

Data Leaks, Data Industry, Social Engineering & OSINT

RZ Pan @ ZJU AAA

Disclaimer

This presentation is not sponsored or endorsed by any real or hypothetical commercial corporation located in Hangzhou, China. The views and opinions expressed are those of the author and do not necessarily reflect that those of any such corporation, should one exist. Only trained professionals that are authorized by local authorities should attempt to review and reproduce the steps or methods provided by the presentation, and the author as well as the Zhejiang University Security Team will not be responsible to any digital, material, physical, or emotional damages caused by such attempts.

Current Situation

<https://www.varonis.com> › ... › Data Security ▾ 翻译此页

98 Must-Know Data Breach Statistics for 2021 | Varonis

2021年4月16日 — How Do **Data Breaches** Occur? **Recent Data Breaches + Statistics; COVID-19 Specific Data Breaches;** Breaches by the Numbers; Cost of a Data ...
How Do Data Breaches Occur? · Data Breach Risk · Historical Data Breach...

<https://www.securitymagazine.com> › articles ▾ 翻译此页

The Top 10 Data Breaches of 2020 | 2020-12-03 | Security ...

2020年12月3日 — **Recent data** from Risk Based Security revealed that the number of records exposed has increased to a staggering 36 billion in 2020.

<https://www.upguard.com> › blog › biggest-dat... ▾ 翻译此页

The 56 Biggest Data Breaches (Updated for 2021) | UpGuard

Our updated list for 2021 ranks the 56 biggest **data breaches** of all time, ... list of the 56 biggest **data breaches** in history, including **recent data breaches** in 2021. ... was the result of a **data leak** on a system run by a **state-owned utility company**.

Aadhaar data breach · Starwood (Marriott) data... · Sociallarks data breach

Entity	Year	Records
Microsoft Exchange servers	2021	unknown
Animal Jam	2020	46,000,000
Betsson Group	2020	unknown
Capcom	2020	350,000
CheckPeople	2020	56,000,000
Clearview AI	2020	unknown (client list)
FireEye	2020	Unknown
Unknown	2020	201,000,000
Instagram	2020	200,000,000

Current Situation

- Does that necessarily affects my own life?
- Yes, of course!

Cu

June 30, 2022, 08:55 AM (This post was last modified: Yesterday, 10:25 AM by Muffin. *Edit Reason: Locked by staff due to all the spam*)

In 2022, the Shanghai National Police (SHGA) database was leaked. This database contains many TB of data and information on Billions of Chinese citizens.

Sell: Shanghai GOV (SHGA.gov.cn) National Police Database

Host: <http://oss-cn-shanghai-shga-d01-a.ops.ga.sh/>

Data leaked from these tables:

----TABLES----

```
person_address_label_info_slave QFpD25bKTJ2eQBxcbe2Aaw 90 0 546148916 0 172.2gb 172.2gb
nb_theme_address_merge_tracks_slave -bUMVB1uRRusUbbqZepEpA 300 0 37483779369 4 22.4tb 22.4tb
nb_theme_address_case_dwd_test 7COIWTt7QU-YPwWub8z_SQ 150 0 22375506 1749307 25.2gb 25.2gb
nb_theme_address_company_dwd-total fpnmEYB9SI6WevHnZIEwIA 150 0 1842856 0 2.8gb 2.8gb
nb_theme_address_case_dwd-total 7X8oNqULQnWFLpzHDaUTbg 150 0 1214119253 0 1tb 1tb
nb_theme_address_company_dwd_test g5f614LGQcGL3oQ6ON2Bbw 150 0 2017931 0 4.3gb 4.3gb
person_address_label_info_master t64pp9WnS3maY9jBjzTtiw 90 0 969830088 0 282.8gb 282.8gb
```

Data Details:

Databases contain information on 1 Billion Chinese national residents and several billion case records, including:

- Name
- Address
- Birthplace
- National ID Number
- Mobile number
- All Crime / Case details

Yesterday, 08:22 AM

UNDERTALEFANS Wrote:

b站请关注嘉然今天吃什么

加载中~
show show way

1.7亿条学生信息遭泄露?学习通报警!曾一年被披露三次漏洞



2022年6月23日 随后，“[学习通](#)”话题一度登上微博热搜第一。当日下午，超星[学习通](#)官微就此事回应表示，尚未发现明确的用户[信息泄露](#)证据，已经向公安机关报案，公安机关已介入调查。但自6月21日以...

北京时间财经 百度快照

网易邮箱又被脱裤啦,大家注意修改密码 - 糯米PHP

2017年6月26日 全球工单系统 [网易邮箱又被脱裤啦,大家注意修改密码](#)这两天一直给我发邮件提示登录失败,上去看了下 IP 发现各地都有。图中香港和第一个四川是我自己登录的,其他...

[糯米PHP](#) 百度快照

网易数据库被脱裤,应该怎么保护好自己的邮箱账号? - 知乎

2015年10月20日 3, 考虑到[脱裤](#)问题, 以及国内企业把密码仅仅md5存储的尿性(那些明文存储请呼叫赵日天, 龙傲天, 福尔康, ...)

知乎 百度快照

A Typical Spam Call

- AI generated audios (TTS)
- Robot-controlled virtual phone
- To mitigate:
- Make use of Google assistant
- Robots can chat with robots
- No living creatures involved
- ...But this can't solve the issue completely

(925) 969-3752



4月 23 日上午 3:50

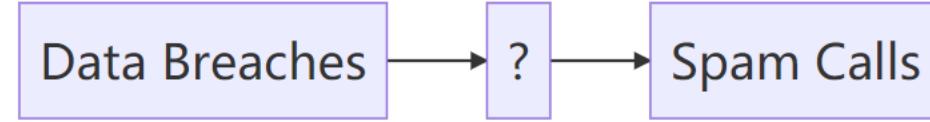
Warranty we sent you several notices in the mail that you have yet to extend your warranty pass the factory cut off, and this is a courtesy call to renew your warranty before we close the file. If you are interested in renewing your auto warranty now, please press 5 now or press 9 to be removed from our list. Hi, this is Suzie calling with the vehicle service department. We are calling about your vehicles manufacturer's warranty. We sent you several notices in the mail that you have yet to extend your warranty pass the factory shut off, and this is a courtesy call to renew your warranty before we close the file. If you are interested in renewing your auto warranty now, please press 5 now or press 9 to be removed from a list. Hi, this is Suzie calling with the vehicle service department. We are calling about your vehicles manufacturer's warranty. We



00:00

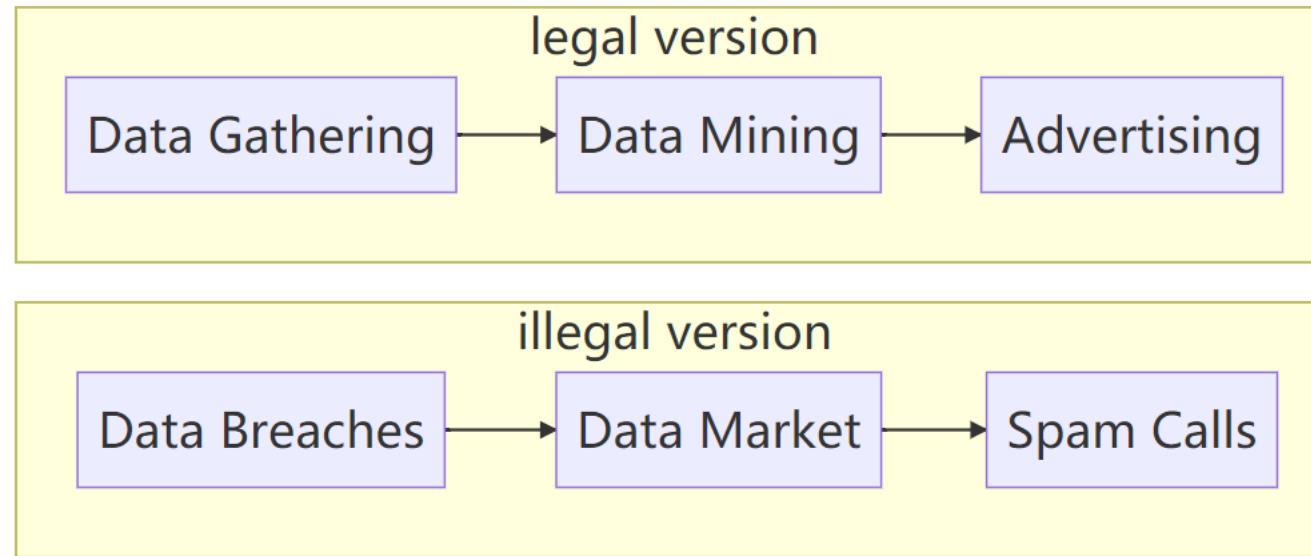
01:36

Model



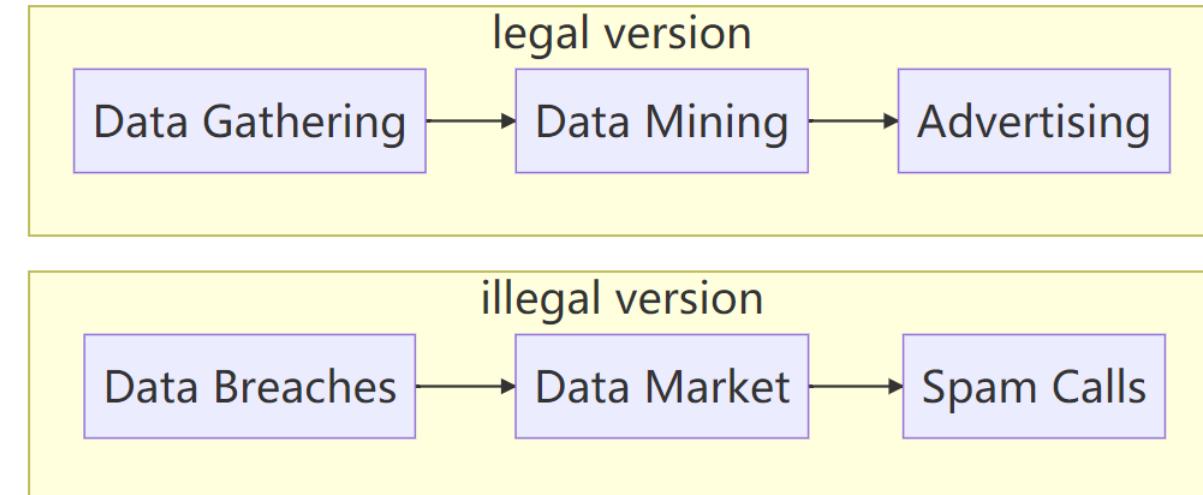
- How did those spammers get my phone numbers?
- Where did those spammers buy leaked data?
- These questions lead to the concept “Data Industry”

