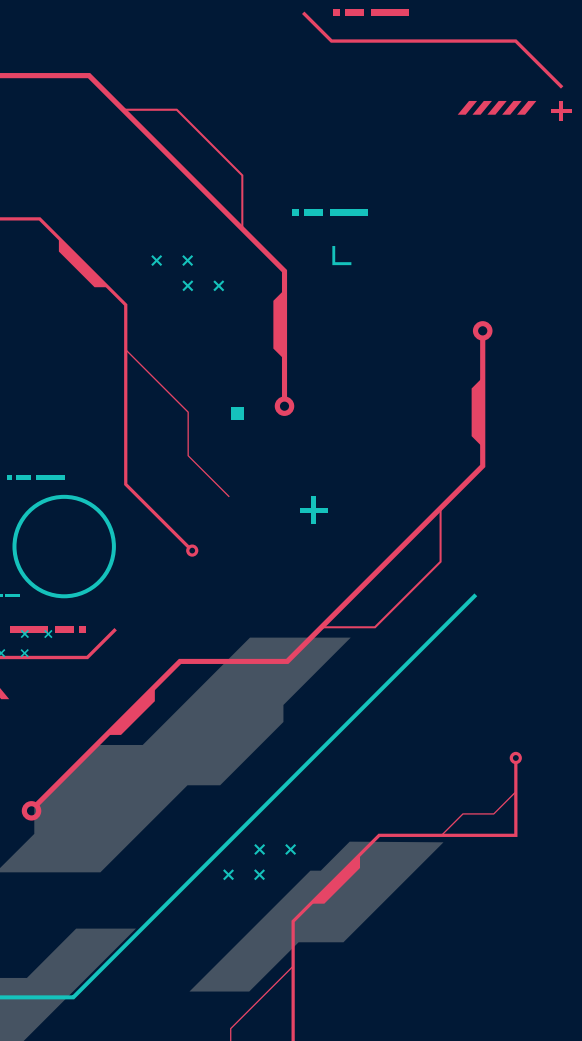
PRIVACY INVASION IN IOS LANGUAGE LEARNING APPLICATIONS

Using open-source tools to detect privacy breaches in language learning apps.

WHOAMI?

- Master’s degree Computer Science
  - Digital Forensic Investigator
- Linux hobbyist (Red Hat enthusiast)



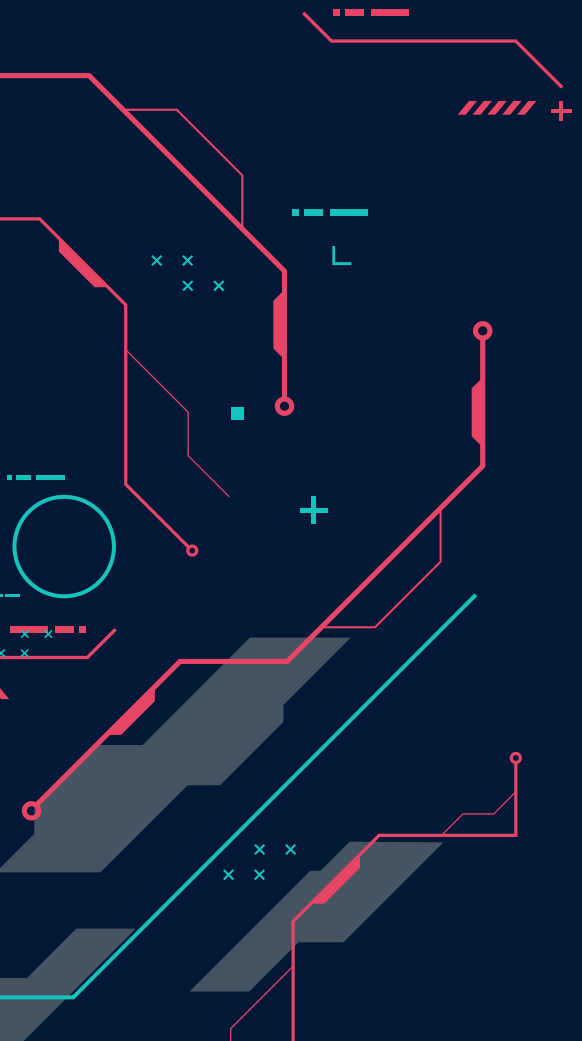
What Is this presentation?

This presentation should serve as a brief explanation of how my Masters research integrated open-source tools for digital forensics and what my research found.

TABLE OF CONTENTS

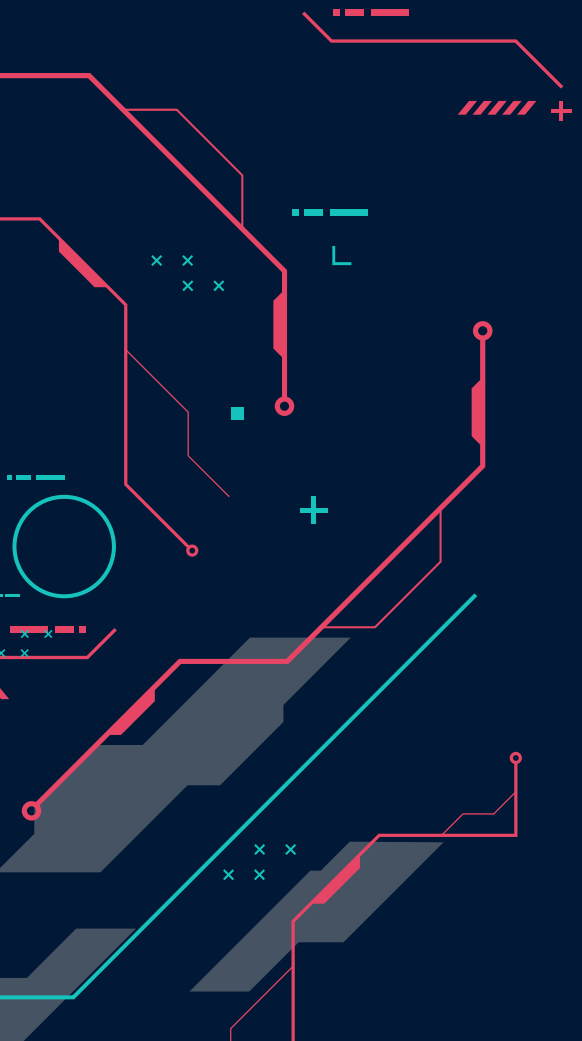
01

BACKGROUND



Background

- I decided to conduct a security analysis of common language learning apps to analyze the differences between free and paid versions.
- After many iterations, I decided to focus on advertising systems and the user information they create and send out.



Overarching question

Is any user-identifying data packaged up and sent in a different way between the free and paid versions of common language-learning apps?

Advertising Terminology

- Demand side client: A client that isrequesting the ad or being advertised to.
- Supply side client: a client that is distributing the ads (i.e. Google Ad Network)
- Real-time Bidding: The split-second process by which advertisers buy the spot that the user will see.
- Interstitial: The property of appearing in between two things.
- CDN: Content distribution network; a series of servers that allows users of a service to query for files of all types.

SAFEGUARDING YOUR ONLINE PRESENCE

Do you know what helps you make your point crystal clear? Lists like this one:

- They’re simple
- You can organize your ideas clearly
- You’ll never forget to buy milk!

And the most important thing: the audience won’t miss the point of your presentation

02

The Apps

Duolingo, Busuu, and Memrise

Duolingo Information

- Most well-known language learning app.
- Free and paid tier.
- Diverse set of features with many languages.
- Has AI voices that aid with pronunciation of languages.
- Has special exercises for learning foreign character sets (i.e. Korean, Greek, Cyrillic, etc)

Duolingo Information

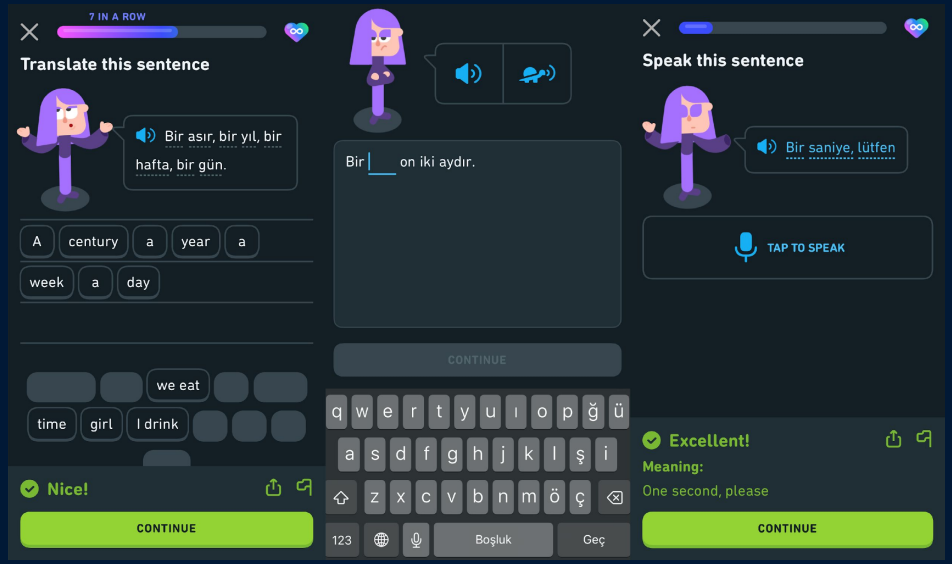


Figure 1a: Duolingo normal exercises.

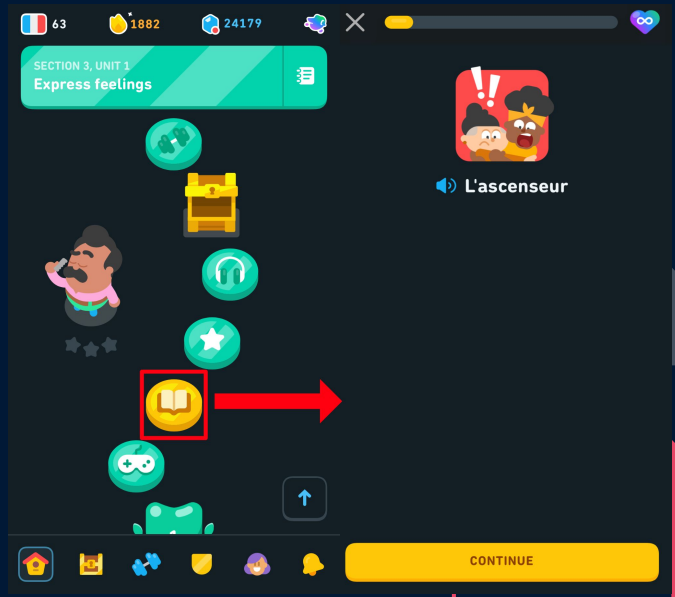


Figure 1b: Duolingo stories.

