



3387.52

Рейтинг

RUVDS.comVDS/VPS-хостинг. Скидка 15% по коду **HABR15**

Подписаться



ru_vds

5 фев в 16:01

Все знают, где ты находишься



Простой



10 мин



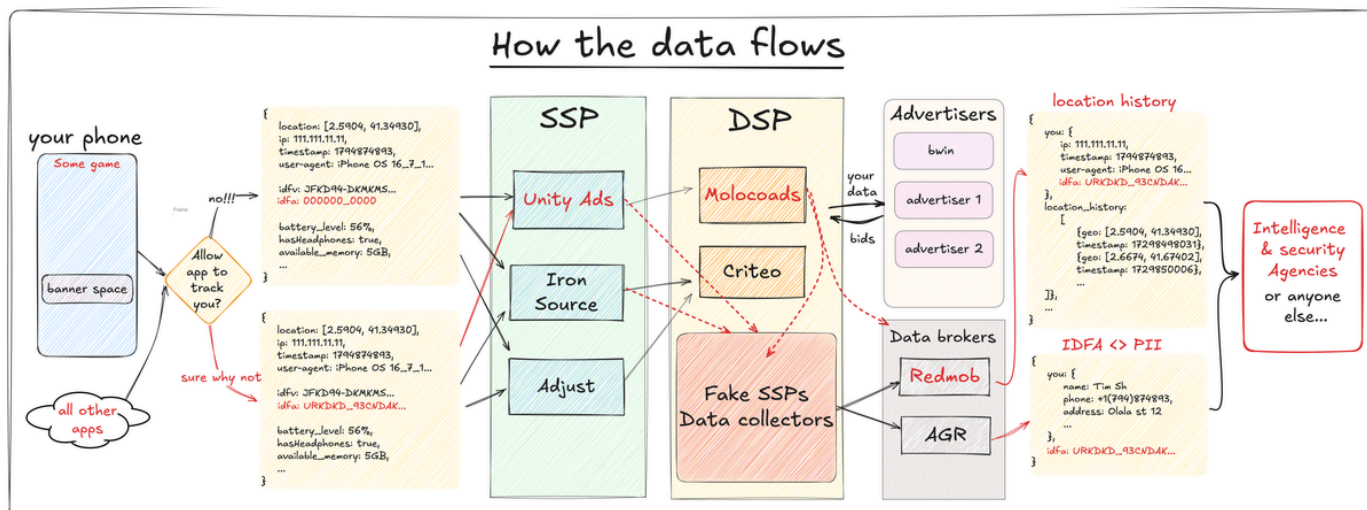
60K

Блог компании RUVDS.com, Аналитика мобильных приложений*, Информационная безопасность*, Монетизация*

Обзор

Перевод

Автор оригинала: Tim



Недавно я прочитал о масштабной утечке геолокационных данных из Gravy Analytics, благодаря которой стало известно, что более двух тысяч приложений из AppStore и Google Play тайно собирали геолокационные данные пользователей без их согласия. И часто об этом не знали даже разработчики.

Я изучил список ([ссылка](#)) и обнаружил как минимум три приложения, установленные на моём iPhone. Проверьте сами!

У меня возникла идея: попробовать отследить себя снаружи, то есть купить свои

геолокационные данные, утёкшие через какое-нибудь приложение.

■ TL;DR

Я потратил пару десятков часов и получил следующие результаты:

1. Я обнаружил пару запросов, отправленных моим телефоном, в которых содержалось **моё точное местоположение** + 5 запросов, через которые утекал **мой IP-адрес**, который при помощи обратного просмотра DNS можно превратить в геолокацию.
2. Я много узнал об аукционах с торгами в реальном времени (RTB, real-time bidding) и протоколе OpenRTB; меня поразили объём и типы данных, отправляемые с заявками к биржам рекламы.
3. Отказался от идеи покупки данных о моём местоположении у брокера данных или сервиса слежения, потому что у меня нет достаточно большой компании, чтобы получить пробный режим или выложить 10-50 тысяч долларов на покупку огромной базы данных с данными миллионов людей + меня.

