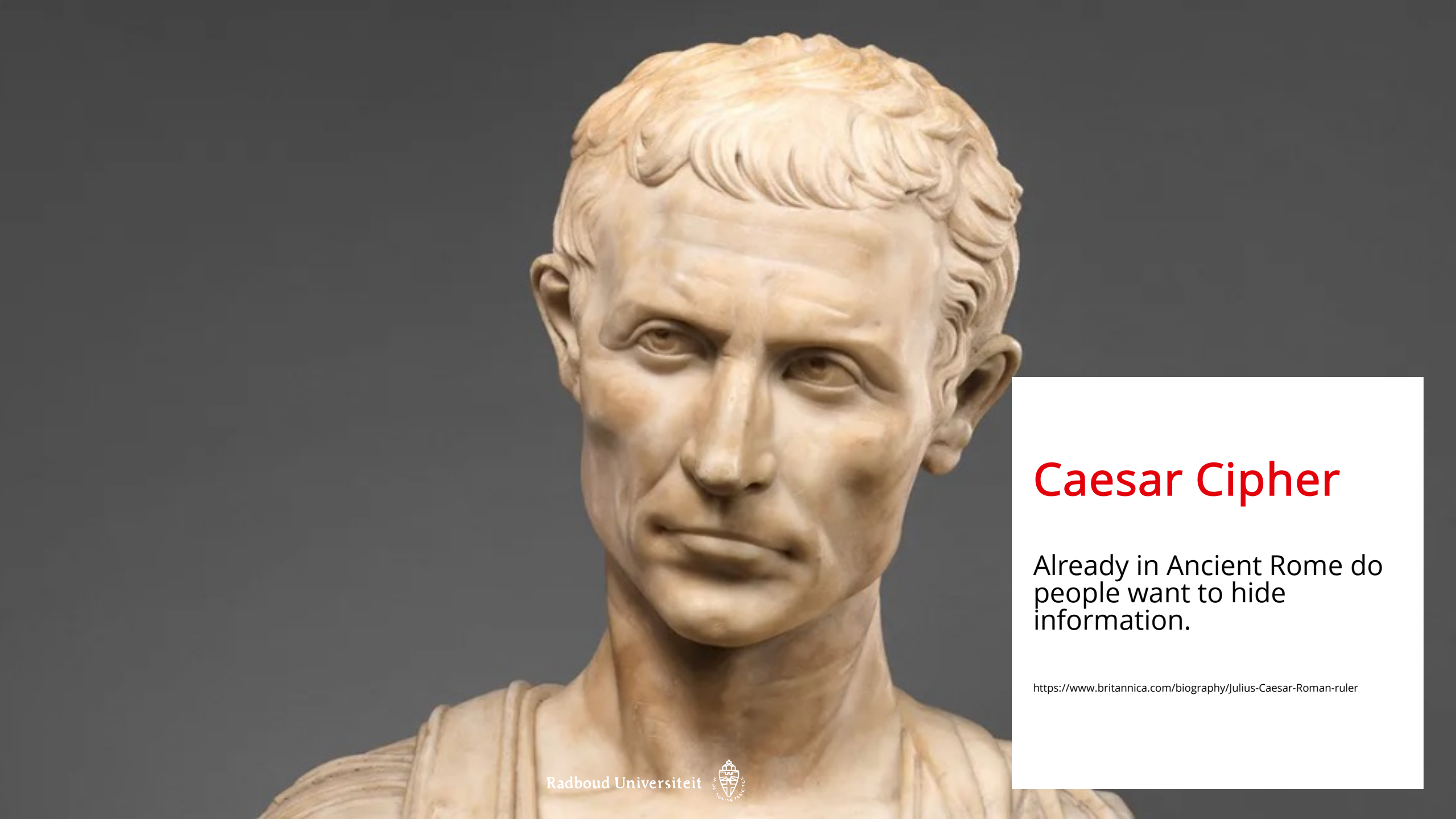
Obfuscation

Artur Wiadrowski
Tianlang Xie

HISTORY OF OBFUSCATION



Caesar Cipher

Already in Ancient Rome do people want to hide information.

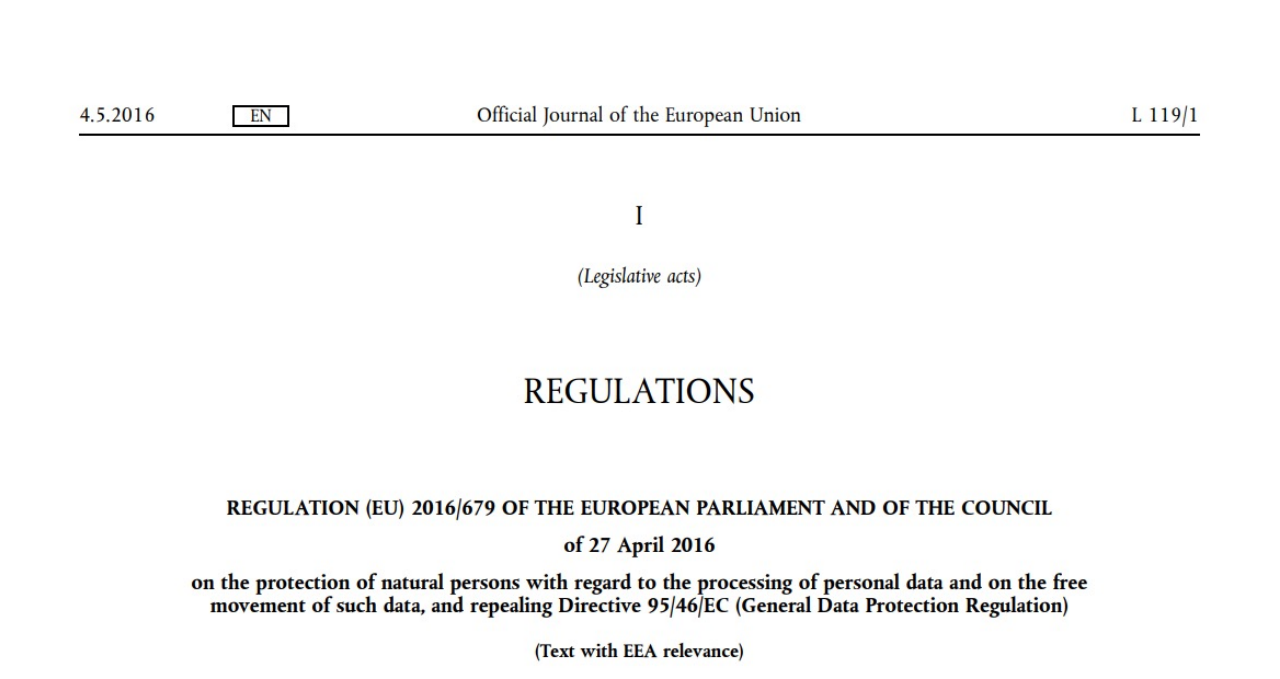
https://www.britannica.com/biography/Julius-Caesar-Roman-ruler

POPULARITY OF THE INTERNET

- Social media takes off, with Facebook launching as early as 2004.
- Users can easily falsify their real names, use pseudonyms.
- Nonetheless, their postings can be linked to them, i.e. with digital fingerprinting.