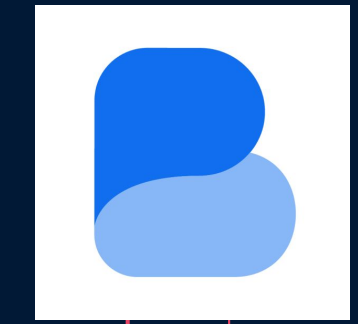
Duolingo Information

Figure 1c: Duolingo Podcast.

Figure 1c: Duolingo Streak and subscription.

Busuu Information

- Another well-known language-learning app.
- Free and paid tier, part of Chegg.
- Grammar, vocab and history lessons.
- Ads come before or after the lesson.
- Features flashcard sections of lessons like Duolingo.



Busuu Information

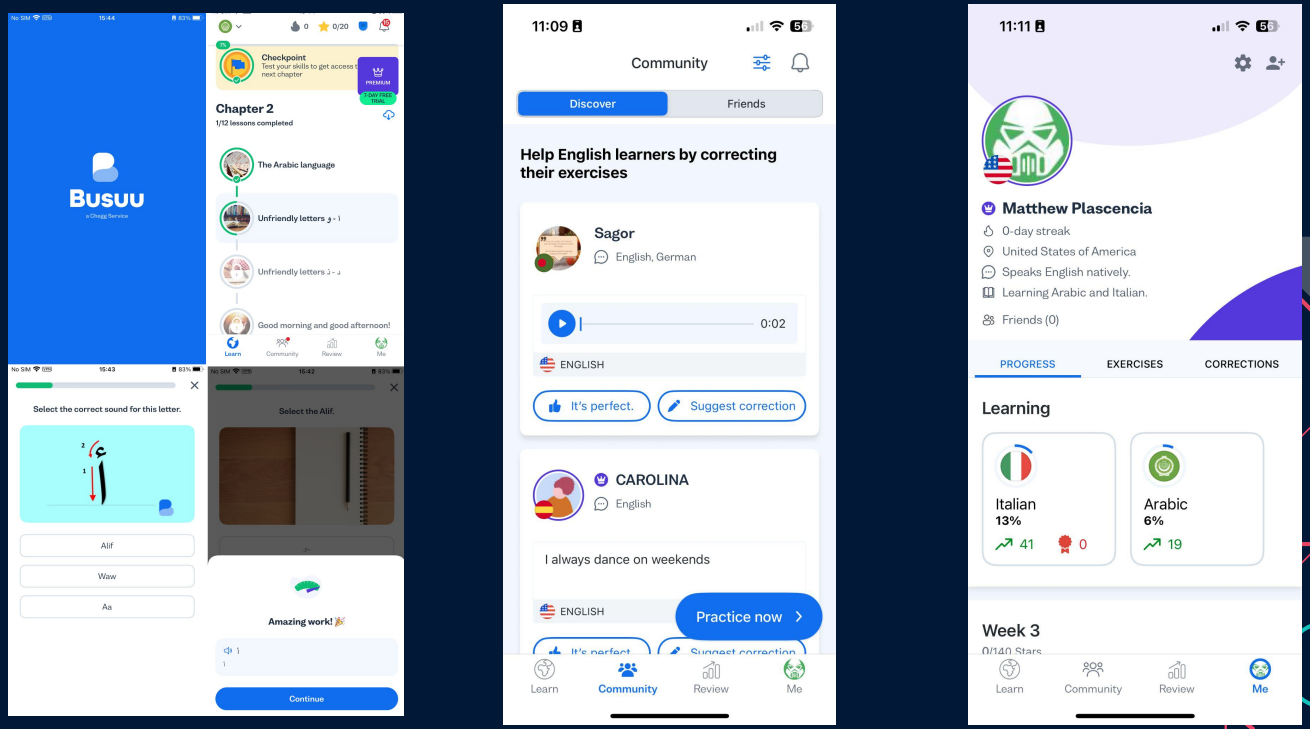


Figure 2: Busuu images.

Memrise Information

- A unique app. Spaced repetition is a noticeable part.
- Free and paid tier.
- Uses a phrasebook format.
- Ads come before or after the lesson.
- Features a flexible platform, including all normal things like the vocab exercises and the grammar ones alongside things like videos and listening exercises.

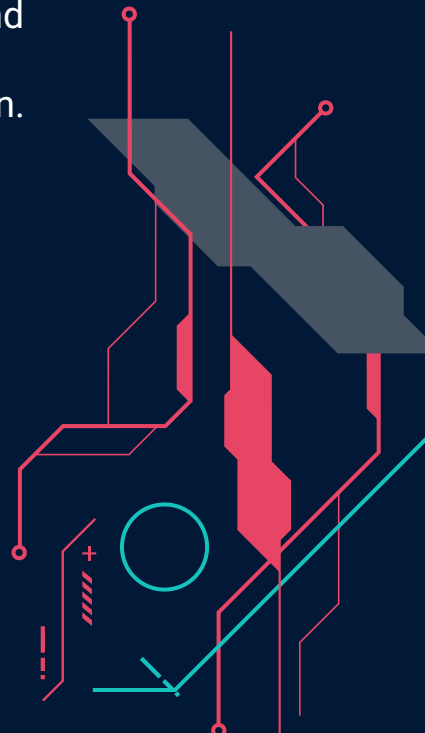
03

The OSS Tools

All of the OSS Tools Uses Throughout the Research.

UFADE (Christian Peter)

- A tool used to pull information from iPhones.
- Supports full file backup (with jailbreak enabled), regular iTunes backup, and logical+ backups.
  - Logical+ backups include iTunes backup and other media information.
- Can also acquire logs of all sorts.
- https://github.com/prosch88/UFADE
- Part of early research in finding what data was on phone.



UFADE (Christian Peter)

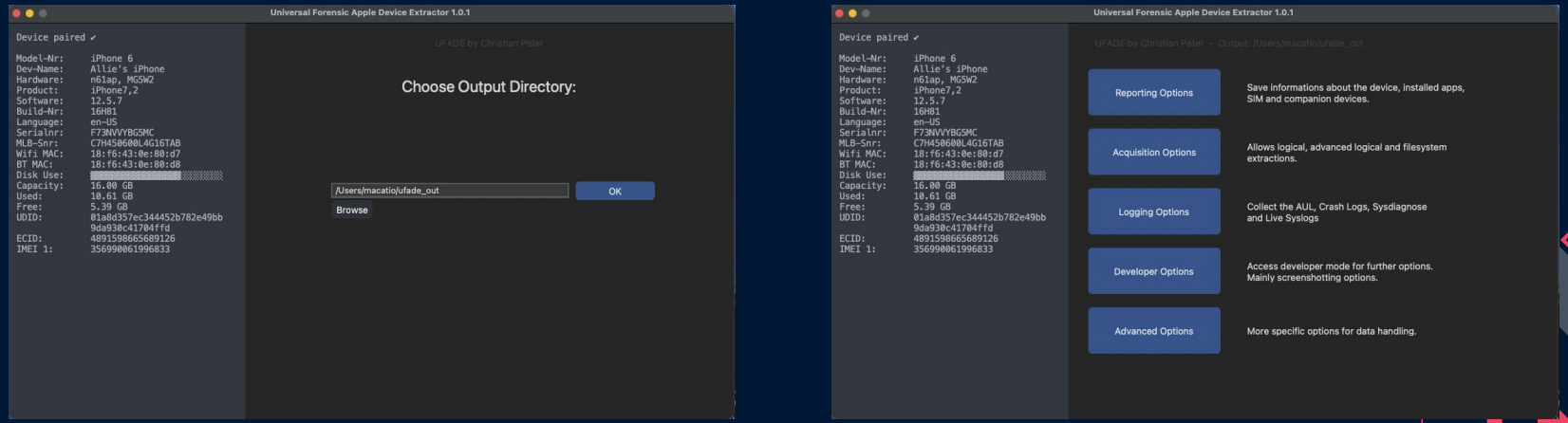
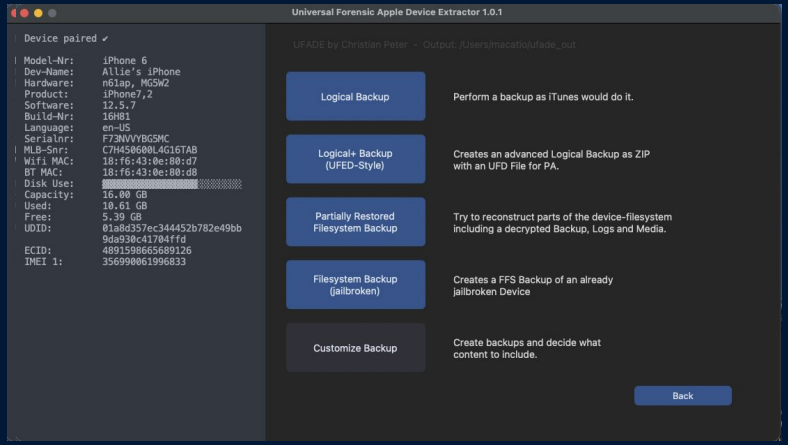


Figure 3a: UFADE Choose Directory and Choose Extraction Option screen.