Ну, на самом деле, деньги-то есть, но такие траты мне кажутся немного нерациональными.

Оказалось, что данные людей из Евросоюза почти самые дорогие.

Но я всё равно знаю, что мои данные о местоположении собираются, и знаю, где их купить!

Начальная точка

Для проведения этого расследования я подготовил следующую систему:

- Свой старый iPhone 11, сброшенный до заводских настроек + новый Apple ID. Мне было слишком некомфортно делать всё это на моём нынешнем телефоне.
- Charles Proxy для записи всего входящего и исходящего трафика. Я настроил на iPhone сертификат SSL для дешифровки всего HTTPS-трафика.
- Простая игра Stack издателя KetchApp — помню, как играл в неё в школе 10-12 лет назад. Почувствовал ностальгию при выборе этой «лабораторной крысы».

К моему удивлению, в списке оказалось много игр KetchApp.



■ Огромное количество запросов

Итак, поехали: мы установили всего одно приложение, если не считать приложений Apple по умолчанию и включённого Charles. Запуск Stack через 3, 2, 1...

Это запросы, которые отправляет приложение в первую минуту после запуска.

Взгляните на время между запросами — они отправляются почти каждую долю секунды.

Давайте изучим содержимое запросов.

Я действительно проверил каждый, но оставил только самые интересные.

■ Unity [ads]

Давайте начнём с самого сочного запроса, отправленного к <https://o.isx.unity3d.com> — в нём первом есть моя геопозиция, хоть я и **отключил Location Services** на iPhone для всех приложений!

Если вы столь же наивны, как и я до этого, то можете удивиться — как 3D-движок Unity связан с рекламой внутри приложения и с отслеживанием местоположения?

Возможно, это просто мониторинговые данные, помогающие совершенствовать движок?

Оказывается, основной источник доходов Unity (в 2023 году компания заработала больше 2 миллиардов долларов) — это Unity Ads, «рекламная сеть для мобильных игр». Звучит любопытно.

Ниже показано тело отправленного Unity Ads запроса в формате JSON. Я оставлю только стоящие упоминания поля — полный запрос содержит больше двухсот ключей.

```
{
  "ts": "2025-01-18T23:27:39Z", // Метка времени
  "c": "ES", // Код страны,
  "d": "sports.bwin.es", // Домен; приложение или веб-сайт, на котором будет отображаться
  "bn": "molocoads-eu-banner", // Что это за moloco ads? Разберёмся ниже!
  "cip": "181.41.[данные удалены]", // Мой IP !!
  "dm": "iPhone12,1",
  "ct": "2", // Тип соединения, например, Wi-Fi
  "car": "Yoigo", // Оператор мобильной сети
  "ifv": "6B00D8E5-E37B-4EA0-BB58-[данные удалены]", // ID для вендора. Мы к этому вернёмся
  "lon": "2.[данные удалены]", // Долгота ...
  "lat": "41.[данные удалены]", // Широта ...
  "sip": "34.227.224.225", // IP-адрес сервера (Amazon AWS в США)
  "uc": "1", // Согласие пользователя на отслеживание = True; чего???
}
```

Итак, мои IP + местоположение + метка времени + какой-то id `ifv` передаются Unity → Moloco Ads → Bwin, после чего я вижу рекламу Bwin в игре.

Великолепно!

Небольшое примечание: отправленное местоположение было не очень точным (но в пределах того же почтового индекса); наверно это вызвано тем, что iPhone был подключён к WiFi и в нём отсутствовала SIM.

Если бы это была LTE, то уверен, что долгота и широта оказались бы гораздо точнее.

■ Привет, Facebook*... А ты тут что делаешь?

Следующим интересным запросом, через который утекли мои IP + метка времени (= данные о геопозиции), был Facebook*.

Что?!

- На этом iPhone не установлено никаких приложений Meta* [Facebook*].
- Я не привязывал ни приложение, ни Apple ID к аккаунту Facebook*.
- Я не давал согласия на то, чтобы Facebook* получал мой IP-адрес!