INTERNET LAWS

- Law starts to catch up with electronic innovations.
- Users become concerned about their privacy.



4.5.2016 EN Official Journal of the European Union L 119/1

I

(Legislative acts)

REGULATIONS

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 27 April 2016

on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

(Text with EEA relevance)

https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

OBFUSCATION TECHNIQUES

- New methods are invented to hide one’s real identity.

INTRODUCTION

Formal definition

Obfuscation is the deliberate addition of ambiguous, confusing, or misleading information to interfere with surveillance and data collection.

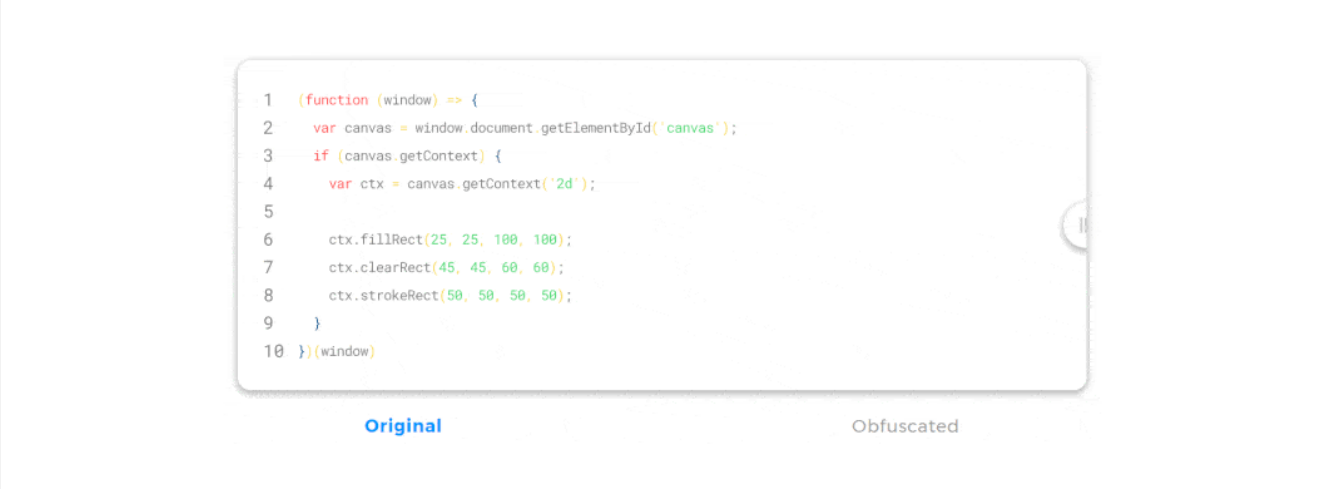
Brunton, F., & Nissenbaum, H. (2015). *Obfuscation: A user's guide for privacy and protest*. MIT Press.

OBFUSCATION — HIDING, NOT SECURING

- **Definition**: Making code/data harder to understand (e.g., renaming variables).
- **Goal**: Deter reverse engineering, not strong security.

Features:
•🔄 **Reversible** (via manual analysis).

•🛠️ Used for code protection, deterring simple copying.



ENCRYPTION — MATHEMATICAL PROTECTION

Definition: Transforming data with algorithms + keys (e.g., AES encryption).

Goal: Ensure confidentiality; block unauthorized access.

Features:

- 🔑 **Key-dependent** (irreversible without the key).
- 🔒 Secures sensitive data (passwords, communications).

```c
int partition(int low, int high)
{
    int tmp,pivotkey;
    tmp = array[low];
    pivotkey = array[low];
    while (low < high)
    {
        while (low < high && array[high] >= pivotkey)
            --high;
        array[low] = array[high];
        while (low < high && array[low] <= pivotkey)
            ++low;
        array[high] = array[low];
    }
    array[low] = tmp;
    return low;
}
```


```
2910b397e9b9d31e756dcafafd9b5b8e1dbd16ec364ad60185cf154b14f093125cc3d18ab15fef3662916705361f31
5c1a7b0b5fdb98b4a8efde82709cd8519bc2e03a83ddb98cf392a7100c8baee6fa92b48f710e9f9e4059a5946b6cb4
10111797836e7489868a902c4f3071a620063b4779f6f2f1b543fca01945e00ffe0393a24d1c942cbab62247040927
a02733e37a5aa27cd0b0ec6fb66d7b7efc0f9dbb9e25fae7aade56d90e2c9c11d5437ddddf2ace27a878f2f7426c30
bb2ac3234d4d71ca101511bb681b1c286bd6f4b68c70bb75fca26aa5868766e48adb162b4e6fde0b74461a659b431c
79dca514d886ee128dbbbf30dbffff5b317bce26fcbad7f78ee0781683e1ab422f6d0cd4e6162f820a98ad89709c4a
f14e58638ad9e5779848090d7ea2c8681ab060f439478e15c6dfdaf495e9c985c159196d9be039a2ec706a816674d6
83b5a7c3b2d713
```

OBFUSCATION VS. ENCRYPTION

Aspect	Obfuscation	Encryption
Security Level	❌ Weak	✅ Strong
Reversibility	No key (human analysis)	Requires key
Primary Use	Code readability	Data confidentiality

CORE CONCEPTS

OBFUSCATION = DIGITAL SMOKE SCREEN

Obfuscation — A Mask, Not a Fortress

- **Metaphor Explanation**:

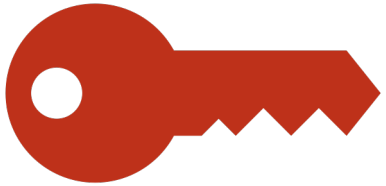
- Smoke/dynamic particles obscure original code/data, making it "visible but unreadable.“

- ⚠️ Smoke can dissipate (reversible via manual analysis).

Smoke hides, but doesn’t protect.

