not gain access to the article, since they have already hit their allowance as an unregistered user (the hitRegiwall boolean is set to true). Figure 64 shows the screen displayed to the user after hitting their unregistered article allowance.

Figure 64. Unregistered user reaching their article allowance in the New York Times application

Localytics is an analytics platform owned by Upland Software. The Localytics SDK is used in the New York Times application to manage push notification services and provide analytics services. Requests are sent to the profile.localytics.com/v1/apps/3bee564c41b04f47d141282-92bac4b0-8ae0-11df-88a5-0038f276d275/profiles/8608e880-b710-44c9-8753-3eab3978f2b endpoint to update the categories of push notifications which are sent to the user. Figure 65 shows a sample of one such request.

In this request, push notifications for breaking-news and morning-briefing have been enabled, and push notifications for evening-briefing have been disabled. The customer_id value in this request is a user identifier which is persistent across sessions of the app, but does not persist across different installations. The push notification settings sent in this request are the default for the New York Times app. The user can change these settings via the notifications menu in the application.

```
POST https://profile.localytics.com/v1/apps/3bee564c41b04f47d141282-92bac4b0-8ae0-11df-88a5-0038f276d275/profiles/8608e880-b710-44c9-8753-3eab3978f2b

Host: profile.localytics.com
Connection: Keep-Alive

{  "changes": [  {    "attr": "Push Channel Enabled: breaking-news",    "op": "assign",    "value": "YES"  },  {    "attr": "Push Channel Enabled: morning-briefing",    "op": "assign",    "value": "YES"  },  {    "attr": "Push Channel Enabled: evening-briefing",    "op": "assign",    "value": "NO"  },  {    "customer_id": "8608e880-b710-44c9-8753-3eab3978f2b",    "database": "app"  }
```

Fig. 65. Localytics request to update user push notification settings

POST requests are made to a.et.nytimes.com/track to track a variety of app usage events and device metrics. Examples of the event data recorded in these requests are shown in figures 66, 67, and 68.

```
"agent": {  "id": "4b00817b-5a22-474d-bb17-0745cabee6c9",  ...
"data": {    "native_app": {      "asset": {        "section": "us",        "type": "section front"      },      "user": {        "type": "anon"      }    }  },  ...
"secure_device_id": "4e1357e4dcdffdd8",  ...
"subject": "page"
```

Fig. 66. Page event sent in request to a.et.nytimes.com/track endpoint

The type of event being logged in these requests is indicated by the value of the subject JSON key, with more detailed event information being stored in the data key. Figure 66 shows a
Fig. 67. Interaction event sent in request to a.et.nytimes.com/track endpoint

![JSON code for interaction event]

Fig. 68. Impression event sent in request to a.et.nytimes.com/track endpoint

![JSON code for impression event]

'page' event, which indicates that a section/article has been accessed. Figure 67 is an 'interaction' event. This event logs a scroll interaction by the user (via the event_data JSON value). Figure 68 is an 'impression' event. An impression refers to the user seeing some specific UI element. This impression event logs that the user has seen a "Subscribe to the New York Times" popup in the application, as indicated by the label JSON value.

All events sent in requests to the a.et.nytimes.com/track endpoint associate two unique identifiers with the user. The id JSON key contains the Google Advertising ID of the device, and the secure_device_id JSON key contains the Android ID.

Table XI shows a list of advertising related domains which were contacted by the New York Times application during this experiment (in which the privacy policy has been accepted by the user).

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Uses Google Advertising ID?</th>
<th>Uses Cookies?</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubleclick.net</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>googleads.com</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

The advertising domains contacted during this experiment are the same as those contacted when the privacy policy is accepted by the user (see Table XI).

3) Reject Privacy/Cookie Policy: Upon rejecting the privacy policy in the New York Times, a POST request is made to samizdat-graphql.nytimes.com which contains the privacy preferences submitted by the user.

The tracking behaviour of this application after rejecting the privacy policy remain unchanged with the previous section (V-G2), in which the policy was accepted by the user.

Table XII shows a list of advertising related domains which were contacted by the New York Times application during this experiment.