Data Industry: Structure



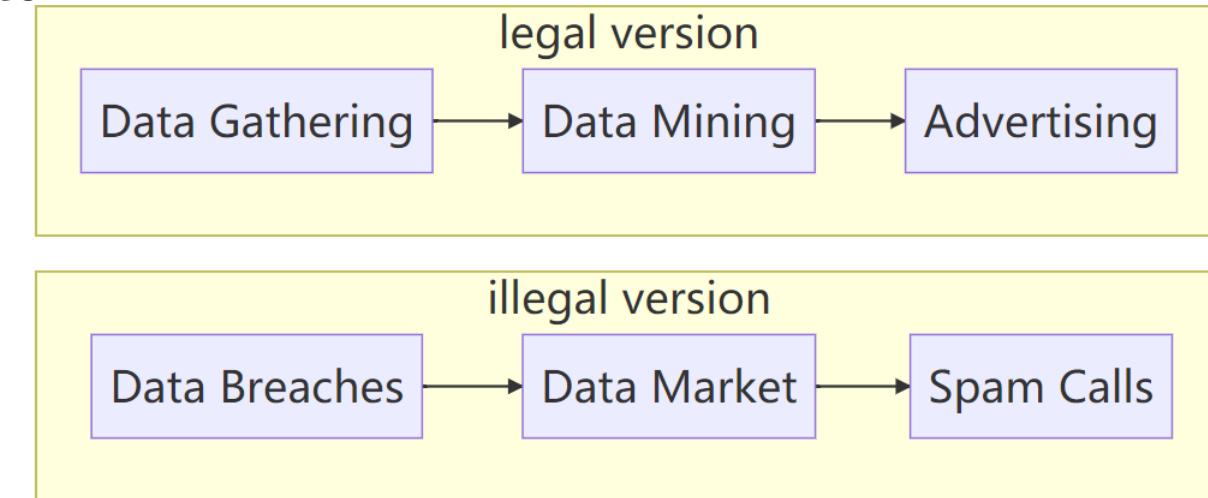
Data Industry: Structure

- Data Gathering:
- Big enterprises
- Gathered from their own services
 - Navigation data
 - Preference data
- User-targeting or anonymized



Data Industry: Structure

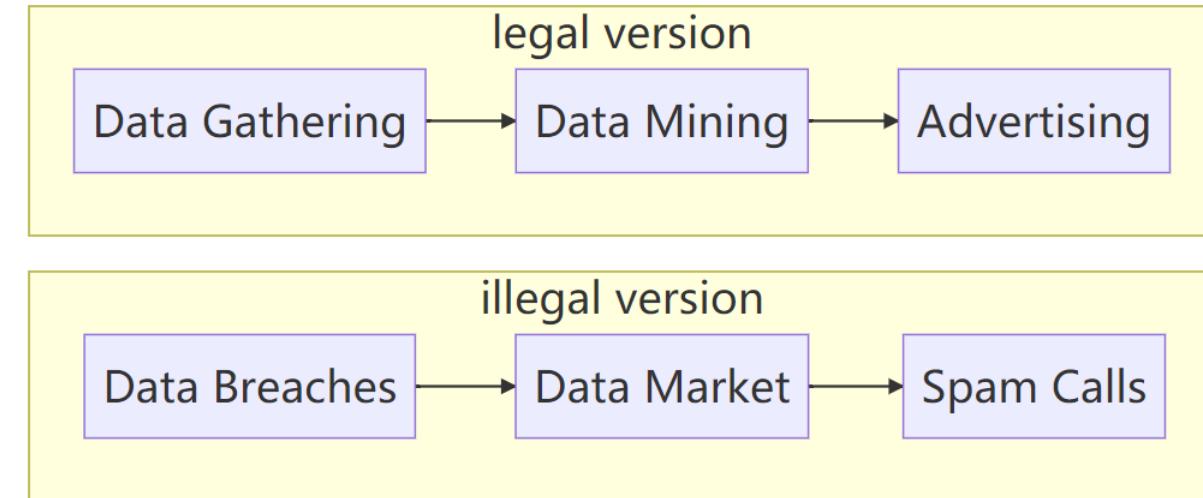
- Data Mining & Advertising:
- Discover and gain profits from data
- Classify your users
- Better timeline waterfall
- Targeted ads
 - Youtube
 - Tiktok



Data Industry: Illegal Parts

- Data Market:
- Place to sell & buy leaked data

- Deep / Dark Web?
- Onion Browser (Tor Browser)?
- Encrypted IM Apps: Telegram, Signal (smaller market)



A Relatively Small Market on TG

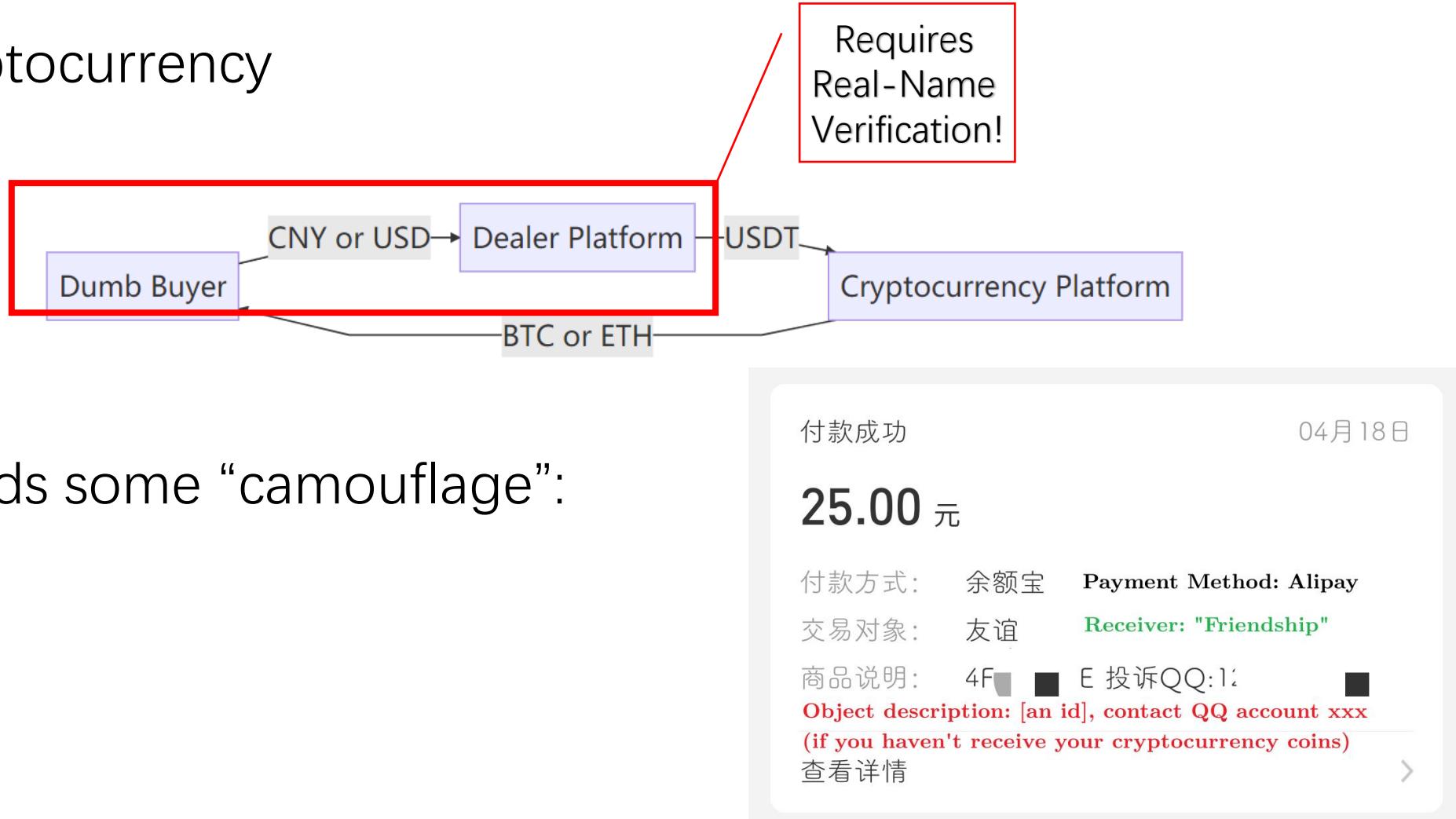


- Operates quite the same as those dark web sites
- Offers example data (~1% of leaked) for buyers to check before payment
- Only accept cryptocurrency



Typical Method of Avoiding Regulation

- Cryptocurrency



- Now that we know where & how to buy leaked data (at least theoretically), let's see what we can do with it

Social Engineering

- Simply pick a random poor guy from a random university
- Let's call this guy "Prof. Yang" from The Random University.
- *Caution: This is actually an **authorized** experiment carried out by a professional pentest team. Trying to reproduce this might lead to violation of relative laws, and you may be sued due to such attempts.*

Social Engineering

- First, gather as much information about him as we can (via search engines):

The screenshot shows a Baidu Wenku search interface with the query "杭州 [REDACTED] 杨 [REDACTED]" entered. Below the search bar, a blue button reads "搜索文档". A "为您推荐:" section displays several blurred document thumbnails.

On the left, a form displays the following information:

姓 名:	杨 [REDACTED]	Name: Yang **
性 别:	男	Gender: Male
出生年月:	1965.	Birth: 1965.*
所在学院(部门):	[REDACTED]	Office: *
所在专业:	[REDACTED]	Major: *
最后学历:	[REDACTED]	
最后学位:	[REDACTED]	
攻读专业:	[REDACTED]	
职 称:	[REDACTED]	Title: *
导师类别:	[REDACTED]	Teacher Type: *

On the right, a document titled "W 杭州 [REDACTED] 导师信息表 - 百度文库 VIP" is shown. It has a rating of "质量 4.5 分". The document content is partially visible, mentioning "高等教育出版社" and "全日制研究生导师信息表". Below the document, it says "2013-03-14 | 共9页 | 100次下载 | 贡献者: [REDACTED] 究生 ...".