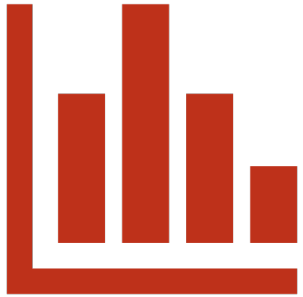
DIFFERENT NOTIONS OF OBFUSCATION

DIFFERENT NOTIONS OF OBFUSCATION



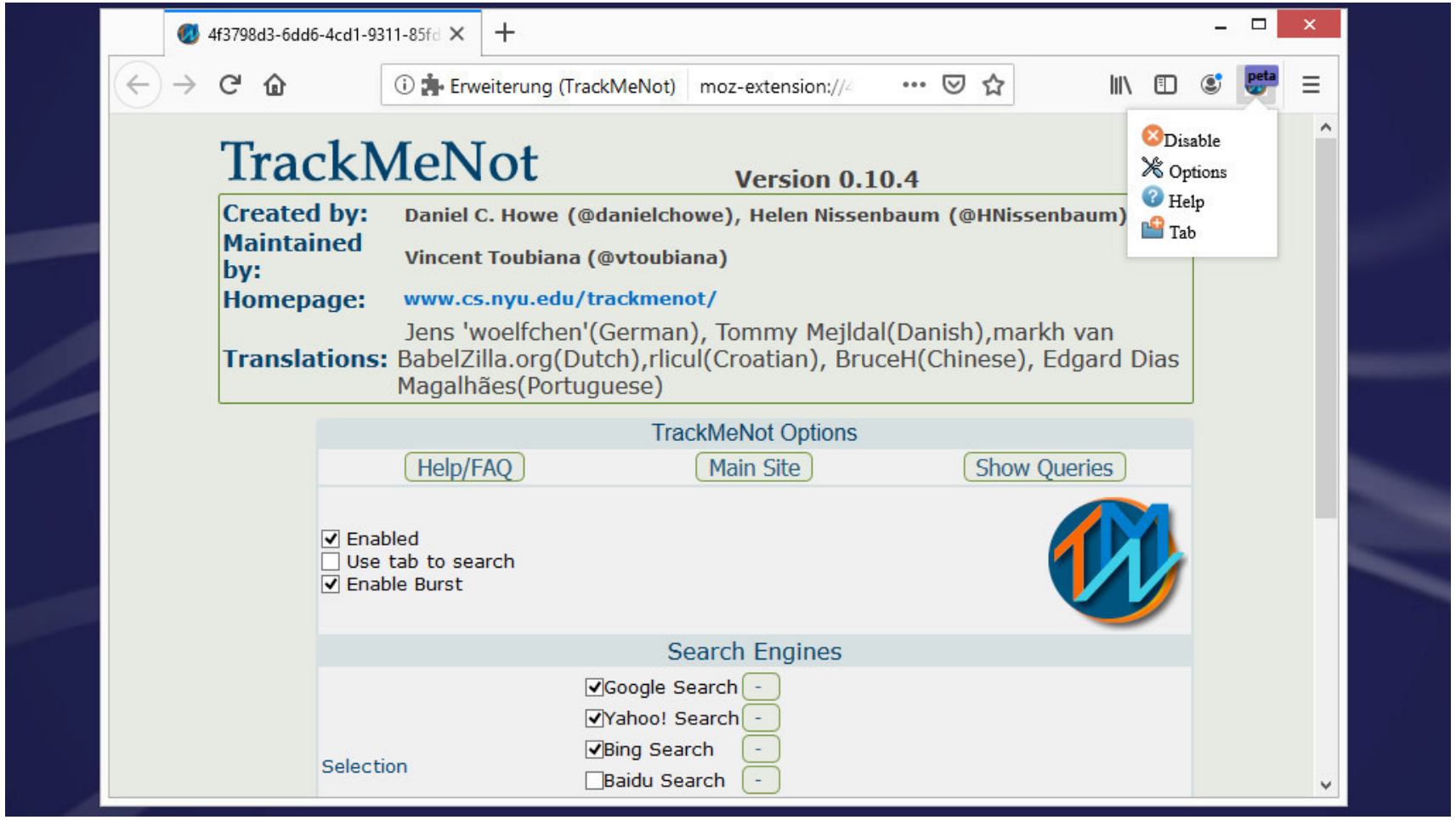
Code obfuscation

Deliberately making source or machine code convoluted to hinder analysis



Data obfuscation

Altering datasets to mask sensitive elements while preserving functional utility.



TrackMeNot

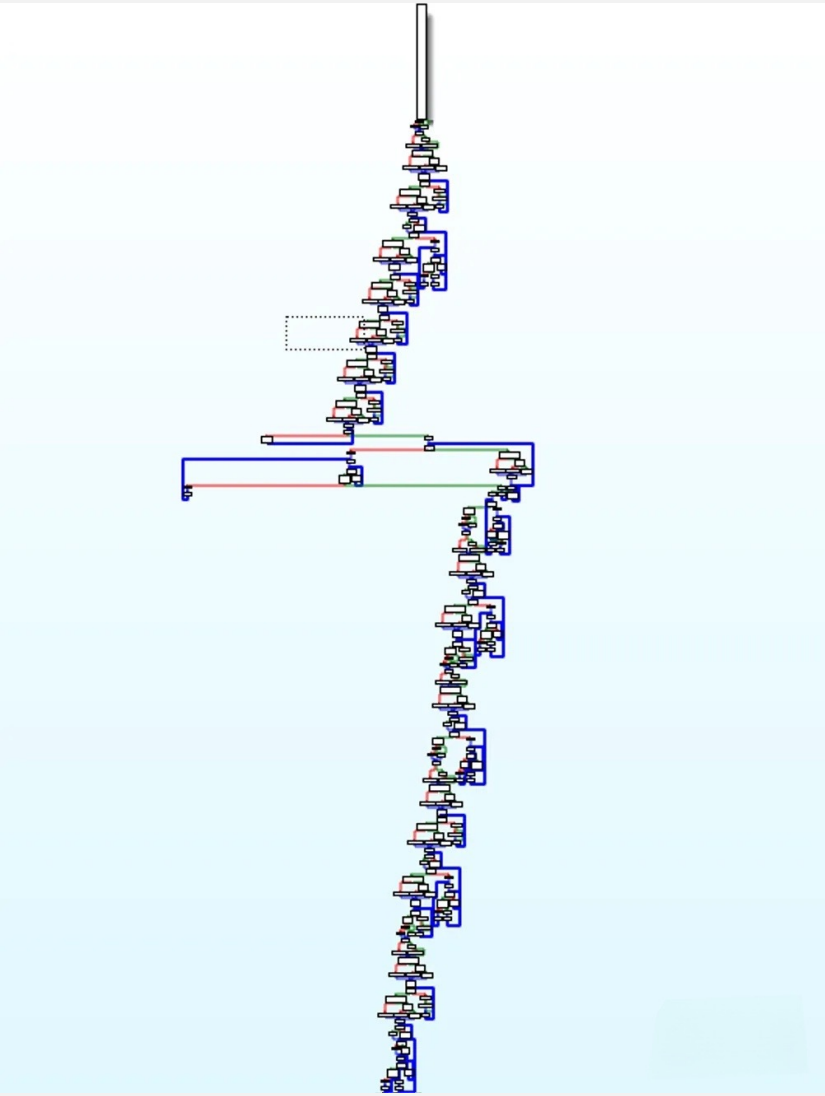
Generation of fake browser queries to disguise interests and behavior patterns.

