Reverse engineering per le masse

O cosa succede quando leggi veramente le privacy policy

Gaetano Priori
staff@reversing.works

Reversing Works
Polimi

April 18, 2024

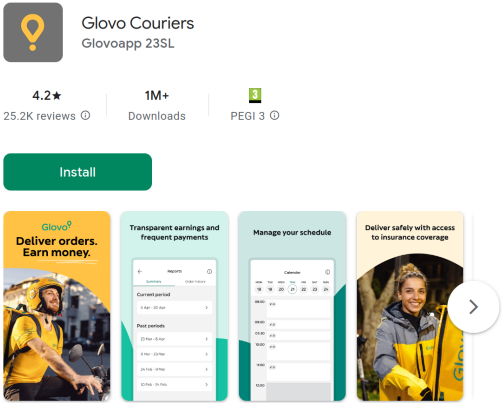
Agenda

1. Introduzione
2. Come è iniziata
3. Il Setup
4. Risultati
5. Prossimi passi

1. Introduzione

Focus

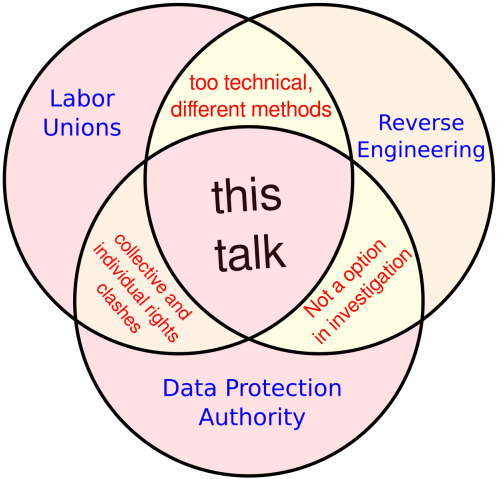
- Possiamo usare i termini ”gig economy” e ”platform workers” in modo intercambiabile.
- Riconosciamo l’impatto significativo che questo fenomeno ha avuto, offrendo nuovi posti di lavoro e creando nuovi mercati.
- In questo intervento sosteniamo che alcune pratiche incorporate in queste applicazioni sono problematiche per concezione.



Chi siamo

- Siamo nati come tracking.exposed nel 2019, e dal 2023 come **reversing.works**
- Siamo nati come un gruppo informale che si è specializzato nell'analisi delle piattaforme digitali.
- Riteniamo che gli algoritmi, i pregiudizi e le politiche contengano complesse dinamiche di potere tecnocratico, per le quali i lavoratori e i loro rappresentanti, come i sindacati, sono spesso impreparati.

Un approccio multidisciplinare



2. Come è iniziata

Una DSAR che non c'è mai stata

- Un rider si è visto **cancellare** il proprio account dall'app Glovo il giorno dopo aver preso parte a uno **sciopero**.
- L'azienda ha dato la colpa a un non meglio precisato "errore tecnico" lato server.
- Per capire perché fosse stato improvvisamente licenziato, il suo avvocato ha presentato una richiesta di accesso dati (**DSAR**) all'azienda.
- L'azienda ha risposto di aver conservato solo i dati contrattuali e di registrazione.

Volevamo capirci qualcosa in più.

O almeno provarci!

Osservando il funzionamento interno dell'applicazione possiamo capire quali dati raccoglie e conserva

Abbiamo trovato un volontario che aveva un account valido e che ha accettato di condividerlo con noi per fare una valutazione completa.

Problemi aggiuntivi

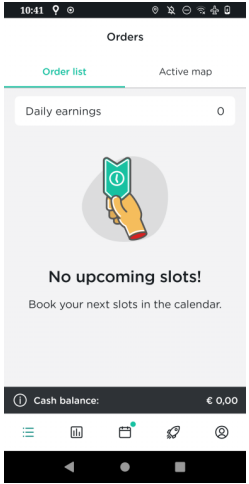
- Non possiamo analizzare il flusso di dati da un'utenza *bannata*.
- Anche riprodurre le stesse condizioni per il nostro volontario è impossibile.
- Abbiamo deciso di fare un'analisi della privacy nel contesto della GDPR.

Fortunatamente esiste un fantastico ecosistema per analizzare le app mobile!