Below the document, the text "E-mail: [REDACTED]@163.com" is displayed. At the bottom, a link to a Google cache page is shown: "webcache.googleusercontent.com/*: 杨** 20[REDACTED] 男 汉 [REDACTED] teacher id birthday citizen id".

Social Engineering

- First, gather as much information about him as we can (via university websites):
- Usually we can only get masked info at this step, i.e.
138****0001

重置密码 Reset Password

特别提醒

1、请使用自己的账号找回，盗取别人账号将追究责任。
2、请注意保护个人密码，以免造成不必要的麻烦。

1.输入用户名 2.选择验证方式 3.提交安全码 4.修改新密码 5.重置结果

i 请输入您需要重置登录密码的用户名

Username 用户名: "Teacher id" is default username

Captcha 验证码: 

You are resetting passwords for user Prof. Yang (his id)

i 你正在为账户 杨 200 重置登录密码，请选择重置方式

手机号码 (135****) By phone sms (no asterisk mask at all!)

邮箱地址 (****@163.com) By email (also no mask)

[下一步](#)

- Phone number:
[call service provider][register location number][4 random numbers]
- Citizen id:
[born city id][province id][birthday][3 random numbers][validation number]

Social Engineering

- Next, try to gain access to his email account *****@163.com
- Use “manual recover”
- Photoshop a fake citizen id card (this is illegal)
- Personal photo leaked from a school card management website
- Name, Birth, Gender leaked from search engine
- Citizen id and home address leaked from TG Market (**Definitely illegal!!! Authorized**)
- Netease Inc. (operates mail.163.com) will manually review this request

Social Engineering

- Netease failed to notice that the id card image was generated
- They decided to reset the email's password for us
- At the same time: We also phone called the Random University's IT department to reset Prof. Yang's another email (school email)'s password (of course this is illegal)
- The IT department was also cheated easily
- Now the team have access to both emails, which contain many important data & forms, and we can reset his passwords on nearly every websites as we controlled his accounts' recovery email
- Result: the team can login to more platforms for further steps (i.e. pentest & attack), the Random University is more vulnerable now

Social Engineering

- Near-source pentesting

我们出发时本来想先挂一个号，然后在大厅里慢慢搞，但计划赶不上变化，到达医院进门后，在大厅里一个人没有，什么缴费，预约啥的都没有人，没人不就更好办了，先在大厅和二楼转了一圈，没发现可以直接插网口的地方，回到一楼的后，发现门口左右两边有两台报告打印机都是连接着网线，一台没开机，一台已开机了，心想这不简单了嘛，网线插入没开机的那台网口就好了；

前进的道路总不会那么平坦，带的网线不够长，苍天绕过谁，没办法，出去买了个两米的网线，这下总该没没问题了吧，并不是想象中那么简单，md，自动获取 IP 还是没网，心里想完了；

不甘心啊啊啊啊啊，跟另外一个同事商量，纠结了一阵子，把那台好的报告打印机给拔了，并且找好了一个理由，有人问我们的话，就说是调整网络的，因为左边那台坏机子网络不通，说整就整，看了下这台打印机的IP，拔掉网线，连入自己的电脑，配置完成后，成功接入到网络中，愉快的开扫

Social Engineering

- Near-source pentesting
 - Fake id card
 - Duplicated RFID / NFC tag
 - As food deliverer, driver, tube repairer, etc.
 - Use photograph to cheat face recognition

Conclusion

- If one has enough information about you, they can easily become you (at least on the Internet). For others, it's hard to distinguish, and this may lead to severe money loss (i.e. your (fake) friends want to borrow money from you).
- Of course, this is illegal; but under the current law framework, seldom can be done to prevent these things from happening, so at least you should know where the threat is coming from.
- Or, maybe you can find some better solutions for this issue!

Self Protection

- Random passwords
- Use a password manager
 - Chrome / Edge
 - Bitwarden, 1Password, etc. (Recommended)



Self Protection

- Check for leaked passwords

✉ Google <no-reply@accounts.google.com>

重要安全提醒

您保存的密码中有一些已在网上外泄

系统检测到，由于您使用的某个网站或应用发生数据泄露事件，您的某些已存密码遭到外泄。不过，您的 Google 帐号未受影响。

为了保障您的帐号安全，Google 密码管理工具建议您立即更改密码。

[检查密码](#)

您也可以访问以下网址查看安全性活动：

<https://myaccount.google.com/notifications>

我们向您发送这封电子邮件，目的是让您了解关于您的 Google 帐号和服务的重大变化。

© 2022 Google LLC, 1600 Amphitheatre Parkway, Mountain View,
CA 94043, USA

21:56

密码安全检查

Self F

- Check



已检查 151 个网站或应用的密码

Google <*no-r*
重要安全提醒

您保存的密码中有
系统检测到，由于
些已存密码遭到外
为了保障您的帐号
检查密码

您也可以访问以下
<https://myaccount.google.com>
我们向您发送这封
务的重大变化。

© 2022 Google LLC
CA 94043, USA

- ! 1 个密码遭外泄
立即更改这些密码
- ! 有 91 个重复使用的密码
创建独一无二的专用密码
- ! 有 75 个帐号使用了安全系数低的密码
创建安全系数高的密码



查看、更改或移除您保存在 Google 帐号中的密
码。转到密码管理器

Self Protection

立即更改这些密码

您保存的密码中有一些已在非 Google 数据泄露事件中外泄。您应立即更改这些密码。[了解详情](#)

以下帐号有安全风险



Spotify: 音乐和播客

[REDACTED]



[更改密码](#)



Self Protection

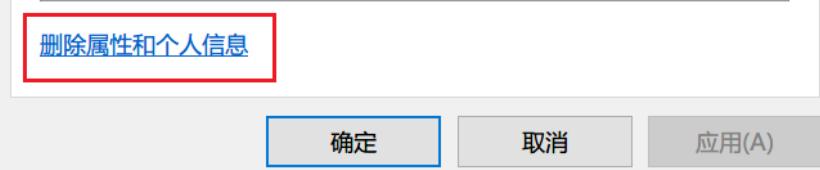
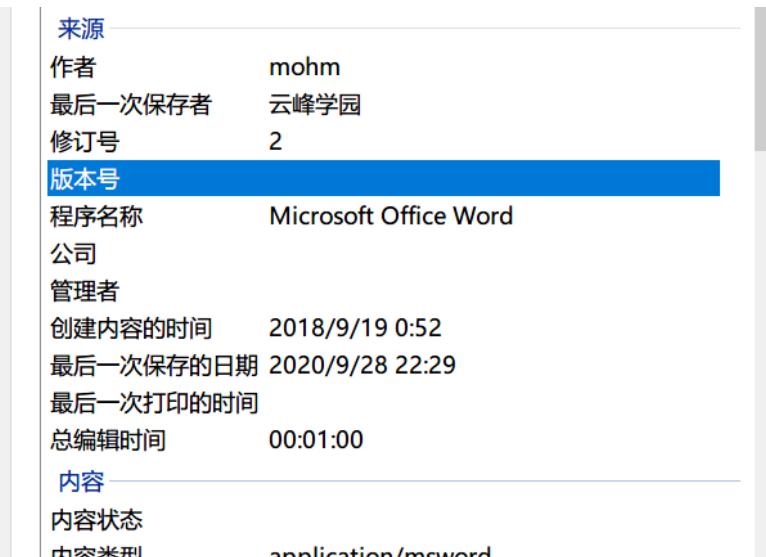
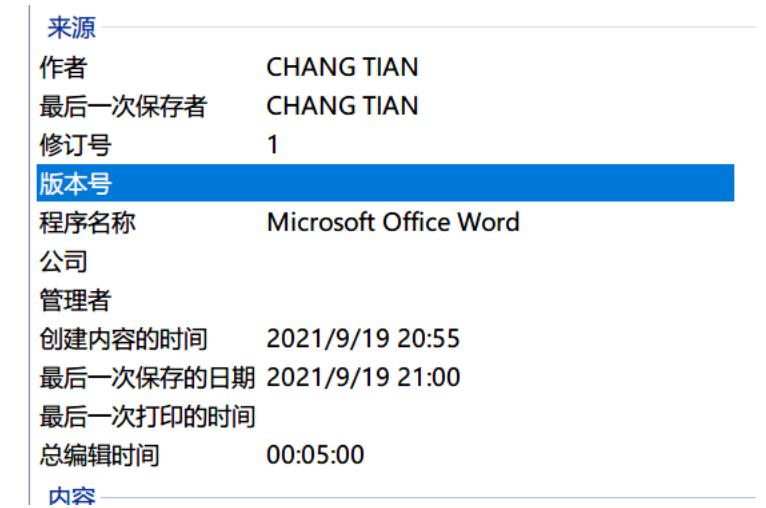
- Avoid accidentally leaking personal related info on the Internet

A Guide to Eliminate Your Personal Info from Your Report

RZ Pan @ ZJU AAA

MS Word/PPT/Excel

- “Author” attribute
- “Last Modified By” attribute
- Right-click on the file, “Attribute (R)”, Click on the “Details” Tab
- Solution: Click “Remove Attribute and Personal Info” which can be found at the bottom of the “Details” Tab
- Don't make further edit, or you'll need to clear these attributes again!