Table XII: Advertising domains contacted by the New York Times application after rejecting privacy policy

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Uses Google Advertising ID?</th>
<th>Uses Cookies?</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubleclick.net</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>googlesyndication.com</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

The advertising domains contacted during this experiment are the same as those contacted when the privacy/cookie policy is accepted by the user.

4) Login: The New York Times application provides login functionality to users. New York Times accounts can be created for free, and can be upgraded to premium accounts by providing payment details in order to subscribe to regular billing. Users must provide an email address in order to create an account. If a user wishes to upgrade to a premium account, they can pay using a credit/debit card (in which case they will need to provide a full name and a postal code), or via PayPal. For the purposes of this experiment, a free account is used.

A unique user ID is generated for each New York Times account.

When logged in, all events which are logged in tracking calls made to the a.et.nytimes.com/track endpoint include the New York Times account ID of the user. Figure 69 shows an example of one such event.

This example shows a 'page' event, which was triggered when the user accessed the arts news section of the application. The account ID of the logged in user is included as the value of the regi_id key. As previously mentioned in Section V-G2, the Google Advertising ID is the value of the id JSON key in these
VI. PRIVACY CONCERNS AND EVALUATION

A. Compliance with user privacy choices

1) Behaviour during fresh-open experiments: The first experiment carried out on each application tested was a ‘fresh-open’. The application is opened for the first time after installation, and left idle for 3 minutes without user interaction. During these experiments, no consent choices have been submitted regarding the privacy/cookie policy. Consequently, the app should behave in accordance with the GDPR and the ePrivacy Directive privacy laws.

<table>
<thead>
<tr>
<th>Application Name</th>
<th>provides privacy/cookie consent form?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irish Times</td>
<td>✗</td>
</tr>
<tr>
<td>Irish Independent</td>
<td>✓</td>
</tr>
<tr>
<td>RTE</td>
<td>✓</td>
</tr>
<tr>
<td>BBC</td>
<td>✓</td>
</tr>
<tr>
<td>Guardian</td>
<td>✗</td>
</tr>
<tr>
<td>CNN</td>
<td>✓</td>
</tr>
<tr>
<td>New York Times</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table XIII shows a breakdown of which of the applications examined during this research provided the user with a privacy/cookie consent form upon opening the app for the first time.

The Irish Times application does not provide users with a privacy consent form on first app open. The form is accessible via the Settings - Privacy Policy screen of the application, which is unlikely to be accessed by the average user of this application. Despite this lack of a conspicuous privacy consent form, the Irish Times application collects a considerable amount of app usage data (see Section V-A), and associates this data with both device identifiers (Google Advertising ID, Android ID), as well as miscellaneous user identifiers set by third-party trackers.

The RTE application does not provide users with a privacy consent form on first app open either. The policy is accessible by navigating to the Account - Privacy Statement screen, or by accessing the ‘Weather’ section. It is unlikely that the average user will navigate to the Privacy Statement screen, and although more likely, there is no guarantee that a user will access the ‘Weather’ section of the RTE application. Similarly to the Irish Times application, the RTE app collects considerable app usage data (see Section V-C) without providing a conspicuous policy form. This collected data is associated with both device identifiers (Google Advertising ID, Android ID), as well as miscellaneous user identifiers set by third-party trackers.

All other applications examined during these experiments provided the user with a privacy form upon opening the app for the first time. The user cannot navigate away from these forms without first submitting their consent choices.

Table XIV shows the applications which used cookies prior to the user submitting any privacy consent choices. According to the ePrivacy Directive, users must be clearly notified regarding the use and purposes of any cookies used by a service.

The Irish Times and RTE News are the only two applications examined during this research which do not present a privacy/cookie policy to the user upon opening the app for the first time.

Both the Irish Times and RTE News applications set cookies via the Xtremepush SDK, the Google Mobile Ads SDK, as well as via first-party domains. In order to comply with the ePrivacy Directive, these applications should notify users of the use and purposes of these cookies upon opening the app.

A number of the applications examined during this research were found to use user tracking services prior to explicit consent from the user. Table XV outlines which applications begin tracking the user prior to the submission of their privacy choices.