3. Il Setup

Il setup

- Eseguirò l'applicazione mobile Glovo su un dispositivo Android pulito che posso controllare, registrando tutto ciò che l'applicazione fa lì.
- Per raccogliere dati sufficienti, lasceremo girare l'app per 48 ore senza interagirci.
- Per farlo ho usato una build di Lineage OS 19 per raspberry pi, che permette di connettersi senza sforzo da remoto (tramite **adb** e **vnc**).



Permessi

Cosa l'app può vedere.

La maggior parte delle interazioni sensibili con i dispositivi è intermediata dal sistema operativo attraverso le autorizzazioni. Qui possiamo vedere un esempio dall'**AndroidManifest**:

```
1    <uses-permission android:name="android.permission.WRITE_EXTERNAL_STORAGE"/>
2    <uses-permission android:name="android.permission.READ_EXTERNAL_STORAGE"/>
3    <uses-permission android:name="android.permission.CAMERA"/>
4    <uses-permission android:name="android.permission.ACCESS_GPS"/>
5    <uses-permission android:name="android.permission.ACCESS_FINE_LOCATION"/>
6    <uses-permission android:name="android.permission.ACCESS_COARSE_LOCATION"/>
7    <uses-permission android:name="android.permission.ACCESS_BACKGROUND_LOCATION"/>
```

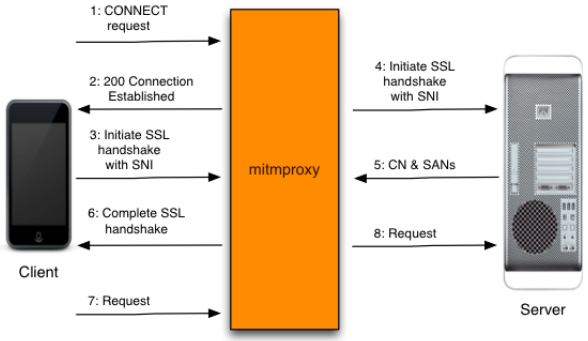
Ma come possiamo sapere quali dati vengono presi, quando e a chi vengono inviati?

Un'analisi preliminare del codice dell'applicazione

- Abbiamo iniziato analizzando staticamente il codice dell'applicazione.
- In generale, è utile per capire quali componenti sono implementati, cosa viene fatto utilizzando degli SDK, ecc.
- Da qui non otterremo tutte le informazioni.

Analisi di rete

- Analisi passiva con Wireshark
- Mitmproxy per ispezionare il traffico TLS.



Dalla documentazione di Mitmproxy

Analisi durante l'esecuzione

- Per registrare ogni accesso alla posizione del dispositivo abbiamo utilizzato Frida, un framework di instrumentazione con un ottimo supporto per le applicazioni mobile.
- In pratica ci permette di iniettare del codice arbitrario nel contesto di esecuzione di un processo. Questo puo' essere scriptato usando i vari binding, es JavaScript con V8.
- Frida supporta gia' le interfaccie offerte dalla JNI e dalla Dalvik, rendendo molto facile l'analisi su Android.
- Per esempio, qui possiamo vedere lo script utilizzato per registrare ogni accesso al servizio di localizzazione

Perché non usare un debugger?

- Un debugger funziona usando la syscall *ptrace* e inserendo un breakpoint (es *int3*)
- Questo comportamento è molto facile da detectare per un programma, es:

```
1 // thread del programma debuggato
2 int3();
3 // es ridirezionare flusso di esecuzione
4 ptrace(PTRACE_GETREGS, pid, NULL, &regs);
5 ptrace(PTRACE_POKEDATA, pid, (void *)regs.rsp,
      CUSTOM_RETURN_ADDR);
6 ptrace(PTRACE_SETREGS, pid, NULL, &regs);
7 ptrace(PTRACE_CONT, pid, NULL, NULL);
```

Un paio di esempi pratici

- In fondo vediamo per esempio il codice usato per monitorare i servizi di localizzazione
- Vediamo poi il funzionamento generico di *frida-trace*
- Non ci dilungheremo ulteriormente su come funziona frida, look at the per un'introduzione consiglio il Frida Handbook o anche la documentazione.

```
Location.getLatitude.implementation = function(){
  console.log( "com.glovoapp.courier fetched Latitude : at: " + getLoggingDateAsString()+"\n");
}

Location.getLongitude.implementation = function(){
  console.log("com.glovoapp.courier fetched longitude : at: " + getLoggingDateAsString()+"\n");
}
```