PDF

- “Author” attribute
- Open the file, and check its attributes/properties in your PDF viewer
- PDFs generated by MS Word:
- Solution: Click “Remove Attribute and Personal Info” before generating PDF
- Solution: Use a PDF metadata editor to delete these metadata
- Solution: Use Typora or LaTeX

标题: 浙江大学实验报告
作者: liulanlan
创建时间: 2022/3/20 23:24:29
修改时间: 2022/3/20 23:24:29
应用程序: Microsoft® Word 2019
PDF 制作程序: Microsoft® Word 2019
PDF 版本: 1.7

Image

- JPEG EXIF –Solution: Ditto
 - Shot on a phone: GPS Info, phone manufacturer, etc.
- Screenshots might contain sensitive information, such as the path of your .exe (shown in cmd title bar)

```
$ strings 1.jpg | head -n 11
Adobe
'JExif
NIKON CORPORATION
NIKON D7200
Adobe Photoshop CC (Windows)
2019:07:10 23:07:41
0221
2019:06:27 17:25:44
2019:06:27 17:25:44
8120437
105.0 mm f/2.8
```

来源

作者

拍摄日期 2019/6/27 17:25

程序名称 Adobe Photoshop CC (Windows)

获取日期

版权

图像

照相机

照相机制造商 NIKON CORPORATION

照相机型号 NIKON D7200

光圈值 f/3

曝光时间 1/160 秒

ISO 速度 ISO-320

曝光补偿 -0.3 步骤

焦距 105 毫米

最大光圈 3.1

测光模式 图案

目标距离

闪光灯模式 无闪光, 强制

闪光灯能量

35mm 焦距 157

Visual Studio

- .sln, .vcxproj, -filters, -user
 - (Possible) IncludePath attribute
- .vs folder
 - (Possible) slnx.sqlite database, can be opened with sqlite3
- **Debug folder**
 - Name.log
 - **Name.pdb**
- Solution: Delete Debug/, double-check .sln & .vs/, etc.

```
<PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Debug|Win32' ">
    <LinkIncremental>true</LinkIncremental>
    <IncludePath>$(VC_IncludePath);$(WindowsSDK_IncludePath);.\include\</IncludePath>
</PropertyGroup>
```

FileSystemEntityId	Name	ParentFileSystemEntityId	IsFile	LastObserved
过滤	过滤	过滤	过滤	过滤
1	1 .vs		NULL 0	637542385807238237
2	2 drawutil.c		NULL 1	637542385807238237
3	3 include		NULL 0	637542385807238237
4	4 libgraphics		NULL 0	637542385807238237
5	5 main.c		NULL 1	637542385807238237
6	6 exception.h		3 1	637542385807238237
7	7 extgraph.h		3 1	637542385807238237
8	8 galloc.h		3 1	637542385807238237

exp07.log - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

main.c

```
d:\work\code\c\graphic\exp07-2\main.c(101): warning C4090: “函数”: 不同的“const”限定符  
d:\work\code\c\graphic\exp07-2\main.c(117): warning C4090: “函数”: 不同的“const”限定符  
d:\work\code\c\graphic\exp07-2\main.c(143): warning C4090: “函数”: 不同的“const”限定符  
d:\work\code\c\graphic\exp07-2\main.c(153): warning C4090: “函数”: 不同的“const”限定符  
exp07.vcxproj -> D:\Work\Code\C\graphic\exp07-2\Debug\exp07.exe
```

17	17 graphics.c	4 1	637542385807238237
18	18 imgui.c	4 1	637542385807238237
19	19 random.c	4 1	637542385807238237
20	20 simpio.c	4 1	637542385807238237
21	21 strlib.c	4 1	637542385807238237
22	22 exp07.vcxproj	NULL 1	637542386318574471

PDB: cvdump.exe

```
PS D:\Work\Code\C\graphic\exp07-2\Debug> .\cvdump.exe .\exp07.pdb | grep Module: | grep -v from
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\main.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\strlib.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\simpio.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\random.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\imgui.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\graphics.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\genlib.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\exception.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\drawutil.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\main.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\strlib.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\simpio.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\random.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\imgui.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\graphics.obj"
** Module: "D:\Work\Code\C\graphic\exp07-2\Debug\genlib.obj"
```

Solution: don't submit a debug build, and don't forget to delete Debug/

Dev C++

- .dev file is just a simple plain-text file

```
[Project]
FileName=expSort.dev
Name=expSort
Type=0
Ver=2
ObjFiles=
Includes="D:\Work\Code\C\graphic\exp08 - 副本\include"
```

- Solution: Double-check .dev file, make sure no absolute path is included

Visual Studio Code

- .vscode folder: *.json, *.log

```
"miDebuggerPath": "C:\\Program Files\\mingw-w64\\x86_64-8.1.0-posix-seh-rt_v6-rev0\\mingw64\\bin\\gdb.exe",
"args": [
    "${workspaceFolder}\\\\task2\\\\libgraphics\\\\exception.c",
    "${workspaceFolder}\\\\task2\\\\libgraphics\\\\genlib.c",
```

- Might contains sensitive info
- Solution: Delete .vscode/ before submitting

.DS_Store (macOS)

```
$ ds_store_exp.py http://hd.zj.qq.com/themes/galaxyw/.DS_Store  
hd.zj.qq.com/
```

```
└── themes  
    └── galaxyw  
        ├── app  
        │   └── css  
        │       └── style.min.css  
        ├── cityData.min.js  
        └── images  
            └── img  
                ├── bg-hd.png  
                ├── bg-item-activity.png  
                └── bg-masker-pop.png
```

- Solution: Delete .DS_Store file before compressing/uploading your file

.md

- If you use specific platforms as your image bed, your personal info might be leaked
 - Retrieve weibo uid (--> user homepage) from image link
 - <https://www.cnblogs.com/idjl/p/9610593.html>
- Or, you use your self-hosted image bed
 - image.me.com --> me.com, blog.me.com,
 - site:me.com inurl:me.com
- Also, .md might leak path info if you import any image by specifying its absolute path
 - Example: ! [Image] (C:\Users\username\Desktop\MyImage\1.jpg)
- Solution: Submit .pdf, not .md

Git

- .git folder
 - upstream url: can get your GitHub/Gitlab account
 - Username & email if you use `git config --local`
 - Also, `git log`
- Solution: Delete .git

Conclusion

- Submit less files
- Use pdf if possible
 - Be careful about metadata in PDFs generated by MS Word
- If you have to submit Word/PPT/Excel, see page 3 for steps to create an anonymous copy

Self Protection

- Sharing image:
- Use proper mask / mosaic, don't send raw copy, and crop out as much as you can

Self Protection

yml 配置 (保密原因自我补全)

```
1 # datahub配置文件
2 public:
3   datahub:
4     # 配置信息
5     config:
6       #控制器
7       startup: false
8       #应用服务器地址
9       endpoint: XXXXXXXXXXXXXXXXXXXXXXXX
10      #用户权限accessId
11      accessId: XXXXXXXXXXXXXXXX
12      #用户权限accessKey
13      accessKey: XXXXXXXXXXXXXXXX
14      servivce:
15        #应用名称
16        - projectName: XXXXXXXX
17        #应用消费—topic名称
18        topicGet: XXXXXXXXXXXXXXXX
19        #应用生产—topic名称
20        topicSet: XXXXXXXXXXXXXXXX
21        #topicid的订阅ID
22        subId: XXXXXXXXXXXXXXXX
```

send raw copy, and crop out as

```
124  /*public static void main(String[] args) {
125    // Endpoint\Region: 华东1为例, 其他Region请按实际情况填写
126    String endpoint = "https://datahub.cn-shanghai-shga-d01.dh.aliyun.ga.sh";
127    String accessId = "0iWV0NCs805VuAAu";
128    String accessKey = "iEwLgpCnXDwT93YMVDb2G60my9ne81";
129    String projectName = "sjc_rwzx";
130    String topicName = "task_center_platform_request";
131    RecordSchema schema = new RecordSchema();
132 }
```

Self Protection

- Sharing image:
- Use proper mask / mosaic, don't send raw copy, and crop out as much as you can
- DingTalk screenshots: phone + name watermark
 - Double-check before sending
- Raw copy contains metadata
 - Recall previous slides
 - Will be eliminated in compressed copies
- Cropping can save your life

照相机

照相机制造商	NIKON CORPORATION
照相机型号	NIKON D7200
光圈值	f/3
曝光时间	1/160 秒
ISO 速度	ISO-320
曝光补偿	-0.3 步骤
焦距	105 毫米
最大光圈	3.1
测光模式	图案
目标距离	
闪光灯模式	无闪光, 强制
闪光灯能量	

OSINT

RZ Pan @ ZJU AAA

What is OSINT

- Open Source INTelligence
- Use publicly accessible data to analyze information
 - Fact check
 - Trace source
 - Save lives

Using OSINT

问下有没有认识2013年入学的[REDACTED]学长/学姐的同学，可能要自杀，想看看能不能拦一下

8163

收藏

显示所有图片

CC98十大热门



现在尝试了很多办法暂时联系不上她

你要不直接报警看看?

好的，我尝试一下

竺院院长已经在杭州报了警了

等等看看吧

Using OSINT

- GitHub id --> GitHub repo --> git log --> name pinyin --> ZJU English news --> ZJU Chinese news --> name --> company --> current name, phone, etc.
- Tools are neutral, human beings are not

How to play with search engines like a pro

- filetype:
 - pdf
 - xls/xlsx Excel
 - doc Word
 - ppt
- site:zju.edu.cn – leaked attachments / files
- inurl:url
- intitle:text
- intext:text

How to play with search engines like a pro

- 13801xxxxxx filetype:xls
 - 水电缴费列表.xls
 - xxx 报名情况统计表.xls
 - etc.
- 13801xxxxxx filetype:docx
 - xxx 学校结题报告.docx
- 张xx site:zju.edu.cn
 - 浙江省 xx 届 xx 竞赛获奖名单
 - 我校学子在第 xx 届“xx 杯”斩获佳绩

Learning OSINT

- Mainly focus on image analyze:
- Image reverse search
- Obtain location from image
- Obtain shot time from image
- Obtain other info from image

Image Reverse Search

- Who post this image online?
- Is this image cropped / edited when spreading?
- Should we consider an image to be reliable?