UFADE (Christian Peter)



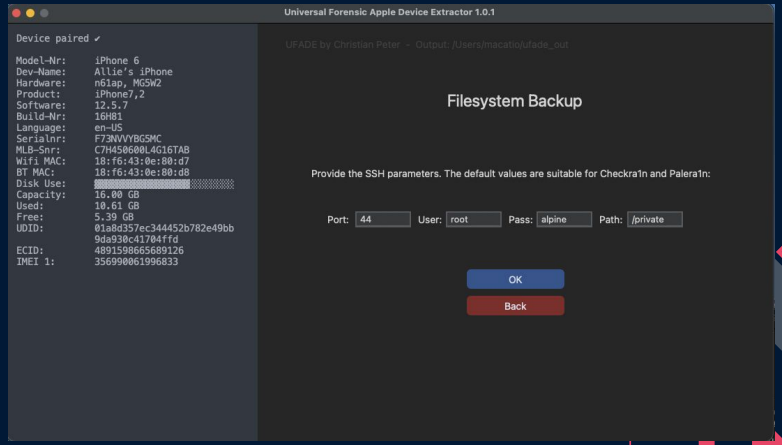
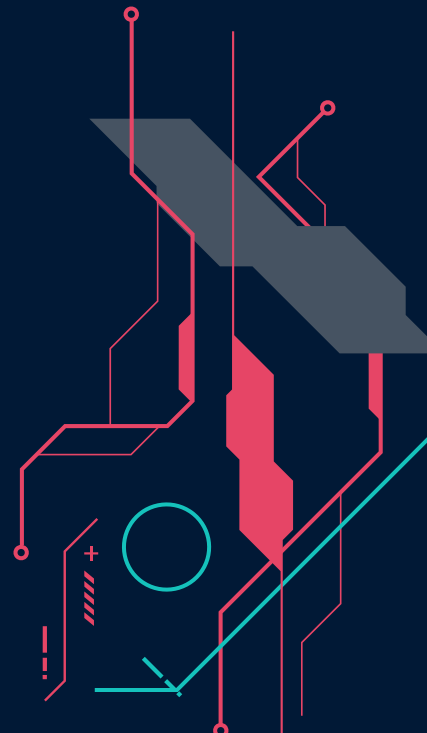


Figure 3b: UFADE selection of Jailbreak acquisition and selection of port options.

Wireshark

- A very common network analysis tool.
- Gathers network traffic from
- Can also acquire logs of all sorts.
- https://wireshark.org
- Part of early research in finding what data was on phone.



Wireshark

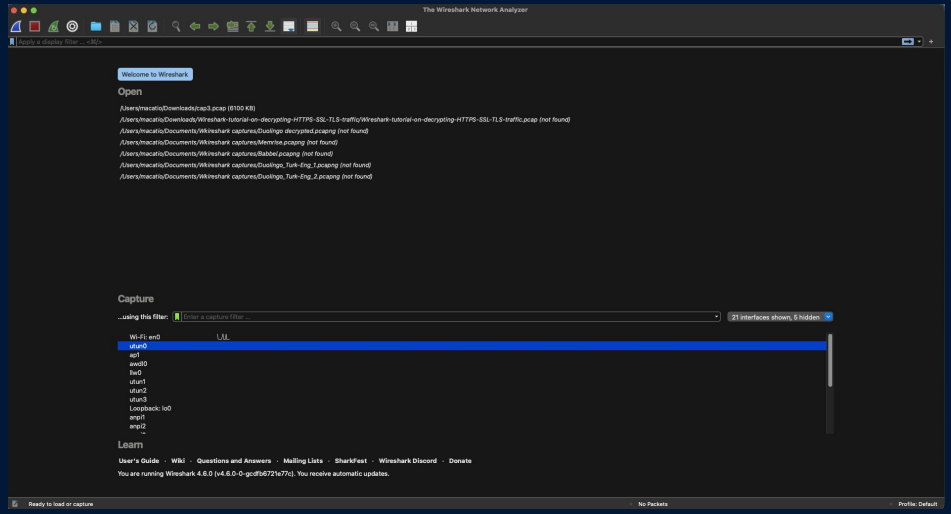


Figure 4a: Wireshark start screen.

Wireshark

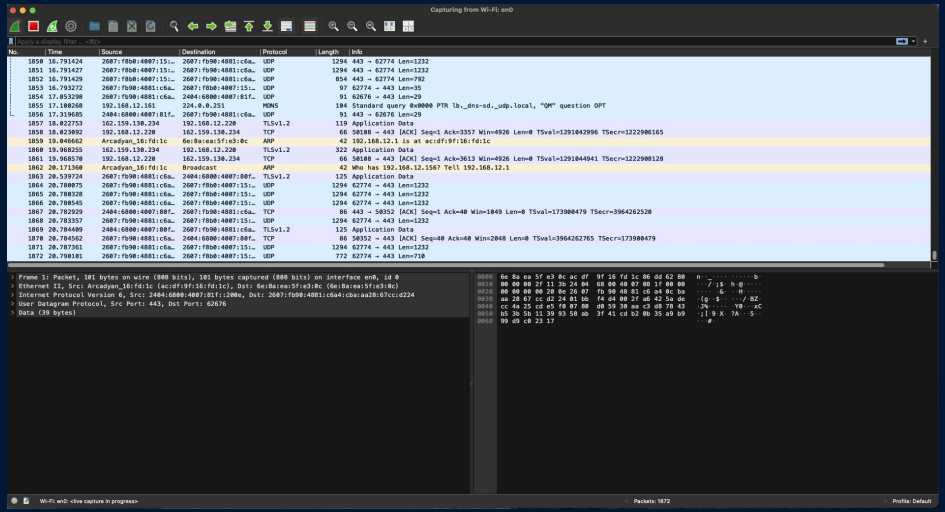
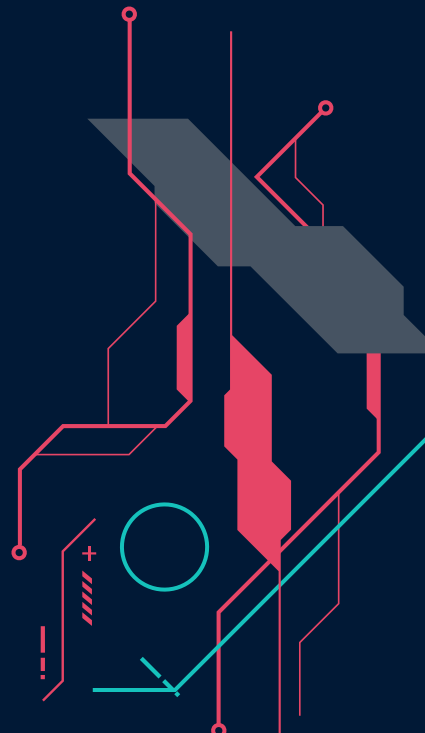


Figure 4b: Wireshark user interface.

WiresharkHelper

- App for MacOS/iOS.
- Helps to bridge the gap between Wireshark and iOS.
- Allows Man in the Middle Proxy so that the Mac can capture the iOS traffic.
- Uses the Remote interface (rvi[x]) to capture the iOS traffic.
- WiresharkHelper is free on iOS and was free on GitHub when I did the research. It is no longer available on GitHub; $10 on Mac App Store.



WiresharkHelper

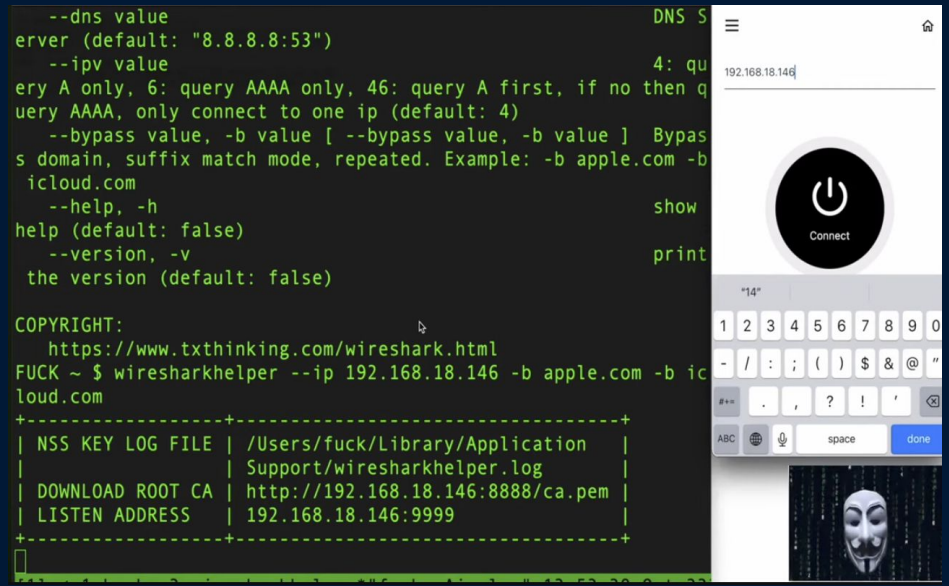


Figure 5a: WiresharkHelper iOS + MacOS

WiresharkHelper

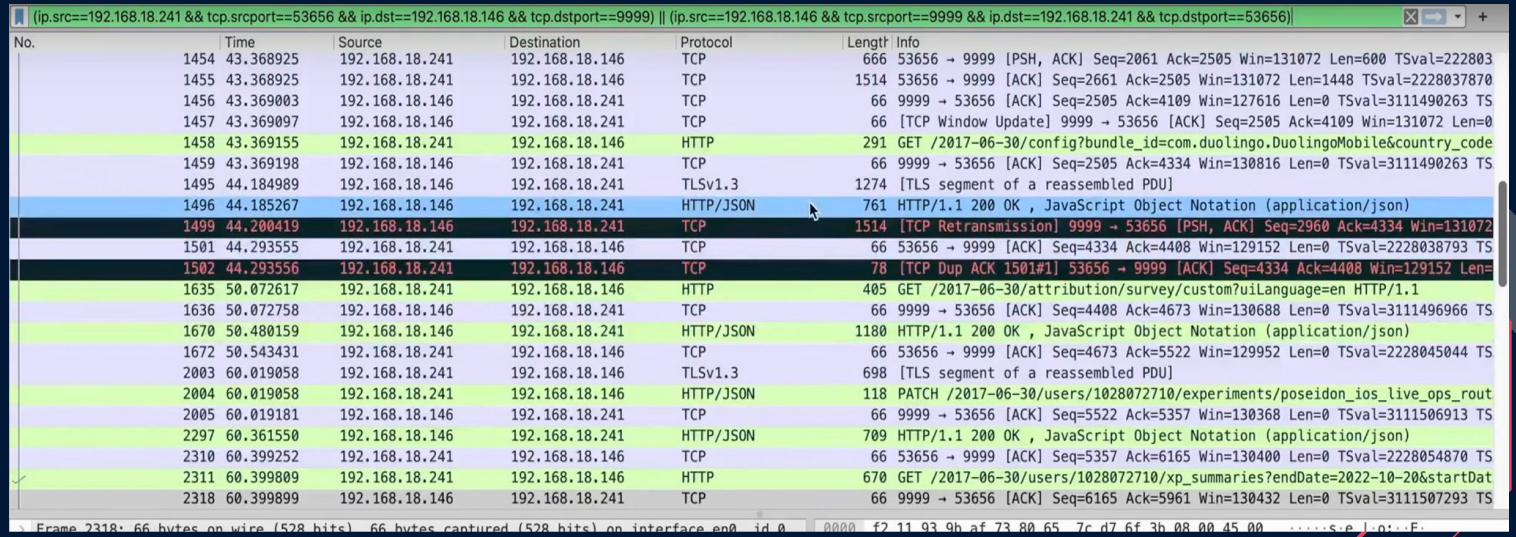
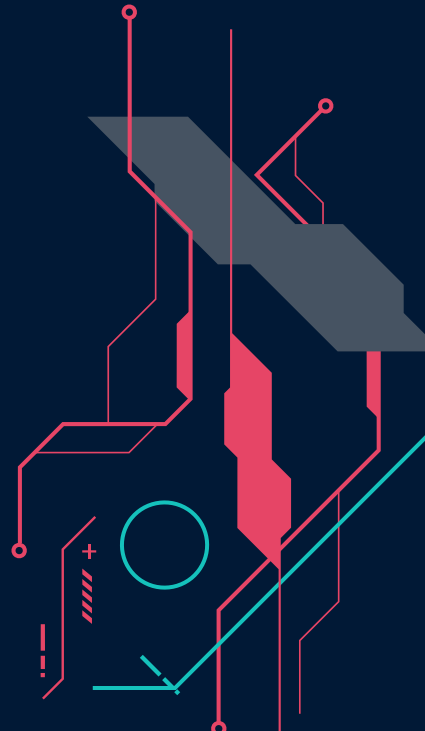