Тем не менее, запрос выглядит так:

```
{
  "bundles": {
    "bidder_token_info": {
      "data": {
        "bt_extras": {
          "ip": "181.41.[данные удалены], // отличные "дополнительные данные", ;
          "ts": 1737244649
        },
        "fingerprint": null
      },
      {
        "куча данных, да просто огромная куча"
      }
    }
  }
}
```

Подробнее мы поговорим об этом в следующем разделе.

■ Зачем вам нужен уровень яркости моего экрана?

Последний заинтересовавший меня запрос был отправлен... снова Unity:

<https://configv2.unityads.unity3d.com> .

Посмотрим, что же столь необходимо Unity в этой конфигурации:

```
{
  "osVersion": "16.7.1",
  "connectionType": "wifi",
  "eventTimeStamp": 1737244651,
```

```
"vendorIdentifier":"6B00D8E5-E37B-[данные удалены]", // снова ifv
"wiredHeadset":false, // что, простите?
"volume":0.5,
"cpuCount":6,
"systemBootTime":1737215978,
"batteryStatus":3,
"screenBrightness":0.34999999403953552,
"freeMemory":507888,
"totalMemory":3550640, // это ОЗУ?
"timeZone":"+0100",
"deviceFreeSpace":112945148
"networkOperator":"6553565535"
"advertisingTrackingId":"00000000-0000....", // интересно...
}
```

Здесь нет «личной информации», но, честно говоря, меня пугает этот объём данных, передаваемый неизвестному списку третьих сторон.

Зачем им знать яркость моего экрана, количество памяти, громкость и то, пользуюсь ли я наушниками?

«Правильный» ответ мне известен — всё это нужно, чтобы компании лучше таргетировали свою аудиторию!

Например, если вы рекламируете мобильное приложение размером 1 ГБ, а у пользователя осталось всего 500 МБ, то какой смысл показывать ему приложение, правда?

Но по этой теме я слышал и много противоречивой информации.

Например, что Uber динамически меняет стоимость такси в зависимости от уровня заряда аккумулятора, потому что вы не будете ждать более дешёвого варианта, когда у вас осталось всего 4% и вы находитесь на улице.

Не знаю, что из этого правда.

Но то, что такие данные доступны рекламодателям, предполагает, что они как минимум могут задуматься об их использовании.

Лично я бы задумался.

Ну да ладно, закончим с запросами.

Мы уже увидели примеры разных утечек IP и геолокации.

Ещё одним «поставщиком» **получившим мой IP** + метку времени был `adjust.com`, но тело запроса оказалось слишком скучным, и я его пропустил.

Давайте поговорим об ID

Должно быть, вы уже заметили в запросах `ifv` и `advertisingTrackingId == IDFA`. Что же это такое?

IFV, или IDFV — это «ID для вендора» («ID for Vendor»).

Это мой уникальный ID для каждого вендора, или разработчика — в данном случае для KetchApp.

Это подтверждается: я установил ещё одну игру KetchApp, записал запросы, и значение `ifv` оказалось для неё таким же.

Advertising Tracking ID — это значение для всех вендоров, оно передаётся приложению, если вы выберете «Allow app to track your activity across ...».

Как мы видим, оно действительно равно `000000-0000...`, потому что я не разрешил приложению выполнять отслеживание.

Я проверил это, вручную отключив и включив опцию слежения для приложения Stack, и сравнив запросы в обоих случаях.

■ И это единственное различие между запретом и разрешением на отслеживание

Я понимаю, что в этом для вас может не быть ничего нового — это особо не скрывается, можно, например, почитать документацию Apple для разработчиков.

Но я считаю, что это не доносится до конечного пользователя корректно: что бесплатные приложения, которые вы устанавливаете и используете, **собирают их точное местоположение** с меткой времени и отправляют их сторонним компаниям.

Единственное, что мешает любому с доступом к заявкам на данные (ещё одному покупающему рекламу агенту, бирже рекламы или датасету, купленному или арендованному у брокера данных, как мы увидим ниже) отследить все ваши перемещения — это IDFA, который не передаётся, когда мы запрещаем приложениям «отслеживать нас между приложениями» с целью «улучшения и персонализации отображения рекламы».