Image Reverse Search

- Baidu: for images that are obviously from Chinese platforms
- Google: for foreign images
- Bing: ditto
- Yandex:
 - Set “attention” area
 - Search for similar images
 - Useful for searching landscape
- TinEye: for finding *exact/y* the same image

Image Reverse Search

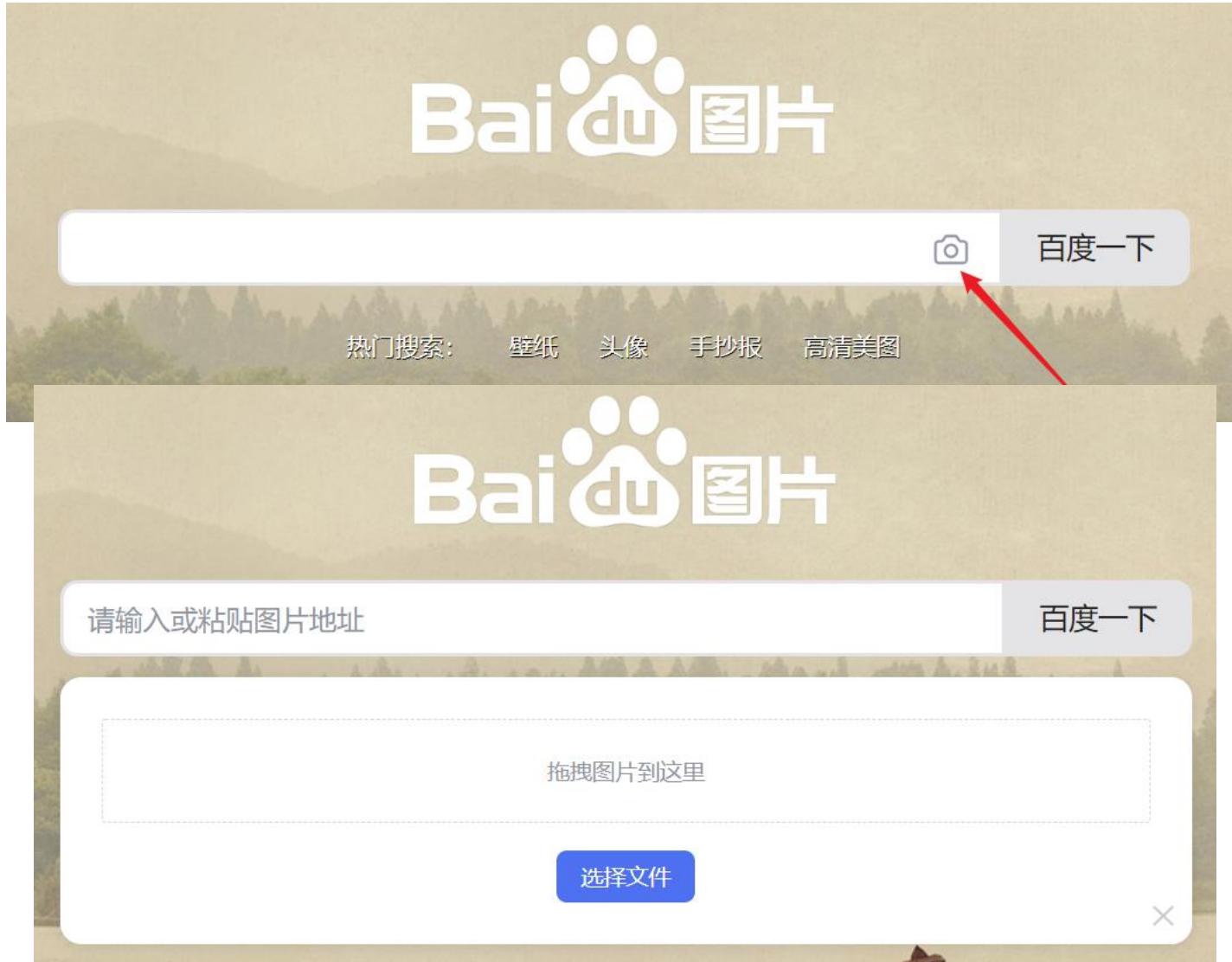


Image Reverse Search: An Example



LIVE

BREAKING NEWS

Russia unleashed more than 500 lions on its streets to ensure that people are staying indoors during this pandemic outbreak.

13:17 VLADMIR PUTIN RELEASED AROUND 500 LIONS TO MAKE PEOPLE STAY INDOORS

A screenshot from a news broadcast showing a lion walking across a city street at night. The scene is dimly lit by streetlights and building lights. A white car is parked on the right side of the street. In the bottom left corner, there is a red banner with white text that reads "LIVE", "BREAKING NEWS", and a main headline: "Russia unleashed more than 500 lions on its streets to ensure that people are staying indoors during this pandemic outbreak.". At the very bottom, there is a yellow banner with black text that reads "13:17 VLADMIR PUTIN RELEASED AROUND 500 LIONS TO MAKE PEOPLE STAY INDOORS".

Image Reverse Search: An Example

- A simple reverse search on Google and “filter by date range” reveals that the news is fake
- But we cannot claim that this lion is from Johannesburg that fast!
- What if the lion image appears even years before 2016?

<https://metro.co.uk> › News › Weird ▾ 翻译此页

Lion casually walks through city centre - Metro UK

2500 × 1482 · 2016年4月15日 — A **lion** was seen casually walking through the city centre in Johannesburg, but was just 'acting' as part of a local production.



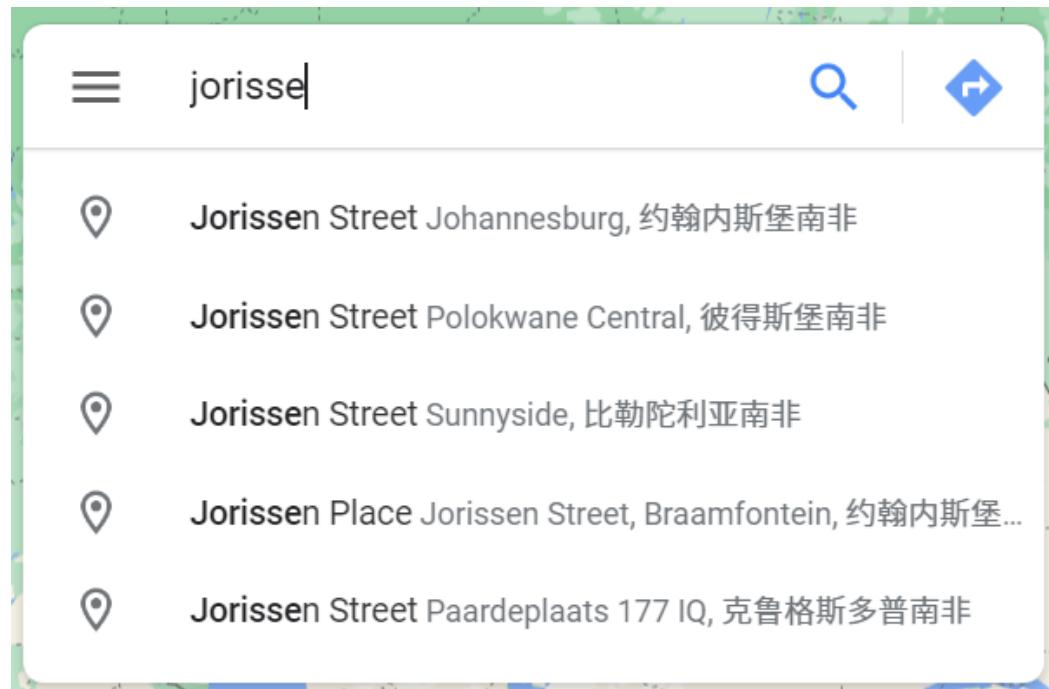
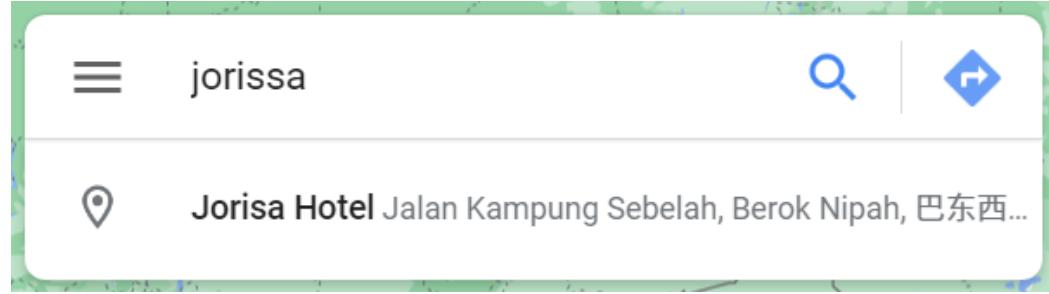
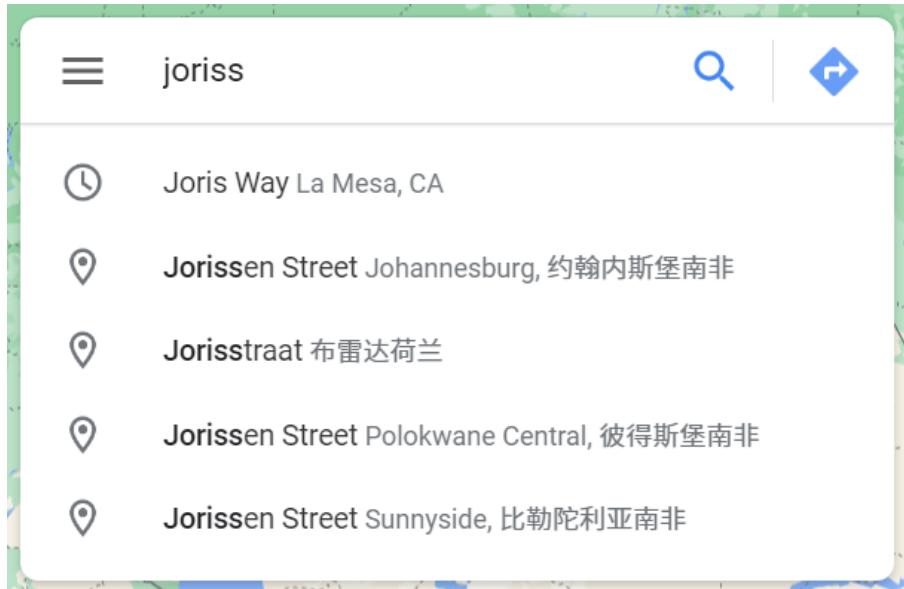
(Cont.)

- Cropped out in previous images!
- Jorissa? Jorisse?
- Not O or S or R, J“O” has round corner, ditto for “S” and “R”
- Probably not B for F, because hard to pronounce “SSB” and “SSF”
- Guessing is important in OSINT!



(Cont.)

- Google Map



(Cont.)

- Use Google Earth to check, one-by-one





© 2020 AfriGIS (Pty) Ltd.

© 2020 Google

US Dept of State Geographer

© 2020 Google

Google Earth

oblem

Imagery Date: 3/2017

26°11'34.16" S

28°02'09.72" E

elev 1770 m

eye alt 1.77 km

Image Reverse Search: An Example

- Some might claim a more simple approach:
- Reverse check, but “city center” is a big area
 - If you are lucky: find the correct road with a few attempts
 - Not lucky: wasting time
- Cannot avoid wasting time in OSINT completely, the key is to use experiences to waste as less time as possible
- Different strategies: when to take risks? Up to you

<https://metro.co.uk> › News › Weird ▾ 翻译此页

[Lion casually walks through city centre - Metro UK](#)

2500 × 1482 · 2016年4月15日 — A lion was seen casually walking through the city centre in Johannesburg, but was just 'acting' as part of a local production.