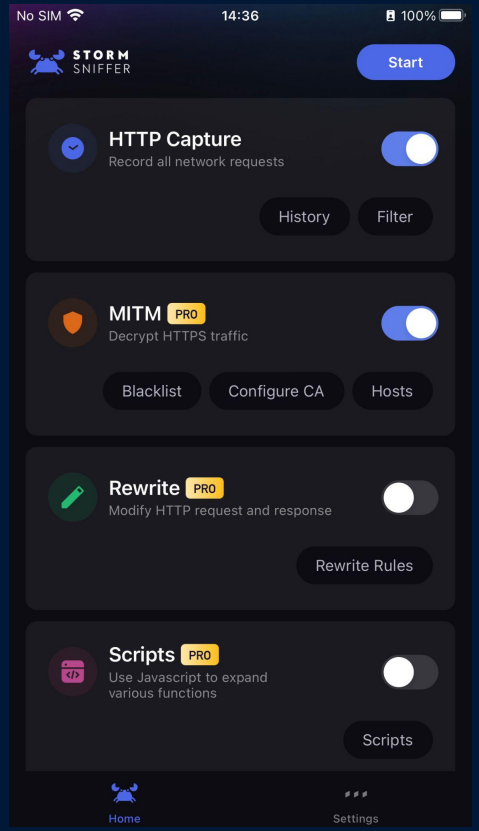
Figure 5b: Wireshark Helper in action: Duolingo packets decrypted.

StormSniffer

- iOS app, (not really open-source).
- All-in-one packet sniffer + HTTPS decryptor.
- Also not fully free: Packet decryptor not completely free.
- User-friendly interface.



StormSniffer



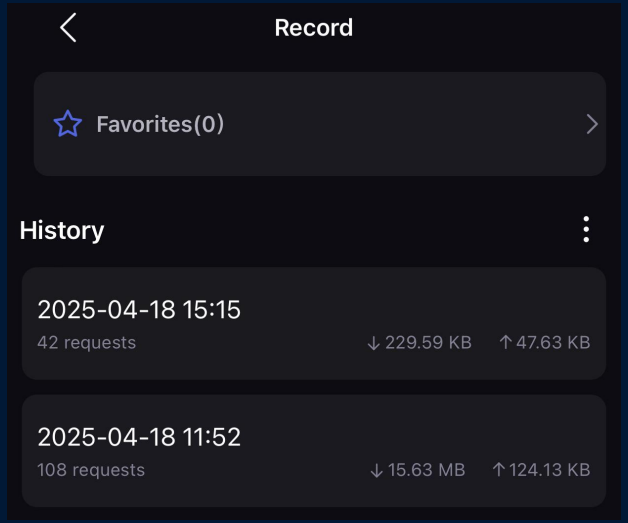
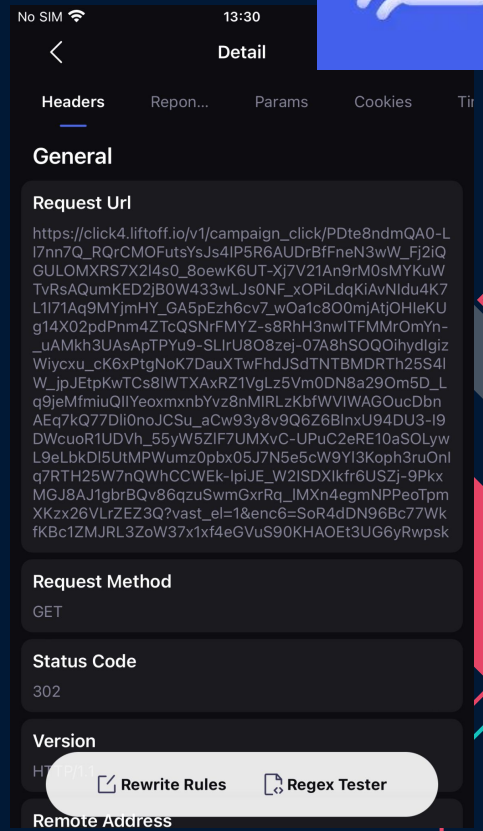


Figure 5: StormSniffer UI



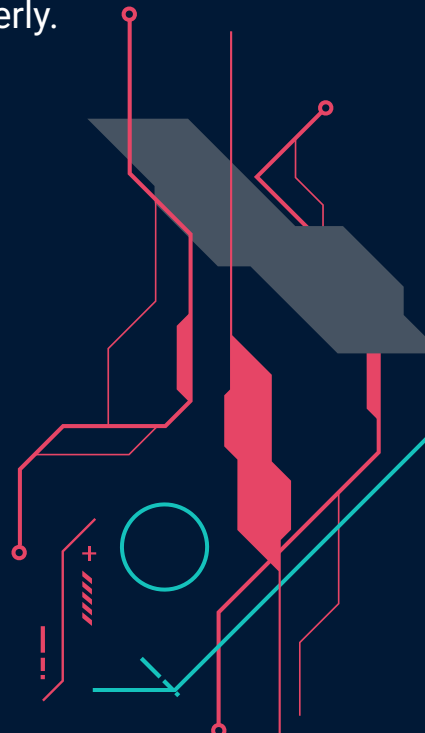
04

Experimentation

Methodologies and Trials + Tribulation

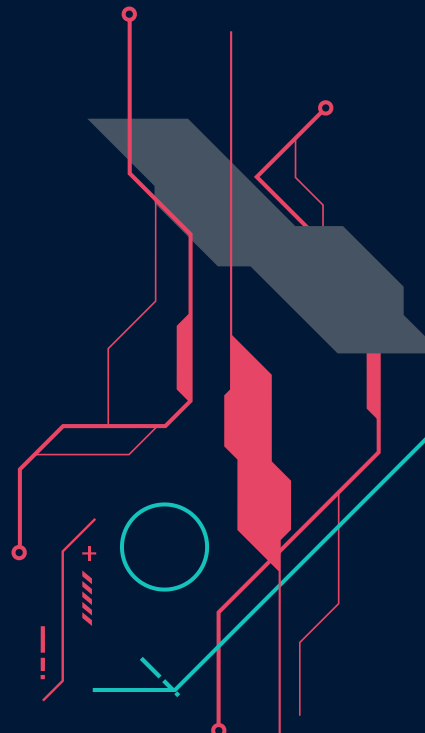
Preparations

- Download all apps on computer/phone.
- Install all requirements and make sure all permissions are configured properly.
- Test everything to see what works.



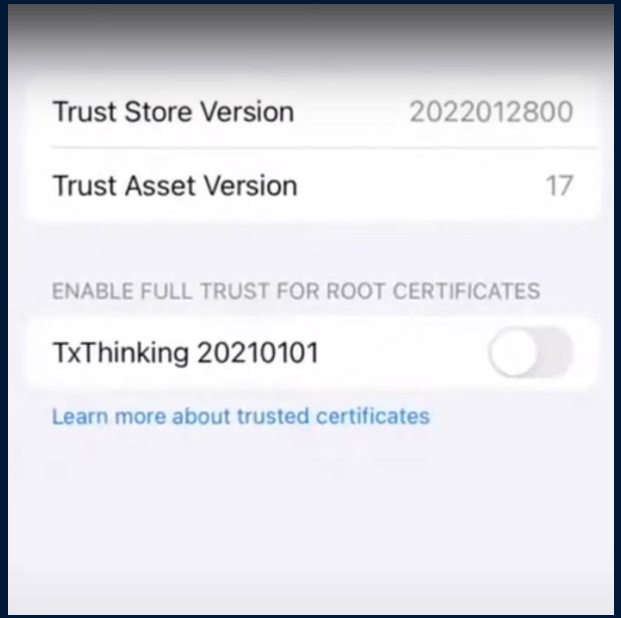
UFADE Usage

- Very beginning phase.
- Extracted full file system after jailbreaking iPhone 8 Plus.
- I wanted to look at what the apps looked like on their file structure levels.
- Found nothing relevant.



Network Sniffing Prep

- First tried Wireshark + WiresharkHelper.
- Had to install and trust a VPN profile for it to work.
- Worked fine for normal traffic; problems for more specific traffic.



Network Sniffing Prep

- Finally tried StormSniffer.
- Had to install and trust a VPN profile for it to work.
- Worked fine for normal traffic; Also checked out for things that WSH did not.