The Irish Times application does not provide the user with a privacy policy upon opening the app, however uses numerous tracking services regardless. Chartbeat API calls to ping.chartbeat.net are made every 15 seconds in this app, recording application usage metrics. The information recorded
includes: the name of the article/section the user is currently viewing, the time spent on the current view, and whether the user is currently reading or idle in the app. Refer back to Section V-A, in which the behaviour of the Chartbeat API is outlined in detail (Figure 11 shows a sample Chartbeat API request). The Chartbeat tracking services are also used in the Irish Independent, BBC, and CNN applications. However, these applications only begin tracking the user via Chartbeat after they have accepted the privacy policy.

Google Analytics tracking requests are also made from the Irish Times application without prior user consent. Figure 12 in Section V-A, shows a list of events sent from the application. Information tracked by Google Analytics in the Irish Times application includes the URLs of articles accessed, and page loading times. Google Analytics uses the Google Advertising ID of the device as a user identifier. Google Analytics is also used by the Irish Independent and Guardian apps, however only once the privacy policies of those applications are accepted.

It is therefore recommended that the Irish Times application includes a privacy policy to users upon opening their application for the first time, and does not begin tracking users via the aforementioned services prior to explicit consent being provided by the user.

The RTE application does not provide the user with a privacy policy upon opening the app, however user tracking services are employed regardless.

Similarly to the Irish Times application, the RTE app makes use of Google Analytics, tracking user interactions without prior consent. Section V-C describes the information tracked by Google Analytics in the RTE app, with Figure 35 showing an excerpt of a Google Analytics event.

The RTE app also uses tracking services of AT Internet. Tracking calls are made to the logws1309.ati-host.net/hit.xiti endpoint. The information tracked in these requests includes device metrics (phone model, network carrier, OS, resolution), and the title of the article/section currently being viewed by the user. Refer back to Section V-C for a detailed explanation of the data tracked in these requests, with Figure 33 showing an example request.

It is recommended that RTE introduce a privacy policy to their application upon first open, and that tracking calls via the aforementioned services are not made until explicit consent is provided by the user.

The New York Times app provides the user with a privacy policy upon opening the app, however begins track-

---

**TABLE XV**

<table>
<thead>
<tr>
<th>Application Name</th>
<th>Tracks user behaviour prior to consent?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irish Times</td>
<td>✓</td>
</tr>
<tr>
<td>Irish Independent</td>
<td>✓</td>
</tr>
<tr>
<td>RTE</td>
<td>✓</td>
</tr>
<tr>
<td>BBC</td>
<td>×</td>
</tr>
<tr>
<td>Guardian</td>
<td>×</td>
</tr>
<tr>
<td>CNN</td>
<td>×</td>
</tr>
<tr>
<td>New York Times</td>
<td>✓</td>
</tr>
</tbody>
</table>

---

2) Behaviour after rejecting privacy policy: The behaviour of these applications, in terms of respecting the privacy choices submitted by the user, varied among the applications studied in this research.

As discussed in Section VI-A1, the Irish Times and RTE applications infringe on the privacy rights of the user by using tracking services without consent, since no privacy policy is provided when opening these apps. The focus of this section is applications which violate the privacy rights of the user after they have rejected the privacy policy.

**TABLE XVI**

<table>
<thead>
<tr>
<th>Application Name</th>
<th>Tracks user behaviour after policy rejection?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irish Times</td>
<td></td>
</tr>
<tr>
<td>Irish Independent</td>
<td>✓</td>
</tr>
<tr>
<td>RTE</td>
<td></td>
</tr>
<tr>
<td>BBC</td>
<td>✓</td>
</tr>
<tr>
<td>Guardian</td>
<td>✓</td>
</tr>
<tr>
<td>CNN</td>
<td></td>
</tr>
<tr>
<td>New York Times</td>
<td></td>
</tr>
</tbody>
</table>

Table XVI shows the applications which utilise tracking functionality after rejection of the privacy policy by the user. It was determined that all applications examined track users via either first, or third-party services, despite the user having rejected the privacy policy.

