TikTok, the evil within

Ivan Arbaliev

Summary

In this research publication I will be exploring the techniques used by TikTok for collecting user data, what is it used for and if it is collected by the Chinese authorities eventually. I will collect information about TikTok's worst and best practices for collecting user data using information I find in my research as well as past research by third party researchers found on the internet.

In the ever-evolving landscape of social media, TikTok has emerged as a global sensation, captivating millions of users with its short-form videos and engaging content. However, behind the viral dances and entertaining challenges lies a complex web of data collection practices that raise questions about user privacy. In this article, we will dive into how TikTok collects data from its users and explore its implications for our digital lives.

How did TikTok become the behemoth it is?

TikTok Is an app for creating and sharing short videos. The videos are tall, not square like on Snapchat or Instagram, and you scroll through them like a feed rather than pressing or swiping from side to side. TikTok's vast reasoning and suggestions make it simple for you to create videos, in addition to the resources it offers users. You can choose from a wide variety of sounds, including snippets from TV shows, YouTube videos, and other TikToks as well as popular song excerpts. You can take part in a dare-like challenge, a dancing meme, or a joke. Alternatively, you might mock each of these things.

Contrary to a common belief, Byte Dance (the company standing behind the most popular application in the app store) is not a social media company. It is an artificial intelligence company first. The AI algorithm is the reason behind TikTok's popularity. It studies its users, creates profiles for each unique account, and chooses which video to show you next. This ensures that the app will not get boring after a while. To a non-IT person, it doesn't look like a lot is happening behind the curtains but to the trained eye of a cyber security analyst, it is raising some serious concerns.

We know that artificial intelligence algorithms work best when provided with a lot of data to crunch. Companies are known for collecting data to ensure a better user experience with their customers, but TikTok has taken it to another level.

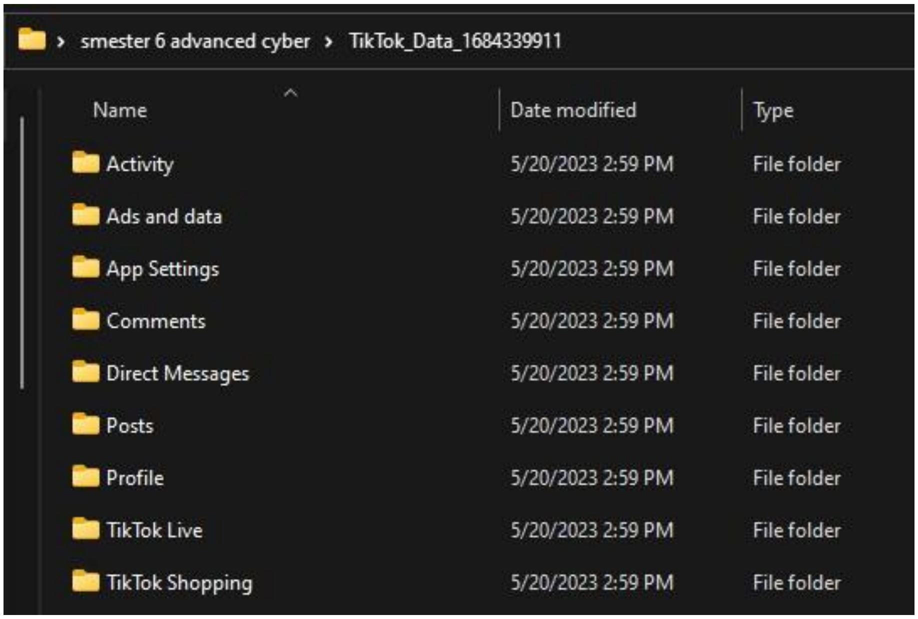
What kind of data is collected from the user?

Doing my research on the topic I found that TikTok's data acquisition strategy can be controversial and sometimes on the border of intrusive to the average user. Thanks to the increasing weight of European laws regarding user data, big companies like Byte Dance, Google, and Facebook are forced to comply with them. This is why in recent times it has become mandatory for big companies that handle lots of valuable personal information to provide a way to download your digital footprint, read how much and what sort of data has been collected from you, and decide whether you want to take part in it or stop using that service.