HOW THE ATTACKERS USE THE OBFUSCATION

Lumma Stealer

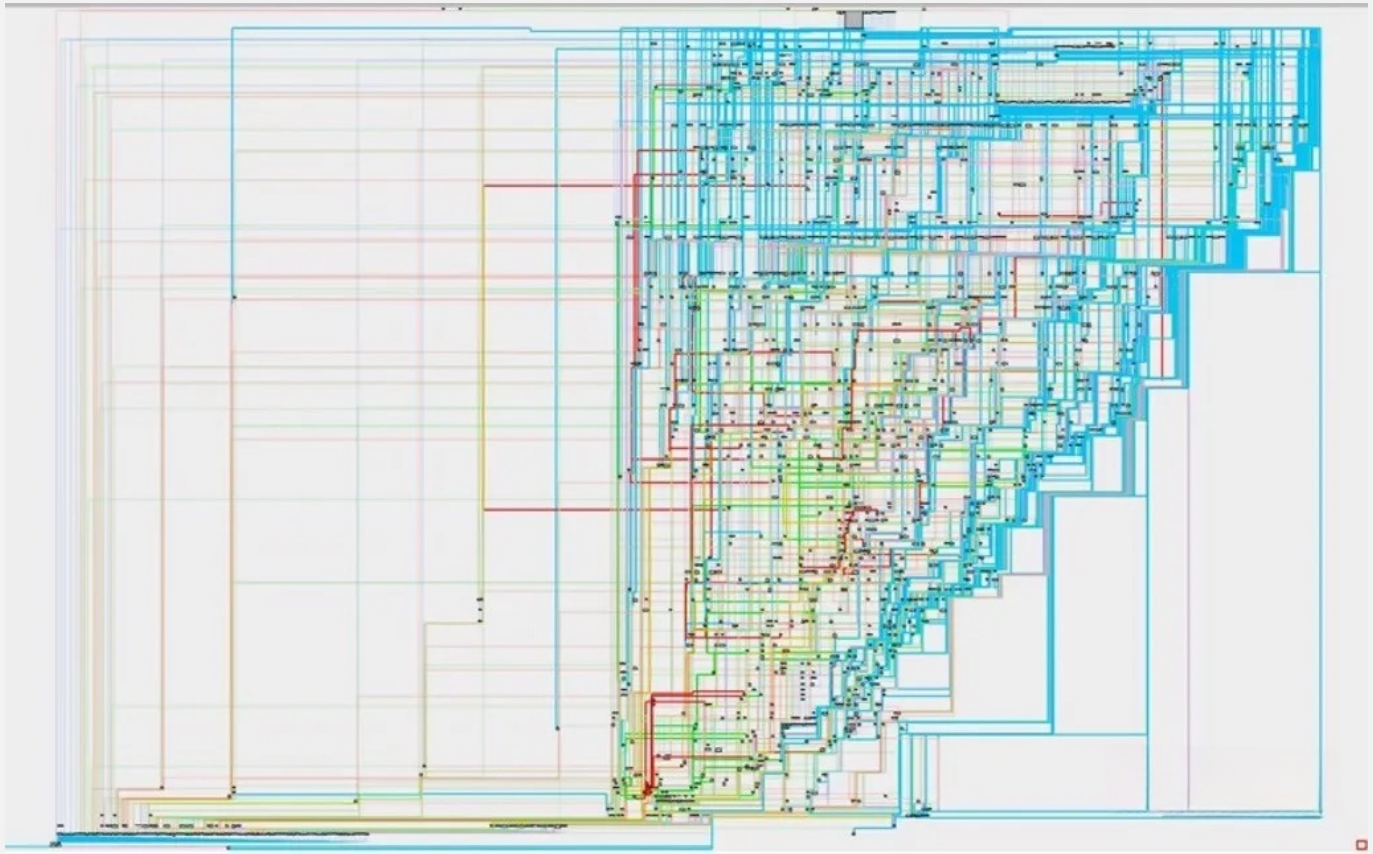
An info-stealing malware that targets browsers, crypto wallets, and apps like Discord.

It collects credentials, cookies, and autofill data, then sends them to attackers.



CODE OBFUSCATION

using obfuscation to hide its decryption process(with thousands of code loops)



DATA OBFUSACTION

```
001ECD7B   83C4 14        add esp,14
001ECD7E   68 A2AA2100    push 21AAA2                    21AAA2:L"Walledx765ets/Eleedx765ctrum"
001ECD83   E8 37670000    call 1F34BF
001ECD88   83C4 04        add esp,4
001ECD8B   89C6           mov esi,eax                    esi:L"Wallets/Electrum", eax:L"%appdata%\\Electrum\\wallets"
001ECD8D   68 D8C22100    push 21C2D8                    21C2D8:L""edx765"
001ECD92   E8 28670000    call 1F34BF
001ECD97   83C4 04        add esp,4
001ECD9A   89C7           mov edi,eax                    eax:L"%appdata%\\Electrum\\wallets"
001ECD9C   68 7CA42100    push 21A47C                    21A47C:L"%appdedx765ata%\\Eledx765ectrum\\waledx765lets"
001ECDA1   E8 19670000    call 1F34BF
001ECDA6   83C4 04        add esp,4
001ECDA9   FF75 E8        push dword ptr ss:[ebp-18]
001ECDAC   53             push ebx
001ECDAD   56             push esi                       esi:L"Wallets/Electrum"
001ECDAE   57             push edi
001ECDAF   50             push eax                       eax:L"%appdata%\\Electrum\\wallets"
001ECDB0   E8 1F45FFFF    call 1E12D4
001ECDB5   83C4 14        add esp,14
001ECDB8   68 DCAA2100    push 21AADC                    21AADC:L"Walledx765ets/Ethedx765ereum"
001ECDBD   E8 FD660000    call 1F34BF
001ECDC2   83C4 04        add esp,4
001ECDC5   89C6           mov esi,eax                    esi:L"Wallets/Electrum", eax:L"%appdata%\\Electrum\\wallets"
001ECDC7   68 CA85210 0   push 21B5CA                    21B5CA:L"keystedx765ore"
001ECDCC   E8 EE660000    call 1F34BF
001ECDD1   83C4 04        add esp,4
001ECDD4   89C7           mov edi,eax                    eax:L"%appdata%\\Electrum\\wallets"
001ECDD6   68 16AB2100    push 21AB16                    21AB16:L"%appdedx765ata%\\Etheedx765reum"
001ECDDB   E8 DF660000    call 1F34BF
001ECDE0   83C4 04        add esp,4
001ECDE3   FF75 E8        push dword ptr ss:[ebp-18]
001ECDE6   53             push ebx
001ECDE7   BB 02000000    mov ebx,2
001ECDEC   56             push esi                       esi:L"Wallets/Electrum"
001ECDED   57             push edi
001ECDEE   50             push eax                       eax:L"%appdata%\\Electrum\\wallets"
001ECDEF   E8 E044FFFF    call 1E12D4
001ECDF4   83C4 14        add esp,14
001ECDF7   FF75 E8        push dword ptr ss:[ebp-18]
001ECDFA   53             push ebx
001ECDFB   68 10A42100    push 21A410                    21A410:L"Wallets/Exodus"
001ECE00   BE 90C32100    mov esi,21C390                 esi:L"Wallets/Electrum"
001ECE05   56             push esi                       esi:L"Wallets/Electrum"
001ECE06   68 EEA32100    push 21A3EE                    21A3EE:L"%appdata%\\Exodus"
001ECE0B   E8 C444FFFF    call 1E12D4
001ECE10   83C4 14        add esp,14
001ECE13   FF75 E8        push dword ptr ss:[ebp-18]
001ECE16   53             push ebx
001ECE17   68 F2B42100    push 21B4F2                    21B4F2:L"Wallets/Ledger Live"
001ECE1C   56             push esi                       esi:L"Wallets/Electrum"
001ECE1D   68 C6B42100    push 21B4C6                    21B4C6:L"%appdata%\\Ledger Live"
001ECE22   E8 AD44FFFF    call 1E12D4
001ECE27   83C4 14        add esp,14
001ECE2A   C745 D8 44915937  mov dword ptr ss:[ebp-28],37599144
001ECE31   E9 3BFDFFFF    jmp 1ECB71
001ECE36   3D 7C478890    cmp eax,9088477C               eax:L"%appdata%\\Electrum\\wallets"
001ECE3B   0F84 26050000  je 1ED367
001ECE41   3D 9423FFA5    cmp eax,A5FF2394               eax:L"%appdata%\\Electrum\\wallets"
001ECE46   0F85 25FDFFFF  jne 1ECB71
001ECE4C   50             push eax                       eax:L"%appdata%\\Electrum\\wallets"
001ECE4D   50             push eax                       eax:L"%appdata%\\Electrum\\wallets"
001ECE4E   89E0           mov eax,esp
001ECE50   8945 CC        mov dword ptr ss:[ebp-34],eax  [ebp-34]:"#/"
001ECE53   83EC 08        sub esp,8
001ECE56   89E0           mov eax,esp
001ECE58   8945 F0        mov dword ptr ss:[ebp-10],eax
001ECE5B   83EC 08        sub esp,8
```