The Irish Independent application does not use third-party tracking services (Chartbeat, Cxense, or Google Analytics) when the privacy policy is rejected. However, tracking calls continue to be made to their first-party domain, mhtr.be37.

**B. Advertiser Behaviour**

Every application which was examined during this research uses the Google Mobile Ads SDK (GMA SDK) to provide advertising services. It is important to note that the behaviour of the GMA SDK is not deterministic across different applications, or even across different sessions of the same application. There is no guarantee that the same advertising domains will be contacted via the SDK on numerous runs of the same app (with the exception of Google’s doubleclick.net, which is contacted by all applications).

---

37The mhtr.be domain is owned by MediaHuis, the media conglomerate which owns the Irish Independent publication.
The Google Mobile Ads SDK was well-behaved with regards to the privacy choices submitted by the user. In all cases, if the privacy policy is rejected, no third-party advertising domains will be contacted via the GMA SDK. Table XVII shows a breakdown of the number of unique third-party domains which were contacted via the GMA SDK.

Table XVII

<table>
<thead>
<tr>
<th>Application Name</th>
<th>When policy accepted</th>
<th>When policy rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irish Times</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Irish Independent</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>RTE</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BBC</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Guardian</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CNN</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>New York Times</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Some applications use advertising services which operate independently of the Google Mobile Ads SDK. These services were seen to be much less accommodating of the privacy choices submitted by the user. The Irish Independent app uses the Outbrain and Teads SDKs. The CNN app also uses the Outbrain SDK, and makes regular connections to the amazon-adsystem.com domain.

Outbrain provide sponsored advertising at the bottom of articles in the Irish Independent and CNN applications. Figure 21 shows an example of an Outbrain block in the Irish Independent application.

An example of an Outbrain advertising request is displayed in Figure 22 (Section V-B). In both applications, these advertising/tracking requests will be sent to Outbrain regardless of the whether the user has accepted, or rejected, the privacy policy of the application. These requests include the article which the user is currently viewing, along with the Google Advertising ID of the device. If a user scrolls to the end of every article they read, Outbrain will receive the name of every article read by the user, and this is associated with the advertising ID of their device.

Outbrain track the browsing habits of users, and can potentially create advertising profiles of users based off these habits. Therefore, advertising requests to Outbrain should not be made in either the Irish Independent or CNN applications, without prior consent from the user.

The Teads SDK is used to provide video advertising in the Irish Independent application. Refer back to Figures 23, 24, and 25, in Section V-B, for examples of advertising/tracking requests made via the Teads SDK. Similarly to Outbrain, Teads track articles accessed by the user, and include the Google Advertising ID of the device in these requests (see Fig.25). Teads also track user engagement with their advertisements. Figure 23 shows a request which logs that a particular Teads advertisement has been displayed to the user. Figure 24 shows a request which logs that a user has reached the midpoint of a video advertisement provided by Teads.

Tracking and advertising requests are made to Teads regardless of whether the user has accepted or rejected the privacy policy of the Irish Independent application. Requests to the r.teads.tv/rich/147398 endpoint track articles accessed by the user, and include the Google Advertising ID of the device. This information could potentially be used to create a targeted advertising profile for the user. Therefore, Teads advertising/tracking requests should not be made in the Irish Independent application prior to explicit consent from the user.

Table XX indicates which applications shared the Google Advertising ID of the device with advertising domains, after accepting versus rejecting the privacy policy of the app. Note that the Irish Times and RTE applications are excluded from this table, since no privacy/cookie policy choices were submitted during their experiments. Nevertheless, the Irish Times app shares the GAID with advertising domains without receiving consent from the user.

The Google Advertising ID is used to deliver targeted advertising content to users, and to build advertising profiles of users. Taking the purpose of this identifier into account, it is recommended that the Irish Times, Irish Independent, and CNN applications cease sharing the GAID with advertising domains, without explicit consent from the user.

Regardless of the privacy choices submitted by the user, the RTE, BBC, and New York Times applications were not seen to share the GAID with any advertising domains during this research.