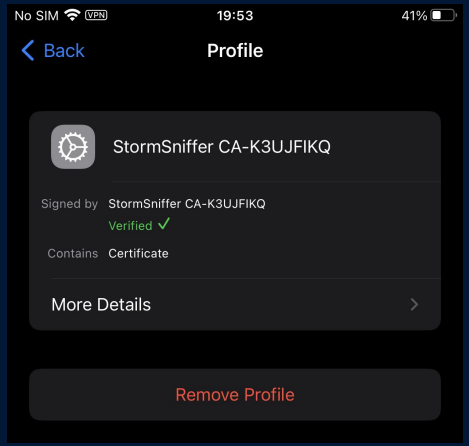
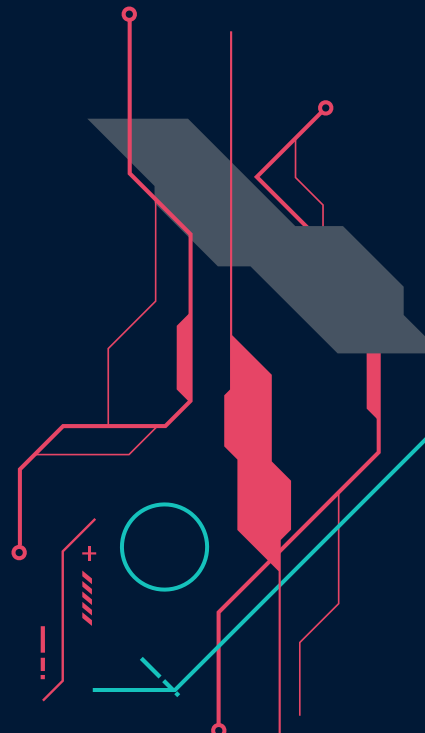


Figure 6: CA Certificate profiles

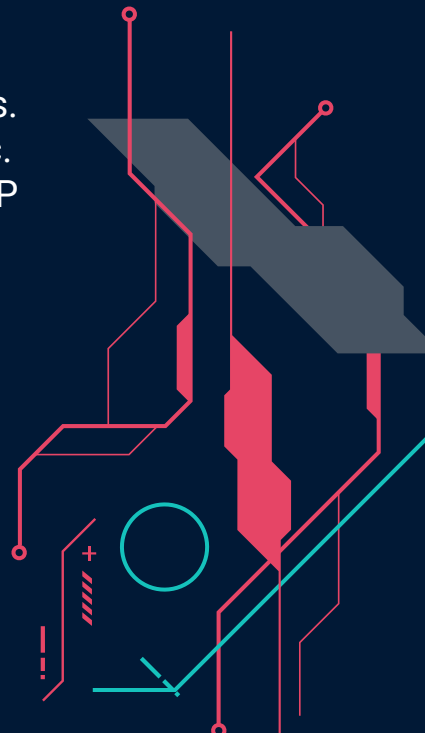
Sniffing: Successes and Failures

- Both solutions captured web traffic well.
- Both caught everything from normal Duolingo/Busuu/Memrise lessons.
- Storm Sniffer picked up audio lessons best.
- WSH failed at moving forward when speaking exercises occurred.
- Both caught ads

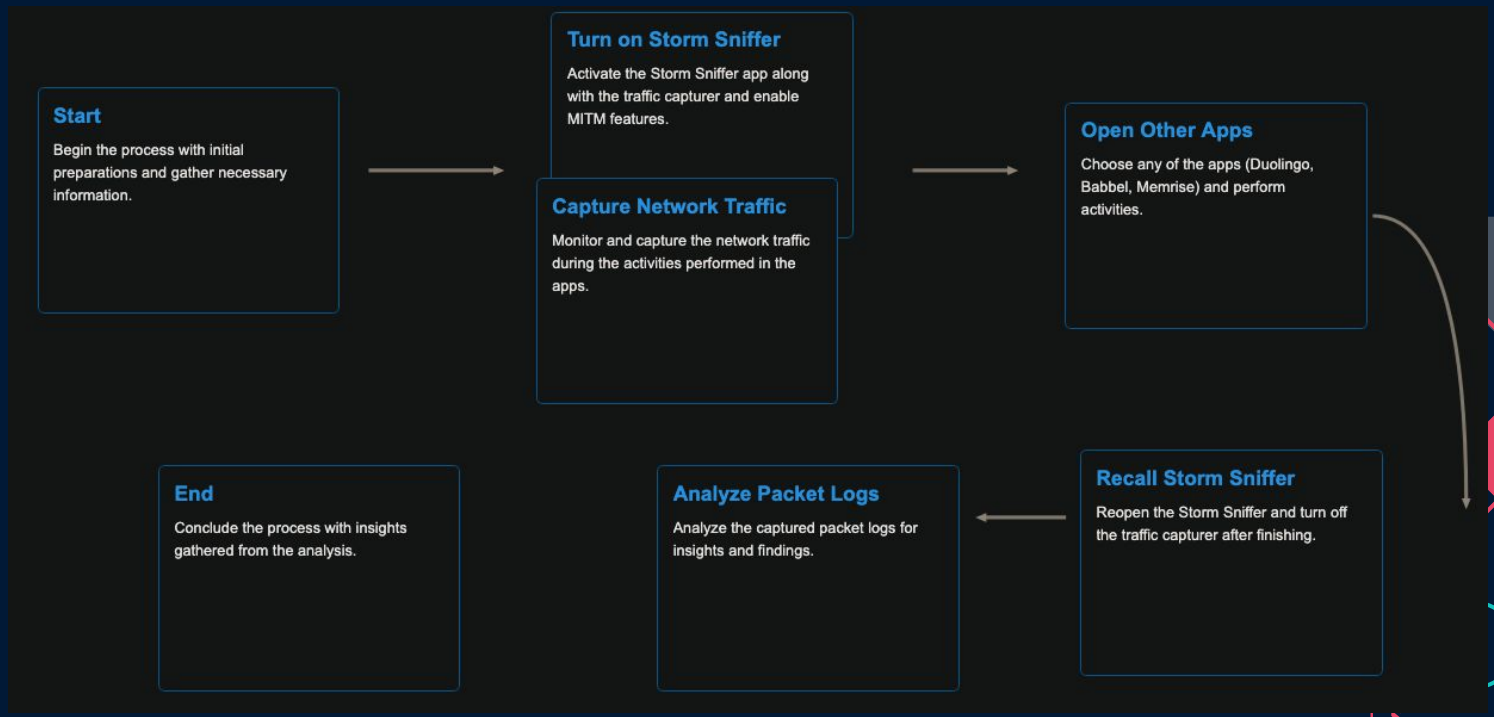


Modifications to Sniffing Process

- When adapted to using tools, modifications were made
  - WSH dropped.
  - Storm Sniffer preferred.
  - Process refined to preferring using the software at the end of lessons.
  - Added in analyzing ad-free traffic after thoroughly analyzing ad traffic.
- Analyzed some traffic files in Wireshark by converting files from HAR2PCAP

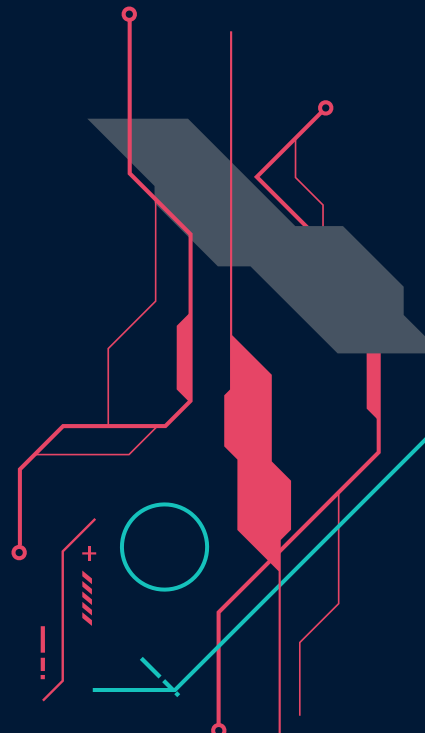


Process Flowchart



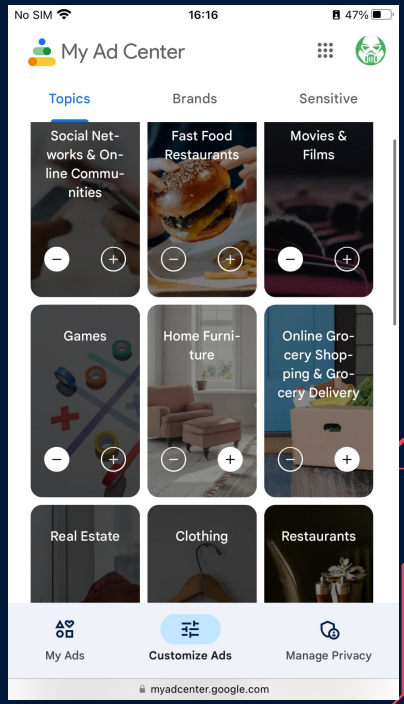
Where does data go on ads?

- *.unityads.unity3d.com
- **o-sdk.mediation.unity3d.com**
- consent-adjust.com
- *.applovin.com
- pubads.g.doubleclick.net



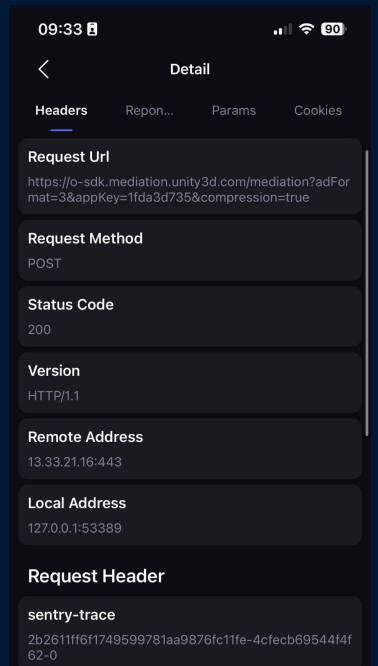
Where are ads from?

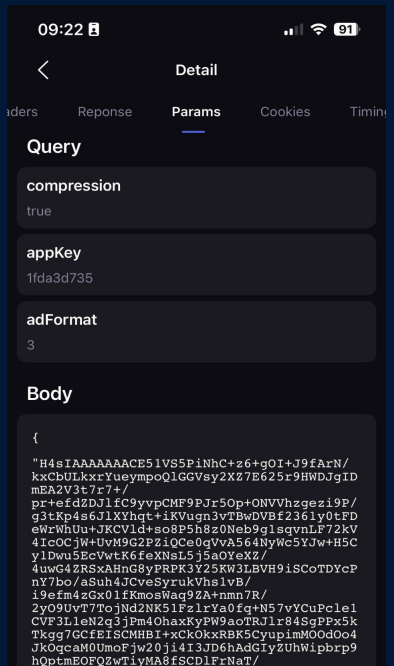
- Unity3d ads
- Google ads net
- Google Ad Center controls what ads you see!



What data goes to ad networks?