The following are directory paths related to sensitive data collection by the trojan. Each folder name includes "dx765" as an obfuscation

```
not found  header not found  file not found  not initialized  Jaxx  Liberty  Nifty  Hisedx765tory  Histedx765ory  Authy  OneKey  DA
pPlay  ZilPay  Nabox  TezBox  Woziedx765lla  Firefe[illegible]ox  ICONex  *.kbdx  *.edx765txt  dp.txt  Systee
dx765m.txt  Software.txt  EnXrypt  Logedx765in  Daedx765ta  Foedx765r  Accedx765ount  Hycon  Lite  Clie
nt  iWlt  Sollet  Binedx765ance  Chaedx765in  Waledx765let  Ronedx765in  Walledx765et  Wombat  %appdata
%\Exodus  Wallets/Exodus  %localappdata%\Coinomi\Coinomi\wallets  %appdedx765ata%\Eledx765ectrum\
waledx765lets  Applications/KeePass  %appdaedx765ta%\Koedx765zilla\Firedx765efox\Profedx765iles
Netwedx765ork\Cookedx765ies  *s  %s\%s  Authenticator  GAuedx765th  Autheedx765nticator  Bedx765OS  A
uthentiedx765cator  Keplr  Tredx765ezor  Passwedx765ord  Manager  alifbnbfobpmeekipheeijimdpnlpgpp
%appdata%\Telegram  Desktop  bcopgchhojmggmffilplmbdicgaihlkp  fhbohimaelbohpjbbldcngcnapnedx765d
odjp  BitClip  dmkamcknogkgcdfhhbddcghachkejeap  Auro  fhmfendgdocmcbmfikdcogofphimnkno  Cyano  Como
do  cphhlgmgameodnhkjdmkpanlelnlohao  logins.json  Terra  Station  Nash  Extension  apedx765p-stoedx7
65re.jsedx765on  nkbihfbeogaeaoehlefedx765nkodbefgpgknn  kpfopkelmapcoipemfendmdcghnegimn  hpglfh
gfnhbgpjdenjgmdgoeiappafln  Steem  Keychain  hcflpincpppdclinealmandijcmnkbgn  Opedx765era  Neoedx7
65n  Walledx765ets/Eleedx765ctrum  Walledx765ets/Ethedx765ereum  %appdedx765ata%\Etheedx765reum  C
bromiedx765um  Phantom  dkdedlpgdmmkkfjabffeganieamfklkm  ejbalbakoplchlghecdaedx765lmeeeajnimhm
nlbmnnijcnlegkjjpcfjclmcfggfefdm  lkcjlnjfpbikmcmbachjpdbijejflpcm  Applications/Telegram  %progr
amfiles%\Steam  Applications/Steam  recentservers.xml  sitemanager.xml  ijmpgkjfkbfhoebgogflfebnme
jmfbml  winhttp.dll  ntdll.dll  crypt32.dll  user32.dll  advapi32.dll  SOFTWARE\Microso
ft\Windows\CurrentVersion\Uninstall  onofpnbbkehpmmoabgpcpmigafmmnjhl  flpiciilemghbmfalicajoolh
kkenfel  ilgcnhelpchnceeipipijaljkblbcobl  oeljdldpnmdbchonielidgobddfffla1  %appdata%\AnyDesk  Ap
plications/AnyDesk  Meedx765taMaedx765sk  Troedx765nLiedx765nk  imloifkgjagghnncjkhggdhalmcnfklk
nhnkbkgjikgcigadomkphalanndcapjk  lodccjjbdhfakaekdiahmedfbieldgik  /c2sock  nanjmdknhkinifnkgdcg
gcfnhdaammmj  kncchdigobghenbbaddojjnnaogfppfj  fihkakfobkmkjojpchpfgcmhfjnmnfpi  Yoroi  Wallets/C
oinomi  bhghoamapcdpbohphigooaddinpkbai  Math  Polymesh  kkpllkodjeloidieedojogacfhpaihoh  amkmjjm
mfldldogmhpjloimipbofnfjih  mnfifefkajgofkcjkemidiaecocnkjeh  Screen.png  aeachknmefphepccionboohc
konoeemg  nkddgncdjgjfcddamfgcmfnlhccnlmig  %programfiles%\Steam\config  Applications/Steam/confi
g  *.conf  jojbfeoedkpkglbfimdfabpdfjaoolaf  Leaf  infrbcajgfhgbjpbeppbkgnabfdkdaf  %appdata%\Ledg
er  Live  Wallets/Ledger  Live  formhistory.sqlite  cookies.sqlite  places.sqlite  \Loceedx765al  Staed
x765te  Coinbedx765ase  keystedx765ore  Brave  Software  Bycne  NeoLine  cjelfplplebdjjenllpjcblmjkfc
ffne  Chredx765ome  DisplayName  Temple  Importedx765ant  Fileedx765s/Froedx765file  %appdedx765ata%
\Opeedx765ra  Softedx765ware\Opedx765era  Staedx765ble  %appdedx765ata%\Opedx765era  Softwedx765ar
e\Opedx765era  GX  Staedx765ble  Opedx765era  Gedx765X  Stabedx765le  cihmoadaighcejopammfbmddcmdekc
je  %appdata%\atomic\Local  Storage  Ededx765ge  %appdaedx765ta%\Binaedx765nce  Walledx765ets/Binan
edx765ce  cnmamaachppnkjgnilldpdmkaakejnhae  Guild  klnaejjgbibmhlephnhpmaofohgkpgkd  jbdaocneliinm
jbjlgalhcelgbejmnid  nlgbhdfgdhgbiamfdfmbikcdghidoadd  nknhiehlkllippfakaeklbeglecifhad  hnfenkno
cfeedx765ofbddgcijnmedx765hnfnkdnaad  ookjlbkiijinhpmnjffcofjonbfbgaoc  CocCoc  afbcbjpbpfadlkmhm
clhkeeodmamcflc  Wallets/Atomic  ibnejdfjmmkpcnlpebklmnkcoihofec  fnjhmkhhmkbedx765jkkabndcnnoga
gogbneec  blnieiiffboillknjnepogjhkgnoapac  gaedmjdfmmahhbjefcbgaolhhanlaolb  ffnbelfdoeiohenkjib
```