Such companies don't like that the GDPR laws are forcing them to make changes in their business plans, that's why they comply but on their own terms. For example, I wanted to request my data from TikTok because I have been using the app for a couple of years and I have accumulated enough data to conduct proper research on their data-collecting methods. I noticed that TikTok made it specifically hard for its users to request their digital footprints. I had to go through menus and submenus to find the "request my data" button and once I clicked it a message appeared, telling me that my request is being evaluated. Once TikTok has approved my request I would be able to download MY data, but there was a catch: The request approval could take up to a week to be approved, and once approved I would be able to download my data only for 4 days until it is locked up again. TikTok states that this practice is done to ensure data privacy and confidentiality but, they are making it increasingly difficult to download user's data with the intention that users will lose interest or give up on it. Is Byte Dance afraid to show you your data?

After my request got approved by TikTok (it took six days) I downloaded my digital footprint and discovered that more data was collected than I originally expected. My whole activity on the app was there. Every video I watched, every comment I posted, and even my direct messages with my friends on the app were there.

Fig1: Data downloaded from TikTok regarding my online activity.

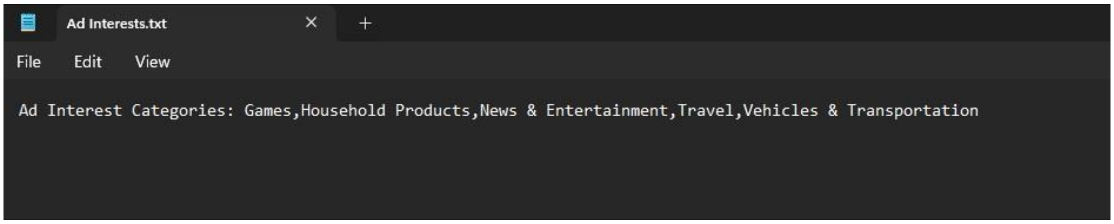


smester 6 advanced cyber > TikTok_Data_1684339911

Name	Date modified	Type
Activity	5/20/2023 2:59 PM	File folder
Ads and data	5/20/2023 2:59 PM	File folder
App Settings	5/20/2023 2:59 PM	File folder
Comments	5/20/2023 2:59 PM	File folder
Direct Messages	5/20/2023 2:59 PM	File folder
Posts	5/20/2023 2:59 PM	File folder
Profile	5/20/2023 2:59 PM	File folder
TikTok Live	5/20/2023 2:59 PM	File folder
TikTok Shopping	5/20/2023 2:59 PM	File folder

I found an interesting file that is directly affecting the algorithm on my feed of videos. I found a file containing my interests according to the algorithm.

Fig2: Screenshot of what TikTok knows about me – Connected to the "interests" algorithm.



```
Ad Interest Categories: Games,Household Products,News & Entertainment,Travel,Vehicles & Transportation
```

Of course, the files I got only represent my online behavior and not the full specter of the data collected. That type of data is collected to make my online experience with the app enjoyable.

In their privacy policy Byte Dance also admits of collecting technical information such as:

- screen resolution
- device type
- time zone
- mobile phone carrier
- battery state
- connected audio devices

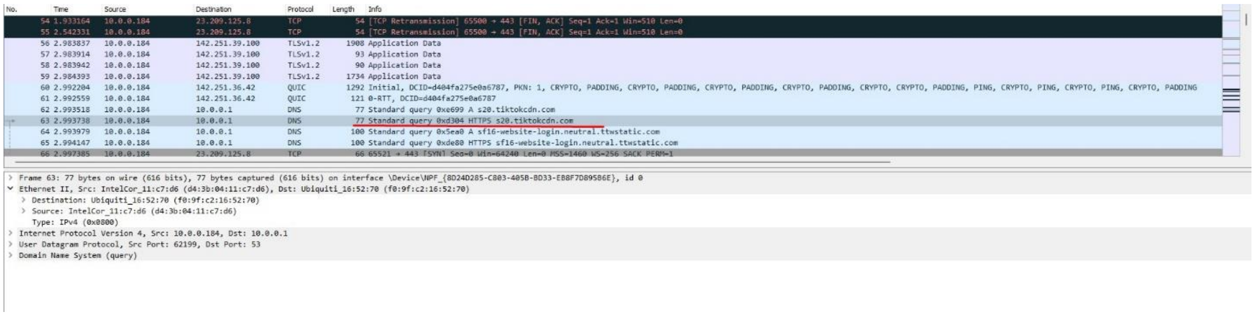
In 2023 an IT company called "Internet 2.0" has examined the app's behavior and snippets of source code. They found that TikTok collects data about what other apps are downloaded on the device, scanning entire hard disks, access contact lists, which is significantly more permission it needs to operate as a social media app.