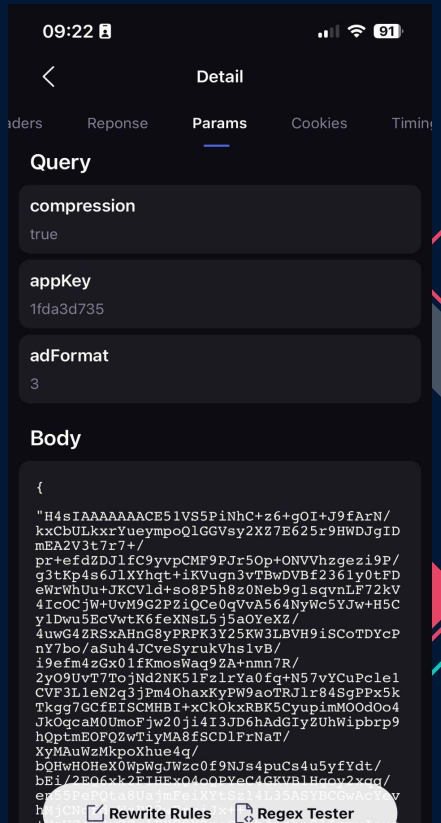
- Data about the device in question is sent towards the ad service and an ad appropriate for that device is returned.





What data goes to ad networks?

- Packet contains Base64 encoded JSON information.
- *nix xystems have the base64 command to decode this
- base64 -Di inputfile | zcat to unroll it.



What does this data look like?

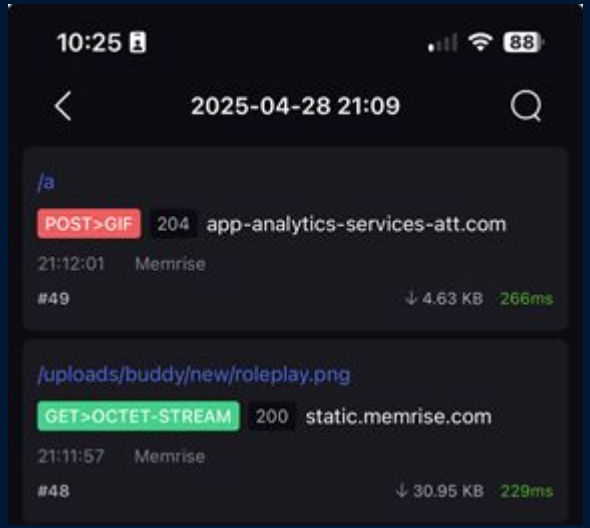

```
{
  "xCodeVersion" : "1600",
  "osVersion" : "16.7.10",
  "connectionType" : "wifi",
  "jb" : "false",
  "bundleId" : "com.duolingo.DuolingoMobile",
  "events" : [
    {
      "eventSessionId" : "9E60256B-C2FD-410A-9E5F-A7FF85533F7E",
      "firstSessionTimestamp" : 1738015217402,
      "timestamp" : 1743644258886,
      "eventId" : 44,
      "rawConnectionType" : "wifi",
      "duration" : 45223
    }
  ],
  "mobileCarrier" : "--",
  "groupNameBN" : "All Countries",
  "abGroup" : "NONE",
  "mt" : "AdMob87000SDK1200iAds510",
  "groupIdRV" : "5814759",
  "abInternal" : "lpmm_abc_0331_disable",
  "tz" : "America\/Los_Angeles",
  "xCodeBuildVersion" : "16A242d",
  "mcc" : 65535,
  "advertisingIdType" : "IDFA",
  "abHashValue" : "0",
  "internalFreeMemory" : 29950,
  "att" : 3,
  "idfv" : "F010F50F-0AC9-48DC-9BE9-21B35021EE78",
  "controllerABValue" : "4",
  "deviceOEM" : "Apple",
  "sdkVersion" : "8.7.0",
  "advertisingId" : "D991D864-A78D-4600-B8E0-1B2D45D43BB5",
  "xCodeSdkBuild" : "22A3362",
  "internalTestId" : {
    "99" : "129"
```

```
  ],
  "mobileCarrier" : "--",
  "groupNameBN" : "All Countries",
  "abGroup" : "NONE",
  "mt" : "AdMob87000SDK1200iAds510",
  "groupIdRV" : "5814759",
  "abInternal" : "lpmm_abc_0331_disable",
  "tz" : "America\/Los_Angeles",
  "xCodeBuildVersion" : "16A242d",
  "mcc" : 65535,
  "advertisingIdType" : "IDFA",
  "abHashValue" : "0",
  "internalFreeMemory" : 29950,
  "att" : 3,
  "idfv" : "F010F50F-0AC9-48DC-9BE9-21B35021EE78",
  "controllerABValue" : "4",
  "deviceOEM" : "Apple",
  "sdkVersion" : "8.7.0",
  "advertisingId" : "D991D864-A78D-4600-B8E0-1B2D45D43BB5",
  "xCodeSdkBuild" : "22A3362",
  "internalTestId" : {
    "99" : "129"
  },
  "groupIdNT" : "5814765",
  "groupNameIS" : "All Countries",
  "groupNameRV" : "All Countries",
  "battery" : 12,
  "serverInitTimestamp" : "1743621446061",
  "gmtMinutesOffset" : -420,
  "mnc" : 65535,
  "isLimitAdTrackingEnabled" : "false",
  "abt" : "A",
  "sessionId" : "9E60256B-C2FD-410A-9E5F-A7FF85533F7E",
  "appKey" : "1fda3d735",
  "deviceOS" : "ios",
  "appVersion" : "7.66.0",
  "bannerSamplingMultiplier" : "10",
```

```
  },
  "groupIdNT" : "5814765",
  "groupNameIS" : "All Countries",
  "groupNameRV" : "All Countries",
  "battery" : 12,
  "serverInitTimestamp" : "1743621446061",
  "gmtMinutesOffset" : -420,
  "mnc" : 65535,
  "isLimitAdTrackingEnabled" : "false",
  "abt" : "A",
  "sessionId" : "9E60256B-C2FD-410A-9E5F-A7FF85533F7E",
  "appKey" : "1fda3d735",
  "deviceOS" : "ios",
  "appVersion" : "7.66.0",
  "bannerSamplingMultiplier" : "10",
  "groupIdNT" : "5814765",
  "groupNameIS" : "All Countries",
  "groupNameRV" : "All Countries",
  "battery" : 12,
  "serverInitTimestamp" : "1743621446061",
  "gmtMinutesOffset" : -420,
  "mnc" : 65535,
  "isLimitAdTrackingEnabled" : "false",
  "abt" : "A",
  "sessionId" : "9E60256B-C2FD-410A-9E5F-A7FF85533F7E",
  "appKey" : "1fda3d735",
  "deviceOS" : "ios",
  "appVersion" : "7.66.0",
  "bannerSamplingMultiplier" : "10",
  "bannerSampling" : "false",
  "timestamp" : 1743644258954,
  "deviceModel" : "iPhone10,2",
  "auid" : "FC84A751-39C2-4BFB-B4B0-3E8582B17142",
  "xCodeSdkName" : "iphoneos18.0",
  "firstSession" : "false",
  "language" : "EN-US",
  "groupIdBN" : "5814763",
  "is_coppa" : "false",
  "groupIdIS" : "5814761"
```

What does this data look like?

- Data going to advertisers look like above. There is also Ad Tracking Transparency.
- This happens for both free and paid situations.
- **POST request to appanalytics-services-att.com**
-



05

Results

What I found and what its impact is.

ATT Ping Contents (Duo Paid)


```
{
  "xCodeVersion" : "1600",
  "osVersion" : "16.7.10",
  "connectionType" : "wifi",
  "jb" : "false",
  "bundleId" : "com.duolingo.DuolingoMobile",
  "events" : [
    {
      "eventSessionId" : "9E60256B-C2FD-410A-9E5F-A7FF85533F7E",
      "firstSessionTimestamp" : 1738015217402,
      "timestamp" : 1743644258886,
      "eventId" : 44,
      "rawConnectionType" : "wifi",
      "duration" : 45223
    }
  ],
  "mobileCarrier" : "--",
  "groupNameBN" : "All Countries",
  "abGroup" : "NONE",
  "mt" : "AdMob87000SDK1200iAds510",
  "groupIdRV" : "5814759",
  "abInternal" : "lpmm_abc_0331_disable",
  "tz" : "America\/Los_Angeles",
  "xCodeBuildVersion" : "16A242d",
  "mcc" : 65535,
  "advertisingIdType" : "IDFA",
  "abHashValue" : "0",
  "internalFreeMemory" : 29950,
  "att" : 3,
  "idfv" : "F010F50F-0AC9-48DC-9BE9-21B35021EE78",
  "controllerABValue" : "4",
  "deviceOEM" : "Apple",
  "sdkVersion" : "8.7.0",
  "advertisingId" : "D991D864-A78D-4600-B8E0-1B2D45D43BB5",
  "xCodeSdkBuild" : "22A3362",
  "internalTestId" : {
    "99" : "129"
```

```
  ],
  "mobileCarrier" : "--",
  "groupNameBN" : "All Countries",
  "abGroup" : "NONE",
  "mt" : "AdMob87000SDK1200iAds510",
  "groupIdRV" : "5814759",
  "abInternal" : "lpmm_abc_0331_disable",
  "tz" : "America\/Los_Angeles",
  "xCodeBuildVersion" : "16A242d",
  "mcc" : 65535,
  "advertisingIdType" : "IDFA",
  "abHashValue" : "0",
  "internalFreeMemory" : 29950,
  "att" : 3,
  "idfv" : "F010F50F-0AC9-48DC-9BE9-21B35021EE78",
  "controllerABValue" : "4",
  "deviceOEM" : "Apple",
  "sdkVersion" : "8.7.0",
  "advertisingId" : "D991D864-A78D-4600-B8E0-1B2D45D43BB5",
  "xCodeSdkBuild" : "22A3362",
  "internalTestId" : {
    "99" : "129"
  },
  "groupIdNT" : "5814765",
  "groupNameIS" : "All Countries",
  "groupNameRV" : "All Countries",
  "battery" : 12,
  "serverInitTimestamp" : "1743621446061",
  "gmtMinutesOffset" : -420,
  "mnc" : 65535,
  "isLimitAdTrackingEnabled" : "false",
  "abt" : "A",
  "sessionId" : "9E60256B-C2FD-410A-9E5F-A7FF85533F7E",
  "appKey" : "1fda3d735",
  "deviceOS" : "ios",
  "appVersion" : "7.66.0",
  "bannerSamplingMultiplier" : "10",
```

```
},
"groupIdNT" : "5814765",
"groupNameIS" : "All Countries",
"groupNameRV" : "All Countries",
"battery" : 12,
"serverInitTimestamp" : "1743621446061",
"gmtMinutesOffset" : -420,
"mnc" : 65535,
"isLimitAdTrackingEnabled" : "false",
"abt" : "A",
"sessionId" : "9E60256B-C2FD-410A-9E5F-A7FF85533F7E",
"appKey" : "1fda3d735",
"deviceOS" : "ios",
"appVersion" : "7.66.0",
"bannerSamplingMultiplier" : "10",
"groupIdNT" : "5814765",
"groupNameIS" : "All Countries",
"groupNameRV" : "All Countries",
"battery" : 12,
"serverInitTimestamp" : "1743644258954",
"deviceModel" : "iPhone10,2",
"auid" : "FC84A751-39C2-4BFB-B4B0-3E8582B17142",
"xCodeSdkName" : "iphoneos18.0",
"firstSession" : "false",
"language" : "EN-US",
"groupIdBN" : "5814763",
"is_coppa" : "false",
"groupIdIS" : "5814761"
```