Кстати: если вы пользуетесь десятью приложениями одного вендора (Playrix, KetchApp или какой-то другой компании с тысячей приложений) и разрешили лишь одному приложению отслеживать вас, то данные, собранные во всех десяти приложениях, пополнят ваш IDFA, который позже может быть заменён на ваши личные данные.

В то же время, в запросах есть не так много данных, как я ожидал, поэтому биржи рекламы вряд ли нашли какой-то хитрый обходной ID, который бы позволил выполнять отслеживание между приложениями без необходимости IDFA.

Я обнаружил примерно двадцать ID наподобие `tid`, `sid`, `device_id`, `uid` (два последних передаются Facebook*) и так далее.

Кстати, настоящее безумие, что Facebook* получил мои IP + метку времени без какого-либо адекватного согласия/связи с приложением с моей стороны.

Думаю, Facebook* с лёгкостью сможет восполнить пробелы и связать мой аккаунт с этими данными, как только я залогинюсь в Instagram* или Facebook* по тому же IP-адресу.

Как выглядит поток данных?

Давайте ненадолго вернёмся к запросу, через который утекло моё местоположение, и взглянем на его трассировку. Обратим внимание на сторону посередине: [stack](#) → [o.isx.unity3d.com](#) → [molocoads](#) → bwin (рекламодатель).

Unity [ads] — это SSP (supply-side platform), который участвует как сборщик данных из приложения с помощью SDK.

Вам, как разработчику приложения, не нужно беспокоиться о сборе нужных данных, регистрации в качестве издателя, биржах рекламы или чём-то подобном — достаточно установить SDK и получать деньги.

Ну ладно, а что насчёт Molocoads?

Incremental growth at immense scale

Tap into the vastness of the open internet via our expansive reach.



mobile apps



6.7B devices



190+ countries

Скриншот лэндинга Molocoads

Moloco ads — это DSP-сеть, перепродающая данные множества SSP (например, Unity, Applovin, Chartboost). По сути, каждого из запрошенных хостов, которые я увидел в Charles Proxy.

Она выполняет некую «смарт-оптимизацию» и связывает пустое место под баннер на экране телефона с рекламодателем.

Похоже, moloco агрегирует кучу данных, и практически любой (*нужно уточнить: любая компания, ставшая рекламным партнёром*) может получать доступ к данным, сделав ставку ниже, чем другие.

Или представьте реальную биржу рекламы, которая делает обычные ставки и попутно собирает все данные в качестве «дополнительного заработка».

По сути, именно так получают свои данные аналитические агентства и брокеры данных.

Тут я начал искать любые упоминания Moloco в Telegram и Reddit, и наткнулся на пост, в котором нашёл ответы на многие свои вопросы:

r/adops • 3 yr. ago
Pubh12

...

ELI5: What is the controversy behind "bidstream data"? Are there really no restraints on who gets this data and what they do with it?

I'm not really familiar with the landscape of RTB advertising. I was hoping people in the biz might be able to shine some light on this for me.

I was a little surprised to learn that so many different companies would get , what is essentially , server log data from every page you visit. Hundreds every time you load up a page. Including IP addresses , user agent , cookie, the exact URL etc.

I just don't understand how this works.

ELI5: Что это за скандал с «bidstream data»? Действительно ли нет никаких ограничений на то, кто получит эти данные и что будет с ними делать? автора u/Pubh12 в сабреддите adops

Я не особо знаком с реалиями RTB-рекламы. Надеюсь, люди из этой сферы смогут меня просветить.

Меня немного удивило то, что множество разных компаний может получать, по сути, данные логов сервера от каждой посещённой пользователем страницы. Сотни элементов данных при каждой загрузке страницы, в том числе IP-адреса, user agent, куки, точный URL и так далее.

Я просто не понимаю, как это работает.

В частности, интересен этот комментарий. процитирую его часть:

Они имеют доступ к данным, если договорятся об интеграции с поставщиком bidstream, то есть с SSP. Сам SSP должен уведомлять вендора о том, кому он даёт доступ к заявкам. Обычно для этого достаточно лишь... сделать заявку.