ETHICAL CONSIDERATIONS

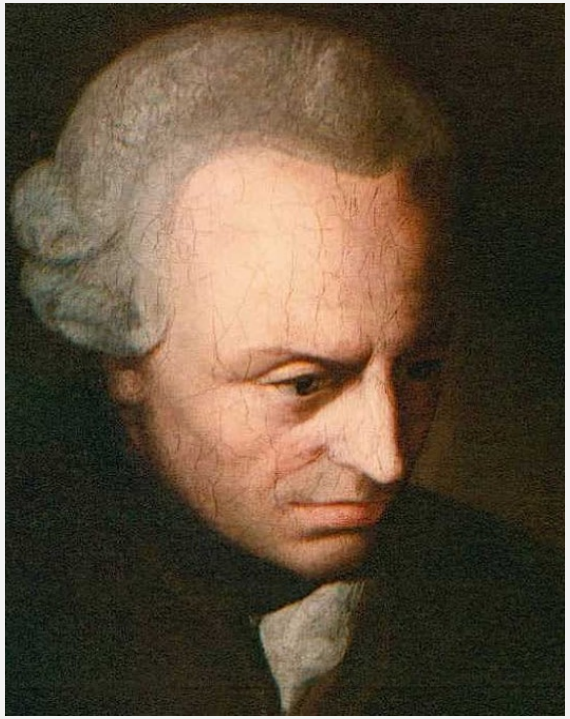
WHAT QUESTIONS DOES OBFUSCATION POSE?

- Dishonesty
- Waste
- “Free-riding”
- Pollution
- Ends and means dichotomy
- Asymmetries of power, knowledge

SYSTEMS OF ETHICS

Ayn Rand

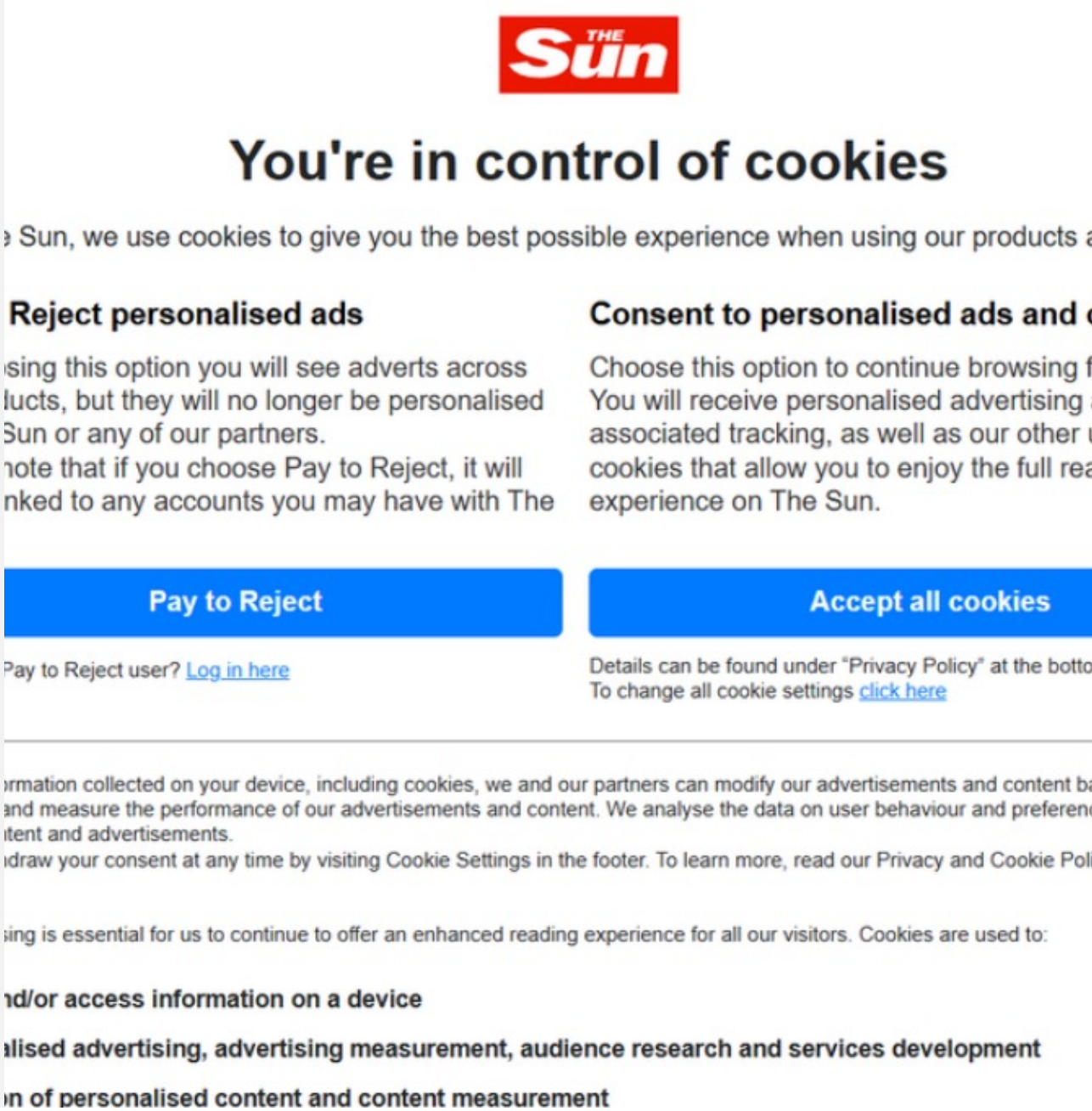
Different theories yield different conclusions as to what is desirable.