ATT Ping Contents (Busuu Paid)

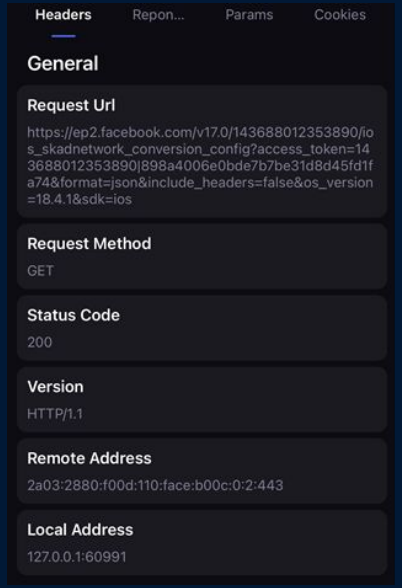

```
"deviceModel" : "iPhone10,2",
"bannerSamplingMultiplier" : "10",
"jb" : "false",
"firstSession" : "false",
"mcc" : 65535,
"groupIdBN" : "3801541",
"bannerSampling" : "false",
"att" : 3,
"bundleId" : "com.busuu.english.app",
"appVersion" : "28.31.0",
"appKey" : "1af944565",
"xCodeSdkName" : "iphoneos18.1",
"deviceOEM" : "Apple",
"is_coppa" : "false",
"abGroup" : "NONE",
"internalTestId" : {

},
"abHashValue" : "0",
"connectionType" : "wifi",
"gmtMinutesOffset" : -420,
"advertisingId" : "D991D864-A78D-4600-B8E0-1B2D45D43BB5",
"mnc" : 65535,
"icc" : "--",
"osVersion" : "16.7.10",
"controllerABValue" : "13",
"deviceOS" : "ios",
"mobileCarrier" : "--",
"advertisingIdType" : "IDFA",
"battery" : 30,
"mt" : "MAX",
"isLimitAdTrackingEnabled" : "false",
"idfv" : "E8BAEEC1-41AC-463D-A603-768C4D66ED12",
"groupIdRV" : "3801537",
"serverInitTimestamp" : "1743458042356",
```

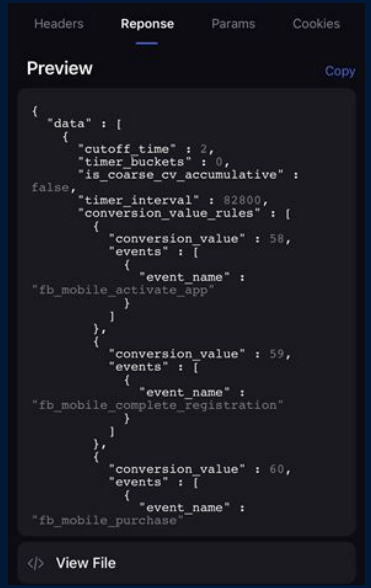
```
"xCodeVersion" : "1610",
"tz" : "America\/Los_Angeles",
"groupIdIS" : "3801539",
"internalFreeMemory" : 30447,
"groupIdNT" : "5124423",
"xCodeSdkBuild" : "22B74",
"groupNameRV" : "All Countries",
"sdkVersion" : "8.1.0",
"xCodeBuildVersion" : "16B40",
"groupNameNT" : "All Countries",
"abt" : "A",
"events" : [
  {
    "eventSessionId" : "7562F58F-209B-4C77-924C-7E79569246D6",
    "timestamp" : 1743458180909,
    "eventId" : 44,
    "firstSessionTimestamp" : 1738972311000,
    "duration" : 70622
  }
],
"sessionId" : "7562F58F-209B-4C77-924C-7E79569246D6",
"groupNameIS" : "All Countries",
"language" : "EN-US",
"auid" : "1D47F676-AF79-4AA4-9647-25D5C5D7B22E",
"groupNameBN" : "All Countries",
"timestamp" : 1743458180981
```

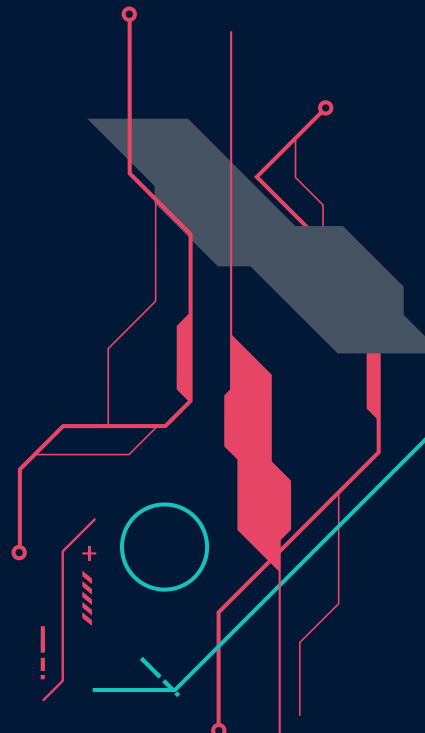
```
{
"d": {
"ac": "4293181",
"ap": "538681648",
"id": "8f34849cea3ed670",
"ti": 1743628952225,
"tk": "4257740",
"tr": "6bd9b11916e81ac97d871adb0053789c",
"ty": "Mobile"
},
"v": [
0,
2
```

ATT Ping Contents (Memrise Paid)

- Busuu did not have any ads mediation or ATT data transferred in the test period.
- Free and Paid versions were the same.
- ep2.facebook.com (facebook ads service) was present.







Impact

- Mobile apps with subscriptions say you are not getting ads when you’re using them on paid mode.
- Categorically true, but they still can sell your data.
- Advertisers can also get a piece if their APIs are still included.
- Even if the data leakage is small, stay away from companies that leak your data when they say they’re not.

Future Work

- See how Android versions of the apps operate
- Do further research into the ad formats on Unity3D mediation
- Find a way to get a copy of what is being POSTed to the ATT servers at the end of lessons
- And much more…

Summary

- Language learning apps are new ways to learn languages.
- Online advertising has many moving parts
- Network sniffers with MITM attacks are useful for reading most web traffic
  - Storm Sniffer is a cheap all-in-one network sniffer.
- Duolingo and Busuu send user info to Unity3D mediation in their paid versions when they’re not supposed to.

THANKS!

Do you have any questions?
plascencia.matt.31@gmail.com
https://substack.com/@matthewplascencia
https://www.youtube.com/@forensicswithmatt

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Fonts & colors used

This presentation has been made using the following fonts:

Orbitron
(https://fonts.google.com/specimen/Orbitron)

Roboto Slab
(https://fonts.google.com/specimen/Roboto)

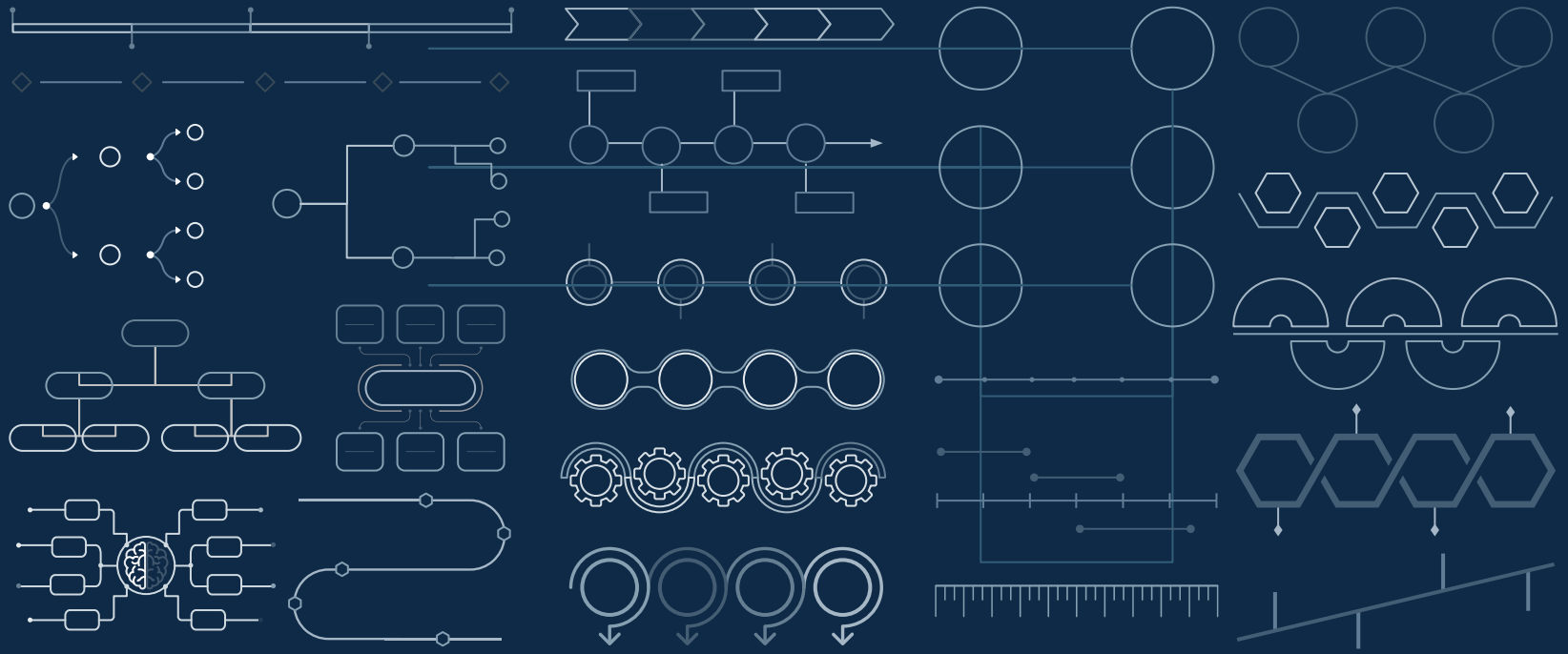
Use our editable graphic resources...