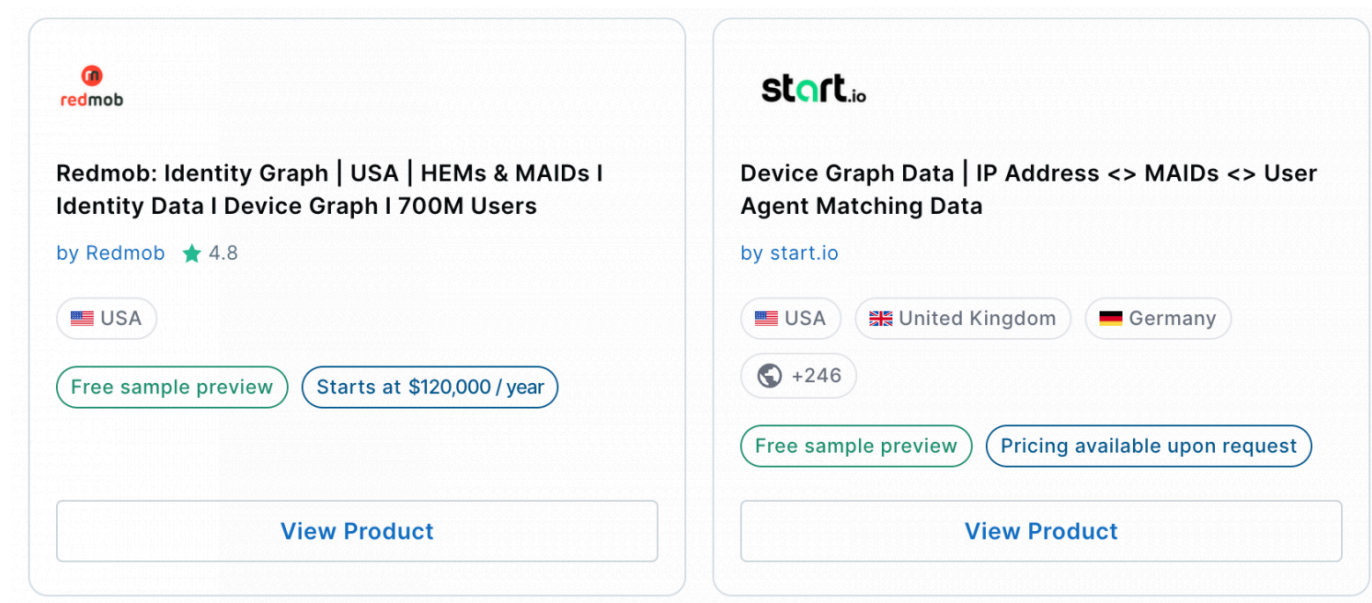
SSP хотят, чтобы вы тратили деньги, именно так их бизнес приносит прибыль. Они могут открывать только часть трафика отдельным вендорам (например, если вы делаете заявки не на весь мир, то не получите всемирный bidstream, а только для тех регионов, в которых ведёте бизнес).

Великолепно.

Брокеры данных

Давайте двигаться дальше. Когда я выяснил, как утекают данные, то начал искать любые места, где их продают. Искать пришлось недолго.