C. User De-Anonymisation

1) Use of Non-Anonymous Identifiers: Many of the device/user identifiers encountered during this work are used to track users anonymously. The account identifiers used when users are logged into these applications however, are not anonymous. The personal data linked to an account ID varies among these applications. The creation of all accounts during this research required at least the provision of an email address, with numerous publications also requiring the user to submit a first/full name, address, and payment details.

The Irish Times application generates an account identifier for each account created. Recall from Section V-A3, that in order to create a premium account, users must provide a full name and phone number, thereby linking the account ID with the real-world identity of the user.
When logged into the Irish Times application, all Google Analytics tracking events are tagged with the account ID of the user. All data collected by Google Analytics therefore becomes personal data, since it can be linked directly to the identity of the user. There is no way to opt out of this behaviour.

Users of the Irish Independent application can log in with a premium account. Recall from Section V-A3, that users must provide a full name and email address in order to create an account. A unique account identifier is generated for any account created with the Irish Independent. The account ID is not an anonymous identifier, since the full name of the user is provided when creating their account.

When a user is logged into the Irish Independent application, and they have accepted the privacy policy, the account ID of the user is included in tracking calls to Google Analytics, Cxense, and MediaHuis.

Both Google Analytics and Cxense track every article accessed by the user in the Irish Independent application, and tag these tracking events with the account ID of the user and the Google Advertising ID of the device.

Significantly, if the logged in user has rejected the privacy policy, the tracking behaviour of Google Analytics, Cxense, and MediaHuis services ceases. Thus, no personally identifiable information is collected in the Irish Independent application if the user does not explicitly consent to such behaviour.

The RTE application provides login functionality, however the user can only login after they have accepted the privacy/cookie policy from the Account - Privacy Statement screen. Recall from Section V-C3, that the data required to create an account with RTE is a first name, and an email.

An account ID is generated when a user creates an RTE account. This account ID is subsequently included in AT Internet tracking calls to logws1309.ati-host.net/hit.xiti, once the user has logged into the RTE application. However, since only a first name is required to create an account with RTE, it is more difficult to link the account ID with the real world identity of the user (when compared with the Irish Times, Irish Independent, and Guardian applications which require a full name in order to create a premium account).

Users can login to the Guardian application with a premium account. To create a premium account, users must provide their full name, address, and email. Optionally, users may also provide a phone number (see Section V-E4).

A unique user ID is generated when a Guardian account is created. When logged into the application, the Guardian user ID will be included in all tracking requests made via their first-party ‘Ophan’ analytics service. Requests to the ophan.theguardian.com/mob endpoint track all articles accessed by the user, user interaction events such as button presses, and A/B test data (see Figures 44, 45, 46, 49). The account ID of the user is not anonymous, since a full name was provided when creating a premium account, thus all events tracked by the Ophan service when logged in can be linked to the real-world identity of the user. These tracking requests are made regardless of whether the user accepts or rejects the privacy policy of the app.

The CNN application provides login functionality to users. Recall from Section V-F4, that an email address is required to create a CNN account.

The CNN app generates a unique account ID to identify logged in users. This ID is included in Snowplow Analytics tracking requests to the collector.cdp.cnn.io/com.snowplowanalytics.snowplow/tp2 endpoint. These requests log all articles/sections accessed by the user, as well as interaction events such as menu button presses (refer to Section V-F2 for more detail regarding Snowplow Analytics tracking behaviour). The CNN account ID is associated only with the email address of the user, since no other personal data is required to create an account. Snowplow Analytics calls can therefore potentially be linked to the email address of the logged in user.

Notably, tracking requests made via the services of Snowplow Analytics are not made when the privacy/cookie policy is rejected by the user.

To create a free account with the New York Times, an email address is required. If a user upgrades to a premium account, and uses credit/debit card as a payment method (as opposed to PayPal), they must also provide their full name, along with a postal code. A unique account ID is associated with each New York Times account.