Disabling TikTok’s permissions is always an option to ensure as much privacy as possible but the developers at Byte Dance have implemented popups that prompt you to enable the permissions back again. They pop up in short intervals to make sure they are as intrusive as possible. Also, it’s important to mention that every TikTok permission is enabled by default.

Does TikTok send data about its users to China?

Capturing network communication from TikTok was straight forward with Wireshark. I managed to capture quite a few packets but unfortunately I didn’t have luck with decrypting them.

Fig3:Wireshark extract showing me connecting to a TikTok server while loading the website

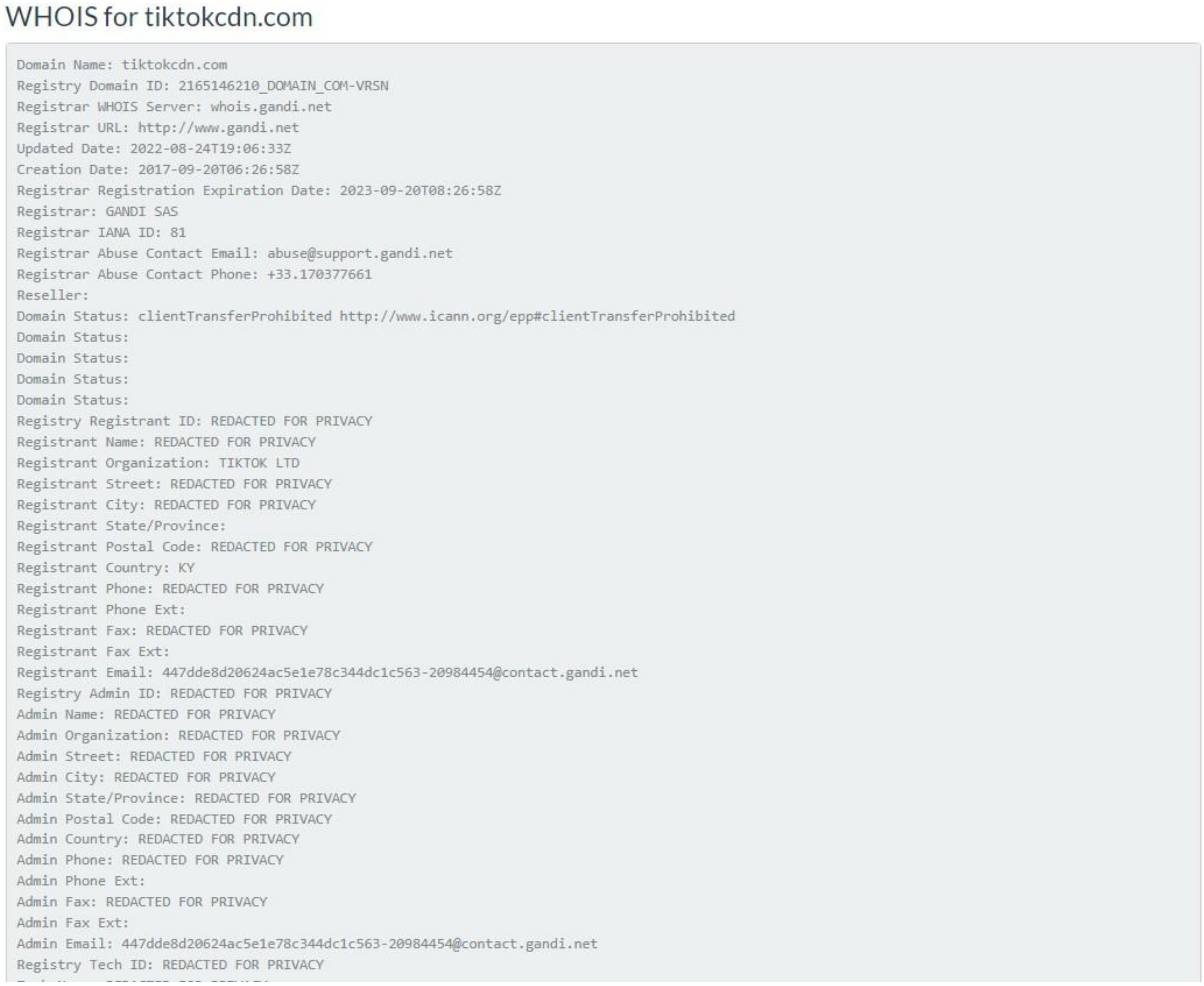


```
No.   Time      Source      Destination     Protocol  Length Info
   54 1.933164  10.0.0.184  23.209.125.8    TCP         54 [TCP Retransmission] 65500 → 443 [FIN, ACK] Seq=1 Ack=1 Win=510 Len=0
   55 2.542331  10.0.0.184  23.209.125.8    TCP         54 [TCP Retransmission] 65500 → 443 [FIN, ACK] Seq=1 Ack=1 Win=510 Len=0
   56 2.983837  10.0.0.184  142.251.39.100  TLSv1.2   1908 Application Data
   57 2.983914  10.0.0.184  142.251.39.100  TLSv1.2     93 Application Data
   58 2.983942  10.0.0.184  142.251.39.100  TLSv1.2     90 Application Data
   59 2.984393  10.0.0.184  142.251.39.100  TLSv1.2   1734 Application Data