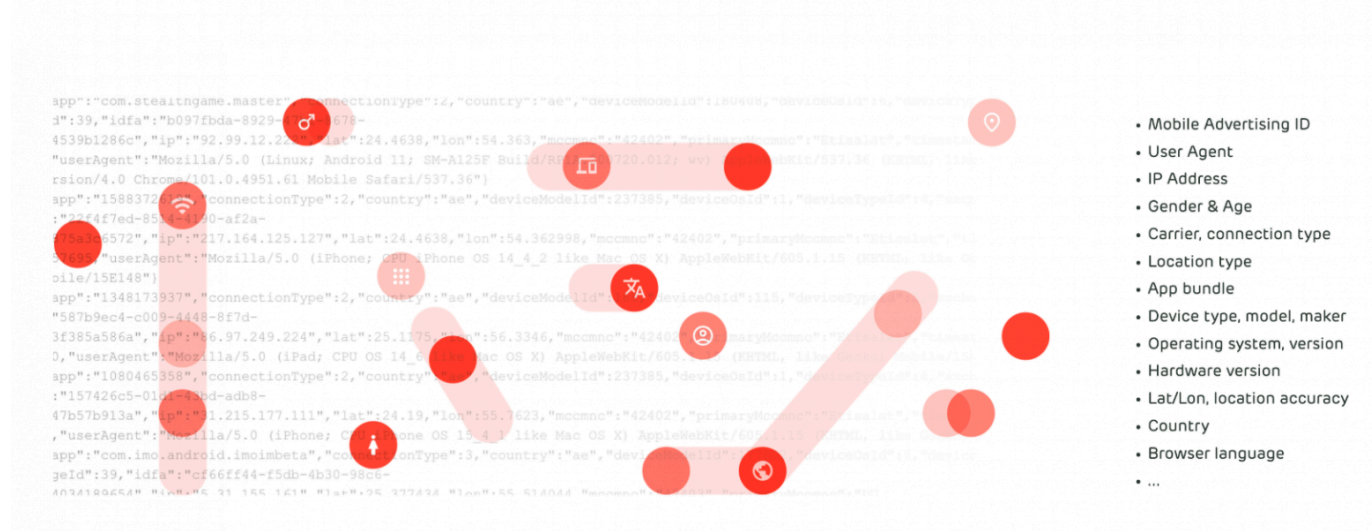
Я обнаружил маркетплейс данных под названием **Datarade**, представляющий собой панель со всевозможными данными. Когда я начал искать данные, относящиеся к MAID, то увидел сотни вариантов на выбор, например, такие:



Цена датасета Redmob удивила меня: 120 тысяч долларов в год... За что?

Давайте посмотрим, что в нём обещают:

Redmob is a data provider focusing on data gathered from smartphones and non-smartphone devices. We unify multiple data sources into a unified data taxonomy that includes insights about the device, carrier, audience behavior, and others. We are dedicated to empowering our customers with the freshest, most accurate, low-latency data for confident decision-making.



Посмотрите на список возможностей справа: не кажутся они вам знакомыми?

Примечание: «low latency» означает, что в датасете есть ваше местоположение за то время, когда его в последний раз передало любое из приложений. То есть может быть даже 5 секунд.

Более того: Redmob предоставляет **бесплатный образец** своих данных.

Я попробовал запросить его на веб-сайте, но образец мне на почту так и не пришёл (наверно, timsh.org не кажется заказчиком с высоким потенциалом).

К счастью, образец публично доступен на Databricks Marketplace с такой аннотацией:

Обогатите свои продукты и сервисы при помощи наших глобальных данных о местоположении, охватывающих более 1,5 миллиарда устройств. Благодаря нашему подробному датасету местоположений вы сможете выявлять скрытые паттерны, выполнять мгновенный анализ и получать глубинное понимание информации.

Также мы предоставляем данные по конкретным регионам (странам Ближнего Востока и Северной Африки, Африке, Азиатско-Тихоокеанскому региону и так далее), исходя из ваших потребностей. В нашей модели ценообразования есть опция ежегодного лицензирования. Кроме того, мы предоставляем бесплатный образец данных, чтобы вы могли самостоятельно оценить качество нашего датасета.