This account ID is transmitted in tracking requests to the a.et.nytimes.com/track endpoint, regardless of the choices made by the user when submitting the privacy/cookie policy of the application. Connections to this endpoint track data such as the sections/articles accessed by the user, as well as UI interactions such as scrolling on a particular page (see Section V-G2 for detailed information on the data sent to this endpoint). If a user is logged in with a free account, this data can be linked to the email of the user, and if a user is logged in with a premium account (in which they use credit/debit card as their payment method), this data can also potentially be linked to the full name and postal code of the user.

As discussed in Section III-A, the long term tracking of user interactions in these applications can reveal detailed personal information about the user. All of the applications which were examined during this research make use of at least one service which logs every article accessed by the user, and these services use long term identifiers which link user behaviour over time. The aggregation of content accessed in these news applications by users over time could potentially reveal hobbies, political affiliations, religious beliefs, gender, among other sensitive personal information. This is not such a privacy concern when this data is being transmitted anonymously, but becomes an issue when the data is linked to the real-world identity of the user who is interacting with the application.

Note that in creating an account with any of the publications examined in this research, the user agrees to the privacy policy of said publication. In doing this, they lawfully agree to their
personal data being handled. It is a recommendation however, for the privacy of the consumer, that these publications provide users with the option to opt-out of long-term tracking, especially in the case when identifiers used in this tracking can be linked to the real-world identity of the user.

Such functionality is already provided by the Irish Independent and CNN applications. The account IDs used in these applications are not used as identifiers in tracking calls if the privacy policy of the application is rejected.

2) Linking of 'anonymous' identifiers with real identity:
Many of the device/user identifiers provided by the Android system, or otherwise, are used to track users anonymously. However, as described in Section III-A, potential exists for these identifiers to become de-anonymised. During the experiments conducted during this research, the potential for identifier de-anonymisation presented itself when the user is logged into an application.

When logged into the Irish Times application, all Google Analytics tracking events are tagged with the account ID of the user. Recall that this account ID is linked to the real-world identity of the user, since a full name is required to create an account. Google Analytics events in this app include both the Irish Times account ID and the Google Advertising ID of the device. The Google Advertising ID is an anonymous identifier. However, by transmitting it alongside the non-anonymous user account ID, the potential exists for the de-anonymisation of the Google Advertising ID.

The potential for identifier de-anonymisation is present when logged into the Irish Independent application via tracking requests of Google Analytics and Cxense. The account ID of an Irish Independent account is linked to the real-world identity of the user, since the full name and email of the user is required to create an account.

Events logged in all Google Analytics requests to the ns.google-analytics.com/batch endpoint include both the Irish Independent account ID, along with the Google Advertising ID of the device. Refer back to Figure 27, which shows a request sent via the Cxense SDK when the user was logged in. The request sends two user identities to Cxense, the Google Advertising ID, and the Irish Independent account ID.

In both of these cases, the anonymous Google Advertising ID is transmitted alongside the Irish Independent account ID, which is not anonymous. This provides the potential for the de-anonymisation of the GAID.

In order to create an account with RTE, the user must provide a first name, and email. When logged in, AT Internet tracking calls to the logws1309.ati-host.net/hit.xiti endpoint are tagged with both the Android ID, and the RTE account ID of the user. Therefore, potential exists for the Android ID to be linked with the first name, and email of the user.

The personal data required to create an account with CNN is an email address. When logged into the CNN application, the account ID of the user is sent alongside the Google Advertising ID in Snowplow Analytics tracking requests (see Figure 58). This allows the potential for the email of the user to be linked with their GAID, since the CNN account ID is directly linked to email address of the user.

To create a free account with the New York Times, an email address is required. If a user upgrades to a premium account, and uses credit/debit card as a payment method, they must also provide their full name and postal code. The New York Times account ID of a free user is linked to their email address, and the account ID of a premium user (who pays with debit/credit card) is also linked to their full name and postal code.

When logged in, tracking requests to the a.et.nytimes.com/track endpoint include the New York Times account ID of the user, the Google Advertising ID, and the Android ID. If the user is logged in with a free account, the GAID and Android ID could potentially be linked to their email. If logged in with a premium account, the GAID and Android ID could potentially be linked to their full name and postal code.