   60 2.992204  10.0.0.184  142.251.36.42   QUIC      1292 Initial, DCID=d404fa275e0a6787, PKN: 1, CRYPTO, PADDING, CRYPTO, PADDING, CRYPTO, PADDING, CRYPTO, PADDING, CRYPTO, PADDING, CRYPTO, CRYPTO, PADDING, PING, CRYPTO, PING, CRYPTO, PING, CRYPTO, PADDING
   61 2.992559  10.0.0.184  142.251.36.42   QUIC       121 0-RTT, DCID=d404fa275e0a6787
   62 2.993518  10.0.0.184  10.0.0.1        DNS         77 Standard query 0xe699 A s20.tiktokcdn.com
   63 2.993738  10.0.0.184  10.0.0.1        DNS         77 Standard query 0xd304 HTTPS s20.tiktokcdn.com
   64 2.993979  10.0.0.184  10.0.0.1        DNS        100 Standard query 0x5ea0 A sf16-website-login.neutral.ttwstatic.com
   65 2.994147  10.0.0.184  10.0.0.1        DNS        100 Standard query 0xde80 HTTPS sf16-website-login.neutral.ttwstatic.com
   66 2.997385  10.0.0.184  23.209.125.8    TCP         66 65521 → 443 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 WS=256 SACK_PERM=1

> Frame 63: 77 bytes on wire (616 bits), 77 bytes captured (616 bits) on interface \Device\NPF_{8D24D285-C803-405B-8D33-EB8F7D89586E}, id 0
v Ethernet II, Src: IntelCor_11:c7:d6 (d4:3b:04:11:c7:d6), Dst: Ubiquiti_16:52:70 (f0:9f:c2:16:52:70)
  > Destination: Ubiquiti_16:52:70 (f0:9f:c2:16:52:70)
  > Source: IntelCor_11:c7:d6 (d4:3b:04:11:c7:d6)
    Type: IPv4 (0x0800)
> Internet Protocol Version 4, Src: 10.0.0.184, Dst: 10.0.0.1
> User Datagram Protocol, Src Port: 62199, Dst Port: 53
> Domain Name System (query)
```