Image Reverse Search: An Example

- Date, location, & other info extracted from image successfully
- --> reverse search is the basic tool in OSINT for image
- Other tools available

Other tools

- Sun, shadow
- Culture specific signs
- Weather
- Other Geography knowledges
- And even Geometry (Yes!)

Sun & Shadow

- SunCalc.org
- SunEarthTools.com
- Make use of reference objects to extract length information!
- e.g.:
- Object : Shadow = 1 : 2.29
- Location is ZJU Zijingang Playground

Cont.

- e.g.:
- Object : Shadow = 1 : 2.29
- Location is ZJU Zijingang Playground
- Date is 2022/7/4

- Use arctan to get sun angle
- Use binary search to narrow time range
- Time will be determined to be around 17:05

Culture Specific Signs

- Buses, trains, road signs, ...
- Game: GeoGuesser!
- Different regions have different:
 - Plates
 - Telegraph pole



South Korea



Singapore



georainbolt

georainbolt

Follow

17 Following 596.9K Followers 16.6M Likes

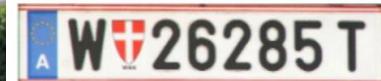
professional google maps player

🔗 georainbolt.com

Kenya



Austria



Monaco



Taiwan



Culture Specific Signs

- Buses, trains, road signs, ...
- Game: GeoGuessr!
- Different regions have different:
 - Plates
 - Telegraph pole
 - Road infra
 - Buses, trains, ...

Yellow phone booth (Jersey)



Snow rocket (Japan)



Colombian Cross

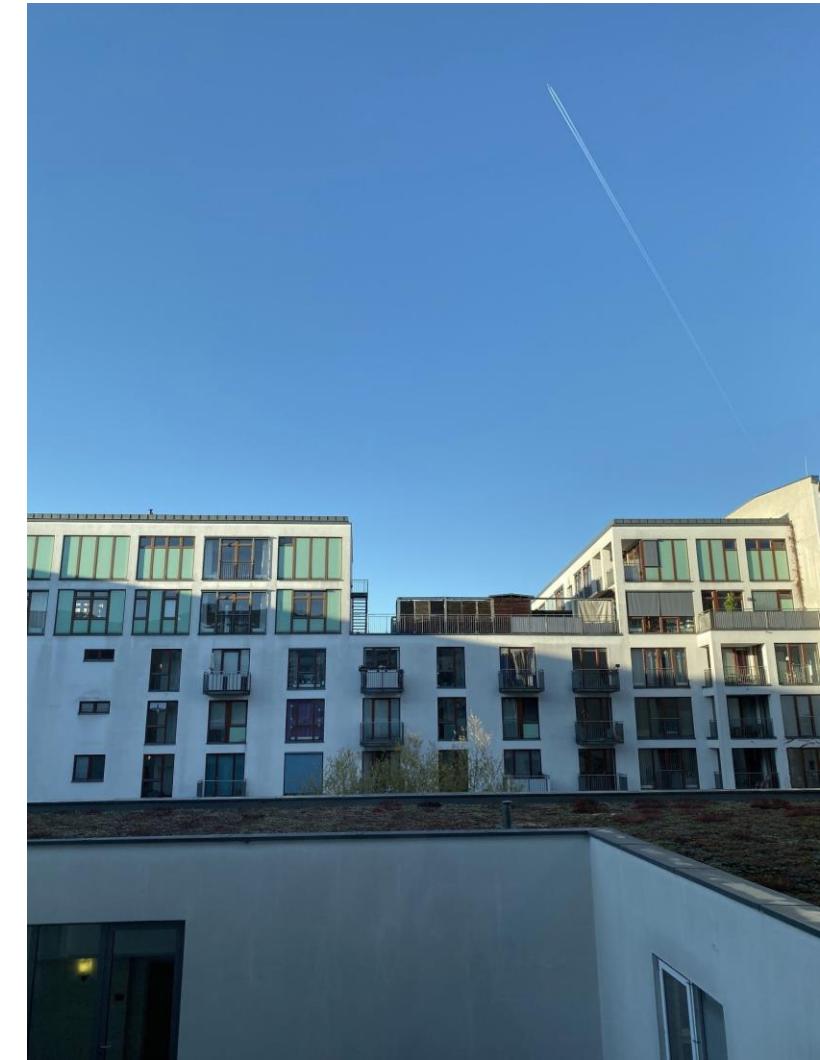


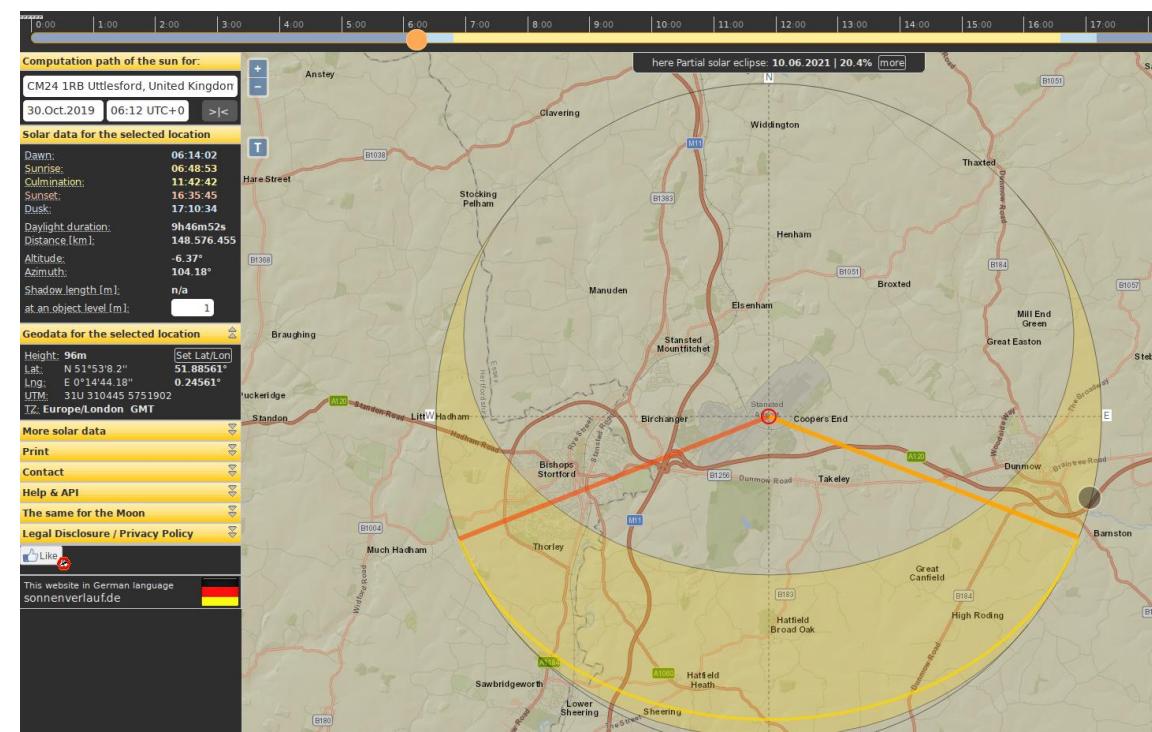
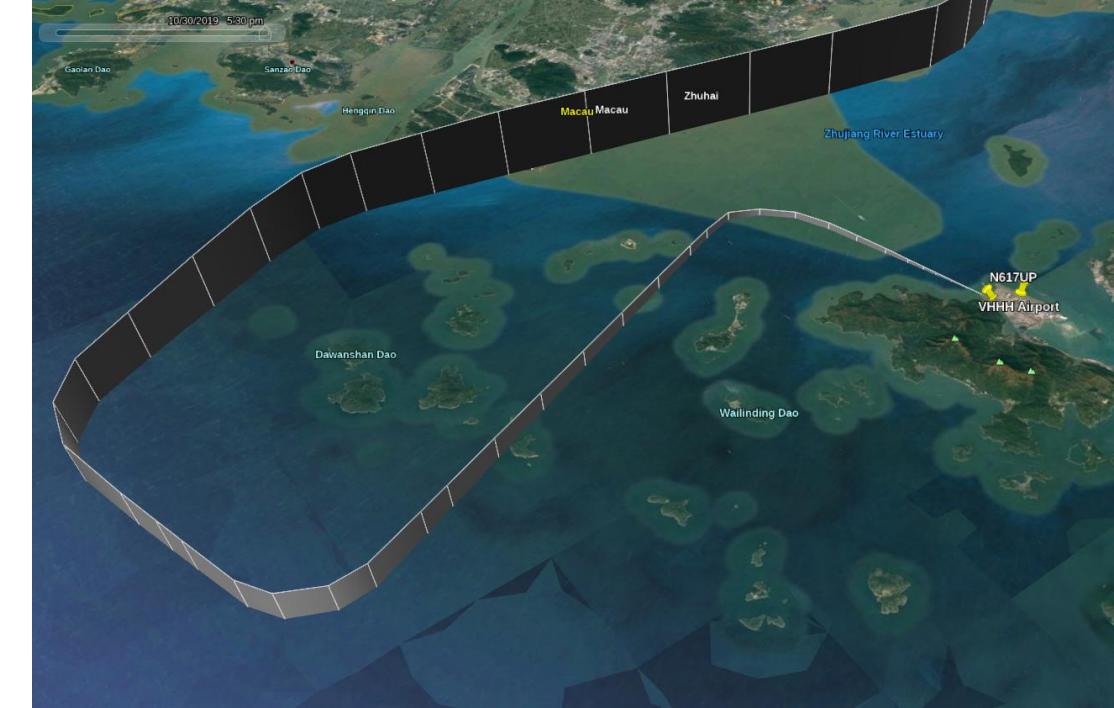
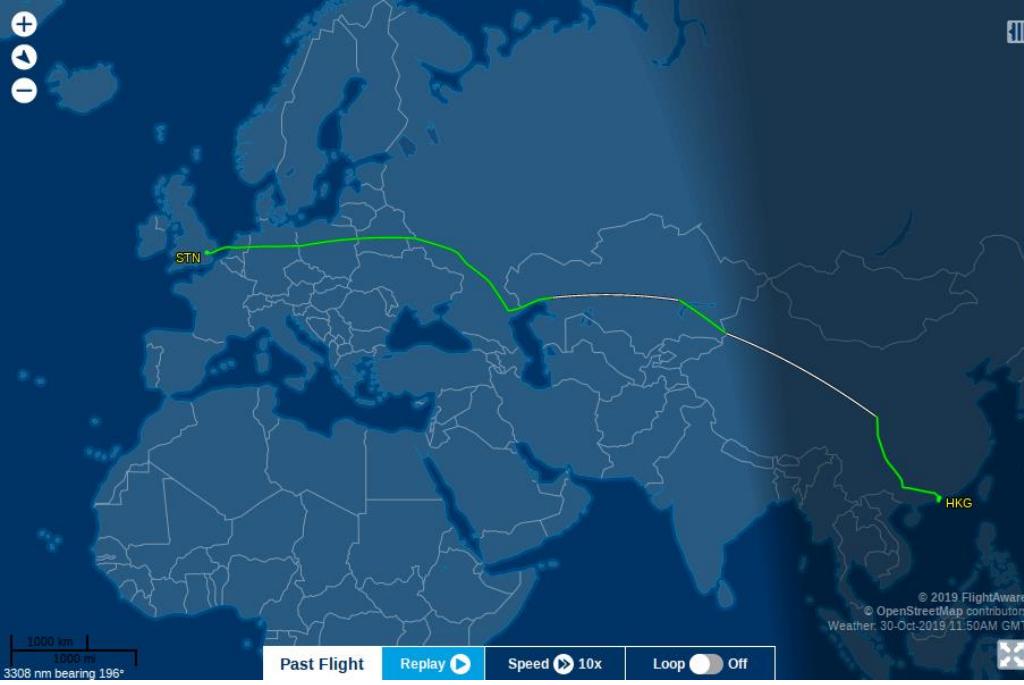
Weather