<table>
<thead>
<tr>
<th>Application Name</th>
<th>De-anonymisation potential of GAID?</th>
<th>De-anonymisation potential of Android ID?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irish Times</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Irish Independent</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>RTE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BBC</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Guardian</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CNN</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>New York Times</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table XIX shows a summary of the de-anonymisation potential of the GAID and Android ID, when users are logged into these applications (after having accepted the privacy/cookie policy).

It is recommended that these applications cease transmitting anonymous identifiers alongside account identifiers which are linked to the real-world identity of the user by name, email, address, etc. This applies especially to the Google Advertising ID of the device. This ID is used extensively across Android applications on a device, for the purposes of targeted advertising, user tracking/profiling, and device identification. Thus, the potential de-anonymisation of this ID presents considerable privacy risks to the user, as the intended anonymous tracking of their behaviour may become attributed to their real-world identity.

D. Cross-application tracking

The potential for cross-application tracking is a privacy concern which was identified during this research. There are a number of services which are used across multiple applications, which potentially facilitate the tracking of user behaviour across independent applications.

1) Google Analytics: The most notable example of a service which could be used in cross-application tracking
is Google Analytics. Google Analytics is used by websites and applications to track user behaviour, and gain insights into how their services are used. Google Analytics is a tremendously popular service. Statistics on the prevalence of Google Analytics usage among Android applications are not available. However, according to BuiltWith\textsuperscript{38}, Google Analytics is used by over 28 million websites\textsuperscript{39}.

Table XX shows a breakdown of which applications make use of the services of Google Analytics. 5 of the 7 applications examined during this research use Google Analytics: Irish Times, Irish Independent, RTE, Guardian, and New York Times.

As described in the experiment results, throughout Section V, Google Analytics is used by these applications to track user behaviour and interactions. The events being tracked by Google Analytics vary between these applications, however one event type shared by all applications is an article access, which logs the URL or name of each article opened by the user.

The cross-application tracking potential of Google Analytics comes from the user identifier used by the service. All events logged by Google Analytics are tagged with the Google Advertising ID of the device. Recall from Section IV-D that the Google Advertising ID of a device is a unique, system-wide value. The value of the GAID is persistent across all Android applications which access it, thus is also persistent as a user identifier for any app which uses the services of Google Analytics. The user behaviour tracked across all applications which use Google Analytics is therefore linked, via this Google Advertising ID. This facilitates potential cross-application tracking through this service.

The use of the GAID as a user identifier by Google Analytics therefore presents privacy risks. The use of the Android ID as an alternative user identifier would mitigate the risk of cross-application tracking. Recall from Section IV-D, that the value of the Android ID is not consistent across applications. The Android ID identifies a unique device-application pairing, thus would result in distinct user identifiers being used by Google Analytics across different applications.

\textsuperscript{38}BuiltWith is a website analysis platform which provides insights into the services used in creating millions of websites around the world.

\textsuperscript{39}https://trends.builtwith.com/analytics/Google-Analytics, accessed 23rd March 2022

2) Facebook SDK: Another service which presents the potential for cross-application tracking is the tracking functionality provided by the Facebook SDK.

<table>
<thead>
<tr>
<th>Application Name</th>
<th>Uses Facebook SDK tracking functionality?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irish Times</td>
<td>✓</td>
</tr>
<tr>
<td>Irish Independent</td>
<td>×</td>
</tr>
<tr>
<td>RTE</td>
<td>✓</td>
</tr>
<tr>
<td>BBC</td>
<td>×</td>
</tr>
<tr>
<td>Guardian</td>
<td>×</td>
</tr>
<tr>
<td>CNN</td>
<td>×</td>
</tr>
<tr>
<td>New York Times</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table XXI shows a breakdown of the applications examined during this research which make use of tracking functionality provided by the Facebook SDK. Tracking capabilities, provided by Facebook, are used in the Irish Times, RTE, and New York Times applications.

Upon opening all of these applications for the first time, a tracking request is sent to the graph.facebook.com domain, logging that the particular application has been opened for the first time on the device, with a \texttt{MOBILE_APP_INSTALL} event. Refer back to Figure 2, which shows an example of this event being logged in the Irish Times application. Included as a user identifier in these requests is the Google Advertising ID of the device. The use of the GAID as a user identifier facilitates the potential for cross-application tracking, since its value is persistent across all applications.