4. Risultati

Primi risultati ottenuti

July 2021

Parte 1: Storico accesso localizzazione

Grazie alla registrazione dell'accesso alla posizione siamo stati in grado di dimostrare che la posizione del rider viene costantemente registrata, anche al di fuori dell'orario di lavoro.

```
com.glovoapp.courier fetched Latitude : 44.4969 at: Wed Jul 28 2021 01:58:44 GMT+0200
com.glovoapp.courier fetched longitude : 11.3515 at: Wed Jul 28 2021 01:58:44 GMT+0200
com.glovoapp.courier fetched Latitude : 44.4969 at: Wed Jul 28 2021 01:58:44 GMT+0200
com.glovoapp.courier fetched longitude : 11.3515 at: Wed Jul 28 2021 01:58:44 GMT+0200
```

Primi risultati ottenuti

Luglio 2021

Parte 2: Dati inviati a glovo, con un parametro *rating* ignoto.

```
2021-07-27 23:59:15 PUT https://api.glovoapp.com/v3/users/glv:courier
```

```
{
   "NIF": "REDACTED",
   "autoAssignentEnabled": true,
   "cityCode": "BOL",
   "description": null,
   "deviceUrn": null,
   "email": "REDACTED@gmail.com ",
   [...]
   "name": "REDACTED",
   "phoneNumber": {
       "countryCode": "IT",
       "number": "+39REDACTED"
   },
   [...]
   "rating": 4.5,
   [...]
}
```

Primi risultati ottenuti

Luglio 2021

Parte 3: Dati inviati a terze parti.

`POST https://sdk.fra-01.braze.eu/api/v3/data`

```
    {
  [...]
  "app_version": "2.95.0",
  "app_version_code": "107830.0.0.0",
  "device_id": "f959377c-REDACTED",
  "events": [{
    "data": {
      "altitude": 282.6000061035156,
      "latitude": 44.4969,
      "ll_accuracy": 13.432000160217285,
      "longitude": 11.3515
    },
    "name": "lr",
    [...]
  }],

```

Risultati ad un anno di distanza

Settembre 2022

- Abbiamo notato che il proxy non era in grado di connettersi a *mParticle*, un tracker di terze parti.
- Sapevamo che l'App Glovo Courier utilizzava il suo SDK
- Esaminando ulteriormente i log di rete di Wireshark abbiamo notato che l'applicazione rispondeva con un RST all'handshake TLS del nostro proxy.

Dare un senso a tutto ciò

Settembre 2022

- L'SDK *mParticle* controllava la firma del certificato SSL fornito dal proxy TLS.
- Questa tecnica è nota come *certificate pinning* e viene implementata utilizzando il **TrustManager**.

```
1 for (Certificate certificate : domain != null ?
      domain.getCertificates() : C4763d.m16361b())
       {
2    keyStore.setCertificateEntry(certificate.
      getAlias(), m16366a(certificateFactory,
      certificate.getCertificate()));
3 }
4 TrustManagerFactory trustManagerFactory =
      TrustManagerFactory.getInstance(
      TrustManagerFactory.getDefaultAlgorithm());
5 trustManagerFactory.init(keyStore);
6 SSLContext sSLContext = SSLContext.getInstance("
      TLS");
7 sSLContext.init(null, trustManagerFactory.
      getTrustManagers(), null);
8 this.f15711c = sSLContext.getSocketFactory();
```

La soluzione

E del perché Frida ritorna utile!

Settembre 2022

Qui vediamo un esempio da manuale di come Frida torna utile: possiamo modificare l'implementazione del metodo che inibisce la richiesta HTTPS.

```
try {
  var array_list = Java.use("java.util.ArrayList");
  var TrustManagerImpl_Activity = Java.use('com.android.org.conscrypt.TrustManagerImpl');
  TrustManagerImpl_Activity.checkTrustedRecursive.implementation = function(certs, ocspData, tlsSctData, host, clientAuth, untrustedChain, trustAnchorChain, used) {
    return array_list.$new();
  };
} catch (err) {
```