Immanuel Kant

PROPERTY RIGHTS

- Posting untrue information on social media can be seen as entitlement to the use of a platform.
- This raises questions about property rights.



POLLUTION

- Obfuscation necessitates processing more data.
- This in turn leads to more environmental pollution caused by the need for more energy.

CAN LYING BE JUSTIFIED?

- Dishonesty can be seen as necessary

- Individuals have less power to control information flow than governments, companies

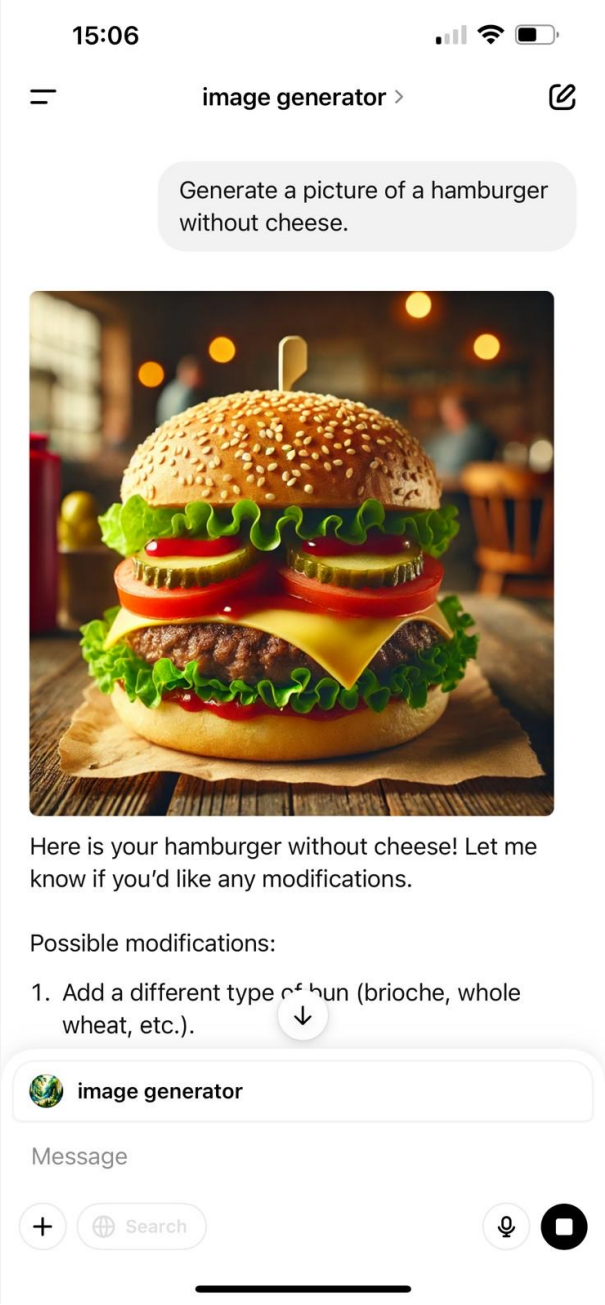
GOOD ASPECTS OF OBFUSCATION

https://www.business-humanrights.org/en/latest-news/what-are-the-panama-papers-a-guide-to-the-biggest-data-leak-in-history/

Individuals fearing retribution for sharing damning information can still do it by hiding their identity.

BAD ASPECTS OF OBFUSCATION

- It leads to data pollution.
- This can skew AI models trained on said data.

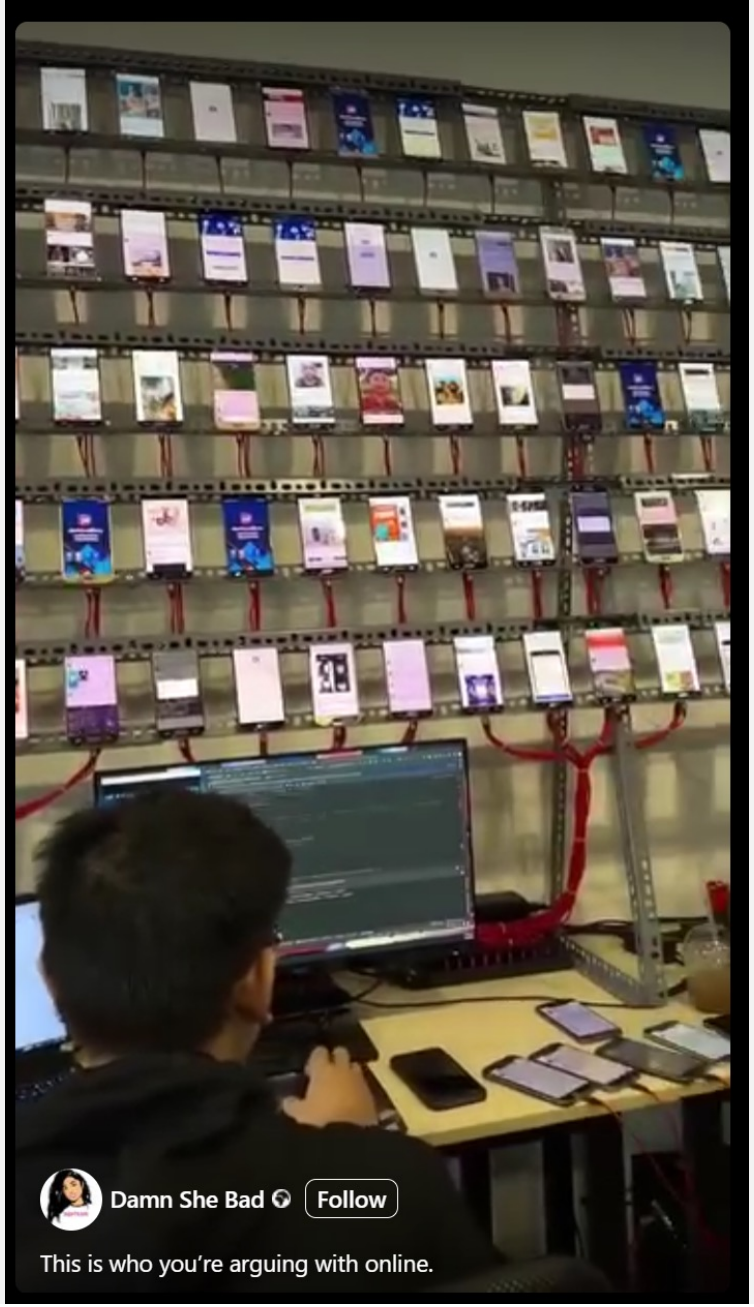


WEAPONIZING OBFUSCATION

Obfuscation is already being deployed on a very large scale, for example to prevent voters from knowing the true views of a political candidate.