I found that I was connecting to a domain called “s20.tiktok.cdn”. I managed to capture only packets from the website version of TikTok. I tried connecting my phone to a proxy on my PC using a custom certificate, but the phone refused to connect.

Fig4: DNS search on a TikTok domain I connected to.



WHOIS for tiktokcdn.com

```
Domain Name: tiktokcdn.com
Registry Domain ID: 2165146210_DOMAIN_COM-VRSN
Registrar WHOIS Server: whois.gandi.net
Registrar URL: http://www.gandi.net
Updated Date: 2022-08-24T19:06:33Z
Creation Date: 2017-09-20T06:26:58Z
Registrar Registration Expiration Date: 2023-09-20T08:26:58Z
Registrar: GANDI SAS
Registrar IANA ID: 81
Registrar Abuse Contact Email: abuse@support.gandi.net
Registrar Abuse Contact Phone: +33.170377661
Reseller:
Domain Status: clientTransferProhibited http://www.icann.org/epp#clientTransferProhibited
Domain Status:
Domain Status:
Domain Status:
Domain Status:
Registry Registrant ID: REDACTED FOR PRIVACY
Registrant Name: REDACTED FOR PRIVACY
Registrant Organization: TIKTOK LTD
Registrant Street: REDACTED FOR PRIVACY
Registrant City: REDACTED FOR PRIVACY
Registrant State/Province:
Registrant Postal Code: REDACTED FOR PRIVACY
Registrant Country: KY
Registrant Phone: REDACTED FOR PRIVACY
Registrant Phone Ext:
Registrant Fax: REDACTED FOR PRIVACY
Registrant Fax Ext:
Registrant Email: 447dde8d20624ac5e1e78c344dc1c563-20984454@contact.gandi.net
Registry Admin ID: REDACTED FOR PRIVACY
Admin Name: REDACTED FOR PRIVACY
Admin Organization: REDACTED FOR PRIVACY
Admin Street: REDACTED FOR PRIVACY
Admin City: REDACTED FOR PRIVACY
Admin State/Province: REDACTED FOR PRIVACY
Admin Postal Code: REDACTED FOR PRIVACY
Admin Country: REDACTED FOR PRIVACY
Admin Phone: REDACTED FOR PRIVACY
Admin Phone Ext:
Admin Fax: REDACTED FOR PRIVACY
Admin Fax Ext:
Admin Email: 447dde8d20624ac5e1e78c344dc1c563-20984454@contact.gandi.net
Registry Tech ID: REDACTED FOR PRIVACY
```

In the article above “Internet 2.0” explains that they have managed to dissect and record the app’s behavior and they have found multiple instances of connections to and from China and Philippines.

Why are countries banning TikTok from governmental devices?

TikTok’s controversial data collection strategy has raised a lot of eyebrows, on how dangerous it is. There have been a lot of debates regarding TikTok’s privacy policy. Of course, Byte Dance denies all accusations and if asked, they will tell you to refer to their overwhelmingly long privacy policy. As we all know TikTok is very heavily sponsored by the Chinese government. Perhaps there is a reason for it? Byte Dance is a Chinese company, and every Chinese company must share the data they have acquired with the Chinese government. Similar tech giants like Google and Facebook collect similar amounts of data about their users but they are in the United States where laws are different, and data is private. The only way for the American government to acquire such data is to have a warrant signed by a judge. Contrary to that Chinese government simply must ask about TikTok's collected data and they simply can't deny. If such data is acquired by a government it can be used to harm other countries, leak classified information, affect key figures in other governments. This is why TikTok has been banned from app stores on governmental devices.

List of countries where TikTok is banned:

- Afghanistan
- Armenia
- Azerbaijan
- Bangladesh
- India
- Indonesia
- Iran
- Jordan
- Pakistan
- Taiwan
- Austria
- Belgium
- Denmark
- Estonia

- France
- Ireland
- Latvia
- Netherlands
- Norway
- UK
- Canada
- US
- Australia

- New Zealand

Conclusion

TikTok Is a record-breaking app, appealing to a wide audience. There are quite a few security concerns regarding user data but nowadays you can’t escape the grasp of the technical giants around us like TikTok, Google, Facebook(meta). If you stop using one service thinking you are protecting your data, you will just use the next one in the list that might do the exact same thing, u was looking to avoid. Overall, there is no real escape from your internet profiling. I am standing behind the TikTok ban from governmental devices because the app might not pose immediate threats to a single user but for a government it could be devastating. If you are using TikTok I would recommend turning off all its permissions and not to text to anyone via the app. Just use a secure app for texting with your friends like WhatsApp, telegram even messenger.

References

Alex Miltsov Researching TikTok: Themes, Methods, and Future Directions 1 Miltsov, A. (2022). “Researching TikTok: Themes, methods, and future directions.” Pp. 664-676 in The SAGE Handbook of Social Media Research Methods, 2nd Edition. - https://www.frontiersin.org/articles/10.3389/fpubh.2021.641673/full

Montag C, Yang H and Elhai JD (2021) On the Psychology of TikTok Use: A First Glimpse from Empirical Findings. Front. Public Health 9:641673. doi: 10.3389/fpubh.2021.641673 - https://www.frontiersin.org/articles/10.3389/fpubh.2021.641673/full

Open Innov. Technol. Mark. Complex. 2022, 8(3), 125; https://doi.org/10.3390/joitmc8030125 - https://www.mdpi.com/2199-8531/8/3/125

Kinnon Ross MacKinnon, Hannah Kia, Ashley Lacombe-Duncan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.12.2021. – https://www.jmir.org/2021/12/e30315/