As was suggested for Google Analytics, the Facebook SDK could use the Android ID instead of the GAID as an user identifier, as this would mitigate the risks of cross-application tracking.

A general recommendation from this research, is that the Google Advertising ID of the device is not used by services for identification purposes other than advertising, due to this potential for cross-application tracking.

Chartbeat is a good example of an user-tracking/analytics service which does not provide the potential for cross-application tracking. The services of Chartbeat were seen to be used by the Irish Times, Irish Independent, BBC, and CNN applications during this research. Chartbeat generates a unique user identifier per application, meaning users activities cannot be linked across different applications. Other services which use system-wide identifiers should adopt a similar technique.

E. Summary of Privacy Concerns

The Irish Times and RTE News applications do not provide users with privacy/cookie policies upon opening the application, and yet make use of both cookies and user tracking services. In doing so they violate the terms of the ePrivacy Directive, as well as GDPR. Both applications should add a privacy/cookie policy to their apps which is displayed to the user upon opening the app for the first time.
All applications make use of the Google Mobile Ads SDK to provide advertising services. This service is well behaved with regards to the users privacy consent choices. The Irish Independent, BBC, and, CNN applications contact third-party advertising services whenever the privacy policy is accepted by the user. However, when all non-optimal permissions of the privacy policy are rejected, no third-party advertising services are contacted by these apps via the Google Mobile Ads SDK.

The Irish Independent app uses the Teads SDK to provide advertising services. Advertising/tracking requests, which include the GAID of the device, are sent to Teads domains regardless of the users privacy/cookie consent choices submitted by the user.

Both the Irish Independent and CNN applications also make use of the Outbrain SDK to provide advertising services. Advertising/tracking requests are sent to Outbrain from these applications, regardless of the users privacy/cookie consent choices. These requests include the GAID of the device as a device identifier.

Non-anonymous account identifiers are included in tracking requests by all applications (except the BBC which does not provide login functionality) when the user is logged in. The data collected in these tracking requests is therefore personal data, since it is linked to the real-world identity of the user. The Irish Independent and CNN applications do not share account identifiers if the user has rejected all non-optimal permissions requested by the privacy/cookie consent form, and other applications should follow suit.

Any application which uses account identifiers, with the exception of the Guardian app, transmits these non-anonymous account IDs alongside other anonymous IDs provided by the Android system (Google Advertising ID, and Android ID). The transmission of non-anonymous identifiers alongside anonymous identifiers should cease, as it allows the potential for the de-anonymisation of identifiers which are intended to be anonymous.

Services used across these applications also present the potential for the cross-application tracking of users. Both Google Analytics, and the tracking functionality provided via the Facebook SDK make use of the GAID as a device identifier. This identifier is persistent across all applications on an Android device, therefore allowing data tagged with this value to be linked across independent applications. To prevent potential cross-application tracking, services should not use system-wide identifiers.

VII. Future Work

The research undertaken during this work could be expanded upon, and there is value in similar research being approached for other mobile applications.

The experiment protocol used in examining these Android news applications could be expanded on, in order to gain an understanding of their tracking behaviour in a wider range of circumstances.

The experiments described in Section IV-E only analyse the behaviour of these applications while they are open on the device. A valuable exercise would be to investigate the user tracking of these applications while they are running in the background of the Android device. Perhaps to determine if the same tracking services used while the application is open, persist while it runs in the background.

Also recall from Section IV-E, that during these experiments, the user did not interact with embedded content (such as YouTube videos or Spotify players) during these experiments. It would be of interest to examine what sort of user data is collected by such embedded content, if any.

Research which examines the tracking services used by Android/iOS applications and determines user behavioural/interaction data collected by these services, is uncommon. There are a huge number of popular applications whose explicit tracking behaviour is still unknown. It would certainly be of value, for the privacy of the consumer, to continue examining mobile applications with the goal of assessing any potential privacy concerns.

REFERENCES