Raw results ▾ +

	id	app	lat	lon	ip	user_agent	yod	country
1	> 54e44147-5657-47d1-9...	> com.onlabgames.Drawingt...	26.972972869873047	81.31372833251953	null	> Android 11 SM-A505F B...	1993	in
2	> 5770e110-a057-45a4-a8...	do.multiple.cloner	12.944018363952637	77.6288070678711	null	> Android 10 SM-J600G B...	1970	in
3	> e8a01bcc-aacb-49e1-9c...	com.tedrasoft.fourpicsoneword	25.639638900756836	88.13812255859375	null	> Android 10 Nokia 6.1 Bu...	1999	in
4	> 466d6e70-8d34-49ef-9b...	com.noomilabs.doublecorks	26.954954147338867	75.7620849609375	null	> Android 11 vivo 1920 B...	1994	in
5	> 57d2b552-81ac-4c5b-b8...	com.skout.android	16.531003952026367	81.52140808105469	null	> Android 11 CPH1937 Bu...	1995	in
6	> c3ab7bf4-6c10-4234-9a...	com.sd2	12.810810089111328	79.88031005859375	null	> Android 10 vivo 1909 B...	2000	in
7	> d9193244-3d51-47d3-8...	multi.parallel.dualspace.cloner	28.677675247192383	77.3235092163086	null	> Android 5.1 A1601 Build...	1987	in
8	> 3e46ab56-74a8-483a-99...	> com.level9.olitrain.transpor...	26.52252197265625	90.5171127319336	null	> Android 9 RMX1811 Buil...	1979	in
9	> ace3e329-c8a0-43f0-83...	com.blued.international	26.299158096313477	73.03114318847656	null	> Android 10 M2006C3LII ...	1980	in
10	> ea03abc3-071a-4ae9-82...	com.taggedapp	17.616336822509766	83.18049621582031	null	> Android 11 Redmi Note ...	1983	in

Часть образца данных для большей наглядности

Самой абсурдной для меня частью стал столбец `app` — очевидный источник данных. Также меня заинтересовал столбец `yod` — если это год рождения, то откуда его берут? Впрочем, ладно, кого интересует год рождения.

Покажи мне персональную информацию!

Ну ладно, допустим, я купил доступ к огромному потоку данных Redmob.

Но моя цель — отслеживать и преследовать людей вроде меня, поэтому мне каким-то образом нужно обмениваться MAID (= ifa) для персональной информации: имени, адреса, номера телефона...

Без проблем! Как ни удивительно, такой датасет тоже присутствует в Datarade.

Взгляните на таблицу из образца с типом MAID <> PII , предоставленным «AGR Marketing Solutions»:

AGR_Mobile_Intent_PII20240903.xlsx (Google Docs)

Внутри находится вся персональная информация (полное имя, почтовый адрес, номер телефона, физический адрес, владение собственностью... и IDFA).

Поздравляю, мы добрались до дна этой кроличьей норы.

Под конец сделаем пару смелых заявлений.

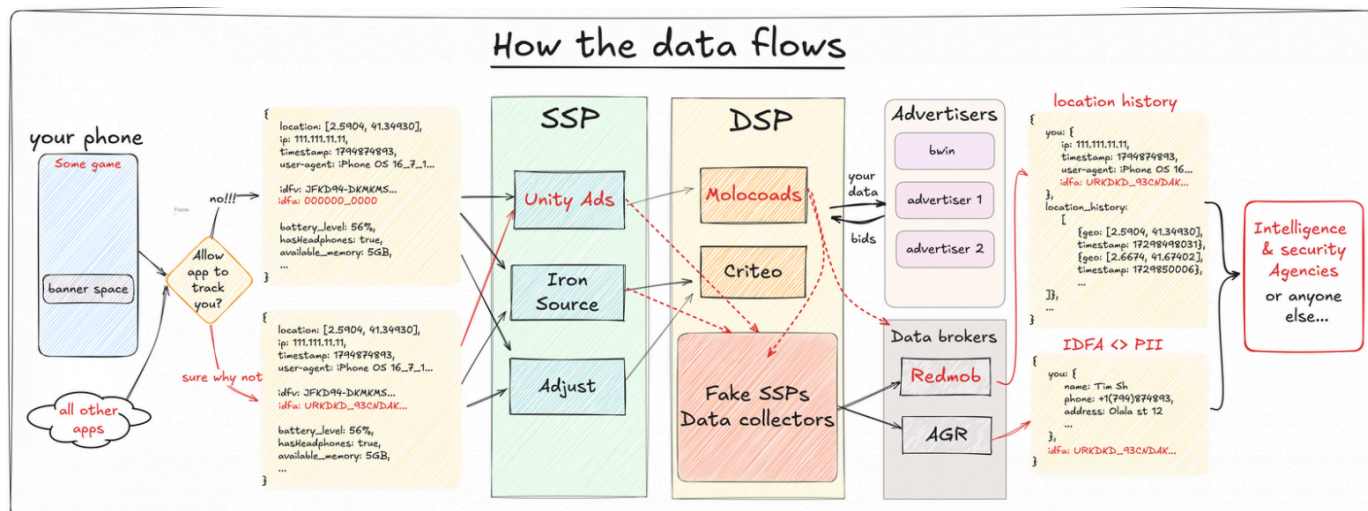
Как себя отследить?