- Can be used for filtering out incorrect possibilities
- Different continents / regions have different weathers
- Can also be used to determine the date if you already have a location
- Many websites store historic weather data

Other Geography Knowledges

- Plane tail track direction
1. The plane in the sky is a Boeing 747 flying from London to Hong Kong
 2. The photo was taken on October 30th 2019.
 3. The photographer was in a hotel.





Other Geography Knowledges

- Directions determined by plane track data & time guess: time should be in the morning
- We can tell that the photographer is at the north side of the track
- Use your two hands to simulate



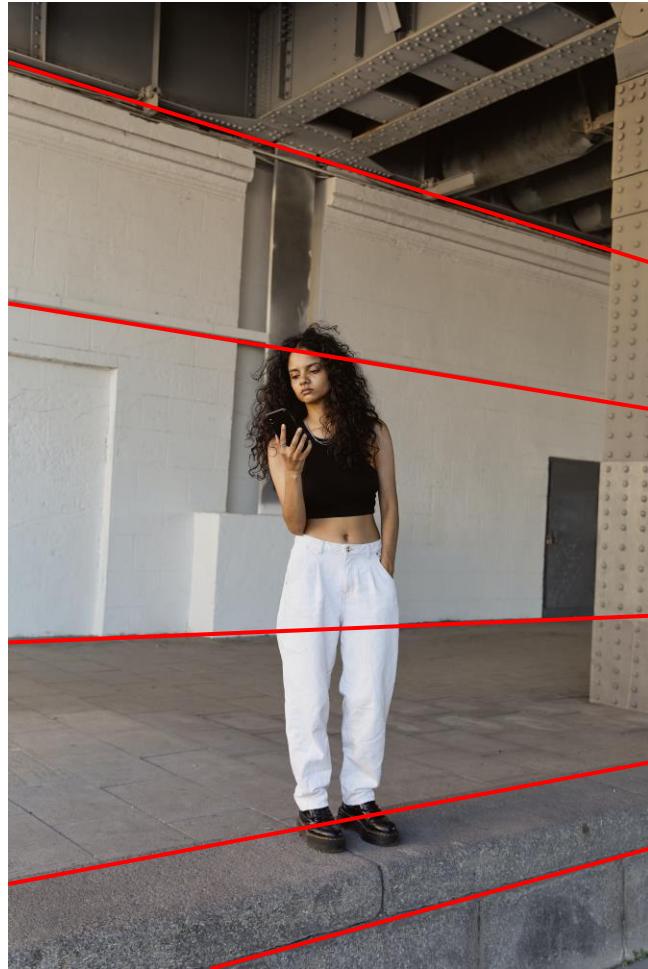
Geometry

- Parallel lines should cross at the horizon line
- Can be used to check whether the image has been edited or not
- Can be used to determine the photographer's height

Geometry



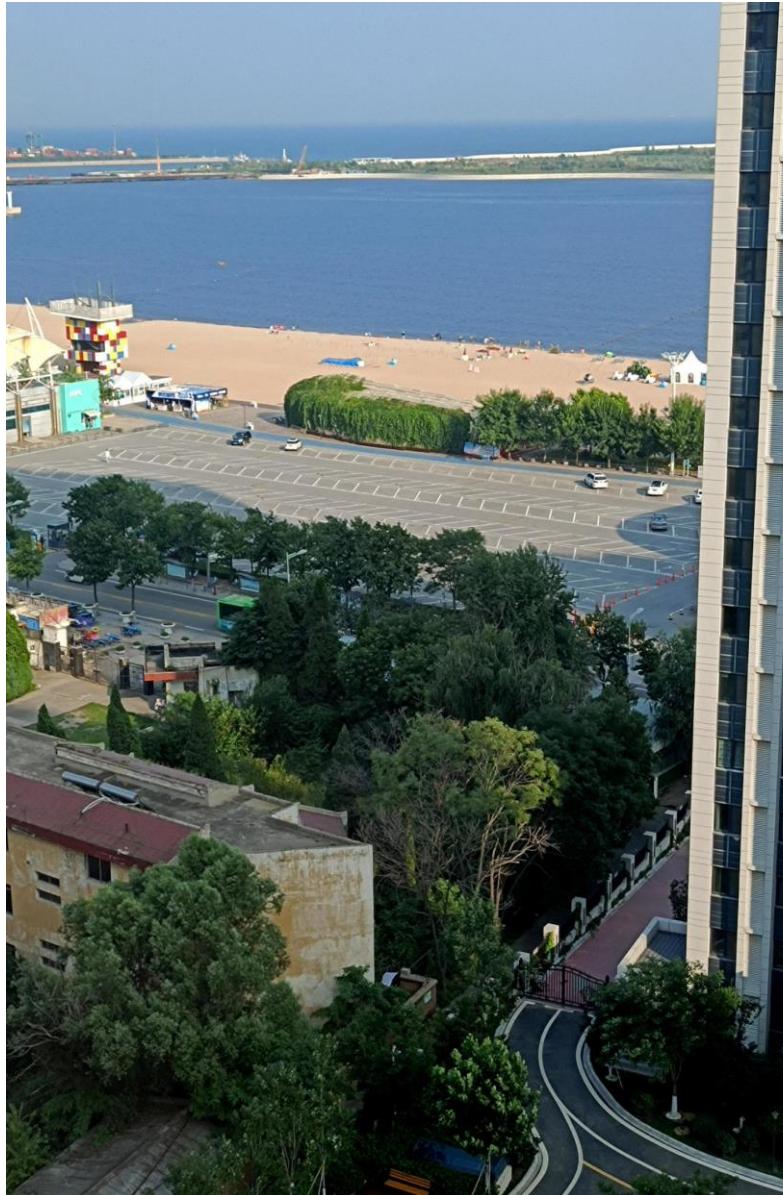
Geometry



Geometry

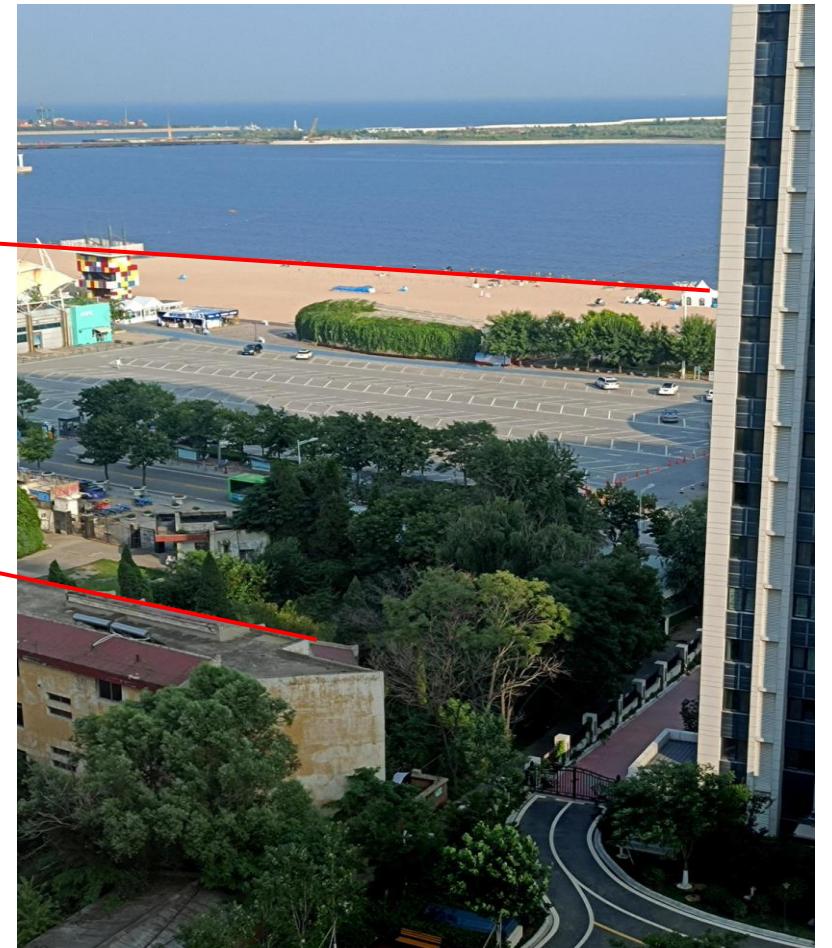
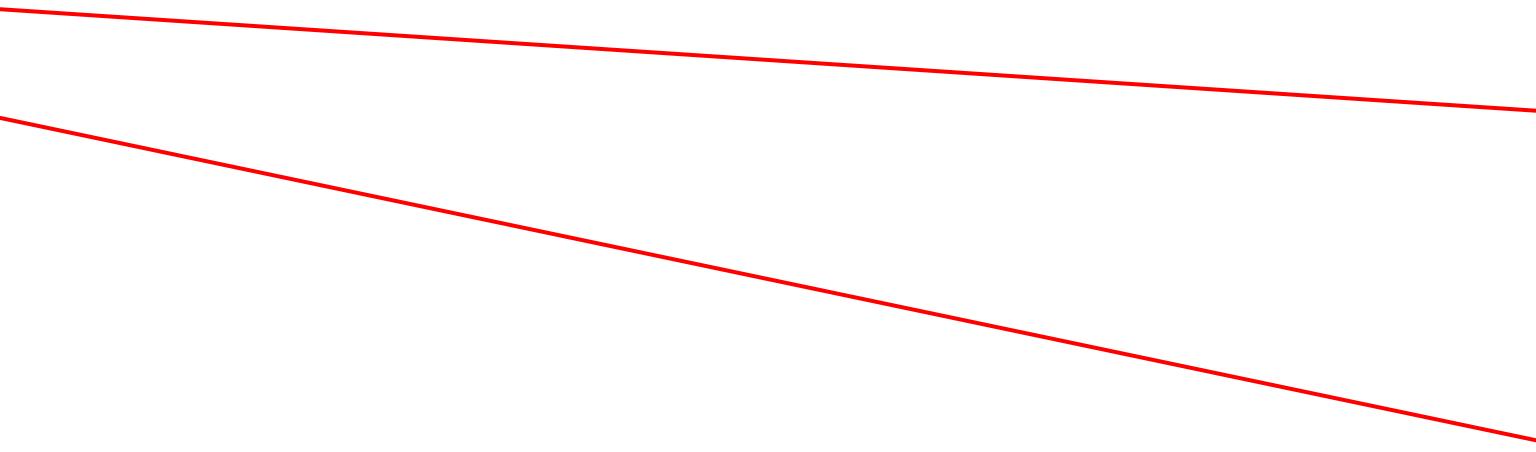