TROLL FARMS

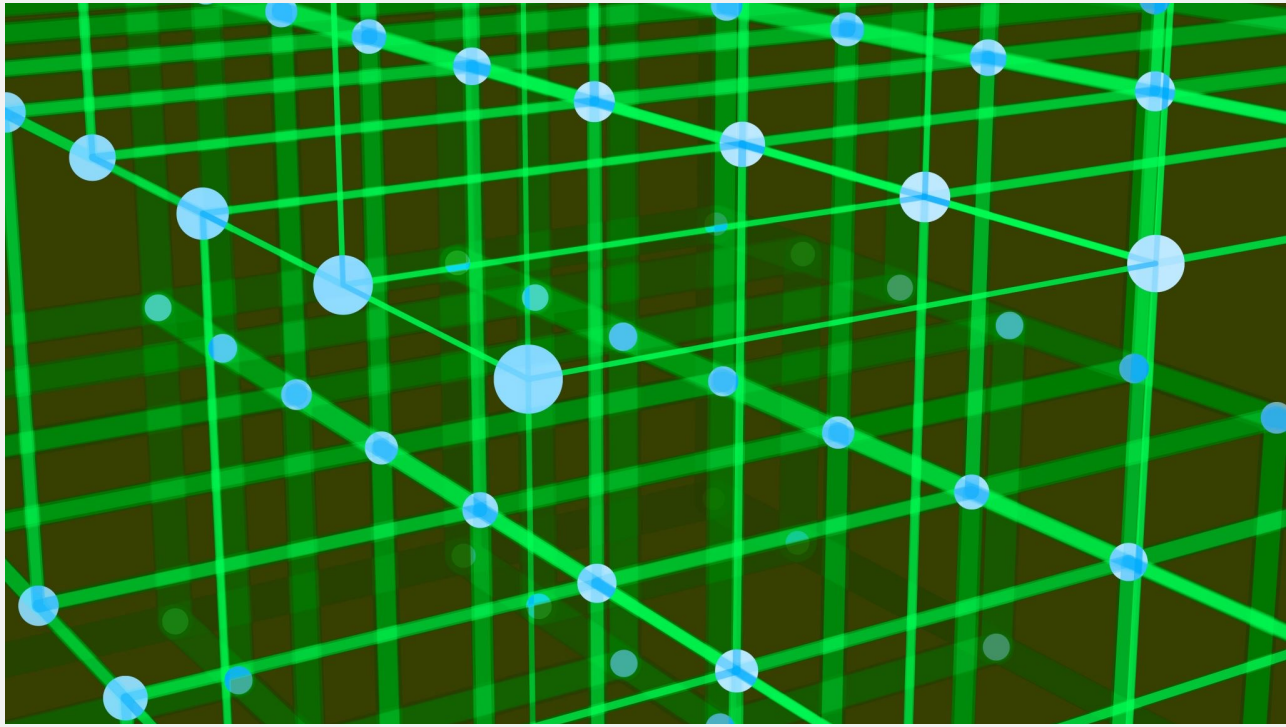
Fake argumentation online designed to radicalize people.



TAKEAWAYS FOR THE FUTURE

GENERATIVE ADVERSARIAL NETWORKS

- Obfuscation data can be generated
- It can also be detected with similar tools.



ARMS RACE, DETECTION TOOLS

BBC

Home News Sport Business Innovation Culture Arts Travel Earth Audio Video Live

Deepfake detection tool unveiled by Microsoft

1 September 2020

Share Save

Leo Kelion
Technology desk editor

https://www.bbc.com/news/technology-53984114

Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints

Ning Yu[1,2] Larry Davis[1] Mario Fritz[3]
[1]University of Maryland, College Park
[2]Max Planck Institute for Informatics
Saarland Informatics Campus, Germany
[3]CISPA Helmholtz Center for Information Security
Saarland Informatics Campus, Germany
ningyu@mpi-inf.mpg.de lsd@cs.umd.edu fritz@cispa.saarland

Abstract

Recent advances in Generative Adversarial Networks (GANs) have shown increasing success in generating photorealistic images. But they also raise challenges to visual forensics and model attribution. We present the first study of learning GAN fingerprints towards image attribution and using them to classify an image as real or GAN-generated. For GAN-generated images, we further identify their sources. Our experiments show that (1) GANs carry distinct model fingerprints and leave stable fingerprints in their generated images, which support image attribution; (2) even minor differences in GAN training can result in different fingerprints, which enables fine-grained model authentication; (3) fingerprints persist across different image frequencies and patches and are not biased by GAN artifacts; (4) fingerprint finetuning is effective in immunizing against five types of adversarial image perturbations; and (5) comparisons also show our learned fingerprints consistently outperform several baselines in a variety of setups [1].

Figure 1. A t-SNE [43] visual comparison between our fingerprint features (right) and the baseline inception features [52] (left) for image attribution. Inception features are highly entangled, indicating the challenge to differentiate high-quality GAN-generated images from real ones. However, our result shows any single difference in GAN architectures, training sets, or even initialization seeds can result in distinct fingerprint features for effective attribution.

At the same time, however, the success of GANs has raised two challenges to the vision community: visual forensics and intellectual property protection.

GAN challenges to visual forensics. There is a widespread concern about the impact of this technology when used maliciously. This issue has also received increasing public attention, in terms of disruptive conse-

https://arxiv.org/abs/1811.08180

FINAL MESSAGE

- Obfuscation is a double-edged sword
- Should be used responsibly

THE END

Sources

[1] Hussam Alkaissi and Samy Mcfarlane. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*, 15, 02 2023.

[2] Stuart Armstrong, Nick Bostrom, and Carl Shulman. Racing to the precipice: a model of artificial intelligence development. *AI & SOCIETY*, 31(2):201–206, 2016.

[3] Anish Athalye, Nicholas Carlini, and David Wagner. Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 274–283. PMLR, 10–15 Jul 2018.

[4] Michael Brennan, Sadia Afroz, and Rachel Greenstadt. Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity. *ACM Trans. Inf. Syst. Secur.*, 15(3), November 2012.

[5] Finn Brunton and Helen Nissenbaum. *Obfuscation: A User's Guide for Privacy and Protest.* The MIT Press, 09 2015.

[6] Michael Cholbi. *Understanding Kant's Ethics.* Cambridge University Press, 2016.

[7] Andrew J Dawson and Martin Innes. How Russia's Internet Research Agency Built its Disinformation Campaign. *The Political Quarterly*, 2019.

[8] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*, PP:1–1, 11 2021.

[9] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110:5802 – 5805, 2013.

[10] Sparsh Mittal. Power management techniques for data centers: A survey. *CoRR*, abs/1404.6681, 2014.

[11] Tara Smith. *Ayn Rand's Normative Ethics: The Virtuous Egoist.* Cambridge University Press, 2006.

[12] Vinicius Luis Trevisan de Souza, Bruno Augusto Dorta Marques, Harlen Costa Batagelo, and João Paulo Gois. A review on Generative Adversarial Networks for image generation. *Computers Graphics*, 114:13–25, 2023.