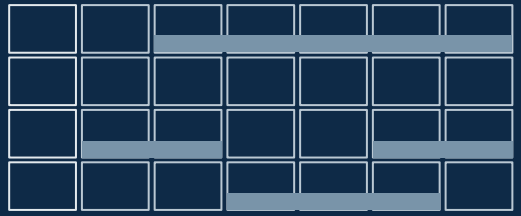
You can easily **resize** these resources without losing quality. To **change the color**, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more **infographics** on Slidesgo.

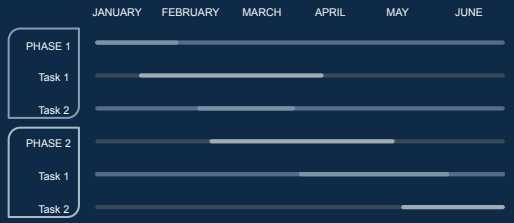
ICON PACK

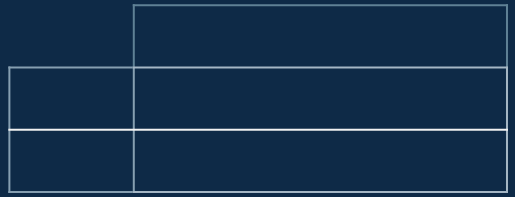


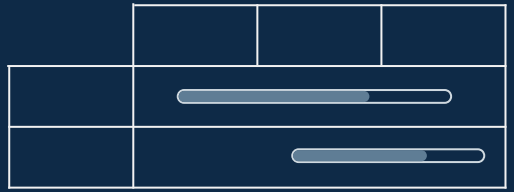


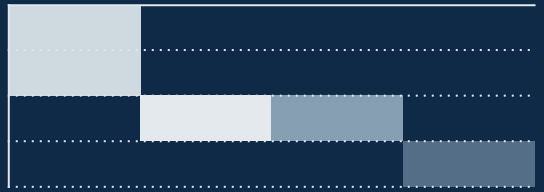



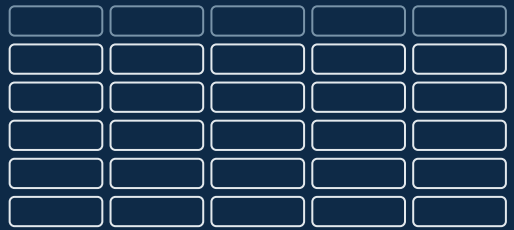


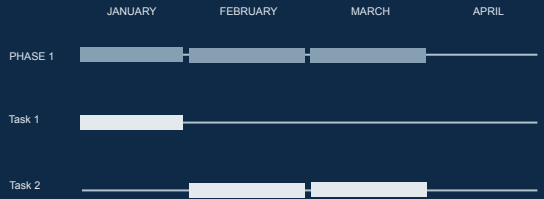

JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
PHASE 1
Task 1
Task 2
PHASE 2
Task 1
Task 2

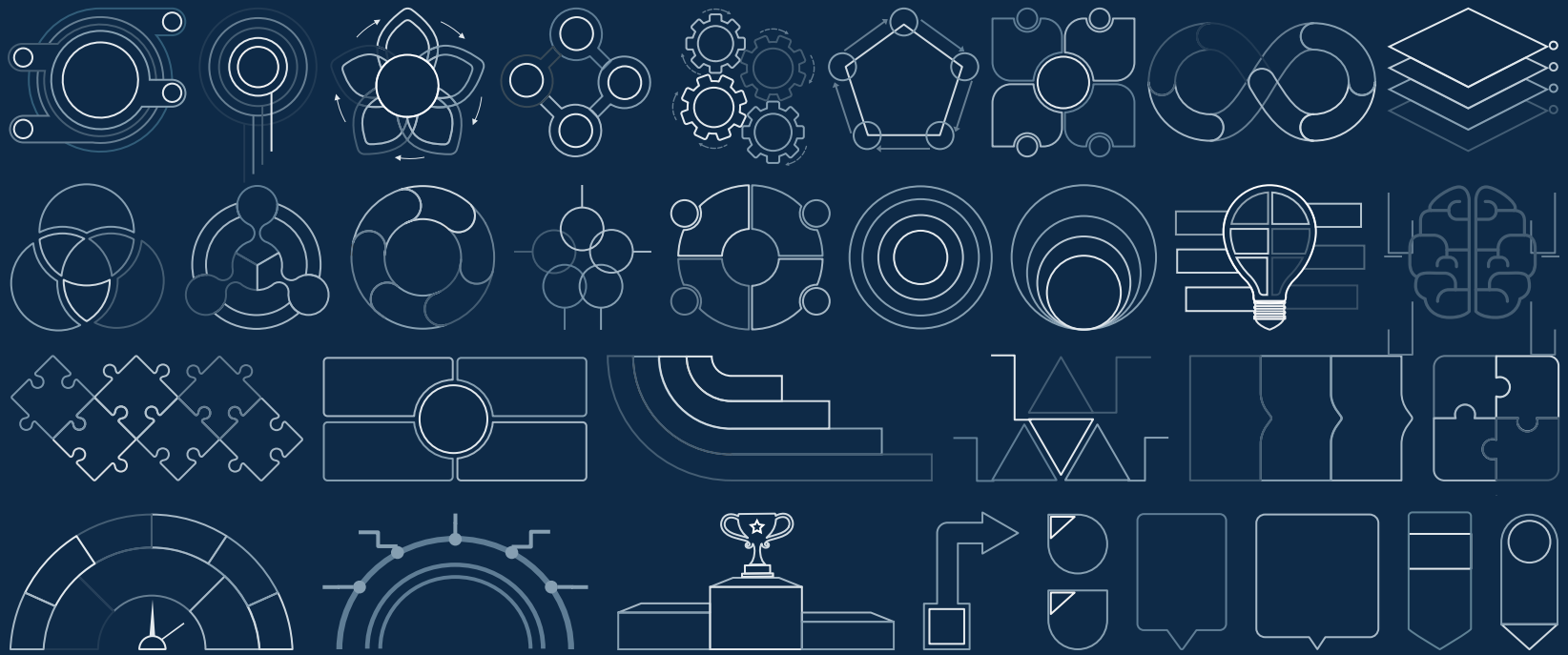


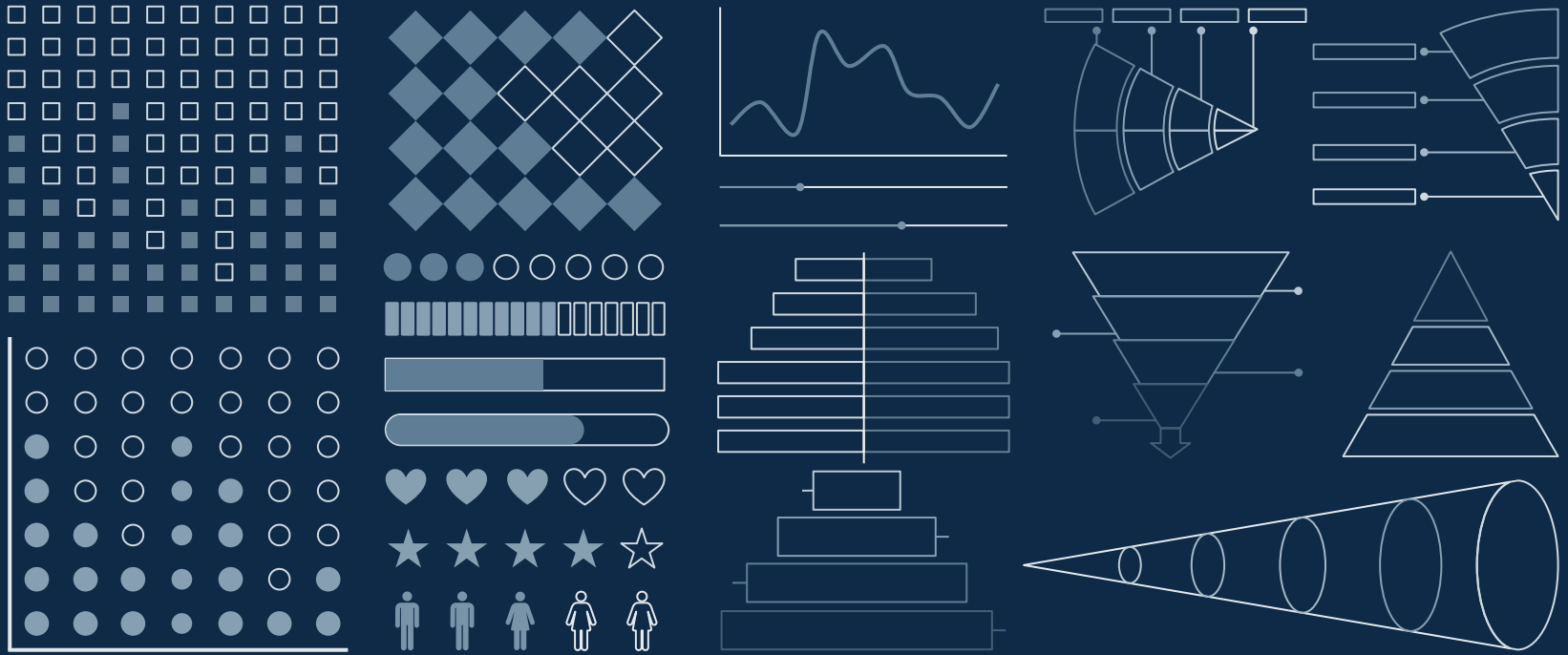







JANUARY
FEBRUARY
MARCH
APRIL
PHASE 1
Task 1
Task 2





...and our sets of editable icons

You can **resize** these icons without losing quality.
You can **change the stroke and fill color**; just select the icon and click on the **paint bucket/pen**.
In Google Slides, you can also use **Flaticon's extension**, allowing you to customize and add even more icons.

Educational Icons

Medical Icons

Business Icons

Teamwork Icons

Help & Support Icons

Avatar Icons

Creative Process Icons

Performing Arts Icons

Nature Icons

SEO & Marketing Icons