Geometry



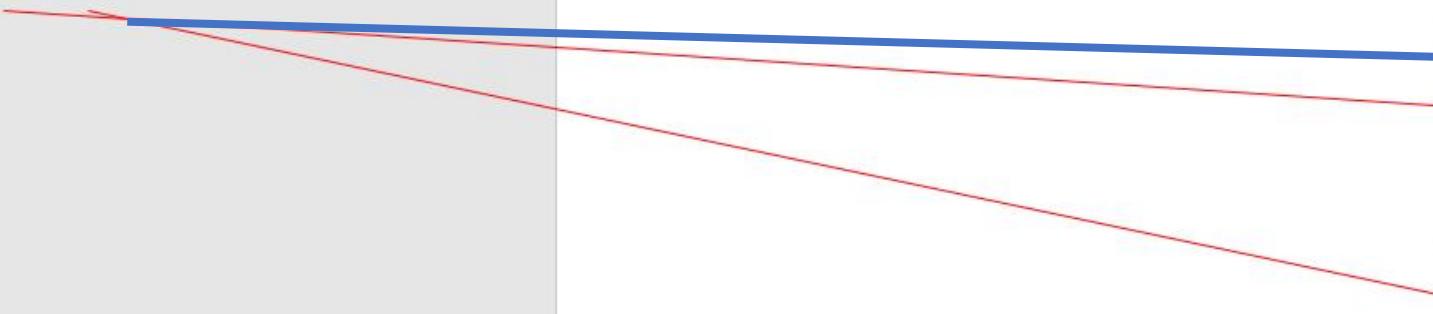
Geometry

Q: Are they really parallel lines?



Geometry

Geometry



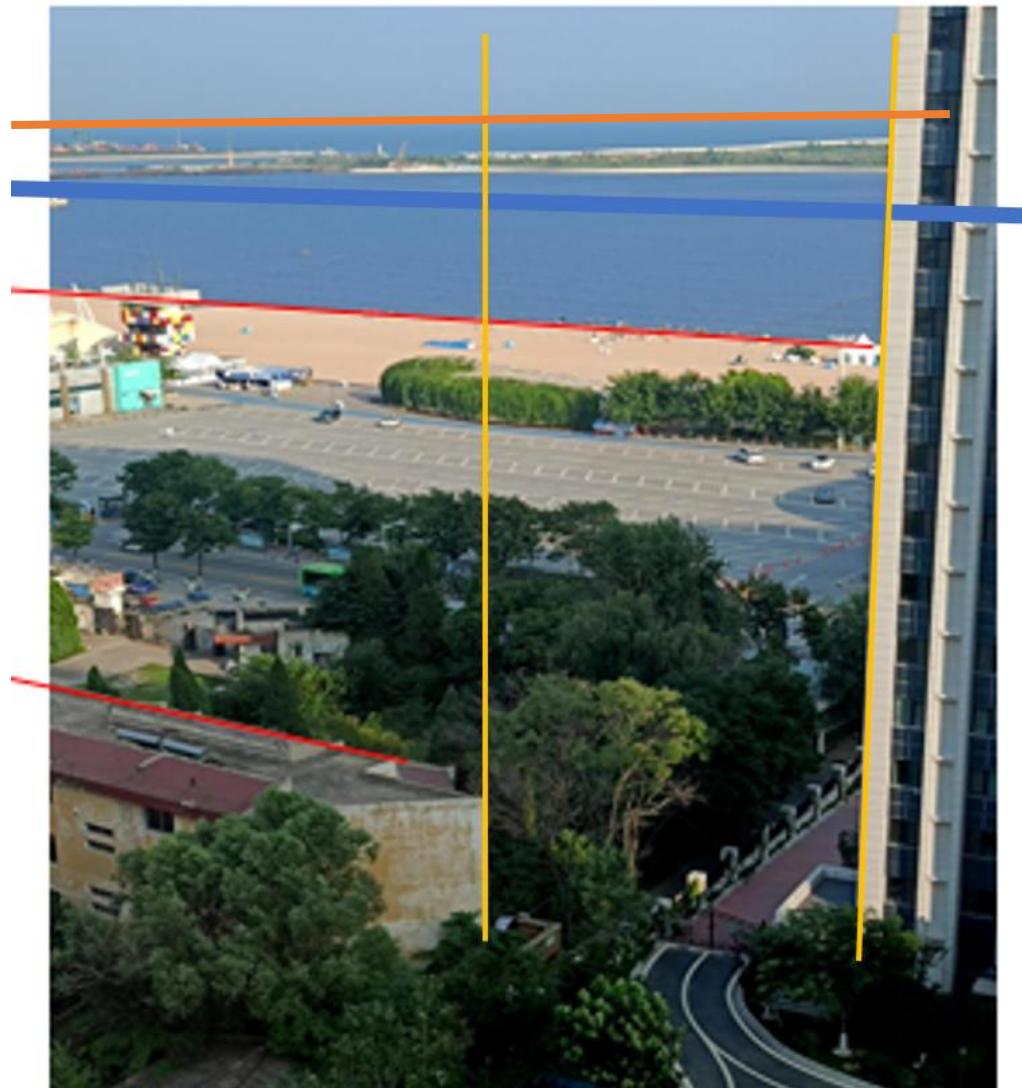
Geometry: Lens Distortion

Correct vertical lens distortion first!

Vertical parallel lines should be parallel in the photo

The horizon line should be perpendicular to all vertical parallel lines: building walls, towers, lamps, ice drops, etc.

Aim: get rid of errors caused by photo rotation



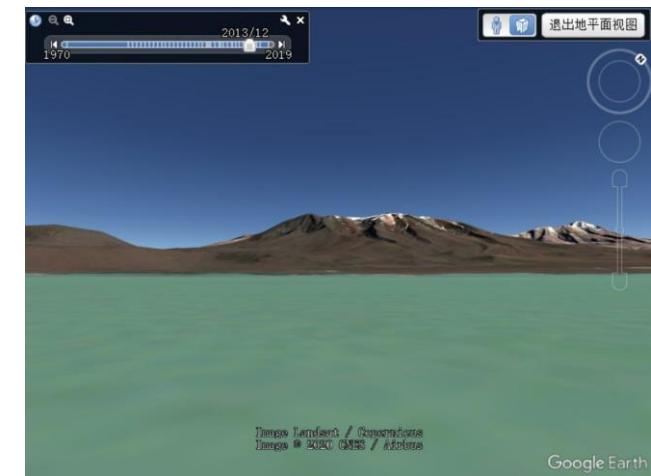
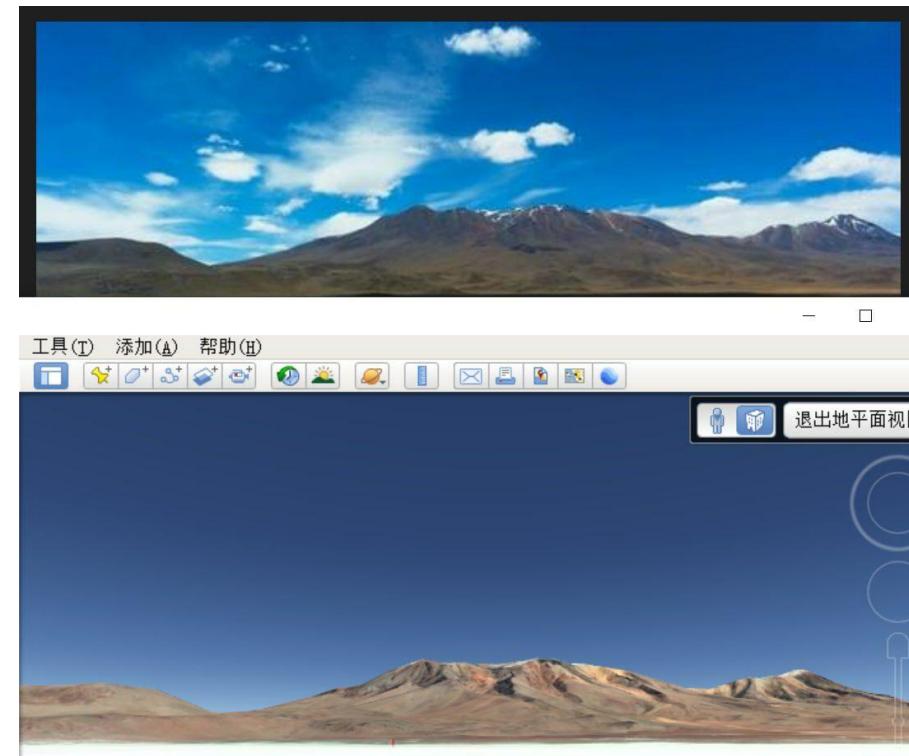
Landscape

- Use Yandex
- Laguna Hedionda, Bolivia



[commons.wikimedia.org](#)

This page was last edited on 7 January 2021, at 18:00.



- Thanks