Очень просто! Достаточно воспользоваться этой инструкцией:

1. Немного попользуйтесь бесплатными приложениями. Перемещайтесь на машине и пешком — это делает геопозиционные данные более ценными.
2. «Разрешите» или «попросите не отслеживать» — комбинация из IP + местоположения + User-agent + геолокации всё равно утечёт сотням «сторонних компаний», вне зависимости от вашего выбора.
3. Подождите несколько секунд, чтобы фальшивые DSP и брокеры данных получили ваши данные.
4. Обменяйте ваше полное имя или номер телефона на IDFA (если он есть), IP-адрес и user-agent через купленные где-нибудь данные MAID <> PII .
5. Теперь зайдите в «Mobility data», содержащие историю геолокаций, и отфильтруйте их по значениям из предыдущего этапа.

Поздравляю! Вы себя нашли.

Я создал диаграмму, в которой содержатся почти все упомянутые выше акторы и данные. Теперь вы можете увидеть, как это всё связано.



Это самое худшее в таких торгах данными, которые постоянно происходят по всему миру — каждая отдельная мелкая часть их легальна (или кажется такой). Пугают они только тогда, когда видишь картину в целом.

Послесловие

На моё расследование сильно повлияли следующие посты и исследования:


The Global Surveillance Free-for-All in Mobile Ad Data

Candy Crush, Tinder, MyFitnessPal: See the Thousands of Apps Hijacked to Spy on Your Location

Under Surveillance. How Location Data Jeopardizes German Security

* Деятельность Meta — соцсети Facebook и Instagram — запрещена в России как экстремистская

Telegram-канал со скидками, розыгрышами призов и новостями IT 

Дарим панель управления  **ispmanager**

Пользуйтесь панелью бесплатно при создании VPS на любом тарифе до конца года

Теги: unity ads, мобильные приложения, мобильная реклама, слежка за пользователями, таргетированная реклама, брокеры данных, ruvds_статьи

Хабы: Блог компании RUVDS.com, Аналитика мобильных приложений, Информационная безопасность, Монетизация мобильных приложений, Разработка мобильных приложений

 +219 339 125

RUVDS.com

VDS/VPS-хостинг. Скидка 15% по коду **HABR15**[Telegram](#) [ВКонтакте](#) [X](#)

584

288.8

Карма

Рейтинг

@ru_vds

Пользователь

[Подписаться](#) Комментарии 125

Публикации

[ЛУЧШИЕ ЗА СУТКИ](#)[ПОХОЖИЕ](#)

DmitryOlkhovoi

21 час назад

Меня заставили повайбкодить



Сложный



18 мин



22K

[Кейс](#) +89 97 93

RationalAnswer

13 часов назад

Тотальный блэкаут на юге Европы, а также чудеса нейро-лизоблюдия от ChatGPT



11 мин



11K

[Дайджест](#)

 +32 12 65**lozhnikov**

22 часа назад

Как доказывали теорему о четырех красках. Часть 1



Простой



11 мин



2.8K

FAQ

 +30 23 1**MrSotnik**

8 часов назад

Новый язык от 1С: Зачем? Кому? Стоит ли лезть?



5 мин



14K

 +29 22 50**alizar**

9 часов назад

Ян Лекун, создатель LeNet, формата DjVu и адвокат опенсорса



Средний



7 мин



1.3K

Обзор

 +26 8 6**EI_Gato_Grande**

9 часов назад

Как ИИ-контент проклял интернет и почему это закономерно



8 мин



3K

 +22 7 5**ru_vds**

5 часов назад

Как защищают фильмы и доставляют их в кинотеатры



Средний



12 мин



1.8K

Обзор

Перевод