Risultati finali

POST nativesdks.mparticle.com/v2/36b7c1298092e74db9a90[...]/ever

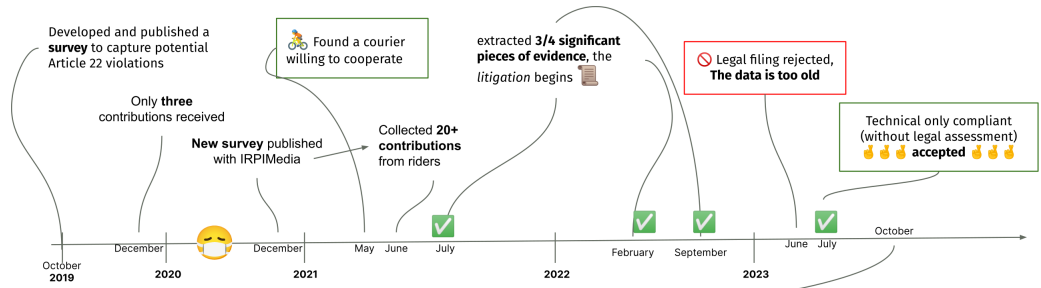
```
{
 [...]
      "dt": "e",
      "et": "Other",
      "n": "Gps state changed on the device",
      "attrs": {
      "lc": {
        "lat": 44.49711678040484,
        "lng": 11.352085079430195,
        "acc": 0
        "city": "BOL",
[...]
}
```

Epilogo

Settembre 2022

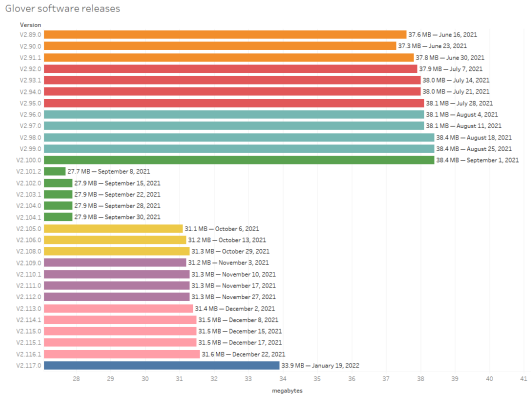
- Utilizzando tecniche di reverse engineering siamo riusciti a dimostrare che l'azienda monitorava costantemente la posizione dei rider, anche al di fuori dell'orario di lavoro, e condivideva i loro dati personali, posizione compresa, con terze parti.
- Questa prova difficilmente sarebbe emersa da un DSAR inviata all'azienda.
- Abbiamo replicato la stessa analisi nel luglio 2021, nel settembre 2022 e nel luglio 2023 ottenendo risultati simili; per avere una prospettiva migliore, potete consultare il paper che abbiamo pubblicato con l'ETUI (Exercising workers' rights in algorithmic management systems).
- Questi risultati hanno portato a una segnalazione al Garante per la protezione dei dati personali.

UNA PROSPETTIVA



Una sfida ulteriore

- Dei risultati troppo vecchi hanno poco significato considerando che una nuova release avviene (circa) **ogni settimana**
- … almeno, questo dimostra che le evidenze non sono frutto di un incidente



Il contesto intorno a noi.

- Nel 2019 Il Garante per la protezione dei dati personali avvia indagini su numerose piattaforme di gig economy operanti in Italia.
- Nel 2021 multa Glovo-Foodinho per 2,6 milioni di euro (multa + misure prescrittive)
- Nel 2022 Glovo vince il ricorso e non paga la multa.
- Nel 2023 la Corte di Cassazione ha respinto il ricorso e Glovo deve pagare la multa.

AI

Italy's DPA fines Glovo-owned Foodinho $3M, orders changes to algorithmic management of riders

Natasha Lomas @riptari / 1:30 PM GMT+1 • July 6, 2021

Comment

5. Prossimi passi

Come rendere tutto questo più accessibile

- **exodus-privacy.org** è un ottimo punto di partenza; https://reports.exodus-privacy.eu.org/en/reports/search/com.glovoapp.courier/
- Privacy International ha rilasciato **Data Interception Environment** https://privacyinternational.org/learn/data-interception-environment
- Come **reversing.works** vogliamo poter aiutare NGO, sindacati e associazioni in queste sfide tecniche.
- Al https://privacycamp.eu abbiamo organizzato un panel

Non siamo soli, ma vorremmo essere di più

Workers Info Exchange
www.workerinfoexchange.org

Organizzare l'aiuto reciproco, le risorse e l'advocacy per migliorare le condizioni di tutte le persone che utilizzano Mechanical Turk di Amazon **turkopticon.net**

The Workers' Algorithm Observatory
wao.cs.princeton.edu

Piattaforma per richiedere i propri dati, per lavoratori di piattaforme come Uber, Uber Eats, Smood.
personaldata.io/uber

Gaetano Priori

Milano, April 18, 2024

staff@reversing.works